Heterogeneity and Aggregate Fluctuations

Minsu Chang*
Georgetown University

Xiaohong Chen
Yale University,
Cowles Foundation

Frank Schorfheide
University of Pennsylvania,
CEPR, PIER, NBER

This Version: July 4, 2023

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 coefficients. 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 and insignificant response of 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. F. Schorfheide: Department of Economics, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu. 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 Bence Bardoczy, Yongsung Chang, Mark Huggett, Dirk Krüger, Víctor Ríos-Rull, Harald Uhlig (co-editor), two anonymous referees, and 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 household or firm heterogeneity are widely used to study distributional effects of macroeconomic policies. In these models heterogeneity evolves dynamically and potentially interacts closely with macroeconomic aggregates like GDP, unemployment, and investment. However, it is an open question to what extent model predictions are consistent with empirical evidence. This paper develops a state-space model that can provide semi-structural evidence about the interaction of aggregate and distributional dynamics at business cycle frequencies. The state-transition equation takes the form of a linear functional vector autoregression (VAR), stacking macroeconomic aggregates and the log density of a cross-sectional distribution. We treat this log density as unobserved and specify a measurement equation for unit-level cross-sectional observations. Our model can be used to trace out the effects of aggregate shocks on cross-sectional distributions; assess whether fluctuations in cross-sectional distributions are important for aggregate fluctuations in the sense of Granger causality; and identify shocks to cross-sectional distributions and examine their effects on aggregate variables.

We represent the log-densities and the transition kernels in the functional VAR by finite-dimensional sieves with fixed basis functions and time-varying coefficients that capture the dynamics. This turns the state-transition equation into a linear VAR for the aggregate variables and the time-varying sieve coefficients. To avoid nonlinear filtering, we effectively linearize the measurement equation for the cross-sectional observation which leads to a convenient two-step estimation procedure. First, one computes maximum likelihood estimates (MLE) of the sieve coefficients from the cross-sectional data, separately for each time period. Second, one estimates a linear state-space model in which the MLEs summarize the cross-sectional data and serve as noisy measures of the sieve coefficients. We implement the second step with Bayesian techniques. A novel feature of our two-step approach is its ability to capture estimation uncertainty about the unobserved cross-sectional densities. For data driven selection of the sieve dimension, the number of VAR lags, and hyperparameters that control the variance of the prior distribution, we derive an approximation of the Bayesian marginal data density (MDD).

In addition to being able to capture density estimation uncertainty, there are multiple benefits associated with our functional approach. First, we model the entire cross-sectional distribution, and not just particular quantiles or inequality measures. Once that has been done, any distributional statistic can be computed by post-processing the initial density

estimates. This is particularly useful when users start out from a broad research question, e.g., how does a structural shock affect earnings inequality, instead of a narrow question of how the 10th and 90th percentiles respond to an aggregate shock. Second, our representation of the cross-sectional distributions through unnormalized log densities, can be coherently embedded in a *linear* autoregressive framework. Unlike quantiles, cumulative distribution functions, or probability density functions, the unnormalized log densities do not have to obey monotonicity or non-negativity that cannot be enforced in a linear law of motion.

Third, the framework facilitates the identification of structural shocks in multiple dimensions. In our empirical application, we study the propagation of novel shocks that are identified by their effect on the cross-sectional distribution. Although not explored in this paper, the inclusion of a cross-sectional distribution in a vector autoregressive law of motion may also assist the shock identification by ensuring that the VAR innovations span the structural shocks, e.g., if a policy maker systematically responds to fluctuations in the cross-sectional distribution, or by generating new sign restrictions that potentially reduce the size of identified sets, e.g., along the lines of Drautzburg and Amir-Ahmadi (2021).

A potential user may be concerned that the implementation costs of the proposed functional approach are substantially higher than including a few cross-sectional statistics, such as percentiles, in a VAR. One of the contributions of our paper is to show that this is not the case. If the user chooses to ignore the uncertainty associated with the cross-sectional density estimates, which is the *modus operandi* in VARs that are augmented by cross-sectional statistics, then the method amounts to stacking macro variables and sieve coefficient MLEs and estimating a VAR with the augmented vector. Our uncertainty quantification step can be easily implemented by estimating a linear Gaussian state-space model instead of a VAR. In addition to software that replicates the simulations and empirical analysis in this paper, we provide a simplified set of codes that implements the first step of the analysis (cross-sectional density estimation) and can be combined with existing VAR and state-space model estimation software.

We present two asymptotic results, assuming that the number of sieve dimension K, the cross-sectional sample size N, and the time series length T tend to infinity. Assuming that $K^3/N = o(1)$ our first Theorem shows that the difference between the exact likelihood and the likelihood under the linearized measurement equation vanishes. This result is confirmed in a computational exercise in which we use an approximately conditionally optimal particle filter, see Herbst and Schorfheide (2015), to numerically evaluate the exact likelihood function of the K-dimensional nonlinear state-space model. Because the error from linearizing

the measurement equation is small and the runtime is reduced by orders of magnitude, we recommend the linearization for empirical work. Our second Theorem shows that the discrepancy between the exact MDD and our MDD approximation (used for data-driven model selection) also vanishes, provided that $KT/N = o(1)$ holds in addition. As a corollary, by maximizing the approximate MDD we also "almost" achieve the maximum value for the exact MDD. Again, we think that the computational gain (in terms of time and precision) from working with an approximation outweighs potential distortions from the approximation error in empirical work.

In a simulation experiment, we generate data from a version of the Krusell and Smith (1998) model, henceforth KS, which we solve using the method proposed by Winberry (2018). The estimates of our functional model can reproduce the evolution of the distribution of asset holdings and the response of the cross-sectional distribution to a technology shock. The MDD shrinks the coefficients that control the spillover from the lagged cross-sectional distribution to current aggregate capital to zero. Thus, according to the estimated functional model, the cross-sectional distribution does not Granger-cause the aggregate capital stock. This finding is a manifestation of the well-known aggregation property of the KS model, which holds exactly in our version of the model. Because of the absence of Granger causality, the estimated response of capital to a technology shock is identical to that from a VAR in the two aggregate variables only, using a single lag as in the functional model.

We conduct a second simulation with a modified KS model in which we break the perfect aggregation property and the cross-sectional distribution Granger-causes aggregate capital. In this case the MDD does not shrink the spillover coefficients to zero, and, fixing the lag length in both models at one, there are discrepancies between the functional and aggregate VAR impulse response function (IRF) of capital. Because the functional model can be rewritten as an infinite-order vector autoregressive system in terms of the two aggregate variables, increasing the number of lags in the aggregate VAR makes this discrepancy vanish.

In an empirical application we fit our model to aggregate total factor productivity (TFP) growth, GDP growth, employment, and cross-sectional data on labor earnings divided by GDP per capita. We obtain the following results. First, based on the degree of shrinkage selected with the MDD criterion, the Granger-causal relationship from the cross-sectional income distribution to the aggregate variables is weak. Estimated impulse responses of the aggregate variables based on the functional model and an aggregate VAR with the same lag length are very similar, mirroring the simulation results from the baseline KS economy. Second, the earnings density responds to a TFP shock in a way that inequality falls. The

implied 10th percentile and the fraction of individuals earning less than per capita GDP rise, whereas normalized earnings at the 50th and 90th percentile and the Gini coefficient fall. Overall, the effects are quantitatively small and largely driven by individuals transitioning between work and unemployment.

Third, the proposed functional model delivers sharper estimates of percentile and inequality measure IRFs than a VAR that combines these statistics directly with the aggregate variables.[1] We obtain the same result based on data simulated from the KS economy. Our functional approach also guarantees that percentiles do not cross in response to shocks and in a forward simulation of the model. Fourth, using maximum share-of-variance identification schemes, we find that IRFs from shocks that maximize the variance share of GDP or employment look very similar to TFP IRFs. Distributional shocks that explain the maximal variance share in the overall cross-sectional distribution or the Gini coefficient do not have a significant effect on GDP.[2]

The methods developed in this paper speak to some but not all aspects of the broader question of whether heterogeneity matters for business cycle fluctuations. The functional model generates direct estimates for responses of cross-sectional distributions to aggregate shocks and for the effects of shocks to the cross-sectional distribution on aggregate outcomes. Quantitative HA models can be evaluated on their ability to replicate these empirical IRFs. Assessing the role that heterogeneity plays for the propagation of aggregate shocks on aggregate outcomes is more delicate. Heterogeneity may matter because it affects the mapping from deep technology and preference parameters into the VAR representation. Our functional model cannot unveil this mapping; it simply generates estimates of the VAR parameters that can track the observed business cycle fluctuations.[3] The proposed method can, however, be used to assess whether cross-sectional distributions Granger-cause aggregates and to examine discrepancies between IRFs from the estimated functional model and a VAR that only uses aggregate data. A well-fitting quantitative heterogeneous agent (HA) model will have to be able to reproduce the empirical Granger causality and IRF mismatch patterns.

The strong overlap between estimated IRFs from the functional model and an aggregate

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

[2] Identification based on the maximum share-of-variance of an observable dates back to Faust (1998) and Uhlig (2003), and has recently been applied by Angeletos, Collard, and Dellas (2020) to study what the authors refer to as the anatomy of business cycles.

[3] There is no analogue in our framework to the popular experiment of comparing HA and RA IRFs by shutting down cross-sectional heterogeneity as in, for instance, Krueger, Mitman, and Perri (2016), Ahn, Kaplan, Moll, Winberry, and Wolf (2018), Kaplan and Violante (2018), Ottonello and Winberry (2020), Bayer, Born, and Luetticke (2020), Cho (2020), and Villalvazo (2021).

VAR is consistent with the HA and RA New Keynesian model estimation results reported in Bayer, Born, and Luetticke (2020). It is also consistent with the notion that many HA models deliver approximate aggregation results; see Chang, Kim, and Schorfheide (2013), Werning (2015) and Berger, Bocola, and Dovis (2019). The latter paper finds that the role of state-dependent preference shocks, which summarize all the information from the cross-section relevant for aggregate dynamics, is quantitatively small. Bilbiie, Primiceri, and Tambalotti (2022) estimate the amplification of business cycle fluctuations due to precautionary savings in a medium-scale New Keynesian model with savers and hand-to-mouth households. The estimated amplification is observationally similar to more volatile shocks in an RA environment. This suggests that their framework will also be consistent with the strong overlap between IRFs from the functional model and the aggregate VAR that we document.

HA models are frequently used to study the effect of aggregate shocks on cross-sectional distributions. 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. Examples of research examining the effect of distributional shocks on aggregate variables are papers by Huggett (1997) and Auclert and Rognlie (2020). The former shows that a redistribution of asset holdings among households while keeping the overall capital stock fixed at its steady state level, triggers a 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.

The structure of the transition equation in our functional model resembles that of HA models solved with linearization techniques. This solution method was initially proposed by Reiter (2009), has been further developed in several papers, including Kaplan and Violante (2018), Childers (2018), and Winberry (2018), and is often used for the likelihood-based estimation of HA models as in Mongey and Williams (2017), Acharya, Chen, Del Negro, Dogra, Matlin, and Sarfati (2019), Bayer, Born, and Luetticke (2020), Cho (2020), and Liu and Plagborg-Møller (2022). In terms of likelihood construction, the Liu and Plagborg-Møller (2022) paper is most closely related to ours, because it also uses the density of

cross-sectional observations. However, in their case the density is part of the HA model solution, whereas we estimate it flexibly from the data.

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). Unlike the majority of papers, we do not assume that the functions are observed without error. Applications of functional data analysis in macroeconometrics are growing steadily. Many of them are related to the yield curve, e.g., Diebold and Li (2006) and Inoue and Rossi (2021). 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 MDD that is used for dimensionality and hyperparameter selection. Theoretical results on the accuracy of the approximations are presented in Section 3. Implementation details such as the choice of basis functions, the choice of prior distributions, and the posterior sampler are discussed in Section 4. Section 5 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 6 and Section 7 concludes. Supplemental derivations, and additional computational details and empirical results are relegated to the Online Appendix.[4]

2 A Functional State Space Model

We now develop the functional model, starting from an infinite-dimensional specification in Section 2.1. Rather than working with an infinite-dimensional model, we consider a

[4]Julia scripts to replicate the simulation experiment and the empirical analysis and to combine the functional analysis with existing VAR and state-space model estimation software are available from the authors' websites.

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 propose a large-sample approximation of the marginal data density that can be used to select the degree of the sieve-approximation, the number of lags, and prior hyperparameters in a data-driven manner.

2.1 An Infinite-Dimensional Model

To fix ideas, we start with an infinite-dimensional model for 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 work with log densities defined as $\ell_t(x) = \ln p_t(x)$. The advantage of log-densities is that they are not subject to non-negativity or monotonicity restrictions that would be difficult to enforce in a linear vector autoregressive law of motion. We decompose Y_t and ℓ_t into a deterministic component $\big(Y_*, \ell_*(x)\big)$ and fluctuations around it. Let

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

For notational convenience we assume 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. The deviations from the deterministic component $\big(Y_t, \ell_t(x)\big)$ evolve jointly according to the following linear functional VAR:

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

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 $\Omega_{yl}(x)$. For now, (2) should be interpreted as a reduced-form functional VAR in which $u_{y,t}$ and $u_{l,t}(x)$ are one-step-ahead forecast errors. The time t density for the

cross-sectional observations x_{it} is given by

$$p_t(x) = \frac{\exp\{\ell_*(x) + \tilde{\ell}_t(x)\}}{\int \exp\{\ell_*(\tilde{x}) + \tilde{\ell}_t(\tilde{x})\}d\tilde{x}}. \tag{3}$$

The numerator ensures that the density $p_t(x)$ integrates to one, which is not automatically guaranteed by the linear law of motion in (2).

2.2 A Collection of Finite-Dimensional Models

Instead of working with the infinite-dimensional model comprising (1), (2), and (3) we estimate a K-dimensional specification, where K is determined in a data-driven manner.[5] Let

$$\ell_t^{(K)}(x) = \sum_{k=1}^{K} \alpha_{k,t}\zeta_k(x) = \big[\zeta_1(x), \ldots, \zeta_K(x)\big] \cdot \begin{bmatrix} \alpha_{1,t} \\ \vdots \\ \alpha_{K,t} \end{bmatrix} = \zeta'(x)\alpha_t, \tag{4}$$

$\ell_*^{(K)}(x) = \zeta'(x)\alpha_*$, and $\tilde{\ell}_t^{(K)}(x) = \zeta'(x)\tilde{\alpha}_t$. 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. 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.

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 (2) as follows:

$$\begin{aligned} B_{ll}^{(K)}(x, \tilde{x}) &= \zeta'(x)B_{ll}\xi(\tilde{x}), \quad B_{yl}^{(K)}(x) = B_{yl}\xi(\tilde{x}) \\ B_{ly}^{(K)}(x) &= \zeta(x)'B_{ly}, \quad u_{l,t}^{(K)}(x) = \zeta'(x)u_{\alpha,t}, \end{aligned} \tag{5}$$

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 $K \times n_y$. Plugging (5) into (2) and assuming that the innovations $u_t' = [u_{y,t}', u_{\alpha,t}']$ are Gaussian yields the following vector autoregressive system for the macroeconomic aggregates and the

[5] We discuss in the Online Appendix how a finite-dimensional approximation can be derived from the infinite-dimensional specification. In this derivation, given the dimension K, the α_ts minimize the Kullback-Leibler divergence from the infinite-dimensional density.

sieve coefficients (omitting K superscripts):

$$\begin{bmatrix} Y_t - Y_* \\ \alpha_t - \alpha_* \end{bmatrix} = \begin{bmatrix} \Phi_{yy} & \Phi_{y\alpha} \\ \Phi_{\alpha y} & \Phi_{\alpha\alpha} \end{bmatrix} \begin{bmatrix} Y_{t-1} - Y_* \\ \alpha_{t-1} - \alpha_* \end{bmatrix} + \begin{bmatrix} u_{y,t} \\ u_{\alpha,t} \end{bmatrix}, \quad u_t \sim \mathcal{N}(0, \Sigma), \tag{6}$$

where $\Phi_{yy} = B_{yy}$, $\Phi_{y\alpha} = B_{yl}C_\alpha$, $\Phi_{ly} = B_{ly}$, $\Phi_{\alpha\alpha} = B_{ll}C_\alpha$, and $C_\alpha = \int \xi(\tilde{x})\zeta'(\tilde{x})d\tilde{x}$.

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 draws x_{it}, $i = 1, \ldots, N$ from the K'th order cross-sectional density $p_t^{(K)}(x)$:

$$x_{it} \sim \text{iid } p_t^{(K)}(x) = \frac{\exp\{\ell_t^{(K)}(x)\}}{\int \exp\{\ell_t^{(K)}(\tilde{x})\}d\tilde{x}}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{7}$$

Write the joint density of the x_{it}s as the product of the marginals and let $\mathcal{L}^{(K)}(\alpha_t|X_t) = \frac{1}{N}\sum_{i=1}^N \ell_t^{(K)}(x_{it})$. Then the measurement equation for the time t cross-sectional observations $X_t = [x_{1t}, \ldots, x_{Nt}]'$ can be expressed as

$$\begin{aligned} p^{(K)}(X_t|\alpha_t) &= \exp\left\{N\mathcal{L}^{(K)}(\alpha_t|X_t)\right\}, \quad \mathcal{L}^{(K)}(\alpha_t|X_t) = \bar{\zeta}'(X_t)\alpha_t - \varphi(\alpha_t) \\ \bar{\zeta}(X_t) &= \frac{1}{N}\sum_{i=1}^N \zeta(x_{it}), \quad \varphi(\alpha_t) = \ln \int \exp\{\zeta'(x)\alpha_t\}\, dx. \end{aligned} \tag{8}$$

Here $\bar{\zeta}(X_t)$ is the K-dimensional vector of sufficient statistics and $\varphi(\alpha_t)$ is the log normalization constant for the density $p_t^{(K)}(x)$. Below we will sometimes use β as a generic parameter for the log likelihood function $\mathcal{L}^{(K)}(\cdot|X_t)$. The assumption of x_{it} being *iid* across i and t is consistent with data sets that comprise repeated cross sections.[6] To the extent that the cross-sectional densities $p_t^{(K)}(x)$ 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.

To summarize, our K-dimensional functional model has a state-space representation with state vector α_t, measurement equation (8), and state-transition equation (6). When estimating the functional model we allow for $p > 1$ lags in the state-transition equation. However, to keep the notation simple, we will for now proceed with a single lag, $p = 1$.

[6] In practice N may vary with t. 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).

2.3 Approximate Filtering

The likelihood function of the functional model can be evaluated using a filter. To simplify the notation, we collect the parameters $(\alpha_*, Y_*, \Phi, \Sigma)$ in the vector θ. Here the matrix Φ comprises Φ_{yy}, $\Phi_{y\alpha}$, $\Phi_{\alpha y}$, and $\Phi_{\alpha\alpha}$. 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:

$$\begin{aligned} p^{(K)}(Y_t, \alpha_t|Y_{1:t-1}, X_{1:t-1}, \theta) &= \int p_G^{(K)}(Y_t, \alpha_t|Y_{t-1}, \alpha_{t-1}, \theta) p^{(K)}(\alpha_{t-1}|Y_{1:t-1}, X_{1:t-1}, \theta) d\alpha_{t-1} \\ p^{(K)}(Y_t, X_t|Y_{1:t-1}, X_{1:t-1}, \theta) &= \int p^{(K)}(X_t|\alpha_t) p^{(K)}(Y_t, \alpha_t|Y_{1:t-1}, X_{1:t-1}, \theta) d\alpha_t \\ p^{(K)}(\alpha_t|Y_{1:t}, X_{1:t}, \theta) &\propto p^{(K)}(X_t|\alpha_t) p^{(K)}(Y_t, \alpha_t|Y_{1:t-1}, X_{1:t-1}, \theta). \end{aligned} \tag{9}$$

The first equation in display (9) iterates the state-transition equation (6) forward and integrates over the hidden state α_{t-1}. We use the G subscript to indicate that the state-transition equation is linear and Gaussian. The second equation generates a forecast of the observables (Y_t, X_t) using the measurement equation (8). The third equation updates the density of the hidden state α_t using Bayes Theorem (here $\propto$ denotes proportionality). The density $p^{(K)}(X_t|\alpha_t)$ defined in (8) does not depend on θ.

Because the measurement equation for X_t is nonlinear and non-Gaussian, the predictive density $p^{(K)}(Y_t, X_t|Y_{1:t-1}, X_{1:t-1}, \theta)$ cannot be calculated analytically. While one could use a sequential Monte Carlo filter to generate a numerical approximation, we approximate the measurement equation $p^{(K)}(X_t|\alpha_t)$ in a way that enables analytical calculations using the Kalman filter recursions. A second-order Taylor series expansion of $\mathcal{L}(\alpha_t|X_t)$ in (8) around the maximum likelihood estimator (MLE)

$$\hat{\alpha}_t = \text{argmax}_{\beta \in \mathcal{A}} \, \mathcal{L}^{(K)}(\beta|X_t) \tag{10}$$

yields ($\mathcal{A}$ is defined in Assumption 1 below)

$$p^{(K)}(X_t|\alpha_t) = \exp\left\{ N\mathcal{L}^{(K)}(\hat{\alpha}_t|X_t) - \frac{N}{2}(\alpha_t - \hat{\alpha}_t)'\hat{V}_t^{-1}(\alpha_t - \hat{\alpha}_t) + N\mathcal{R}(\alpha_t) \right\}. \tag{11}$$

$\hat{V}_t$ is the negative inverse Hessian associated with $\mathcal{L}(\beta|X_t)$ evaluated at $\hat{\alpha}_t$, and $\mathcal{R}(\alpha_t)$ is the

remainder term from the second-order Taylor series approximation. Define

$$\begin{aligned} p_{KF}^{(K)}(X_t|\alpha_t) &= \exp\left\{N\mathcal{L}^{(K)}(\hat{\alpha}_t|X_t) - \frac{N}{2}(\alpha_t-\hat{\alpha}_t)'\hat{V}_t^{-1}(\alpha_t-\hat{\alpha}_t)\right\} \\ &= p_{pen}^{(K)}(X_t|\hat{\alpha}_t)\times p_G^{(K)}(\hat{\alpha}_t|\alpha_t), \end{aligned} \tag{12}$$

where

$$p_{pen}^{(K)}(X_t|\hat{\alpha}_t) = \exp\left\{N\mathcal{L}^{(K)}(\hat{\alpha}_t|X_t)\right\}\left(\frac{2\pi}{N}\right)^{K/2}|\hat{V}_t|^{1/2}.$$

We use the *pen* subscript to indicate that the expression can be interpreted as the maximized likelihood function of X_t penalized by $(2\pi/N)^{K/2}|\hat{V}_t|^{1/2}$. The $p_G^{(K)}(\hat{\alpha}_t|\alpha_t)$ term in (12) is the Gaussian density corresponding to the "measurement" equation[7]

$$\hat{\alpha}_t = \alpha_t + N^{-1/2}\eta_t, \quad \eta_t \sim \mathcal{N}\big(0,\hat{V}_t\big). \tag{13}$$

Replacing the exact measurement equation $p^{(K)}(X_t|\alpha_t)$ by the approximation $p_{KF}^{(K)}(X_t|\alpha_t)$ the filter iterations in (9) become

$$\begin{aligned} p_{KF}^{(K)}(Y_t,\alpha_t|Y_{1:t-1},X_{1:t-1},\theta) &= \int p_G^{(K)}(Y_t,\alpha_t|Y_{t-1},\alpha_{t-1},\theta)p_{KF}^{(K)}(\alpha_{t-1}|Y_{1:t-1},X_{1:t-1},\theta)d\alpha_{t-1} \\ p_{KF}^{(K)}(Y_t,X_t|Y_{1:t-1},X_{1:t-1},\theta) &= p_{pen}^{(K)}(X_t|\hat{\alpha}_t)\int p_G^{(K)}(\hat{\alpha}_t|\alpha_t)p_{KF}^{(K)}(Y_t,\alpha_t|Y_{1:t-1},X_{1:t-1},\theta)d\alpha_t \\ &= p_{pen}^{(K)}(X_t|\hat{\alpha}_t)p_{KF}^{(K)}(Y_t,\hat{\alpha}_t|Y_{1:t-1},X_{1:t-1},\theta) \\ p_{KF}^{(K)}(\alpha_t|Y_{1:t},X_{1:t},\theta) &\propto p_G^{(K)}(\hat{\alpha}_t|\alpha_t)p_{KF}^{(K)}(Y_t,\alpha_t|Y_{1:t-1},X_{1:t-1},\theta). \end{aligned} \tag{14}$$

Because the state-transition density $p_G^{(K)}(Y_t,\alpha_t|Y_{t-1},\alpha_{t-1},\theta)$ is Normal under the assumption that the shock vector u_t is normally distributed, all the densities in (14) are Gaussian – we initialize the filter with a Gaussian density for α_0 – and the Kalman filter (KF) recursions can be used to track means and covariance matrices. It can be verified by induction that $X_{1:t-1}$ affects the densities only through $\hat{\alpha}_{1:t-1}$. Thus, we can replace $X_{1:t-1}$ by $\hat{\alpha}_{1:t-1}$ in the conditioning sets in (14).

To summarize, our 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

[7]This is not a statement about the asymptotic sampling distribution of the MLE $\hat{\alpha}_t$.

measures of the latent α_ts.

2.4 Bayesian Model Selection Criterion

In empirical work the likelihood function is combined with a prior distribution $p^{(K)}(\theta|\lambda)$ which depends on a vector λ of hyperparameters that control the variance of the prior. We use the Bayesian marginal data density (MDD) to determine λ, the sieve order K, and the number of lags in the state-transition equation (2).[8] The MDD is defined as

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

This expression is very difficult to compute. As we discussed in Section 2.3 the computation of $p^{(K)}(Y_t, X_t|Y_{1:t-1}, X_{1:t-1}, \theta)$ requires a nonlinear filter. Moreover, the high-dimensional parameter vector θ needs to be integrated out numerically.

We simplify the problem in two steps. First, we replace the conditional density of (Y_t, X_t) by the KF approximation and define

$$p_{KF}^{(K)}(Y_{1:T}, X_{1:T}|\lambda) = \left(\prod_{t=1}^{T} p_{pen}^{(K)}(X_t|\hat{\alpha}_t) \right) \int \left(\prod_{t=1}^{T} p_{KF}^{(K)}(Y_t, \hat{\alpha}_t|Y_{1:t-1}, \hat{\alpha}_{1:t-1}, \theta) \right) p^{(K)}(\theta|\lambda) d\theta. \tag{16}$$

Second, we set the measurement error variance in (13) to zero which will allow us to replace the KF density $p_{KF}^{(K)}(Y_t, \hat{\alpha}_t|Y_{1:t-1}, \hat{\alpha}_{1:t-1}, \theta)$ by the Gaussian state transition density $p_G^{(K)}(Y_t, \alpha_t = \hat{\alpha}_t|Y_{t-1}, \alpha_{t-1} = \hat{\alpha}_{t-1}, \theta)$ such that we obtain the following MDD approximation:

$$\begin{aligned} & p_*^{(K)}(Y_{1:T}, X_{1:T}|\lambda) \\ = & \left(\prod_{t=1}^{T} p_{pen}^{(K)}(X_t|\hat{\alpha}_t) \right) \int \left(\prod_{t=1}^{T} p_G^{(K)}(Y_t, \alpha_t = \hat{\alpha}_t|Y_{t-1}, \alpha_{t-1} = \hat{\alpha}_{t-1}, \theta) \right) p^{(K)}(\theta|\lambda) d\theta. \end{aligned} \tag{17}$$

The integral on the right-hand side of the equation is the MDD associated with the VAR in (6) where the latent α_ts are replaced by the MLEs $\hat{\alpha}_t$. There exists a large literature on how to evaluate VAR MDDs and we will choose a conjugate prior distribution $p^{(K)}(\theta|\lambda)$ that allows us to compute $p_*^{(K)}(Y_{1:T}, X_{1:T}|\lambda)$ analytically.

[8] MDD based hyperparameter selection or averaging has been widely used in the VAR literature; see, for instance, Del Negro and Schorfheide (2004) and Giannone, Lenza, and Primiceri (2015).

3 Approximation Error Bounds and Model Selection

In this section we formally characterize the accuracy of the KF approximation and the MDD approximation. We begin in Section 3.1 by defining some notation and stating a basic assumption and a lemma. We present an approximation error bound for the KF state distribution and likelihood increment in Section 3.2 and provide a bound for the MDD approximation in Section 3.3. Model selection is discussed in Section 3.4. Throughout this section we write $|f(N,K;\omega)| \lesssim \alpha_{K,N}$ to mean that there is a constant C such that for every $\epsilon > 0$ we can find (K_0, N_0) such that $\mathbb{P}^\omega\{|f(N,K;\omega)| \leq C\alpha_{K,N}\} \geq 1-\epsilon$ for $K > K_0$ and $N > N_0$. Moreover, we write $f(N,K;\omega) \asymp 1 \pm \alpha_{K,N}$ to denote that $|f(N,K;\omega) - 1| \lesssim \alpha_{K,N}$.

3.1 Preliminaries

Cross-sectional Densities. The derivatives of $\mathcal{L}^{(K)}(\beta|X)$ with respect to β of order two and higher only depend on derivatives of $\varphi(\beta)$; see (8). Let $\mathcal{L}^{(K)(2)}(\beta)$ be the matrix of second derivatives and $\mathcal{L}^{(K)(4)}_{hklm}(\beta)$ be the fourth derivatives, where the subscript indicates the elements of β with respect to which we are differentiating.

Assumption 1 *The fourth derivatives $\mathcal{L}^{(K)(4)}_{hklm}(\beta)$ are uniformly bounded by a constant M for $\beta \in \mathcal{A} \subseteq \mathbb{R}^K$, where $\mathcal{A} = \{\beta \in R^K : \varphi(\beta) \leq C < \infty\}$. The constant does not depend on K or X. Moreover, the matrix $[-\mathcal{L}^{(K)(2)}(\beta)]$ is positive definite at $argmax_{\beta\in\mathcal{A}}\, \mathcal{L}^{(K)}(\beta|X_t)$.*

In our application we use cubic splines and assume that $x \in [\underline{x}, \overline{x}]$. Thus, Assumption 1 is satisfied. To state the following lemma we add a (K) superscript to the remainder $\mathcal{R}(\cdot)$ and the densities $p(X_t|\cdot)$, and a 1:N subscript to X_t to highlight the dependence of these objects on (K, N).

Lemma 1 *Suppose Assumption 1 is satisfied. Then, there is a finite positive constant C and an $N_o(C)$ such that uniformly for all $\alpha_t \in int(\mathcal{A})$, for all $t = 1, ..., T$, and $N \geq N_0 > K^3$*

$$\exp\left\{N\mathcal{R}^{(K)}(\alpha_t)\right\} \asymp 1 \pm \sqrt{\frac{K^3}{N}}, \quad \frac{p^{(K)}(X_{1:N,t}|\alpha_t)}{p^{(K)}_{KF}(X_{1:N,t}|\alpha_t)} \asymp 1 \pm \sqrt{\frac{K^3}{N}}.$$

3.2 Kalman Filter Approximation Error

Our first set of approximation results is about the relationship between the density of the latent state α_t given time t information under the exact filter in (9) and the KF in (14), and the likelihood increments under the two filters. Throughout the remainder of this section we will assume that $K^3/N = o(1)$.

Theorem 1 *Suppose Assumption 1 is satisfied. Then, there exists a constant C and an $N_o(C)$ that do not depend on α_t or θ such that for all $N \geq N_0 \geq K^3$ and $t = 1, \ldots, T$*

$$\frac{p^{(K)}(\alpha_t|Y_{1:t}, X_{1:N,1:t}, \theta)}{p^{(K)}_{KF}(\alpha_t|Y_{1:t}, X_{1:N,1:t}, \theta)} \asymp 1 \pm \sqrt{\frac{K^3}{N}}.$$

and

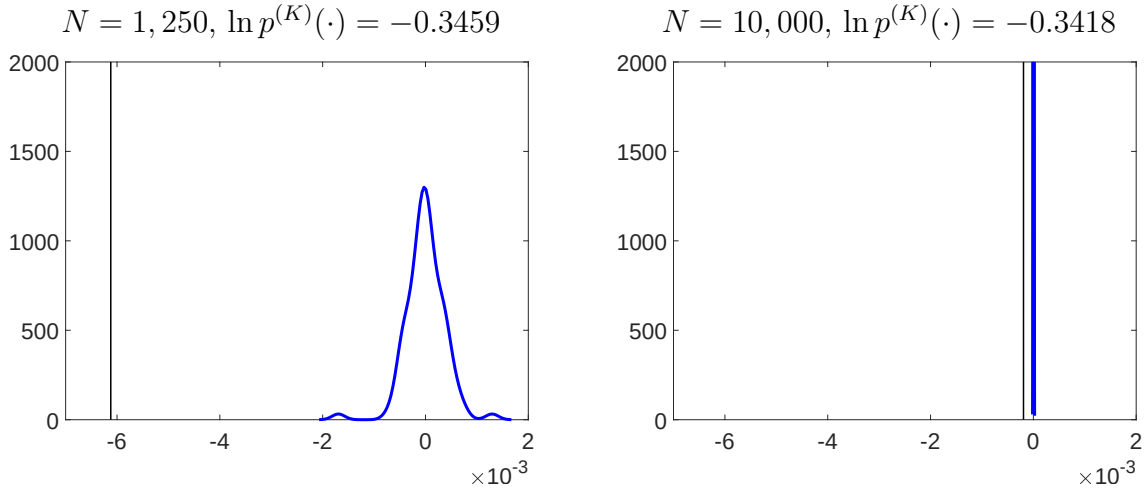
$$\frac{p^{(K)}(Y_t, X_{1:N,t}|Y_{1:t-1}, X_{1:N,1:t-1}, \theta)}{p^{(K)}_{KF}(Y_t, X_{1:N,t}|Y_{1:t-1}, X_{1:N,1:t-1}, \theta)} \asymp 1 \pm \sqrt{\frac{K^3}{N}}.$$

A proof of the theorem is provided in the Online Appendix. To provide a numerical illustration, we generate a data set $(Y_{1:T}, X_{1:N,1:T})$ from the calibrated heterogeneous agent model of Section 5 and evaluate the likelihood function at the posterior mean estimate θ_* of the functional model. The exact likelihood increments $p^{(K)}(Y_t, X_{1:N,t}|Y_{1:t-1}, X_{1:N,1:t-1}, \theta)$ are computed with a particle filter (PF), which was first developed by Gordon and Salmond (1993). A PF approximates the distribution of $\alpha_t|(Y_{1:t}, X_{1:N,1:t}, \theta)$ through a swarm of particles $\{\alpha_t^j, W_t^j\}_{j=1}^M$. The approximation is stochastic and its accuracy depends crucially on the mutation of the particle values between periods $t-1$ and t. We use the linearized measurement equation (13) to derive an approximately conditionally optimal proposal density for the α_{t-1}^j particle values. The general implementation of the PF follows Herbst and Schorfheide (2015) and details are provided in the Online Appendix. Conditional on θ_* we compute PF approximations of

$$\ln p^{(K)}(Y_{1:T}, X_{1:N,1:T}|\theta_*) = \sum_{t=1}^{T} \ln p^{(K)}(Y_t, X_{1:N,t}|Y_{1:t-1}, X_{1:N,1:t-1}, \theta_*)$$

for $K = 8$ and $T = 25$ and two choices of N. We use $M = 5,000$ particles. We run the stochastic filter $n_{sim} = 100$ and plot kernel density estimates for the n_{sim} log likelihood values.

Figure 1: Kalman Filter Log Likelihood Approximation Errors



Notes: Log likelihood values $\ln p^{(K)}(Y_{1:T}, X_{1:N,1:T}|\theta_*)/(N+2)$ in deviations from the "exact" value. Vertical lines correspond to KF approximations and densities represent distribution of PF log likelihood values across multiple runs for $M = 5,000$ particles.

Results are plotted in Figure 1. Because the PF likelihood approximations are unbiased, we first calculate the mean of the likelihood values across the n_{sim} runs of the filter and regard the log of the mean as the "exact" log-likelihood value. To make results comparable for $N = 1,250$ and $N = 10,000$ we compute $\ln p^{(K)}(Y_{1:T}, X_{1:N,1:T}|\theta_*)/(N+2)$ where $n_y = 2$ is the number of aggregate series. The log likelihood values shown in the figure are in deviations of the "exact" values, which are -0.3467 for $N = 1,250$ and -0.3418 for $N = 10,000$. The KF log likelihood is represented by a vertical line and the distribution of PF log likelihood approximations by a density. Deviations of the KF value from zero can be interpreted as approximation errors of the linear filter. A comparison between the left and the right panel indicates that the approximation error vanishes as N increases, as Theorem 1 predicts.[9]

3.3 MDD Approximation Error

We propose to use the MDD approximation $p_*^{(K)}(Y_{1:T}, X_{1:N,1:T}|\lambda)$ in (17) to select the number of lags $p \leq \bar{p}$, the sieve dimension $K \in \mathcal{K}_N$, and the prior hyperparameters $\lambda \in \Lambda$. We provide a bound for the discrepancy between the $p_*^{(K)}(\cdot)$ and the $p^{(K)}(\cdot)$ MDD based on the nonlinear measurement equation, defined in (15). Because the MDD-based lag length selection under

[9]The dispersion of PF log likelihood values is larger for $N = 1,250$ than for $N = 10,000$ because θ_* is the posterior mean based on the $N = 10,000$ sample and the precision of the PF is sensitive to the fit of the model/parameter values; e.g., see Herbst and Schorfheide (2015).

the assumption that the upper bound $\bar{p}$ does not depend on the sample size is standard, we focus on the (K, λ) selection in the remainder of this section. We assume that the set $\mathcal{K}_N$ from which K is selected does not grow too fast. Let $\bar{K}_N = \max\{K \mid K \in \mathcal{K}_N\}$ and $\bar{K}_N = o(N^{1/3})$.

Theorem 2 *Suppose Assumption 1 is satisfied and* $T\bar{K}_N/N = o(1)$. *Then,*

$$\max_{(K,\lambda)\in\mathcal{K}_N\times\Lambda} \left| \ln p^{(K)}(Y_{1:T}, X_{1:N,1:T}|\lambda) - \ln p_*^{(K)}(Y_{1:T}, X_{1:N,1:T}|\lambda) \right| \lesssim \left(\sqrt{\frac{\overline{K}_N^3}{N}} + \frac{T}{2N}\overline{K}_N \right).$$

A proof of the theorem is provided in the Online Appendix. The rate $\sqrt{\overline{K}_N^3/N}$ reflects the discrepancy between $\ln p^{(K)}(\cdot)$ and $\ln p_{KF}^{(K)}(\cdot)$. The second rate, $T\overline{K}_N/N$, comes from replacing $\ln p_{KF}^{(K)}(\cdot)$ by $\ln p_*^{(K)}(\cdot)$.[10]

3.4 Model Selection

Let $(\hat{K}, \hat{\lambda})$ and (K_*, λ_*) be the maximizers of $\ln p^{(K)}(\cdot|\lambda)$ and $\ln p_*^{(K)}(\cdot|\lambda)$, respectively. We can deduce the following corollary from Theorem 2:

Corollary 1 *Suppose Assumption 1 is satisfied and* $T\bar{K}_N/N = o(1)$. *Then*

$$\ln p^{(\widehat{K})}(Y_{1:T}, X_{1:N,1:T}|\hat{\lambda}) - \ln p^{(K_*)}(Y_{1:T}, X_{1:N,1:T}|\lambda_*) \lesssim 2\left(\sqrt{\frac{\overline{K}_N^3}{N}} + \frac{T}{2N}\overline{K}_N \right)$$

$$\ln p_*^{(K_*)}(Y_{1:T}, X_{1:N,1:T}|\lambda_*) - \ln p_*^{(\widehat{K})}(Y_{1:T}, X_{1:N,1:T}|\hat{\lambda}) \lesssim 2\left(\sqrt{\frac{\overline{K}_N^3}{N}} + \frac{T}{2N}\overline{K}_N \right).$$

The corollary establishes a weak form of equivalence between model selection based on the exact MDD $\ln p^{(K)}(\cdot|\lambda)$ and the approximation $\ln p_*^{(K)}(\cdot|\lambda)$: the argmax of $\ln p_*^{(K)}(\cdot|\lambda)$,

[10] In the empirical application in Section 6 N varies with t. The minimum value is 12,411. The time series length is $T = 110$. Most of the calculations are based on $K = 10$, which leads to $K^3/N = 0.08$, $TK/(2N) = 0.06$. For MDD comparisons we consider $\bar{K} = 22$ which leads to $\bar{K}^3/N = 0.86$ and $T\bar{K}/(2N) = 0.10$. Because the law of motion for α_t induces a prior for the spline coefficient, we allowed $\bar{K}$ to be slightly larger than the rate restriction suggests.

(K_*, λ_*), almost maximizes the exact MDD, and vice versa.[11] We show in the Online Appendix that, under a version of the drifting prior in (18) below and a log likelihood function that is quadratic in θ, the hyperparameter estimates based on the exact and approximate MDD are within ϵ distance for each $K \in \mathcal{K}_N$ as $(N,T) \longrightarrow \infty$. A strong equivalence statement regarding the K selection requires assumptions on a DGP $\mathbb{P}^o_{N,T}(\cdot)$ and the notion of a pseudo-optimal dimension $K_{oN} \in \mathcal{K}_N$. A proof is left for future research.

To understand the trade-offs in choosing λ and K, we now decompose $\ln p_*^{(K)}(\cdot)$ into goodness-of-fit and penalty terms. For ease of exposition we assume that the log likelihood function and prior density are both quadratic in θ. In our model this assumption is exactly satisfied for the regression coefficients Φ, and holds asymptotically for large T for the parameters in Σ. This assumption avoids having to keep track of standard remainder terms. To keep the hyperparameter selection asymptotically non-trivial we balance the informational contents of the prior distribution and the likelihood function by rescaling the hyperparameter. We assume that

$$\theta|(K,\omega) \sim \mathcal{N}\left(\underline{\theta}, \frac{\omega}{T}\underline{\mathcal{V}}\right). \tag{18}$$

The local hyperparameter $\omega \in \Omega = [0,\infty)$ is related to the precision hyperparameter λ through $\lambda = T/\omega$.

Rather than conducting a Laplace approximation around the MLE $\hat{\theta}_*$, we first compute the posterior mean/mode and variance to capture the effect of the prior distribution in the MDD formula. In fact, under the assumption of a log-quadratic likelihood and prior density the following formulas are exact. The posterior of θ is Gaussian with mean and variance (scaled by T)

$$\bar{\theta}_* = \bar{\mathcal{V}}_*\left(\hat{\mathcal{V}}_*^{-1}\hat{\theta}_* + \frac{1}{\omega}\underline{\mathcal{V}}^{-1}\underline{\theta}\right), \quad \bar{\mathcal{V}}_* = \left(\hat{\mathcal{V}}_*^{-1} + \frac{1}{\omega}\underline{\mathcal{V}}^{-1}\right)^{-1},$$

where $\hat{\mathcal{V}}_*$ is the negative inverse Hessian associated with $\ln p_G^{(K)}(Y_t, \alpha_t = \hat{\alpha}_t|\cdot,\theta)$, evaluated

[11] In the context of selecting the dimension of a linear model Shao (1997) refers to the weak notion as loss efficiency. He compares the behavior of various model selection procedures under finite-dimensional and infinite-dimensional DGPs.

at the MLE $\hat{\theta}_*$. This leads to

$$\begin{aligned}
&\frac{1}{NT}\ln p_*^{(K)}(Y_{1:T}, X_{1:N,1:T}|\omega) \qquad (19)\\
&= \frac{1}{T}\sum_{t=1}^{T}\mathcal{L}^{(K)}(\hat{\alpha}_t|X_{1:N,t}) - \frac{K}{2N}\ln(N)\\
&\quad -\frac{1}{2N}\left(\hat{\theta}_*'\hat{\mathcal{V}}_*^{-1}\hat{\theta}_* + \frac{1}{\omega}\underline{\theta}'\underline{\mathcal{V}}^{-1}\underline{\theta} - \bar{\theta}_*'\bar{\mathcal{V}}_*^{-1}\bar{\theta}_*\right) - \frac{1}{2NT}\ln\left|\omega\underline{\mathcal{V}}\hat{\mathcal{V}}_*^{-1} + I_d\right|\\
&\quad +\frac{1}{NT}\sum_{t=1}^{T}\ln p_G^{(K)}(Y_t, \alpha_t = \hat{\alpha}_t|Y_{t-1}, \alpha_{t-1} = \hat{\alpha}_{t-1}, \hat{\theta}_*) + \frac{K}{2N}\ln(2\pi) + \frac{1}{2}\frac{1}{NT}\sum_{t=1}^{T}\ln|\hat{V}_t|\\
&= (A_1 - A_2) + (-A_3 - A_4) + (A_5 + A_6 + A_7),
\end{aligned}$$

say. Note that the dimension of θ and is $O(K^2)$.

The $O_p(1)$ term $A_1 = \frac{1}{T}\sum_{t=1}^{T}\mathcal{L}^{(K)}(\hat{\alpha}_t|X_{1:N,t})$ captures the goodness-of-fit of the cross-sectional density estimates and is non-decreasing in K. It is penalized by the Schwarz criterion (BIC) penalty $A_2 = K/(2N)\ln N$, meaning that every additional sieve term is penalized by $1/(2N)\ln N$.[12] The selection of ω conditional on K depends on A_3 and A_4.[13] For finite values of ω the term A_3 is greater than zero and can be interpreted as an in-sample goodness-of-fit adjustment to the maximized VAR likelihood in A_5, due to the use of a prior. As $\omega \longrightarrow \infty$, meaning the prior becomes "flat," the difference between posterior mean and MLE vanishes and A_3 converges to zero. A_4 penalizes the choice of the hyperparameter ω. The penalty diverges as $\omega \longrightarrow \infty$. A_4 is minimized for $\omega = 0$ when θ is forced to be equal to the prior mean $\underline{\theta}$. The remaining terms, A_5, A_6, and A_7, are of the smaller order $O_p(K/N)$.

To summarize, dimensionality and hyperparameter selection based on the approximate MDD $p_*^{(K)}(Y_{1:T}, X_{1:N,1:T}|\lambda)$ trades off goodness-of-fit and model complexity captured by the penalty terms A_2 and A_4. By maximizing the approximate MDD the user can speed up computations considerably, because it simplifies a high-dimensional integration problem. In our application we have an analytical formula for the $p_*^{(K)}(\cdot|\lambda)$ MDD. The $p^{(K)}(\cdot|\lambda)$ MDD value attained by this computational strategy is almost as high, in the sense of Corollary 1 as that obtained by maximizing the exact MDD $p_*^{(K)}(Y_{1:T}, X_{1:N,1:T}|\lambda)$.

[12] In fact, Kooperberg and Stone (1992) recommend for log-spline density models K selection by BIC.

[13] A_4 is $O_p(K^2/NT)$. A_3 is also of order $O_p(K^2/NT)$ if two conditions hold. First, $K^2/T = o(1)$ which delivers a $\sqrt{T}$ consistent MLE. Second, the "true" θ parameter is sampled from the drifting prior (18), which implies that the MLE behaves asymptotically as $\hat{\theta}_* = \underline{\theta} + \hat{\vartheta}_*/\sqrt{T}$, where $\hat{\vartheta}_* = O_p(1)$.

4 Implementation Details

Basis Functions. A convenient basis for the log density is a spline which is a piecewise polynomial function with knots x_k, $k = 1, \ldots, K-1$; see 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.[14] In our two applications we restrict x to the interval $[0, \bar{x}]$. For the estimation based on simulated data in Section 5 we also use a cubic function for the last segment of the spline:

$$\begin{aligned} \zeta_1(x) &= x \\ \zeta_k(x) &= \left[\max\{x - x_{k-1},\, 0\}\right]^3, \quad k = 2, \ldots, K. \end{aligned} \tag{20}$$

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 (8). For the empirical application in Section 6 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

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

For the simulated data (20) 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. As a diagnostic, we recommend to visually examine the fit of the estimated cross-sectional densities, as in Figures 4 and 8. In addition, the MDD $p_*^{(K)}(Y_{1:T}, X_{1:T}|\lambda)$ can also be used to select among basis functions, which is what we did to choose between (20) and (21) in the simulation experiment of Section 5 and the empirical analysis in Section 6.

Sieve Coefficients. The first step in the estimation of the functional model is the computation of the MLEs $\hat{\alpha}_t$ for $t = 1, \ldots, T$. In this step we also generate the sequence of covariance

[14] In general, the methods developed in this paper are not restricted to cubic splines. Other basis functions such as polynomials (including trigonometric, Hermite, Laguerre) or wavelets could be used.

matrices $\hat{V}_t$, and the MDD components $p_{KF}^{(K)}(X_t|\alpha_t)$, $t = 1, \ldots, T$. The Online Appendix discusses three adjustments that may be required to implement the empirical analysis: (i) If the data are top-coded, the likelihood needs to be adjusted. We assume that the censoring point as well as the number of censored observations are observed. $p_t(x)$ is assumed to describe the uncensored density. (ii) 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 of the cross-sectional distributions are varying over time. If there are perfect collinearities in $\hat{\alpha}_t$, we remove them. Even if there are none, we normalize the $\hat{\alpha}_t$s. (iii) The cross-sectional observations may exhibit seasonality that leads to seasonal variation in the $\hat{\alpha}_t$s. In our application in Section 6 the aggregate data are seasonally adjusted. Thus, we remove deterministic seasonalities from the $\hat{\alpha}_t$s before combining them with the aggregate variables in a VAR.

Priors and Posteriors. The second step of the analysis consists of the estimation of the state-space model, based on the sequence $(\hat{\alpha}_t, \hat{V}_t)$, $t = 1, \ldots, T$, generated in the first step. In our implementation we condition on estimates of the deterministic components, $\hat{Y}_*$ and $\hat{\alpha}_*$. Recall that the measurement equation is given by (13) and that we assume that $Y_t - Y_*$ is observed without error. Define the $n_w \times 1$ vector $W_t = [(Y_t - Y_*)', (\alpha_t - \alpha_*)']'$ and write the state transition (6) more compactly as

$$W_t = \sum_{j=1}^{p} \Phi^{(j)} W_{t-j} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma). \tag{22}$$

We now also allow for multiple lags. Define $\Phi = [\Phi^{(1)}, \ldots, \Phi^{(p)}]'$ and note that (Φ, Σ) do not enter the measurement equation.

Following Carter and Kohn (1994), posterior inference for the state-space model is based on a Gibbs sampler that iterates over the following two conditional posterior distributions:

$$\alpha_{1:T}|(Y_{1:T}, X_{1:T}, \Phi, \Sigma), \quad (\Phi, \Sigma)|(Y_{1:T}, X_{1:T}, \alpha_{1:T}).$$

Because we are using the linearized measurement equation (A.67) the latent states $\alpha_{1:T}$ can be sampled using a Kalman filter / simulation smoother. We adjust the standard algorithm to account for the aggregate variables Y_t being directly observed and multiple lags, i.e., $p > 1$. Details are provided in the Online Appendix. Conditional on $\alpha_{1:T}$, the derivation of the posterior distribution of (Φ, Σ) follows the Bayesian analysis of a VAR. Because the vector α_t can be potentially large, we use the VAR parameterization and prior proposed

by Chan (2022). His specification is suitable for high-dimensional settings because it leads to equation-by-equation estimation while allowing for some asymmetry of the prior across equation.

Starting point for the specification of the prior is the assumption that the elements of Φ are an Inverse-Gamma scale mixtures of Normals with mean zero. The overall prior precision is controlled by the hyperparameter λ_1. The hyperparameter λ_2 controls the relative precision of coefficients that capture the effect of lagged α_ts on current Y_ts. As $\lambda_2 \longrightarrow \infty$, the cross-sectional density does not Granger-cause the aggregate variables and the system becomes block-triangular. Because Chan's (2022) prior is conjugate, the posterior distribution has a convenient Inverse-Gamma Normal form and draws can be generated by direct sampling. Moreover, the MDD in (17) can be evaluated analytically. Further details on the prior distribution, the derivation of the posterior, and the MDD formula are provided in the Online Appendix.

Recovering Cross-Sectional Densities. Based on the estimated state-transition equation (22) we can generate forecasts and impulse response functions for the compressed sieve coefficients. However, the dynamics of these coefficients in itself are not particularly interesting. Thus, we have to convert them back into densities using

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

Further Shortcuts. We generally recommend to estimate the functional model using the approximate state-space representation that accounts for the estimation of the α_ts with the measurement equation (13). However, as the calculations underlying Theorem 2 indicate, the large (N, K, T) approximation error calculations would also justify the estimation of the VAR parameters θ by simply equating the latent α_t with $\hat{\alpha}_t$, i.e., setting the measurement error in (13) to zero, and estimating the VAR parameters based on

$$\prod_{t=1}^{T} p_G^{(K)}(Y_t, \alpha_t = \hat{\alpha}_t | Y_{t-1}, \alpha_{t-1} = \hat{\alpha}_{t-1}, \theta), \tag{23}$$

in the notation of (17). In the notation of (22) this amounts to replacing W_t by $\hat{W}_t = [(Y_t - Y_*)', (\hat{\alpha}_t - \alpha_*)']'$. To facilitation this short-cut, we provide Julia code that (i) estimates $\hat{\alpha}_t$ based on X_t for $t = 1, \ldots, T$; and (ii) converts vectors α into cross-sectional densities $p^{(K)}(x|\alpha)$. Empirical researchers can then easily combine this code with their preferred VAR

estimation routine.

Mixed-Frequency Considerations. In some applications, the cross-sectional observations x_{it} might be observed at a lower frequency, say, annually instead of quarterly, than the aggregate time series. Suppose that x_{it} corresponds to household earnings and the length of the period t is one quarter. It is important to distinguish two cases. Suppose at the end of every year the household is asked: "how much did you earn in the fourth quarter of the current year?" In this case, we observe draws x_{it} from $p_t(x)$ in periods $t = 4, 8, 12, \ldots$ and we can write (13) as

$$\hat{a}_t = \begin{cases} \alpha_t + N^{-1/2}\eta_t & \text{for } t = 4, 8, 12, \ldots \\ \emptyset & \text{in other periods} \end{cases}$$

The estimation can be easily handled within the current state-space setup using the approach in Schorfheide and Song (2015). Alternatively, suppose that at the end of every year the household is asked: "how much did you earn in the current year?" Thus, the survey generates observations of the form $\bar{x}_{it} = x_{it} + x_{it-1} + x_{it-2} + x_{it-3}$ where now i refers to the same household in the four periods. Our model does not have enough structure to recover the distribution of $\bar{x}_{it}$ from the sequence $p_t(x), p_{t-1}(x), p_{t-2}(x), p_{t-3}(x)$. Additional assumptions would be required and we leave a careful treatment of this case for future research.

5 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 estimate it based on artificial data simulated from a Krusell and Smith (1998) economy. We show that in the perfect aggregation environment of the KS model, estimated responses of aggregate variables to an aggregate shock from the functional model are identical to those from a VAR in aggregate variables only. The MDD hyperparameter selection shrinks the estimates of the parameter block that captures the effect of the lagged distribution onto current aggregates to zero. If the DGP is modified to break the perfect aggregation structure, some discrepancies between the two IRF estimates emerge, but considerable overlap between the credible bands remain. We also document that the estimated functional model can reproduce the distributional responses of the DGP.

5.1 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

$$x_{jt+1} + c_{jt} = (1-\tau) W_t \epsilon_{jt} + b W_t (1 - \epsilon_{jt}) + x_{jt}(1 + R_t), \quad x_{jt+1} \geq \underline{x}. \tag{24}$$

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. The government is assumed to balance its budget 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. Households face the borrowing constraint $x_{jt+1} \geq \underline{x}$. Finally, a representative firm produces output Y_t according to

$$Y_t = \exp\{z_t\} K_t^{\alpha} L^{1-\alpha}, \quad z_t = \rho_z z_{t-1} + \sigma_z \omega_t, \quad \omega_t \sim \mathcal{N}(0,1), \tag{25}$$

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. In equilibrium $K_t = \int x_{jt} dj$.

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. It is convenient to use $\mu_{t,\epsilon}$ to denote the conditional distribution of assets given the employment status ϵ_t. To solve the model, we use the approach by Winberry (2018). Details are provided in the Online Appendix. The solution represents $\mu_{t,\epsilon}$ as a mixture of a discrete and continuous part:

$$q_{t,\epsilon}(x) = \widehat{m}_{t,\epsilon} \Delta_{\underline{x}}(x) + (1 - \widehat{m}_{t,\epsilon}) p_{t,\epsilon}(x). \tag{26}$$

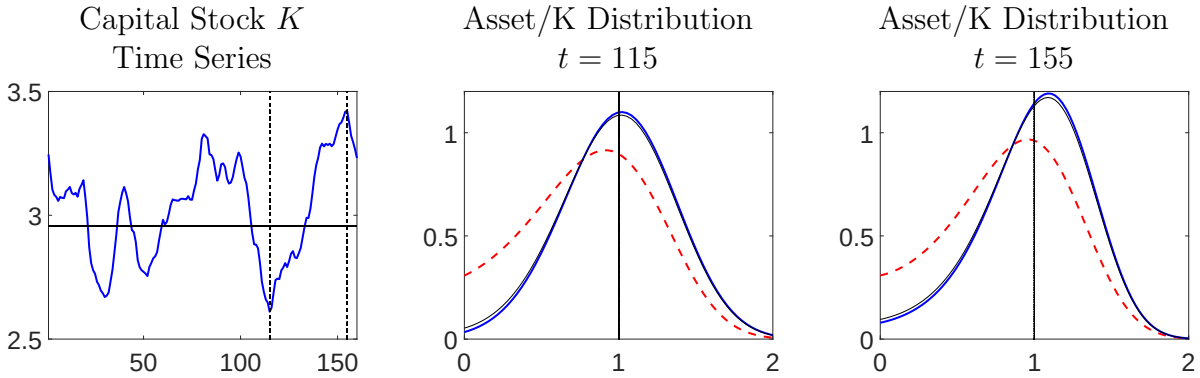
Here, $\widehat{m}_{t,\epsilon}$ is the mass of individuals for whom the borrowing constraint $\underline{x}$ is binding and $\Delta_{\underline{x}}(x)$ is the Dirac function.[15] The continuous part of the asset distribution is parameterized

[15] The Dirac function has the property that $\Delta_{\underline{x}}(x) = 0$ for $x \neq \underline{x}$ and $\int \Delta_{\underline{x}}(x) dx = 1$.

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

Figure 2: Features of Simulated Data (Baseline)



Notes: Aggregate capital attains its minimum in period $t = 115$ and its maximum in period $t = 155$. Each density is normalized so that it integrates to one. Distribution of unemployed is red dashed, distribution of employed is blue solid, and aggregated distribution is solid black.

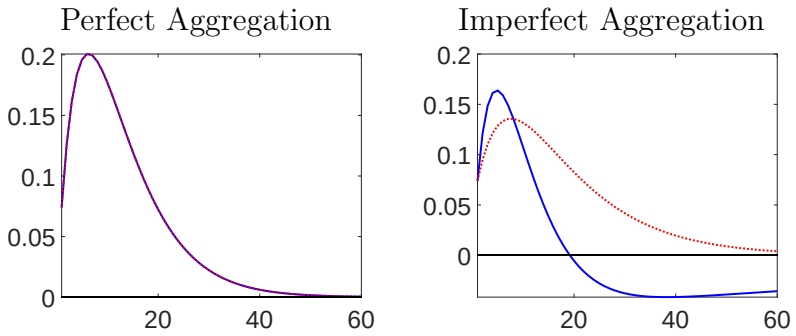
in terms of the time-varying centralized moments $m_{t,\epsilon,k}$ and represented as

$$p_{t,\epsilon}(x) = \exp\left\{\gamma_{t,\epsilon,0} + \gamma_{t,\epsilon,1}(x - m_{t,\epsilon,1}) + \sum_{k=2}^{3} \gamma_{t,\epsilon,k}\left[(x - m_{t,\epsilon,1})^k - m_{t,\epsilon,k}\right]\right\}. \tag{27}$$

The coefficients $\gamma_{t,\epsilon,k}$s are determined such that the density integrates to one and is consistent with the moments $m_{t,\epsilon,k}$.

We calibrate the KS economy to loosely match features of annual U.S. data. The parameterization is summarized in Table 1. The Winberry (2018) solution generates a vector autoregressive law of motion for technology z_t, capital K_t, the mass $\widehat{m}_{t,\epsilon}$ of employed and unemployed households at the borrowing constraint, and the moments $m_{t,\epsilon,k}$. We simulate this law of motion for 160 periods and draw *iid* cross-sectional observations from (26), where $p_{t,\epsilon}(x)$ is defined in (27). The left panel of Figure 2 depicts the time series of the aggregate capital stock from $t = 1$ to $t = 160$. The capital stock reaches a trough in period $t = 115$

Figure 3: Capital Stock Response to TFP Shock



Notes: Capital responses to a TFP shock: heterogeneous agent model solution (blue), population approximation in terms of aggregate variables (red). Left panel: baseline KS model with perfect aggregation. Right panel: modified KS solution.

and peaks in period $t = 155$.

We standardize household-level asset holdings by the aggregate capital stock and focus on the cross-sectional distributions of a/K. The densities for the unemployed, the employed households, and the two types combined are plotted in the center panel and the right panel of Figure 2. Because the unemployment rate is only 7%, the overall asset distribution is very similar to the distribution of employed households. We condition on $x > \underline{x} = 0$ and plot $p_{t,\epsilon}(x)$ given in (27). The 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 mode of the asset/capital distribution shifts to the right. Due to the normalization, the mean is equal to one in both periods.

Our version of the KS model features perfect aggregation, meaning that the VAR(1) law of motion for $\big(z_t, K_t, \{\widehat{m}_{t,\epsilon}\}, \{m_{t,\epsilon,k}\}\big)$ has a lower triangular structure in which (z_t, K_t) does not depend on $\big(\{\widehat{m}_{t-1,\epsilon}\}, \{m_{t-1,\epsilon,k}\}\big)$ and thus the cross-sectional distribution does not Granger-cause the aggregate variables. In the left panel of Figure 3 we compare the response of capital in the baseline KS model to a TFP shock with responses generated from a VAR(1) in the aggregate variables (z_t, K_t), henceforth aggregate VAR(1).[16] Under the baseline KS model the two IRFs are identical, meaning that aggregate capital perfectly summarizes the feature of the cross-sectional asset distribution that is relevant for aggregate dynamics. In the right panel of Figure 3 we consider a case in which aggregation is imperfect. Rather than specifying an alternative heterogeneous agent model, we simply modify the baseline KS

[16]Rather than explicitly solving a representative agent model, we simply derive a VAR(1) law of motion for (z_t, K_t) from the VAR solution of the KS model using population OLS.

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=160.

solution by letting the lagged second moment of the asset distribution affect capital.[17] Under this modification the response of the (modified) heterogeneous agent model now differs from the response computed from the aggregate VAR(1).

5.2 Functional Model Estimation

We now estimate the functional model based on the simulated time series for (z_t, K_t) and cross-sectional data.[18] Unless otherwise noted, results are generated from the baseline KS model that features perfect aggregation. The econometrician observes z_t, K_t, and iid draws from the cross-sectional distribution $x_{it} \sim p_{t,\epsilon=1}(x)$, where $p_{t,\epsilon=1}(x)$ was defined in (27). The empirical model differs from the data generating process (DGP) in the following important dimension: the representation of the densities in the DGP is based on the moments $m_{t,\epsilon,k}$ whereas the empirical model is based on a VAR that includes the spline coefficient vector α_t. 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. We document below that the estimated functional model is successful in capturing the joint dynamics of the cross-sectional asset distribution and the technology shock for data generated from the KS model.

Model Selection. The first step in the analysis is the determination of the approximation order K, the lag length p, and the hyperparameter λ based on the MDD approximation $p_*^{(K)}(Y_{1:T}, X_{1:T}|\lambda)$ in (17). We use the spline basis in (20) and place the knots at pre-determined percentiles of the empirical distribution of the simulated x_{it}s, pooled across

[17] Kaplan and Violante (2022) survey a variety of structural mechanisms that break the perfect aggregation in the KS model.

[18] Due to precautionary savings behavior the mass of employed individuals with zero assets is essentially zero and does not vary over time. Hence, we exclude it from the list of aggregate variables.

Table 3: Model Selection

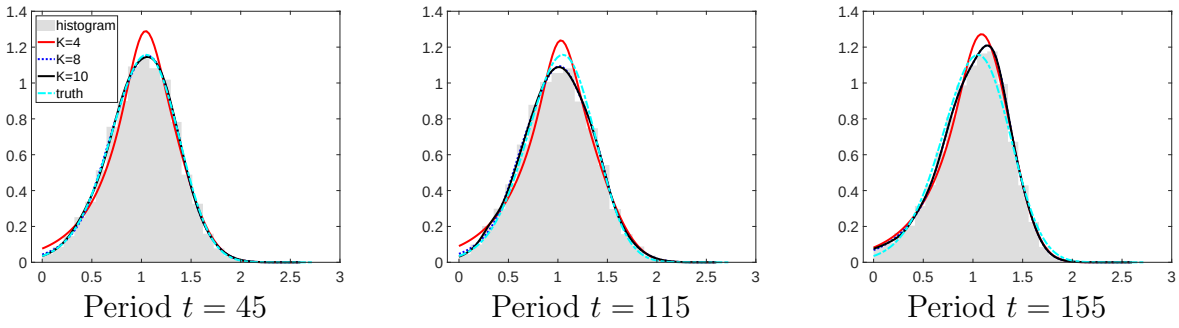
	T=90				T=150			
N	$\hat{p}$	$\hat{K}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{p}$	$\hat{K}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
Baseline Specification: Perfect Aggregation								
1,000	1	6	7.39	148	1	6	2.72	5E8
5,000	1	8	7.39	5E8	1	8	7.39	5E8
10,000	1	8	7.39	5E8	1	8	7.39	5E8
Alternative Specification: Imperfect Aggregation								
1,000	1	6	20.08	7.39	2	6	1.00	1096
5,000	1	8	20.08	2.27	1	8	20.08	2.72
10,000	1	8	20.08	2.72	1	8	7.39	2.72

a large simulation with $N_{sim} = 10,000$ and $T_{sim} = 160$. 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 consider two choices for T, namely $T = 90$ and $T = 150$ and three choices of N: 1,000, 5,000, and 10,000. Conditional on N we evaluate the approximate marginal data density $p_*^{(K)}(Y_{1:T}, X_{1:T}|\lambda, p)$ in (17) over a grid of lag lengths p, sieve dimensions K, and λ values. For λ_1 and λ_2, respectively, we consider 31 equally-spaced values of $\ln \lambda_j$ between -10 and 20.

In Table 3 we report $\hat{p}$, $\hat{K}$, $\hat{\lambda}_1$, and $\hat{\lambda}_2$ for data from the baseline specification (perfect aggregation) and the modified specification (imperfect aggregation). In eleven out of twelve cases the selected lag length is $\hat{p} = 1$, which is consistent with the first-order autoregressive structure of the DGP. For $N = 1,000$ the estimated spline order is $\hat{K} = 6$, whereas it is $\hat{K} = 8$ for $N = 5,000$ and $N = 10,000$. Thus, as the cross-sectional dimension of the sample increases, the goodness-of-fit gain in the MDD criterion through additional sieve terms outweighs the 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.

A comparison between the baseline specification and the modified specification indicates that $\hat{\lambda}_2$, the estimated degree of relative shrinkage (to zero) for the VAR coefficients that control the effect of a_{t-1} on Y_t, $\Phi_{y\alpha}$ in (6), is much higher under perfect aggregation than under imperfect aggregation. This is consistent with the perfect aggregation specification being block triangular, i.e., $\Phi_{y\alpha} = 0$. The hyperparameter $\hat{\lambda}_1$ is in eleven out of twelve cases

Figure 4: Cross-sectional Fit ($N = 10,000, T = 150, K = 8$ 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.

weakly smaller for $T = 150$ than for $T = 90$. The larger the time dimension, the smaller the need to add prior information to reduce the posterior variance.

Cross-sectional Fit. In Figure 4 we compare the fit of the estimated cross-sectional densities (various K) for $N = 10,000$

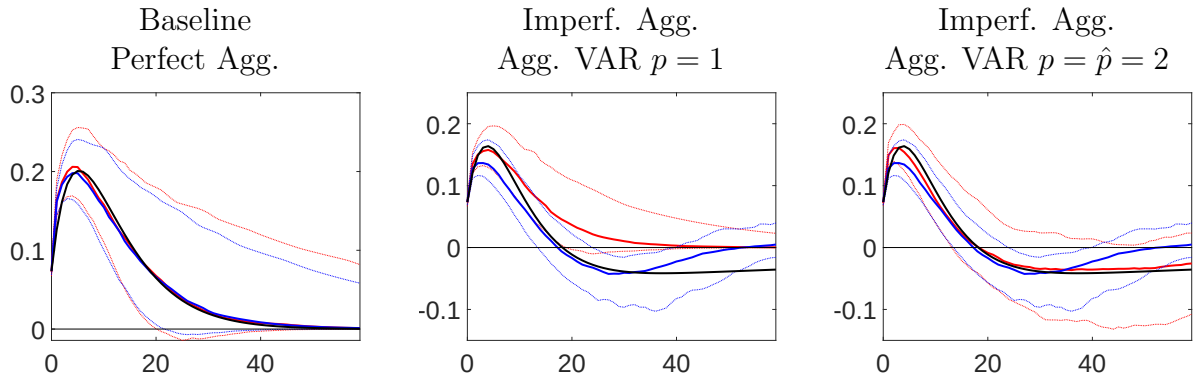
$$p^{(K)}(x|\hat{\alpha}_t) = \exp\left\{\zeta'(x)\hat{\alpha}_t - \ln \int \exp\left\{\zeta'(\tilde{x})\hat{\alpha}_t\right\} d\tilde{x}\right\} \tag{28}$$

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 2 that the capital stock reaches a trough in period $t = 115$ and peaks in period $t = 155$. The estimated densities for $K = 8$ and $K = 10$ are virtually indistinguishable and provide a smooth approximation of the histograms constructed from the observations. Because there is no improvement in fit, the MDD criterion selects $K = 8$. The $K = 4$ density shows noticeable differences from the histograms in all periods. In period $t = 45$ the density estimate for the selected $\hat{K} = 8$ almost perfectly matches the true cross-sectional density, whereas there are some small discrepancies in the modal area for the other two periods. Overall, the density estimation works well in this simulation.

Aggregate IRFs. We proceed by estimating the state-space representation of the functional model based on the ($N = 10,000$, $T = 150$) sample for the model specification reported in (3).[19] In addition to the functional model, we also estimate a VAR in the aggregate variables (z_t, K_t) only, using the same prior as for the functional model. Figure 5 shows the impulse response of capital to a 3-standard deviation technology shock. We compute "true"

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

Figure 5: Capital Response from Functional Model and Aggregate VAR



Notes: 10th, 50th, and 90th percentiles of the posterior distribution of IRFs. Blue lines correspond to functional VAR IRFs and red lines to IRFs from the aggregate VAR. Left panel: estimation based on data from baseline specification. Center and right panels: estimation based on data from the alternative specification with imperfect aggregation.

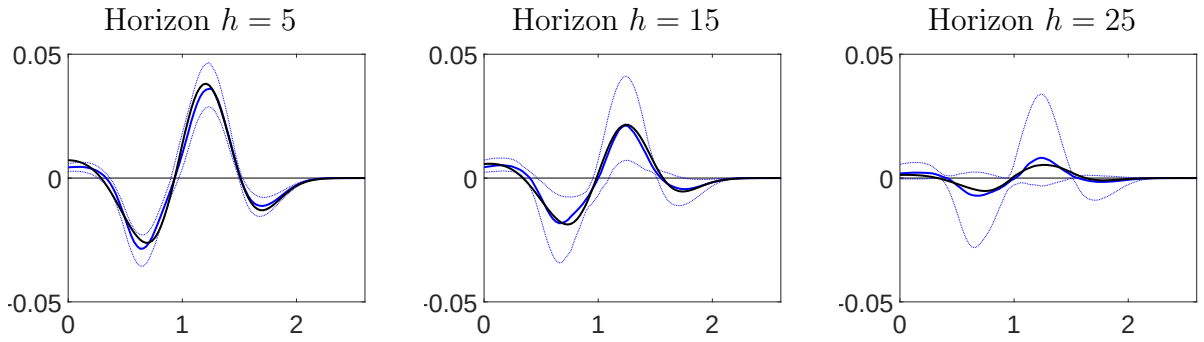
responses from the solution of the KS economy and estimated responses (10th, 50th, and 90th percentiles of the posterior distribution) from the functional model and the aggregate VAR. Because z_t is exogenous in the KS economy, its innovation can be easily identified using a Cholesky factorization of Σ.

The left panel of Figure 5 is based on data from the baseline perfect aggregation specification. Here the IRFs from the functional VAR(1) and the aggregate VAR(1) in (z_t, K_t) line up almost perfectly. This is expected in view of the perfect match of the population responses in the left panel in Figure 3. The cross-sectional distribution does not Granger-cause the aggregate variables, and, as a consequence, no additional lags are required in the aggregate VAR to recover the capital stock IRF.

The IRFs in the center and right panels of Figure 5 are obtained from estimates based on imperfect aggregation data. In the center panel we are forcing $p = 1$, which is the lag order of the functional VAR, and observe a divergence of the IRFs from functional and aggregate model. The divergence is in line with the discrepancies in the right panel of Figure 3. If the cross-sectional distribution Granger-causes the aggregate variables, then (z_t, K_t) no longer evolves as a VAR(1). Instead, the bivariate process takes the form of a VARMA process, which, strictly speaking, leads to a VAR(∞).[20] In finite samples, the infinite-order process has to be approximated by a finite-dimensional VAR. These considerations suggest that the lag length in the aggregate VAR needs to be increased to recover the capital stock IRF

[20]Here the VARMA process is restricted and implies that z_t evolves according to an exogenous AR(1) process. Thus, the identification of the technology shock as z_t innovation remains unchanged.

Figure 6: Density Differential Response



Notes: Impulse responses are generated from the estimated VAR. Results are based on $T = 150$, $N = 10,000$, $K = 8$, and $\hat{\lambda}(K)$ (see Table 3). "True" response from the KS model appears in black, responses from the estimated functional state-space model are in blue (pointwise median is solid, 5th and 95th percentiles are dotted)

of the functional VAR. In the right panel we generate the IRFs from the aggregate model based on the number of lags $\hat{p} = 2$, estimated by the MDD. In this case functional and aggregate responses line up well. An additional lag of the aggregate variables substitutes for the missing cross-sectional information.[21]

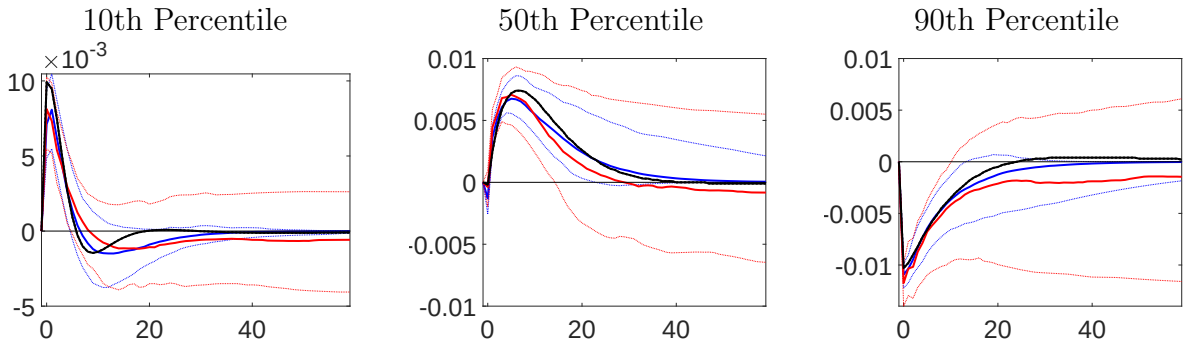
More generally, we recommend reporting IRFs of aggregate variables from the functional model, an aggregate VAR with the same number of lags as the functional model, and possibly an aggregate VAR with data-driven lag choice. To assess heterogeneous agent models, one can compare the discrepancy between functional and aggregate IRFs based on model generated and actual data. If the two discrepancies differ, then the aggregation implications of the heterogeneous agent model are inconsistent with the data.

Density IRFs. For each posterior draw (Φ, Σ) we generate an impulse response function for α_t by iterating (22) forward and then reconstruct the density $p^K(x|\alpha_{t+h})$. Figure 6 shows the difference between the shocked and the steady state densities at horizons $h = 5$, $h = 15$, and $h = 25$.[22] We refer to this difference as the response of the density. The true responses from the KS model are plotted in black. At all horizons the posterior median response aligns well with the true response, indicating that the estimated functional model can reproduce the cross-sectional dynamics of the KS model. According to the responses probability mass

[21] In the Online Appendix, we report results from a second simulation experiment, based on a stylized HANK model due to Auclert and Rognlie (2020). This model can generate imperfect aggregation based on countercyclical income risk. However, empirically the effect is weak. Based on data generated from this model, the estimated IRFs from the functional model and the aggregate VAR are very similar, even for identical lag lengths.

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

Figure 7: Percentile Response from Functional VAR and Percentile VAR



Notes: 10th, 50th, and 90th percentiles of the posterior distribution of IRFs. Blue lines correspond to the functional VAR IRFs and red lines to IRFs from the alternative VAR that includes percentiles of the asset/K distribution.

shifts from $a/K < 1$ to $a/K > 1$ while at the same time the mass for $a/K > 1.5$ decreases, keeping the mean of the cross-sectional distribution at one.

Percentile IRFs. Once we have obtained the IRFs for the cross-sectional densities, we can compute IRFs for statistics derived from these densities. Here we consider percentiles of the cross-sectional distribution. We compare the percentile IRFs derived from the functional model to those obtained by estimating a VAR that combines the aggregate data with sample percentiles (10th, 20th, 50th, 80th, and 90th) computed from the cross-sectional observations. Results are plotted in Figure 7. Two observations stand out. First, the estimated percentile responses from the functional model line up well with the true responses. The 10th percentile and the median of the asset/K distribution respond positively, whereas the 90th percentile falls. Thus, the asset share of low-wealth households increases. In turn, the asset share of high-wealth households has to drop. Second, the bands from the percentile VAR are much wider than the bands from the functional VAR.

We provide some intuition for the imprecision of the percentile VAR using a local projection argument and making a number of simplifying assumptions. A formal analysis in the state-space / VAR setting is beyond the scope of this paper. The functional model implies that the information from the repeated cross-sections $X_{1:N,t}$ can be summarized in the K-dimensional vectors $\hat{\alpha}_t$, which are noisy measures of the α_t vectors. The latter evolve jointly with the aggregate variables Y_t according to a VAR. Suppose the lag order $p = 1$ and the object of interest is the contemporaneous response to the technology innovation $\epsilon_{z,t}$. Then, write

$$\alpha_t = \Psi \epsilon_{z,t} + \Phi_{\alpha y} Y_{t-1} + \Phi_{\alpha\alpha} \alpha_{t-1} + \tilde{u}_{\alpha,t}, \tag{29}$$

where $\tilde{u}_{\alpha,t}$ is a vector of forecaster errors generated by shocks other than $\epsilon_{z,t}$. We proceed with the following simplifying assumptions: (i) $\hat{\alpha}_t = \alpha_t$ (no uncertainty in regard to cross-sectional densities), (ii) the technology shock innovations $\epsilon_{z,t}$ are observed (they can be recovered from the observed TFP process in the KS model and in our application), (iii) the quantiles $q_t = M'\alpha_t$ are a linear function of the α_ts, (iv) the dimension of q_t is less than K. A linear projection approach based on the functional model would estimate Ψ in (29) by OLS, using (Y_{t-1}, α_{t-1}) as controls. Giving this estimator a quasi-Bayesian interpretation, the posterior variance for the impact response is given by $M'\big(\Sigma_{\tilde{u}_\alpha} \otimes (\sigma^2_{\epsilon_z})^{-1}\big)M$, because $\epsilon_{z,t}$ is orthogonal to (Y_{t-1}, α_{t-1}), where $\Sigma_{\tilde{u}_\alpha}$ is the covariance matrix of $\tilde{u}_{\alpha,t}$ and $\sigma^2_{\epsilon_z}$ is the variance of $\epsilon_{z,t}$.

Define $M_\perp$ as a matrix with columns that are orthogonal to the columns of M and let $W = ([M\, M_\perp]')^{-1}$ such that we can write the linear projection regression associated with the percentile model as

$$\begin{aligned} q_t &= M'\Psi\epsilon_{z,t} + M'\Phi_{\alpha y}Y_{t-1} + M'\Phi_{\alpha\alpha}Wq_{t-1} + M'\Phi_{\alpha\alpha}WM'_\perp\alpha_{t-1} + M'\tilde{u}_{\alpha,t} \qquad (30) \\ &= M'\Psi\epsilon_{z,t} + M'\Phi_{\alpha y}Y_{t-1} + M'\Phi_{\alpha\alpha}Wq_{t-1} + \nu_{\alpha,t}. \end{aligned}$$

The key difference between (30) and (29) is that the omission of $M'_\perp\alpha_{t-1}$ as additional control variable inflates the regression error variance, generates serial correlation in the error term $\nu_{\alpha,t}$, and thereby raises the estimation uncertainty associated with the impact response $M'\Psi$. In practice, the magnitude of the inefficiency depends on the extent to which the percentiles q_t span the information in α_t. But the larger the dimension of q_t, the more likely it is that quantiles may cross in response to shocks, which is clearly an undesirable feature of a "many quantiles" specification. The inefficiency and the incoherence of a "many quantiles" model are avoided by our functional approach.

6 Empirical Analysis

The empirical analysis focuses on the joint dynamics of total factor productivity, real per-capita GDP, and employment at the aggregate level, and the cross-sectional distribution of earnings. Estimation results for the functional state-space model are presented in Section 6.1. In Section 6.2 we report impulse responses of aggregate variables, the cross-sectional distribution, and inequality measures derived from the cross-sectional distribution to aggregate

shocks. Section 6.3 compares IRFs from the functional model to IRFs from VARs that include inequality statistics directly. Finally, in Section 6.4 we report responses to shocks that explain the largest share of the variance of particular variables or statistics.

6.1 Data and Model Estimation

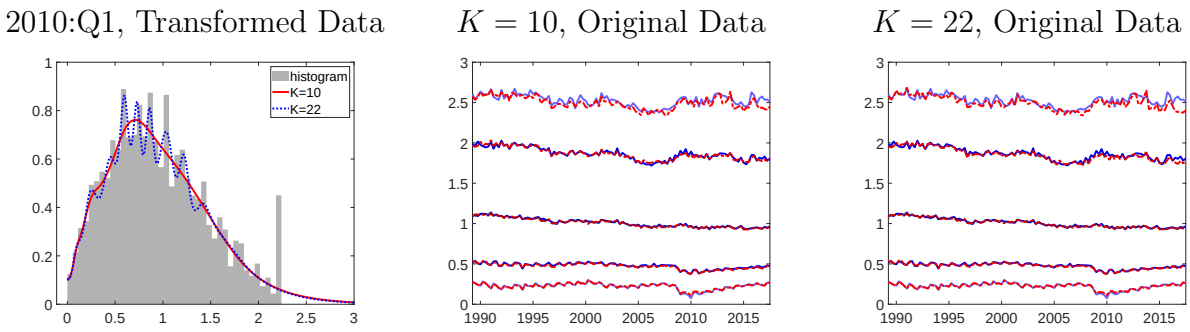
Data. We use three macroeconomic aggregates in our empirical analysis: total factor productivity (TFP), real per-capita GDP, and the employment rate. In addition, we use cross-sectional data on earnings. Real per-capita GDP (*A939RX0Q048SBEA*) is provided by the Federal Reserve Bank of St. Louis' FRED 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 remainder of this paper we standardize individual-level earnings by 2/3 of nominal per-capita GDP to remove (most of) the common trend from the cross-sectional data. The factor 2/3 is a rule-of-thumb number for the labor share. After this standardization an individual earning "1" receives the labor share of GDP per capita. Rather than taking a logarithmic transformation of the standardized earnings data, we apply an inverse hyperbolic sign transformation $x = g(z)$, where x corresponds to the transformed data for which we fit the log-spline densities and z to the original earnings/GDP data. For small values of z the function is approximately equal to z and for large values of z it is equal to $\ln(z) + \ln(2)$. Further details are provided in the Online Appendix.

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. As in the simulation study in Section 5, we consider different approximation orders K. We use the spline basis in (21) 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 left panel of Figure 8 we show three types of density estimates for the transformed earnings in 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 log-spline density estimates are constructed to extrapolate the top-coded

Figure 8: Fitted Densities and Percentiles of Earnings/GDP Distribution



Notes: Left panel: fitted cross-sectional densities and histograms. Center and right panels: sample percentiles (10%, 20%, 50%, 80%, and 90%) are blue, percentiles from estimated densities are red.

income values (spike in the right tail of the histogram). In the center and right panels 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 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 $x = g(z)$ into actual earnings z.

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 plot in the left panel. 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 $\{\hat{\alpha}_t^{(K)}\}$. 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_* and α_* 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:Q2 as period

Table 4: Hyperparameter Estimates and Log MDD Differentials

	Optimal				$\lambda_2 = 1$		
K	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{p}$	MDD	$\hat{\lambda}_1$	$\hat{p}$	MDD
4	54.6	1	4	0	54.6	4	0
6	54.6	2.7	4	8,483	54.6	4	8,481
8	148.4	1	4	9,302	148.4	4	9,302
10	*54.6*	*1096.6*	*1*	*9,347*	148.4	1	9,330
14	148.4	4.9E8	1	9,668	148.4	1	9,650
18	403.4	20.1	1	9,711	403.4	1	9,704
22	**403.4**	**7.4**	**1**	**12,740**	403.4	1	12,730

Notes: The log MDD differentials are computed with respect to $K = 4, \lambda_1 = \hat{\lambda}_1, \lambda_2 = \hat{\lambda}_2, p = \hat{p}$. For each K we maximized the MDD with respect to λ and p to obtain $\hat{\lambda}_j(K)$ and $\hat{p}(K)$.

$t = 1$ and use the earlier observations to initialize lags in the VAR law of motion.

Hyperparameter Selection and Granger Causality. We proceed by evaluating the (approximate) log MDD as a function of the model dimension $K \in \{4, 6, 8, 10, 14, 18, 22\}$, the lag length $p \in \{1, 2, 3, 4\}$, and the hyperparameters λ_1 and λ_2. For each $\ln \lambda_j$ we consider 31 equally-spaced values on the interval $[-10, 20]$. Results are summarized in Table 4. For each value of K we report the optimal $\hat{\lambda}_1$ and $\hat{\lambda}_2$ (columns 2-3), the number of lags (column 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 8. 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. The only exception is the transition from $K = 8$ to $K = 10$ when the lags drop from 4 to 1. Recall that λ_2 controls the relative precision of the prior for the submatrices $\Phi_{ya}^{(j)}$ that capture spillovers from the lagged distributions to the current aggregate variables. The optimal values $\hat{\lambda}_2(K)$ show additional shrinkage of the Φ_{ya} blocks to zero. The degree of relative shrinkage is quite sensitive to K, and, except for the $K = 14$ case, not as strong as in the KS simulation in Section 5.2. The last column of the table reports log MDD differentials obtained by setting $\lambda_2 = 1$. In this case the degree of shrinkage for the off-diagonal block Φ_{ya} is same as for the other coefficients. The MDDs deteriorate, except for $K = 4$ and $K = 8$ when $\hat{\lambda}_2 = 1$, but the change in MDD is smaller than for changes in K, because λ_2 affects primarily the fit in the time series dimension, instead of the cross-sectional dimension.

Overall, we conclude that the Granger-causal relationship from the cross-sectional income distribution to the aggregate variables is weak, albeit not fully absent.[23]

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 8 – 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. Results for $K = 22$ are presented in the Online Appendix. 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 $K = 22$.

6.2 Effects of Aggregate Shocks

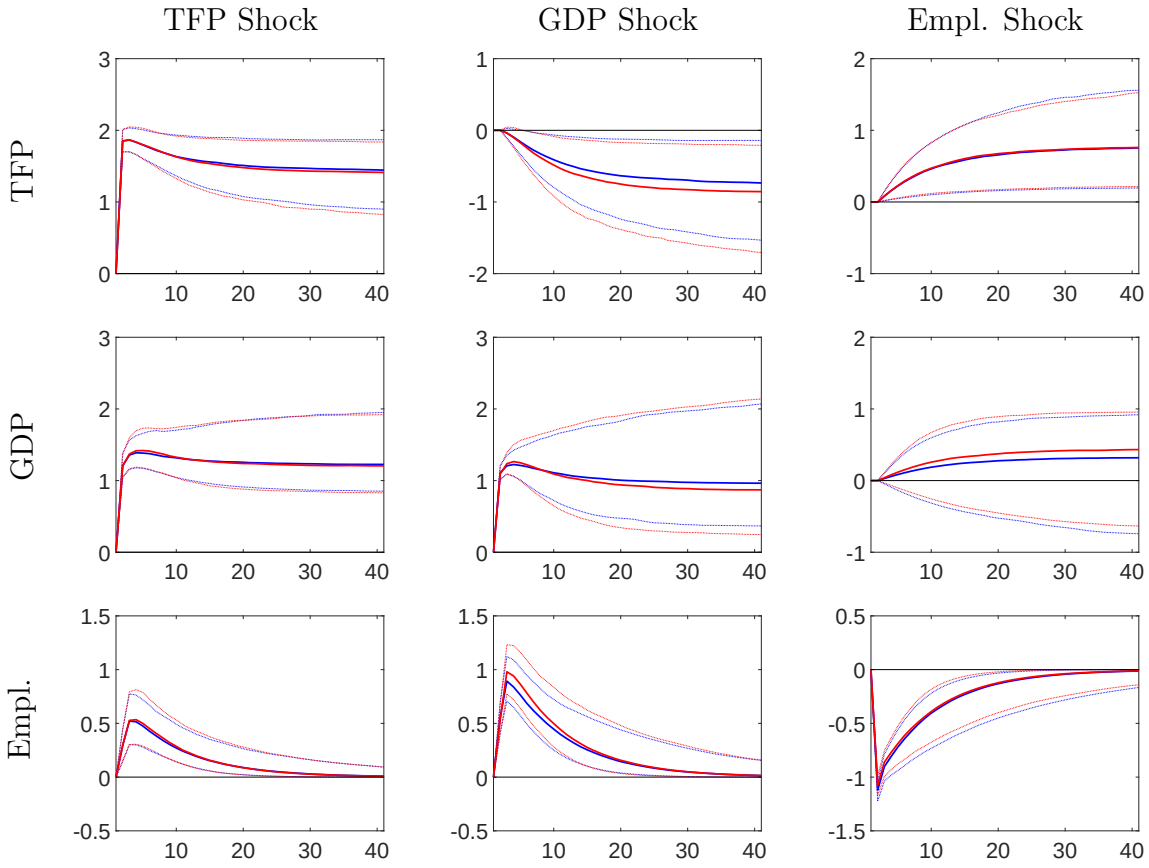
Identification. In the vector Y_t we order TFP growth first, GDP growth second, and the employment rate third. Let Σ_{tr} be the lower-triangular Cholesky factor of Σ such that $\Sigma = \Sigma_{tr}\Sigma_{tr}'$ and let Ω be an orthonormal matrix. The relationship between the reduced-form innovations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

$$u_t = \Sigma_{tr}\Omega\epsilon_t. \tag{31}$$

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 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.

[23] If we restrict the lag lengths to $p = 1$ for $K = 4, 6, 8$ the relative shrinkage on the off-diagonal blocks increases to 148.4, 54.6, and 20.1, respectively.

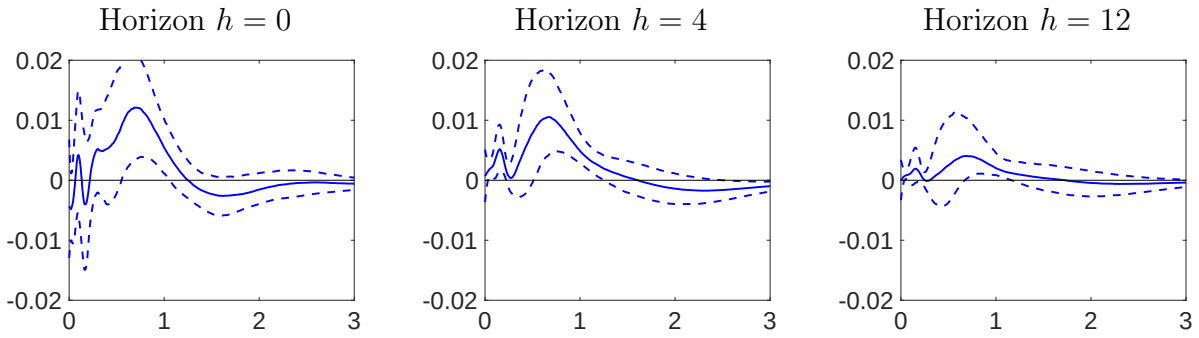
Figure 9: Responses of Aggregate Variables to Aggregate Shocks



Notes: 10th, 50th, and 90th percentiles of the posterior distribution of IRFs for three-standard deviation aggregate shocks (orthogonalized via Cholesky factorization; see (31)). Panels depict responses of the log level of TFP and GDP, scaled by 100, and responses of the employment rate in percent. Solid blue responses are based on functional VAR(1) with $K = 10, \hat{\lambda}_1 = 54.6, \hat{\lambda}_2 = 1096.6$; dashed red responses are based on aggregate VAR(1) with $\hat{\lambda}_1 = 20.1$.

Response of Aggregate Variables. Percentiles (10th, 50th, 90th) of the posterior distribution of IRFs are plotted in Figure 9. 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 the functional model against responses obtained by estimating a VAR in the three aggregate variables only, henceforth aggregate VAR. We set $p = 1$ to match the lag length of the functional model and choose the hyperparameter λ_1 by MDD maximization. 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). Second, the GDP growth shock raises GDP permanently, creates a temporary employment boom and a drop in measured

Figure 10: Earnings Density Differential (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. We depict differences between the shocked and the steady state cross-sectional density at various horizons.

total factor productivity in the long run. Finally, the third shock leads to a drop in the employment rate and it raises TFP and GDP with a one-period delay.

The estimated responses from the functional and the aggregate model are essentially identical. The degree of IRF overlap is as strong as in the left panel of Figure 5, which was obtained from the baseline specification of the KS model that features perfect aggregation. This is consistent with the strong shrinkage of $\Phi_{y\alpha}$ to zero ($\hat{\lambda}_2 = 1,096.6$) reported in Table 4. Our empirical result is consistent with Bayer, Born, and Luetticke (2020) who find that including or excluding the inequality data does not change the estimates and variance decompositions of their HANK model. In addition, the aggregation results in Werning (2015) and Berger, Bocola, and Dovis (2019) are compatible with the strong overlap we find.

Distributional Responses. Figure 10 illustrates the response of the density of transformed earnings to a three-standard deviation TFP innovation. As in Figure 6, we show the differentials 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 horizons $h = 4$ 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

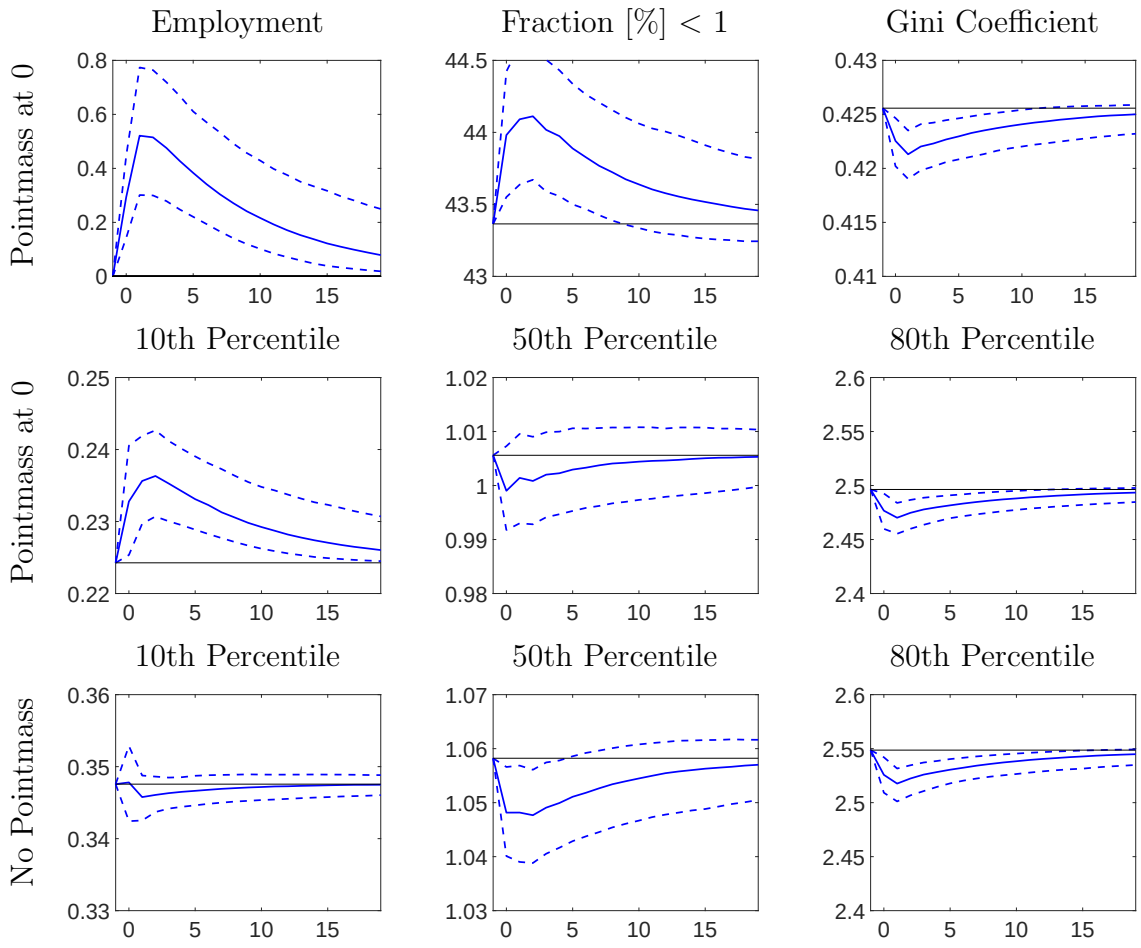
zero labor earnings, and, second, individuals whose earnings do not rise as strongly as GDP per capita. The first effect is consistent with a model in which individuals are heterogeneous with respect to their idiosyncratic productivity and only individuals whose productivity exceeds a state-dependent 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 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.

Inequality Measures. 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 11 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. The upper left panel reproduces the employment responses. The statistics in the top two rows are computed after assigning zero earnings to the unemployed individuals. The statics in the bottom row are computed from the continuous part of the cross-sectional distribution only.

We first consider the "Pointmass at 0" case. According to the posterior median of the IRF, overall the fraction of individuals earning less than per-capita GDP increases from 43.5% to 44%. While the density differential response for $h = 0$ depicted in Figure 10 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 0.4225. Earnings at the 10th percentile of the cross-sectional distribution also increase from 0.224 to 0.233, at the median they fall slightly from 1.005 to 0.999, with a band that ranges upon impact from approximately 0.99 to 1.01. Earnings at the 90th percentile fall very slightly, from 2.496 to 2.477. This compression of the earnings distribution is consistent with the decrease in the Gini coefficient.

A comparison between the percentile reponses in the second and third row shows the roles played by the extensive margin, that is individuals moving into and out of unemploy-

Figure 11: 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.

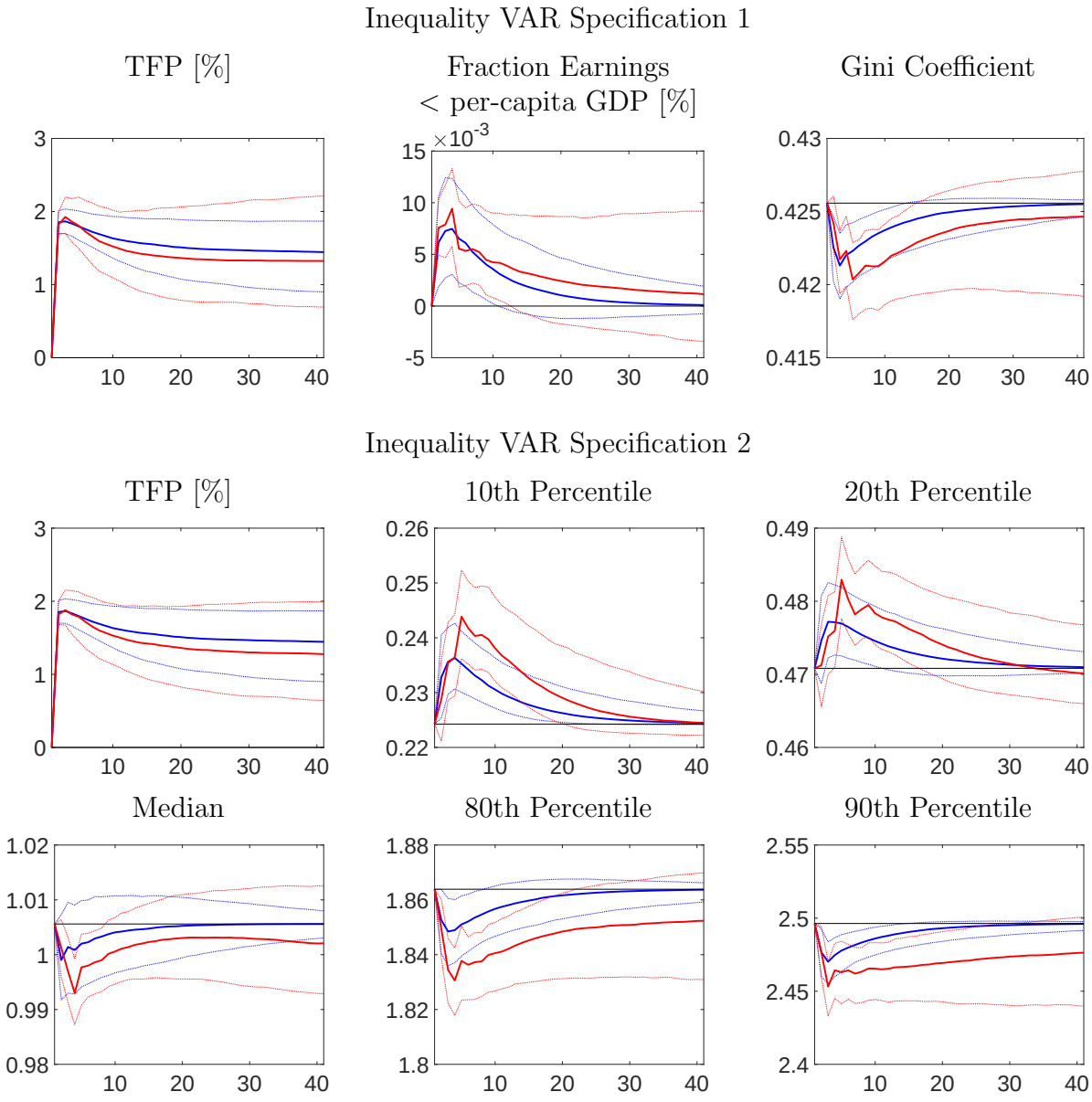
ment. Ignoring the pointmass at zero raises the baseline level of the percentiles. If the pointmass at zero is included, the 10th percentile rises (posterior median) from its baseline level 0.225 to 0.237. If we condition on employment then the posterior median of the 10th percentile response falls slightly below the baseline level. This is consistent with individuals transitioning into employment after a positive TFP shock, are generally low productivity individuals. The responses of the 50th and 90th percentile are not sensitive to the extensive margin. As a consequence, not accounting for a pointmass at zero dampens the responses of the Gini coefficient and the fraction of units earning less than GDP per capita (not shown in the figure).

6.3 Comparison to VARs with Inequality Statistics

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. As a robustness check, we estimate two VARs that combine the three aggregate variables with inequality statistics. We refer to these VARs as inequality VARs. Specification 1 combines the aggregate variables with the fraction of individuals earning less than GDP per capita and the Gini coefficient series directly computed from the cross-sectional observations. Specification 2 augments the aggregate variables with the 10th, 20th, 50th, 80th and 90th percentiles. The inequality VARs also take the form of (22), where W_t is redefined to replace α_t by the relevant distributional statistics. The VARs are estimated with the same prior as the functional model. The MDD criterion selects a lag length of $\hat{p} = 4$ and a relative shrinkage hyperparameter $\hat{\lambda}_2 = 4.9E8$, meaning that the estimated system is block-triangular and the inequality statistics do not Granger-cause the aggregate variables. In fact the $\hat{\lambda}$ pattern is similar to the one obtained from the functional VAR applied to data simulated from the KS economy. In a similar vein, Berger, Bocola, and Dovis (2019) construct preference shocks that summarize all the information from the cross-section relevant for aggregate dynamics. These shocks feedback to aggregate variables is found to be quantitatively small.

Figure 12 overlays impulse responses to a three standard deviation technology shock for TFP from the functional model and the inequality VAR. The top row contains results for Specification 1, and the remaining two rows results for Specification 2. The TFP responses between the functional model and the inequality VARs are well aligned, albeit at the posterior median, the long-run response under the former is slightly lower than under the latter. All other responses match in shape, but at the posterior mean the inequality VAR responses are of larger magnitude than the responses computed from the functional model. Most notably, the credible bands for the inequality VARs are much wider than those for the functional model, in particular in the long run. Thus, 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. The empirical findings in regard to the width of the credible bands are consistent with the results from the simulation experiment reported in Figure 7.

Figure 12: Functional Model vs. Inequality VAR



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(4). Percentile responses are computed for the original data.

6.4 Maximum-Share-of-Variance Shocks

We now consider four shocks that explain the largest share of the variance of a particular variable or statistic. This approach dates back to Faust (1998) and Uhlig (2003) and has been recently used, for instance, to study the anatomy of business cycles in Angeletos, Collard,

and Dellas (2020). Recall the relationship between one-step-ahead forecast errors u_t and structural shocks ϵ_t in (31). Denote the first column of Ω by q. Then, $\Sigma_{tr}q$ is the impact of a one-standard deviation shock $\epsilon_{1,t}$. For the first two shocks considered, we choose the value q_* such that the shock $\epsilon_{1,t}$ explains the largest share of the variance of a targeted element of W_t, denoted by $M'W_t$, where M' is a row vector that either picks output growth or employment. Let $\Sigma_W(q;\Phi,\Sigma)$ denote the unconditional covariance matrix of W_t explained by $\epsilon_{1,t}$. Then

$$q_*(M) = \text{argmax}_{\|q\|=1}\ M'\Sigma_W(q;\Phi,\Sigma)M. \tag{32}$$

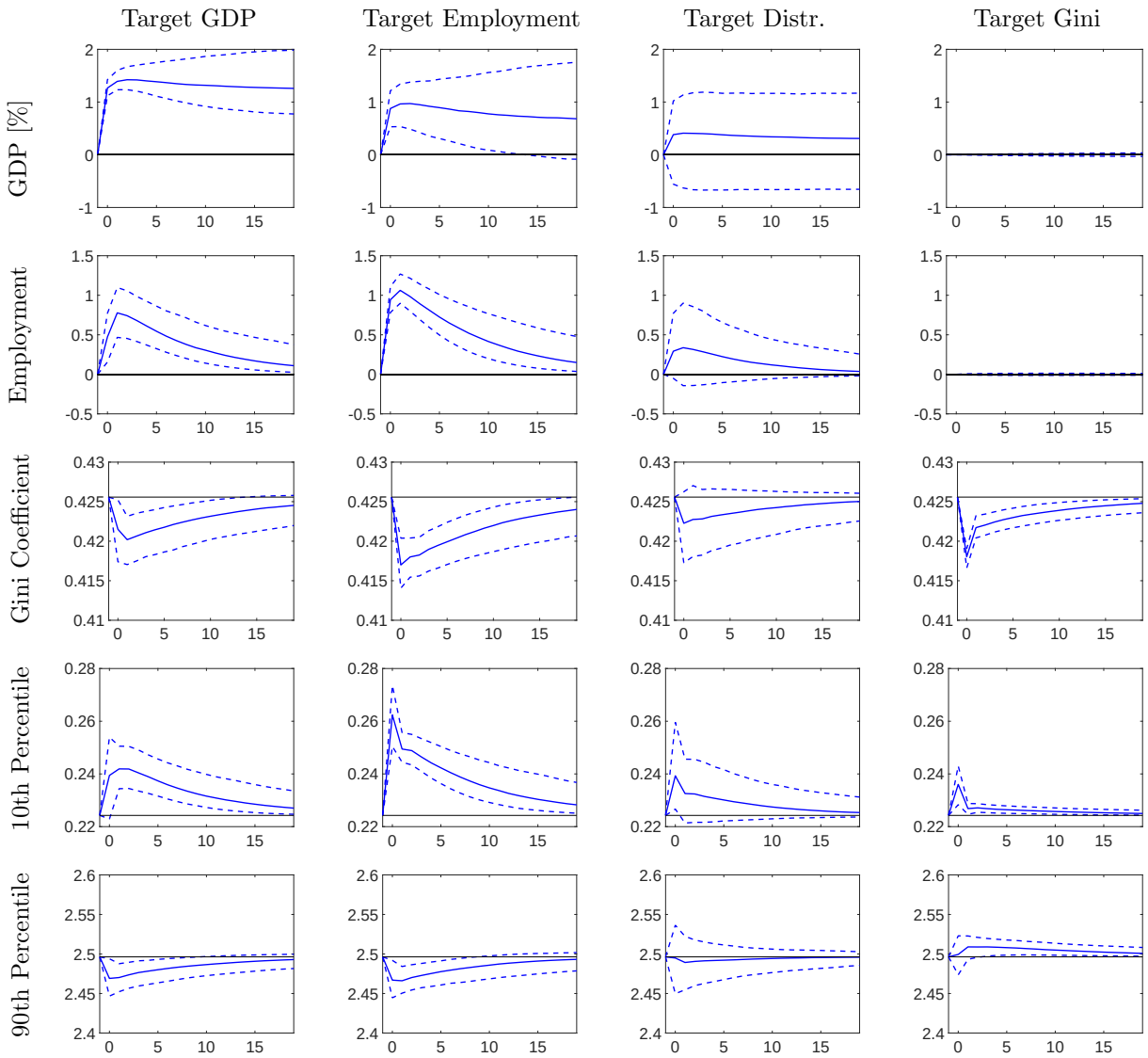
The third shock, henceforth "distribution" shock, targets the variation in the cross-sectional distribution. The unnormalized log density of the cross-sectional observations is given by $\ell_t^{(K)}(x) = \zeta'(x)\alpha_t$. We denote the covariance kernel of this function by $\Omega_\ell(x,\tilde{x}) = \zeta'(x)\Sigma_\alpha(q)\zeta(\tilde{x})$, where $\Sigma_\alpha(q)$ is defined in the same way as $M'\Sigma_W(q)M$ above. Now M' is a matrix that selects the α_t elements from W_t. We discretize the domain of x and then maximize the sum of the eigenvalues of the covariance matrix for the discretized x.[24] The fourth shock, henceforth "Gini" shock, is constructed as distributional shock that conditional on (Φ,Σ) maximizes the variation in the Gini coefficient without affecting the aggregate variables upon impact. We restrict the candidate values of q to $q = [0', q_\alpha']'$, where the partition of q conforms with the partition of $u_t' = [u_{y,t}', u_{\alpha,t}']$ and q_α is a K-dimensional unit-length vector.

We plot the IRFs for the four maximum-share-of-variance shocks in Figure 13. The responses to the shocks that target GDP and employment variance look qualitatively and quantitatively very similar. They mimic the responses to a TFP shock in Figure 9. Thus, at the aggregate level most of the fluctuations in GDP and employment are driven by a TFP-like shock that generates positive comovements between GDP and employment. This shock spills over into the cross-sectional distribution of earnings as more individuals become employed and transition from zero to positive earnings. Earnings relative to GDP rise at the 10th percentile and fall at the 90th percentile and overall inequality as measured by the Gini coefficient falls.

The IRFs for the shocks that target the share of variance of features of the cross-sectional distribution have similar effects on the cross-sectional distribution as the aggregate target shocks: the Gini coefficient falls and income at the 10th percentile rise. In response to the distribution shock employment rises, but not as strongly as in response to the GDP and

[24] Maximizing the trace of $\Omega_\ell(x,\tilde{x})$ produced similar IRFs.

Figure 13: Business Cycle Anatomy



Notes: 10th, 50th, and 90th percentiles of posterior distributions of IRFs. Sign normalizations: positive response upon impact of target variable for "GDP" and "employment" shock; positive response upon impact of 10th percentile to "distribution" shock and negative response of Gini coefficient to "Gini" shock. The range of the y-axis is identical within rows.

employment target shocks. The response of output at the posterior median is positive, but the credible bands are wide and the sign of the response is ambiguous. Finally, the fourth shock is a pure distributional shock that has no effect on aggregate GDP and employment.

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. The absence of an aggregate response to the Gini shock is consistent with the response to a distributional shock that Auclert and Rognlie (2020) find for standard neoclassical models. The quantitatively small effect of a distributional shock to aggregate variables is also consistent with the results of Berger, Bocola, and Dovis (2019) and Bilbiie, Primiceri, and Tambalotti (2022). A thorough assessment of the extent to which the current generation of HA models can reproduce the impulse response patterns estimated with our functional model is left for future research.

7 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 aggregate and distributional 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 hardly affects the estimated 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 the responses of aggregate output to distributional shocks are not significant. 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. Extensions left for future work include the introduction of time-varying volatility and allowing for a more general mixed-frequency structure.

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.

Angeletos, G.-M., F. Collard, and H. Dellas (2020): "Business-Cycle Anatomy," *American Economic Review*, 110, 3030–3070.

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

Barron, A., and C.-H. Sheu (1991): "Approximation of Density Functions by Sequences of Exponential Families," *Annals of Statistics*, 19(3), 1347–1369.

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

Berger, D. W., L. Bocola, and A. Dovis (2019): "Imperfect Risk-Sharing and the Business Cycle," *NBER working paper*, 26032.

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

Bilbiie, F., G. Primiceri, and A. Tambalotti (2022): "Inequality and Business Cycles," *Working Paper*.

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.

Chan, J. C. C. (2022): "Asymmetric Conjugate Priors for Large Bayesian VARs," *Quantitative Economics*, 13(3), 1145–1169.

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.

Chang, Y., S.-B. Kim, and F. Schorfheide (2013): "Labor Market Heterogeneity, Aggregation, and the Policy-(In)variance of DSGE Model Parameters," *Journal of the European Economic Association*, 11(S1), 193–220.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, chap. 76, pp. 5549–5632. Elsevier, New York.

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.

Drautzburg, T., and P. Amir-Ahmadi (2021): "Identification and Inference with Ranking Restrictions," *Quantitative Economics*, 12, 1–39.

Faust, J. (1998): "The Robustness of Identified VAR Conclusions about Money," *Carnegie-Rochester Conference Series on Public Policy*, 49, 207–244.

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

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

Gordon, N., and D. Salmond (1993): "A Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEEE Proceedings-F*, 140, 107–113.

Herbst, E., and F. Schorfheide (2015): *Bayesian Estimation of DSGE Models*. Princeton University Press, Princeton.

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 (2021): "The Effects of Conventional and Unconventional Monetary Policy: A New Approach," *Quantitative Economics*, 12(4), 1085–1138.

Kaplan, G., and G. L. Violante (2022): "The Marginal Propensity to Consume in Heterogeneous Agent Models," *NBER Working Paper*, 30013.

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.

——— (1992): "Logspline Density Estimation for Censored Data," *Journal of Computational and Graphical Statistics*, 1(4), 301–328.

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 (2022): "Full-Information Estimation of Heterogeneous Agent Models Using Macro and Micro Data," *Quantitative Economics*, Forthcoming.

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.

Portnoy, S. (1988): "Asymptotic Behavior of Likelihood Methods for Exponential Families when the Number of Parameters Tends to Infinity," *Annals of Statistics*, 16(1), 356–366.

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.

Shao, J. (1997): "An Asymptotic Theory for Linear Model Selection," *Statistica Sinica*, 7, 221–264.

Stone, C. J. (1990): "Large-Sample Inference For Log-Spline Models," *Annals of Statistics*, 18(2), 717–741.

Uhlig, H. (2003): "What Moves Real GNP," *Manuscript, University of Chicago*.

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*.

Werning, I. (2015): "Incomplete Markets and Aggregate Demand," *NBER working paper*, 21448.

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

Zellner, A. (1983): "Applications of Bayesian Analysis in Econometrics," *The Statistician*, 32, 23–34.

Online Appendix: Heterogeneity and Aggregate Fluctuations

Minsu Chang, Xiaohong Chen, and Frank Schorfheide

The Online Appendix consists of the following parts:

A. Proofs and Derivations for Section 3

B. Implementation Details

C. Solution of the KS Model

D. A Second Simulation Experiment

E. More About the Empirical Analysis

A Proofs and Derivations for Section 3

A.1 Relationship Between Infinite- and Finite-Dimensional Model

Suppose we start from the infinite-dimensional model of Section 2.1 in the main paper. For each time $t = 1, ..., T$, let the data $X_t = [x_{1t}, ..., x_{Nt}]'$ be an i.i.d. draw from the true unknown density $p_t^o(\cdot)$ of $x_{it} \in [\underline{x}, \overline{x}]$. Let $p\left(x_{it}|\alpha_t, K\right)$ is the best KL projection of true unknown density $p_t^o(\cdot)$ of x_{it} onto the K-dimensional closed linear subspace (of $L^2(p_t^o)$) generated by the spline sieve basis $\left\{\zeta(\cdot)'\beta : \beta \in \mathbb{R}^K\right\}$:

$$\alpha_t = \arg\min_{\beta \in \mathbb{R}^K} \left\{ \int_{z \in [\underline{x}, \overline{x}]} p_t^o(z) \ln \left(\frac{p_t^o(z)}{p\left(z|\beta, K\right)} \right) dz \right\}, \tag{A.1}$$

where for any candidate parameter $\beta \in \mathbb{R}^K$,

$$p\left(z|\beta, K\right) = \frac{\exp\left\{\zeta'(z)\beta\right\}}{\int_{z \in [\underline{x}, \overline{x}]} \exp\left\{\zeta'(z)\beta\right\} dz} = \exp\left\{ \zeta'(z)\beta - \ln\left(\int_{z \in [\underline{x}, \overline{x}]} \exp\left\{\zeta'(z)\beta\right\} dz \right) \right\}.$$

Write the K'th order representation of the density of $X_t = [x_{1t}, ..., x_{Nt}]'$:

$$\begin{aligned} p^{(K)}\left(X_t|\alpha_t\right) &\equiv \prod_{i=1}^{N} p\left(x_{it}|\alpha_t, K\right) = \exp\left\{N\mathcal{L}^{(K)}(\alpha_t|X_t)\right\}, \\ \mathcal{L}^{(K)}(\alpha_t|X_t) &= \bar{\zeta}'(X_t)\alpha_t - \varphi(\alpha_t), \quad \varphi(\alpha_t) = \ln \int_{[\underline{x}, \overline{x}]} \exp\left\{\zeta'(z)\alpha_t\right\} dz. \end{aligned} \tag{A.2}$$

Then,

$$\alpha_t = \text{argmax}_{\beta\in\mathcal{A}}\, \mathbb{E}^o[\mathcal{L}^{(K)}(\beta|X_t)] = \text{argmax}_{\beta\in\mathcal{A}}\, \mathbb{E}^o[\bar{\zeta}'(X_t)\beta - \varphi(\beta)] \tag{A.3}$$

and satisfies the population first-order condition

$$\mathbb{E}^o[\zeta(x_{it})] - \varphi^{(1)}(\alpha_t) = 0. \tag{A.4}$$

Let $\hat{\alpha}_t$ be the maximum likelihood estimator (MLE) using K dimensional sieve basis:

$$\hat{\alpha}_t = \text{argmax}_{\beta\in\mathcal{A}}\, \mathcal{L}^{(K)}(\beta|X_t), \tag{A.5}$$

which is the unique solution to the score equation

$$\bar{\zeta}(X_t) - \varphi^{(1)}(\hat{\alpha}_t) = 0. \tag{A.6}$$

Suppose that $\ln p_t^o(\cdot)$ belongs to a Sobolev or a Hölder ball with smoothness $s > 1/2$. Then we have for an r-th order spline basis with $r > s$:

$$\int p_t^o(x) \ln\left(\frac{p_t^o(x)}{p^{(K)}(x|\alpha_t)}\right) dx \lesssim K^{-2s}, \quad \int p_t^o(x) \ln\left(\frac{p^{(K)}(x|\alpha_t)}{p^{(K)}(x|\hat{\alpha}_t)}\right) dx = O_p\left(\frac{K}{N}\right). \tag{A.7}$$

see, e.g., Stone (1990), Barron and Sheu (1991), and Chen (2007). For B-splines and wavelet sieves we obtain the following bounds on the "bias" (or approximation error) and the standard deviation (discrepancy between $\hat{\alpha}_t$ and α_t) uniformly in t,

$$\begin{aligned} \left\|\ln[p_t^o(\cdot)] - \ln[p^{(K)}(\cdot|\alpha_t)]\right\|_{L^\infty([\underline{x},\overline{x}])} &\lesssim K^{-s}, \\ \left\|\ln[p^{(K)}(\cdot|\hat{\alpha}_t)] - \ln[p^{(K)}(\cdot|\alpha_t)]\right\|_{L^\infty([\underline{x},\overline{x}])} &= O_p\left(\sqrt{\frac{K\ln K}{N}}\right). \end{aligned} \tag{A.8}$$

Suppose $s > 1$, then we can define the set $\mathcal{K}_N$ such that the bias is of strictly smaller order than the standard derivation part, say

$$\mathcal{K}_N \equiv \left\{K \in \mathbb{N} : \underline{K} \le K \le \overline{K},\ \underline{K} = \underline{c} \times [N\ln N]^{\frac{1}{1+2s}},\ \overline{K} = \overline{c} \times \frac{N^{1/3}}{(\ln N)^2},\ 0 < \underline{c}, \overline{c} < \infty\right\}.$$

Then, uniformly over t and $K \in \mathcal{K}_N$,

$$\begin{aligned} \left\|\ln[p_t^o(\cdot)] - \ln[p^{(K)}(\cdot|\hat{\alpha}_t)]\right\|_{L^2([\underline{x},\overline{x}])} &= O_p\left(\sqrt{\frac{K}{N}}\right) = o_p(N^{-\frac{1}{3}}); \\ \left\|\ln[p_t^o(\cdot)] - \ln[p^{(K)}(\cdot|\hat{\alpha}_t)]\right\|_{L^\infty([\underline{x},\overline{x}])} &= O_p\left(\sqrt{\frac{K\ln K}{N}}\right) = o_p(N^{-\frac{1}{3}}); \\ \int p_t^o(x)\ln\left(\frac{p_t^o(x)}{p^{(K)}(x|\hat{\alpha}_t)}\right)dx &= O_p\left(\frac{K}{N}\right) = o_p(N^{-\frac{2}{3}}) = o_p(N^{-\frac{1}{2}}). \end{aligned} \tag{A.9}$$

Having defined the K-dimensional process α_t in (A.1) based on $p_t^o(\cdot)$ from the infinite-dimensional model comprising (1), (2), and (3) in the main text, one can define pseudo-optimal values for Y_*, α_*, Φ_{yy}, $\Phi_{y\alpha}$, $\Phi_{\alpha y}$, $\Phi_{\alpha\alpha}$, and Σ in (6) of the main text by population regression. Let $\mathbb{P}^o$ denote the probability measure that governs the law of motion of (Y_t, p_t^o). Under the assumption that $(Y_t, \alpha_t^{(K)})$ is covariance stationary under $\mathbb{P}^o$, we can define the pseudo-optimal VAR parameters (Φ, Σ) by population MLE under $\mathbb{P}^o$ given K. Setting $p = 1$ and dropping the intercepts, we would obtain (superscripts omitted):

$$\begin{aligned} (\Phi_o, \Sigma_o) = \text{argmax}_{\Phi,\Sigma} & -\frac{1}{2}|\Sigma| \\ & -\frac{1}{2}\mathbb{E}^o\left\{\left(\begin{bmatrix} Y_t' & \alpha_t' \end{bmatrix} - \begin{bmatrix} Y_{t-1}' & \alpha_{t-1}' \end{bmatrix}\Phi\right)\Sigma^{-1}\left(\begin{bmatrix} Y_t \\ \alpha_t \end{bmatrix} - \Phi'\begin{bmatrix} Y_{t-1} \\ \alpha_{t-1} \end{bmatrix}\right)\right\}, \end{aligned} \tag{A.10}$$

where $\mathbb{E}^o[\cdot]$ is the expectations operator associated with $\mathbb{P}^o$. Using the definition of $C_\alpha = \int \xi(\tilde{x})\zeta'(\tilde{x})d\tilde{x}$, we can define $B_{yl} = \Phi_{y\alpha}C_\alpha^{-1}$ and $B_{ll} = \Phi_{\alpha\alpha}C_\alpha^{-1}$. This generates finite-dimensional pseudo-optimal kernels in (5) in the main text. We note that the density $p^{(K)}(X_t|\alpha_t)$ defined in (A.2) does not depend on θ which collects the parameters $(\alpha_*, Y_*, \Phi, \Sigma)$.

It is beyond the scope of the paper, to track the errors of approximating the infinite-dimensional model by a K dimensional model formally. In the main text, for any $K \in \mathcal{K}_N$, we ignore the sieve approximation error and treat $p^{(K)}(x|\alpha_t)$ as the density generates the data x_{it} and assume that (6) and (7) hold exactly.

A.2 Derivatives of the Log Likelihood of X_t

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

evaluated at any candidate parameter value $\beta = (\beta_1, ..., \beta_K)'$ has the form

$$\mathcal{L}(\beta|X_t) = \sum_{k=1}^{K} \beta_k \bar{\zeta}_k(X_t) - \varphi(\beta), \qquad \text{with} \quad \varphi(\beta) \equiv \ln \int \exp\left\{\sum_{k=1}^{K} \beta_k \zeta_k(z)\right\} dz.$$

The first-order derivatives of $\mathcal{L}(\beta|X_t)$ with respect to β_k for $k = 1, \ldots, K$ are given by

$$\begin{aligned} \mathcal{L}_k^{(1)}(\beta|X_t) &= \bar{\zeta}_k(X_t) - \varphi_k^{(1)}(\beta), \\ \varphi_k^{(1)}(\beta) &= \frac{\int \zeta_k(z) \exp\left\{\sum_{k=1}^{K} \beta_k \zeta_k(z)\right\} dz}{\int \exp\left\{\sum_{k=1}^{K} \beta_k \zeta_k(z)\right\} dz} \equiv \int \zeta_k(z) p(z|\beta, K) dz, \end{aligned}$$

where

$$p(z|\beta, K) \equiv \frac{\exp\left\{\sum_{k=1}^{K} \beta_k \zeta_k(z)\right\}}{\int \exp\left\{\sum_{k=1}^{K} \beta_k \zeta_k(z)\right\} dz} \equiv \exp\left\{\sum_{k=1}^{K} \beta_k \zeta_k(z) - \varphi(\beta)\right\},$$

and

$$\frac{\partial \ln p(z|\beta, K)}{\partial \beta_k} = \zeta_k(z) - \varphi_k^{(1)}(\beta).$$

The sieve MLE $\hat{\alpha}_t$ is unique and satisfies

$$\bar{\zeta}_k(X_t) - \varphi_k^{(1)}(\hat{\alpha}_t) = 0, \quad k = 1, \ldots, K.$$

The second-order derivatives of $\mathcal{L}(\beta|X_t)$ with respect to (β_k, β_l) are given by

$$\begin{aligned} \mathcal{L}_{kl}^{(2)}(\beta|X_t) &\equiv H_{kl}(\beta) = -\varphi_{kl}^{(2)}(\beta) \\ &= -\int \zeta_k(z) \frac{\partial \ln p(z|\beta, K)}{\partial \beta_l} p(z|\beta, K) dz \\ &= -\int \zeta_k(z) \left[\zeta_l(z) - \varphi_l^{(1)}(\beta)\right] p(z|\beta, K) dz \\ &= -\int \left[\zeta_k(z) - \varphi_k^{(1)}(\beta)\right] \left[\zeta_l(z) - \varphi_l^{(1)}(\beta)\right] p(z|\beta, K) dz. \end{aligned}$$

The third-order derivatives of $\mathcal{L}(\beta|X_t)$ with respect to $(\beta_k, \beta_l, \beta_m)$ take the form

$$\begin{aligned} \mathcal{L}_{klm}^{(3)}(\beta|X_t) &= -\varphi_{klm}^{(3)}(\beta) \\ &= -\int \zeta_k(z) \left(\left[\zeta_l(z) - \varphi_l^{(1)}(\beta)\right] \left[\zeta_m(z) - \varphi_m^{(1)}(\beta)\right] - \varphi_{lm}^{(2)}(\beta) \right) p(z|\beta, K) dz \\ &= -\int \left[\zeta_k(z) - \varphi_k^{(1)}(\beta)\right] \left[\zeta_l(z) - \varphi_l^{(1)}(\beta)\right] \left[\zeta_m(z) - \varphi_m^{(1)}(\beta)\right] p(z|\beta, K) dz. \end{aligned}$$

The 4th-order derivatives of $\mathcal{L}(\beta|X_t)$ with respect to $(\beta_k, \beta_l, \beta_m, \beta_j)$ take the form

$$\begin{aligned}
&\mathcal{L}^{(4)}_{klmj}(\beta|X_t) \\
&= -\varphi^{(4)}_{klmj}(\beta) \\
&= -\int \zeta_k(z)\left(\left[\zeta_l(z)-\varphi^{(1)}_l(\beta)\right]\left[\zeta_m(z)-\varphi^{(1)}_m(\beta)\right]-\varphi^{(2)}_{lm}(\beta)\right)\left[\zeta_j(z)-\varphi^{(1)}_j(\beta)\right]p(z|\beta,K)dz \\
&\quad -\int \zeta_k(z)\left(-\varphi^{(2)}_{lj}(\beta)\left[\zeta_m(z)-\varphi^{(1)}_m(\beta)\right]-\left[\zeta_l(z)-\varphi^{(1)}_l(\beta)\right]\varphi^{(2)}_{mj}(\beta)-\varphi^{(3)}_{lmj}(\beta)\right)p(z|\beta,K)dz \\
&= -\int \left[\zeta_k(z)-\varphi^{(1)}_k(\beta)\right]\left[\zeta_l(z)-\varphi^{(1)}_l(\beta)\right]\left[\zeta_m(z)-\varphi^{(1)}_m(\beta)\right]\left[\zeta_j(z)-\varphi^{(1)}_j(\beta)\right]p(z|\beta,K)dz \\
&\quad +\varphi^{(2)}_{kj}(\beta)\varphi^{(2)}_{lm}(\beta)+\varphi^{(2)}_{lj}(\beta)\varphi^{(2)}_{km}(\beta)+\varphi^{(2)}_{mj}(\beta)\varphi^{(2)}_{kl}(\beta).
\end{aligned}$$

The derivatives $\mathcal{L}^{(2)}_{kl}(\hat{\alpha}_t|X_t) \equiv H_{kl}(\hat{\alpha}_t)$, $\mathcal{L}^{(3)}_{klm}(\hat{\alpha}_t|X_t)$, $\mathcal{L}^{(4)}_{klmj}(\hat{\alpha}_t|X_t)$ depend on data X_t only through the sieve MLE $\hat{\alpha}_t$.

A.3 Proof of Lemma 1

Recall that $\mathcal{R}(\alpha_t)$ is the remainder term from the second-order Taylor series approximation in (11), which is

$$\mathcal{R}(\alpha_t) = \frac{1}{6}\sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{m=1}^{K}\mathcal{L}^{(K)(3)}_{klm}(\overline{\beta}_{klm,t})(\alpha_{kt}-\hat{\alpha}_{kt})(\alpha_{lt}-\hat{\alpha}_{lt})(\alpha_{mt}-\hat{\alpha}_{mt}),$$

where $\overline{\beta}_{klm,t}$ lies in between α_t and $\hat{\alpha}_t$ elementwise. A formula for $\mathcal{L}^{(K)(3)}_{klm}(\beta)$ is provided in Section A.2. We can deduce from Assumption 1:

$$\left|\exp\left\{N\mathcal{R}(\alpha_t)\right\}-1\right| \leq CN\left\|\alpha_t-\hat{\alpha}_t\right\|^3.$$

We note that (see Portnoy (1988) and Stone (1990)) $\|\alpha_t-\hat{\alpha}_t\| \lesssim \sqrt{K/N}$. Thus, the first statement of the Lemma follows. The second part of the Lemma follows from

$$\frac{p(X_t|\alpha_t)}{p_{KF}(X_t|\alpha_t)} = \exp\{N\mathcal{R}(\alpha_t)\}. \quad \blacksquare$$

A.4 Proof of Theorem 1

The theorem is proved recursively. We start with a recursive assumption for period $t-1$ and will verify that it remains satisfied after one iteration of the filters. Let $D_{1:t} = (Y_{1:t}, X_{1:t})$.

Recursive Assumption. We start from the assumption that

$$\frac{p(\alpha_{t-1}|D_{1:t-1},\theta)}{p_{KF}(\alpha_{t-1}|D_{1:t-1},\theta)} \asymp \left(1 \pm \sqrt{\frac{K^3}{N}}\right). \tag{A.11}$$

Initialization. We initialize both the KF and the exact filter in period 0 with $p(\alpha_0|Y_0,\hat{\alpha}_0,\theta)$, which is assumed to be Gaussian:

$$\alpha_0|(Y_0,\hat{\alpha}_0,\theta) \sim \mathcal{N}\big(\hat{\alpha}_0, P_{0|0}(\alpha\alpha)\big), \quad P_{0|0}(\alpha\alpha) = \hat{V}_0/N.$$

To simplify the notation, we will subsequently drop $(Y_0,\hat{\alpha}_0)$ from the conditioning set. Because we use the same initialization for KF and exact filter, the recursive assumption is trivially satisfied for $t = 1$.

For each time period, the filtering consists of a forecasting and updating step. We will provide expressions for these steps using the approximate KF, and the exact filter. Recall that the state-transition equation for the KF and the exact filter are identical.

Time t Forecasting. We begin by forecasting (Y_t,α_t) with the KF and the exact filter:

$$\begin{aligned} p_{KF}(Y_t,\alpha_t|D_{1:t-1},\theta) &= \int p_G(Y_t,\alpha_t|Y_{t-1},\alpha_{t-1},\theta)p_{KF}(\alpha_{t-1}|D_{1:t-1},\theta)d\alpha_{t-1} \\ p(Y_t,\alpha_t|D_{1:t-1},\theta) &= \int p_G(Y_t,\alpha_t|Y_{t-1},\alpha_{t-1},\theta)p(\alpha_{t-1}|D_{1:t-1},\theta)d\alpha_{t-1}. \end{aligned}$$

Then,

$$\begin{aligned} &\Big|p(Y_t,\alpha_t|D_{1:t-1},\theta) - p_{KF}(Y_t,\alpha_t|D_{1:t-1},\theta)\Big| \\ &\leq \int p_G(Y_t,\alpha_t|Y_{t-1},\alpha_{t-1},\theta)\Big|p(\alpha_{t-1}|D_{1:t-1},\theta) - p_{KF}(\alpha_{t-1}|D_{1:t-1},\theta)\Big|d\alpha_{t-1} \\ &= \int p_G(Y_t,\alpha_t|Y_{t-1},\alpha_{t-1},\theta)p_{KF}(\alpha_{t-1}|D_{1:t-1},\theta)\left|\frac{p(\alpha_{t-1}|D_{1:t-1},\theta)}{p_{KF}(\alpha_{t-1}|D_{1:t-1},\theta)} - 1\right|d\alpha_{t-1} \\ &\leq C\sqrt{\frac{K^3}{N}}p_{KF}(Y_t,\alpha_t|D_{1:t-1},\theta). \end{aligned}$$

In turn, we can write

$$\frac{p(Y_t,\alpha_t|D_{1:t-1},\theta)}{p_{KF}(Y_t,\alpha_t|D_{1:t-1},\theta)} \asymp \left(1 \pm \sqrt{\frac{K^3}{N}}\right). \tag{A.12}$$

We now turn to forecasting (Y_t,X_t) conditional on $(D_{1:t-1})$. The predictive densities for

KF and exact filter are given by

$$\begin{aligned} p_{KF}(D_t|D_{1:t-1},\theta) &= p_{pen}(X_t|\hat{\alpha}_t)\int p_G(\hat{\alpha}_t|\alpha_t)p_{KF}(Y_t,\alpha_t|D_{1:t-1},\theta)d\alpha_t \\ p(D_t|D_{1:t-1},\theta) &= p_{pen}(X_t|\hat{\alpha}_t)\int p_G(\hat{\alpha}_t|\alpha_t)\exp\{N\mathcal{R}(\alpha_t)\}p(Y_t,\alpha_t|D_{1:t-1},\theta)d\alpha_t \end{aligned}$$

We can now use Lemma 1(i) to deduce that

$$\begin{aligned} &\left|p(D_t|D_{1:t-1},\theta) - p_{KF}(D_t|D_{1:t-1},\theta)\right| \\ &= p_{pen}(X_t|\hat{\alpha}_t)\int \left|p_G(\hat{\alpha}_t|\alpha_t)\right| \exp\{N\mathcal{R}(\alpha_t)\}p(Y_t,\alpha_t|D_{1:t-1},\theta) \\ &\qquad\qquad - p_{KF}(Y_t,\alpha_t|D_{1:t-1},\theta)\Big|d\alpha_t \\ &= p_{pen}(X_t|\hat{\alpha}_t)\int p_G(\hat{\alpha}_t|\alpha_t)p_{KF}(Y_t,\alpha_t|D_{1:t-1},\theta) \\ &\qquad\qquad \times \left|\exp\{N\mathcal{R}(\alpha_t)\}\frac{p(Y_t,\alpha_t|D_{1:t-1},\theta)}{p_{KF}(Y_t,\alpha_t|D_{1:t-1},\theta)} - 1\right| d\alpha_t. \end{aligned}$$

It follows that

$$\frac{p(D_t|D_{1:t-1},\theta)}{p_{KF}(D_t|D_{1:t-1},\theta)} \asymp \left(1 \pm \sqrt{\frac{K^3}{N}}\right). \tag{A.13}$$

Note that we can also exchange numerator and denominator in (A.13):[25]

$$\frac{p_{KF}(D_t|D_{1:t-1},\theta)}{p(D_t|D_{1:t-1},\theta)} \asymp \left(1 \pm \sqrt{\frac{K^3}{N}}\right). \tag{A.14}$$

Time t Updating. The updating is based on Bayes Theorem. For the KF and the exact filter, respectively, it takes the form

$$\begin{aligned} p_{KF}(\alpha_t|D_{1:t},\theta) &= \frac{p_{pen}(X_t|\hat{\alpha}_t)p_G(\hat{\alpha}_t|\alpha_t)p_{KF}(Y_t,\alpha_t|D_{1:t-1},\theta)}{p_{KF}(Y_t,X_t|D_{1:t-1},\theta)} \\ p(\alpha_t|D_{1:t},\theta) &= \frac{p_{pen}(X_t|\hat{\alpha}_t)p_G(\hat{\alpha}_t|\alpha_t)\exp\{N\mathcal{R}(\alpha_t)\}p(Y_t,\alpha_t|D_{1:t-1},\theta)}{p(Y_t,X_t|D_{1:t-1},\theta)}. \end{aligned}$$

[25]Note that for any fixed (K,N) all densities are non-zero because of the functional form assumptions. The state-transition equation for (Y_t,α_t) is linear Gaussian, and the density of the $x_{it}|\alpha_t$ is constructed by exponentiating $\zeta'(x_{it})\alpha_t$ which does not take the value $-\infty$. Thus, the density ratios are well defined.

Thus,

$$\begin{aligned}\frac{p(\alpha_t|D_{1:t},\theta)}{p_{KF}(\alpha_t|D_{1:t},\theta)} &= \exp\{N\mathcal{R}(\alpha_t)\}\frac{p_{KF}(Y_t,X_t|D_{1:t-1},\theta)}{p(Y_t,X_t|D_{1:t-1},\theta)}\frac{p(Y_t,\alpha_t|D_{1:t-1},\theta)}{p_{KF}(Y_t,\alpha_t|D_{1:t-1},\theta)} \\ &\asymp \left(1\pm\sqrt{\frac{K^3}{N}}\right)\left(1\pm\sqrt{\frac{K^3}{N}}\right)\left(1\pm\sqrt{\frac{K^3}{N}}\right) \\ &\asymp \left(1\pm\sqrt{\frac{K^3}{N}}\right)\end{aligned} \tag{A.15}$$

and the recursive assumption for $t+1$ is satisfied. ■

A.5 Particle Filter to Compute "Exact" Likelihood Function

The exposition follows Herbst and Schorfheide (2015) and we focus on the case of $p = 1$. We abstract from the compression/standardization step discussed in Section 4, use α_t for the basis function coefficients, and ignore the deterministic components α_* and Y_*.

A particle filter represents the density $p(\alpha_t|Y_{1:t}, X_{1:t})$ through a swarm of particles $\{\alpha_t^j, W_t^j\}_{j=1}^M$. We use $h(\cdot)$ to denote a function of α_t for which expectations are supposed to be evaluated. We start with the recursive assumption that

$$\frac{1}{M}\sum_{j=1}^{M} h(\alpha_{t-1}^j)W_{t-1}^j \approx \mathbb{E}[h(\alpha_{t-1})|Y_{1:t-1}, X_{1:t-1}] \tag{A.16}$$

and rewrite $p(y_t, X_t|Y_{1:t-1}, X_{1:t-1})$ as follows:

$$\begin{aligned}&p(y_t, X_t|Y_{1:t-1}, X_{1:t-1}) \\ &= \int_{\alpha_t} p(X_t|\alpha_t)p(Y_t,\alpha_t|y_{1:t-1}, X_{1:t-1})d\alpha_t \\ &= \int_{\alpha_t} p(X_t|\alpha_t)\left[\int_{\alpha_{t-1}} p(y_t,\alpha_t|Y_{t-1},\alpha_{t-1})p(\alpha_{t-1}|Y_{1:t-1}, X_{1:t-1})d\alpha_{t-1}\right]d\alpha_t \\ &= \int_{\alpha_{t-1}}\left[\int_{\alpha_t} p(X_t|\alpha_t)p(\alpha_t|y_t, Y_{t-1},\alpha_{t-1})d\alpha_t\right]p(y_t|Y_{t-1},\alpha_{t-1})p(\alpha_{t-1}|Y_{1:t-1}, X_{1:t-1})d\alpha_{t-1}.\end{aligned} \tag{A.17}$$

Each iteration of the filter will start with particles from $p(\alpha_{t-1}|Y_{1:t-1}, X_{1:t-1})$ and draw new particles from $p(\alpha_t|y_t, Y_{t-1}, \alpha_{t-1})$. The particle filter can be implemented using the following algorithm:

Algorithm 1 (Generic Particle Filter)

1. **Initialization.** *Draw the initial particles from the distribution* $\alpha_0^j \overset{iid}{\sim} p(\alpha_0)$ *and set* $W_0^j = 1$, $j = 1, \ldots, M$.

2. **Recursion.** *For* $t = 1, \ldots, T$*:*

   (a) **Forecasting** α_t**.** *Draw* $\tilde{\alpha}_t^j$ *from density* $g_t(\tilde{\alpha}_t|\alpha_{t-1}^j)$ *and define the importance weights*

   $$\omega_t^j = \frac{p(\tilde{\alpha}_t^j|y_t, Y_{t-1}, \alpha_{t-1}^j)}{g_t(\tilde{\alpha}_t^j|\alpha_{t-1}^j)}. \tag{A.18}$$

   (b) **Forecasting** y_t**.** *Define the incremental weights*

   $$\tilde{w}_t^j = p(X_t|\tilde{\alpha}_t^j)p(y_t|Y_{t-1}, \alpha_{t-1}^j)\omega_t^j. \tag{A.19}$$

   The predictive density $p(y_t, X_t|Y_{1:t-1}, X_{1:t-1}, \theta)$ *can be approximated by*

   $$\hat{p}(y_t, X_t|Y_{1:t-1}, X_{1:t-1}) = \frac{1}{M}\sum_{j=1}^{M} \tilde{w}_t^j W_{t-1}^j. \tag{A.20}$$

   (c) *Define the normalized weights*

   $$\tilde{W}_t^j = \frac{\tilde{w}_t^j W_{t-1}^j}{\frac{1}{M}\sum_{j=1}^{M} \tilde{w}_t^j W_{t-1}^j}. \tag{A.21}$$

   (d) **Selection.** *Resample the particles, for instance, via multinomial resampling. Let* $\{\alpha_t^j\}_{j=1}^M$ *denote* M *iid draws from a multinomial distribution characterized by support points and weights* $\{\tilde{\alpha}_t^j, \tilde{W}_t^j\}$ *and set* $W_t^j = 1$ *for* $j =, 1 \ldots, M$*. An approximation of* $\mathbb{E}[h(\alpha_t)|Y_{1:t}]$ *is given by* $\bar{h}_{t,M} = \frac{1}{M}\sum_{j=1}^{M} h(\alpha_t^j)W_t^j$.

3. **Likelihood Approximation.** *The approximation of the log-likelihood function is given by*

   $$\ln \hat{p}(Y_{1:T}, X_{1:T}|\theta) = \sum_{t=1}^{T} \ln\left(\frac{1}{M}\sum_{j=1}^{M} \tilde{w}_t^j W_{t-1}^j\right). \tag{A.22}$$

In order to evaluate the incremental weights $\tilde{\omega}_t^j$, one has to compute the densities $p(X_t|\tilde{\alpha}_t^j)$ and $p(y_t|Y_{t-1}, \alpha_{t-1}^j)$. The conditional log density of the cross-sectional observations is given by

$$\ln p(X_t|\tilde{\alpha}_t^j) = N\mathcal{L}(\tilde{\alpha}_t^j|X_t). \tag{A.23}$$

For the density of the aggregate observations define

$$y_{t|t-1}^j = \Phi_{yy} y_{t-1} + \Phi_{y\alpha}\alpha_{t-1}^j$$

such that

$$\ln p(y_t|Y_{t-1}, \alpha_{t-1}^j) = -\frac{n_y}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_{yy}| - \frac{1}{2}(y_t - y_{t|t-1}^j)'\Sigma_{yy}^{-1}(y_t - y_{t|t-1}'j).$$

Approximately Conditionally Optimal Proposal. The most important choice in the configuration of the algorithm is the proposal density $g_t(\tilde{\alpha}_t|\alpha_{t-1}^j)$. The conditional optimal choice is the conditional posterior

$$g_t^*(\tilde{\alpha}_t|\alpha_{t-1}^j) = p(\tilde{\alpha}_t|X_t, Y_t, \alpha_{t-1}^j, Y_{t-1}) \propto p(X_t|\tilde{\alpha}_t)p(\tilde{\alpha}_t|Y_t, Y_{t-1}, \alpha_{t-1}^j). \tag{A.24}$$

Because of the form in which $\tilde{\alpha}_t$ enters the density $p(X_t|\tilde{\alpha}_t)$ it is not possible to directly sample from posterior. Instead, we will construct a proposal density based on the linearized measurement equation

$$\hat{\alpha}_t(X_t) = \alpha_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \hat{V}_t).$$

Using a partitioned version of the state-transition equation, we define

$$\begin{aligned} Y_{t|t-1}^j &= \Phi_{yy}Y_{t-1} + \Phi_{y\alpha}\alpha_{t-1}^j \\ \alpha_{t|t-1}^j &= \Phi_{\alpha y}Y_{t-1} + \Phi_{\alpha\alpha}\alpha_{t-1}^j. \end{aligned}$$

Because conditional on s_{t-1}^j there is no uncertainty about the lagged state, uncertainty about the time t state is solely due to the innovation vector u_t. We denote the partitions of the innovation covariance matrix Σ that conform with the partitions Y_t and α_t by Σ_{ab}. We factor the joint distribution of (Y_t, α_t) into a marginal distribution of y_t and a conditional distribution of $\alpha_t|y_t$:

$$\begin{aligned} Y_t|(Y_{1:t-1}, \alpha_{t-1}^j) &\sim \mathcal{N}\big(Y_{t|t-1}^j, \Sigma_{yy}\big) \\ \alpha_t|(Y_t, Y_{1:t-1}, \alpha_{t-1}^j) &\sim \mathcal{N}\big(\alpha_{t|y,t-1}^j, P_{t|y,t-1}^*(\alpha\alpha)\big), \end{aligned} \tag{A.25}$$

where

$$\begin{aligned} \alpha_{t|y,t-1}^j &= \alpha_{t|t-1}^j + \Sigma_{\alpha y}\Sigma_{yy}(Y_t - Y_{t|t-1}^j) \\ P_{t|y,t-1}^*(\alpha\alpha) &= \Sigma_{\alpha\alpha} - \Sigma_{\alpha y}\Sigma_{yy}^{-1}\Sigma_{y\alpha}. \end{aligned}$$

The distribution of $\hat{\alpha}_t$ conditional on Y_t and s_{t-1}^j information is

$$\hat{\alpha}_t|(Y_t, Y_{1:t-1}, \hat{\alpha}_{1:t-1}) \sim N(\hat{\alpha}_{t|y,t-1}^j, F_{t|y,t-1}^*), \tag{A.26}$$

where

$$\begin{aligned}\hat{\alpha}^j_{t|y,t-1} &= \alpha^j_{t|y,t-1} \\ F^*_{t|y,t-1} &= P^*_{t|y,t-1}(\alpha\alpha) + V_t.\end{aligned}$$

We can use the standard Kalman filter updating formulas, substituting in the means and variances that are conditional on y_t:

$$\begin{aligned}\alpha^j_{t|t} &= \alpha^j_{t|y,t-1} + P^*_{t|y,t-1}(\alpha\alpha)\big[F^*_{t|y,t-1}\big]^{-1}(\hat{\alpha}_t - \hat{\alpha}^j_{t|y,t-1}) \\ P^*_{t|t}(\alpha\alpha) &= P^*_{t|y,t-1}(\alpha\alpha) - P^*_{t|y,t-1}(\alpha\alpha)\big[F^*_{t|y,t-1}\big]^{-1}P^*_{t|y,t-1}(\alpha\alpha).\end{aligned}$$

In sum, for each particle j, we sample

$$\tilde{\alpha}^j_t \sim \mathcal{N}\big(\alpha^j_{t|t}, P^*_{t|t}(\alpha\alpha)\big). \tag{A.27}$$

This leaves us with the computation of the importance weights ω^j_t in (A.18):

$$\begin{aligned}\ln p(\tilde{\alpha}^j_t|Y_t, Y_{t-1}, \alpha^j_{t-1}) &= -\frac{K}{2}\ln(2\pi) - \frac{1}{2}\ln|P^*_{t|y,t-1}(\alpha\alpha)| \\ &\quad -\frac{1}{2}(\tilde{\alpha}^j_t - \alpha^j_{t|y,t-1})'\big[P^*_{t|y,t-1}(\alpha\alpha)\big]^{-1}(\tilde{\alpha}^j_t - \alpha^j_{t|y,t-1}) \qquad \text{(A.28)} \\ \ln g_t(\tilde{\alpha}^j_t|\alpha^j_{t-1}) &= -\frac{K}{2}\ln(2\pi) - \frac{1}{2}\ln|P^*_{t|t}(\alpha\alpha)| \\ &\quad -\frac{1}{2}(\tilde{\alpha}^j_t - \alpha^j_{t|t})'\big[P^*_{t|t}(\alpha\alpha)\big]^{-1}(\tilde{\alpha}^j_t - \alpha^j_{t|t}). \qquad \text{(A.29)}\end{aligned}$$

A.6 Proof of Theorem 2

We subsequently use a notation that is slightly different from the main text. In particular, we will denote the observations by $D^N_{1:T} = (Y_{1:T}, X_{1:N,1:T})$ and instead of using the superscript (K) we will add K to the conditioning set of the various densities. Moreover, let

$$\begin{aligned}L_*(D^N_{1:T}|\theta, K) &= \prod_{t=1}^{T} p_G(Y_t, \alpha_t = \hat{\alpha}_t|Y_{t-1}, \alpha_{t-1} = \hat{\alpha}_{t-1}, \theta, K); \\ m_*(D^N_{1:T}|K, \lambda) &= \int L_*(D^N_{1:T}|\theta, K) f_T(\theta|\lambda) d\theta.\end{aligned}$$

Likewise define

$$\begin{aligned} L_{KF}(D^N_{1:T}|\theta,K) &= \prod_{t=1}^{T}\int p_G(\hat{\alpha}_t|\alpha_t)p_{KF}(Y_t,\alpha_t|D^N_{1:t-1},\theta,K)d\alpha_t; \\ m_{KF}(D^N_{1:T}|K,\lambda) &= \int L_{KF}(D^N_{1:T}|\theta,K)f_T(\theta|\lambda)d\theta. \end{aligned}$$

Then

$$\begin{aligned} \ln p_*(D^N_{1:T}|K,\lambda) &= \ln\left(\prod_{t=1}^{T}p_{pen}(X_t|\hat{\alpha}_t,K)\right)+\ln\left(m_*(D^N_{1:T}|K,\lambda)\right); \\ \ln p_{KF}(D^N_{1:T}|K,\lambda) &= \ln\left(\prod_{t=1}^{T}p_{pen}(X_t|\hat{\alpha}_t,K)\right)+\ln\left(m_{KF}(D^N_{1:T}|K,\lambda)\right). \end{aligned}$$

We bound the difference between $\ln p(D^N_{1:T}|K,\lambda)$ and $\ln p_*(D^N_{1:T}|K,\lambda)$ as follows:

$$\begin{aligned} &\left|\ln p(D^N_{1:T}|K,\lambda)-\ln p_*(D^N_{1:T}|K,\lambda)\right| \\ &\leq \left|\ln p(D^N_{1:T}|K,\lambda)-\ln p_{KF}(D^N_{1:T}|K,\lambda)\right|+\left|\ln p_{KF}(D^N_{1:T}|K,\lambda)-\ln p_*(D^N_{1:T}|K,\lambda)\right| \\ &= B_1(K,\lambda)+B_2(K,\lambda). \end{aligned} \tag{A.30}$$

In the subsequent steps we will bound $B_1(K,\lambda)$ and $B_2(K,\lambda)$.

Step 1. We construct a bound for $B_1(K,\lambda)$, starting from:

$$\begin{aligned} &\left|p(D^N_{1:T}|K,\lambda)-p_{KF}(D^N_{1:T}|K,\lambda)\right| \\ &\leq \int\left|\prod_{t=1}^{T}p(D^N_t|D^N_{1:t-1},\theta)-\prod_{t=1}^{T}p_{KF}(D^N_t|D^N_{1:t-1},\theta)\right|p(\theta|K,\lambda)d\theta \\ &= \int\left|\prod_{t=1}^{T}\frac{p(D^N_t|D^N_{1:t-1},\theta)}{p_{KF}(D^N_t|D^N_{1:t-1},\theta)}-1\right|\times\prod_{t=1}^{T}p_{KF}(D^N_t|D^N_{1:t-1},\theta)p(\theta|K,\lambda)d\theta \\ &\leq \max_\theta\left|\prod_{t=1}^{T}\frac{p(D^N_t|D^N_{1:t-1},\theta)}{p_{KF}(D^N_t|D^N_{1:t-1},\theta)}-1\right|\times\int\prod_{t=1}^{T}p_{KF}(D^N_t|D^N_{1:t-1},\theta)p(\theta|K,\lambda)d\theta \\ &= \max_\theta\left|\prod_{t=1}^{T}\frac{p(D^N_t|D^N_{1:t-1},\theta)}{p_{KF}(D^N_t|D^N_{1:t-1},\theta)}-1\right|\times p_{KF}(D^N_{1:T}|K,\lambda) \\ &\leq C\sqrt{\frac{\overline{K}^3}{N}}p_{KF}(D^N_{1:T}|K,\lambda), \end{aligned}$$

where the last inequality follows from Theorem 1. Thus, we obtain uniformly over (K, λ),

$$\left| \frac{p(D_{1:T}^N | K, \lambda)}{p_{KF}(D_{1:T}^N | K, \lambda)} - 1 \right| \leq C \sqrt{\frac{\overline{K}_N^3}{N}}.$$

We deduce that

$$\max_{K,\lambda} \; B_1(K, \lambda) \lesssim \sqrt{\frac{\overline{K}_N^3}{N}} = o(1). \tag{A.31}$$

Step 2. Before bounding $B_2(K, \lambda)$ we conduct some intermediate calculations. Recall that the KF likelihood is constructed from Gaussian densities of the form $\mathcal{N}(\mu_t^{KF}(\theta), \Sigma_t^{KF}(\theta))$. Write the KF log likelihood as

$$\begin{aligned}
&\ln L_{KF}(D_{1:T}^N | \theta, K) \qquad\qquad (A.32)\\
&= -\frac{T(n_y + K)}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \ln \left| \Sigma_t^{KF}(\theta) \right| \\
&\quad -\frac{1}{2} \sum_{t=1}^{T} (D_t^N - \mu_t^{KF}(\theta))' (\Sigma_t^{KF}(\theta))^{-1} (D_t^N - \mu_t^{KF}(\theta)) \\
&= -\frac{T(n_y + K)}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \ln |\Sigma^*(\theta) + N^{-1} \Delta_t^{\Sigma}(\theta)| \\
&\quad -\frac{1}{2} \sum_{t=1}^{T} \left(D_t^N - \mu_t^*(\theta) - N^{-1} \Delta_t^{\mu}(\theta) \right)' \left(\Sigma^*(\theta) + N^{-1} \Delta_t^{\Sigma}(\theta) \right)^{-1} \left(D_t^N - \mu_t^*(\theta) - N^{-1} \Delta_t^{\mu}(\theta) \right) \\
&= \ln L_*(D_{1:T}^N | \theta, K) + \sum_{t=1}^{T} \frac{F_t(\theta)}{2N},
\end{aligned}$$

with $F_t(\theta)/N$ defined as

$$\begin{aligned}
\frac{F_t(\theta)}{N} &= \left(\ln |\Sigma^*(\theta)| - \ln |\Sigma^*(\theta) + N^{-1} \Delta_t^{\Sigma}(\theta)| \right) \\
&\quad + \Bigg(\left(D_t^N - \mu_t^*(\theta) \right)' \left(\Sigma^*(\theta) \right)^{-1} \left(D_t^N - \mu_t^*(\theta) \right) \\
&\quad - \left(D_t^N - \mu_t^*(\theta) - N^{-1} \Delta_t^{\mu}(\theta) \right)' \left(\Sigma^*(\theta) + N^{-1} \Delta_t^{\Sigma}(\theta) \right)^{-1} \left(D_t^N - \mu_t^*(\theta) - N^{-1} \Delta_t^{\mu}(\theta) \right) \Bigg)
\end{aligned}$$

Notice that $F_t(\theta)$ is the sum of log determinants of $(n_y + k) \times (n_y + k)$ matrices and the inner products of a $(n_y + k)$ vectors. Applying Lemma 2 (see below), we have uniformly in $K \leq \overline{K} = o(N^{1/3})$,

$$\sup_{\theta \in \Theta(K), t} \left| \frac{F_t(\theta)}{N} \right| = O\left(\frac{K}{N} \right). \tag{A.33}$$

Step 3. Now consider

$$\begin{aligned} m_{KF}(D^N_{1:T}|K,\lambda) &= \int L_{KF}(D^N_{1:T}|\theta,K) f_T(\theta|\lambda) d\theta \\ &= \int \exp\left\{ \ln L_*(D^N_{1:T}|\theta,K) + \sum_{t=1}^{T} \frac{F_t(\theta)}{2N} \right\} f_T(\theta|\lambda) d\theta \\ &= \int \frac{L_*(D^N_{1:T}|\theta,K) f_T(\theta|\lambda)}{f(\theta|D^N_{1:T},\lambda)} \exp\left\{ \sum_{t=1}^{T} \frac{F_t(\theta)}{2N} \right\} f(\theta|D^N_{1:T},\lambda) d\theta \\ &= \int m_*(D^N_{1:T}|K,\lambda) \exp\left\{ \sum_{t=1}^{T} \frac{F_t(\theta)}{2N} \right\} f(\theta|D^N_{1:T},\lambda) d\theta. \end{aligned} \tag{A.34}$$

Taking logs, we obtain

$$\begin{aligned} &\ln(m_{KF}(D^N_{1:T}|K,\lambda)) \\ &= \ln(m_*(D^N_{1:T}|K,\lambda)) + \ln\left[\int \exp\left\{ \sum_{t=1}^{T} \frac{F_t(\theta)}{2N} \right\} f(\theta|D^N_{1:T},\lambda) d\theta \right] \end{aligned} \tag{A.35}$$

Thus,

$$\begin{aligned} B_2(K,\lambda) &= \left| \ln p_{KF}(D^N_{1:T}|K,\lambda) - \ln p_*(D^N_{1:T}|K,\lambda) \right| \\ &= \left| \ln\left[\int \exp\left\{ \sum_{t=1}^{T} \frac{F_t(\theta)}{2N} \right\} f(\theta|D^N_{1:T},K,\lambda) d\theta \right] \right| \\ &= \left| \ln\left(\mathbb{E}^\theta_* \left[\exp\left\{ \sum_{t=1}^{T} \frac{F_t(\theta)}{2N} \right\} \;\middle|\; D^N_{1:T},K,\lambda \right] \right) \right| \end{aligned}$$

where $\mathbb{E}^\theta_*[\cdot|D^N_{1:T},K,\lambda]$ is the expectation associated with the posterior density $\theta|D^N_{1:T} \sim f(\theta|D^N_{1:T},K,\lambda)$.

We proceed by expanding the exponential function of the $L_*(D|\theta,K)$ likelihood around zero (assuming that $T\overline{K}_N/N = o(1)$):

$$\exp(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \ldots.$$

Thus,

$$\exp\left\{ \frac{T}{2N}\frac{1}{T}\sum_{t=1}^{T} F_t(\theta) \right\} = 1 + \frac{T}{2N}\frac{1}{T}\sum_{t=1}^{T} F_t(\theta) + \frac{1}{2}\left(\frac{T}{2N}\frac{1}{T}\sum_{t=1}^{T} F_t(\theta) \right)^2 + \ldots.$$

In turn,

$$\begin{aligned}\mathbb{E}_*^\theta\left[\exp\left\{\sum_{t=1}^T \frac{F_t(\theta)}{2N}\right\} \;\middle|\; D_{1:T}^N, K, \lambda\right] &= 1 + \frac{T}{2N}\frac{1}{T}\sum_{t=1}^T \mathbb{E}_*^\theta\big[F_t(\theta)|D_{1:T}^N, K, \lambda\big] + \ldots, ; \\ \ln\left(\mathbb{E}_*^\theta\left[\exp\left\{\sum_{t=1}^T \frac{F_t(\theta)}{2N}\right\} \;\middle|\; D_{1:T}^N, \lambda\right]\right) &= \frac{T}{2N}\frac{1}{T}\sum_{t=1}^T \mathbb{E}_*^\theta\big[F_t(\theta)|D_{1:T}^N, K, \lambda\big] + \ldots\end{aligned}$$

Using (A.33) and the rate Assumption in the theorem, we deduce that

$$\begin{aligned}\max_{(K,\lambda)} B_2(K,\lambda) &= \max_{(K,\lambda)\in\mathcal{K}_N\times\Omega}\left|\frac{T}{2N}\frac{1}{T}\sum_{t=1}^T \mathbb{E}_*^\theta\big[F_t(\theta)|D_{1:T}^N, \lambda\big] + \ldots\right| \\ &\lesssim \frac{T}{2N}\overline{K}_N = o(1).\end{aligned} \tag{A.36}$$

Step 4. We can now combine (A.31) and (A.36) which leads to the final bound on the MDD discrepancy:

$$\begin{aligned}&\max_{(K,\lambda)\in\mathcal{K}_N\times\Omega}\left|\frac{\ln p(D_{1:T}^N|K,\lambda) - \ln p_*(D_{1:T}^N|K,\lambda)}{NT}\right| \\ &\leq \max_{(K,\lambda)\in\mathcal{K}_N\times\Omega}\frac{1}{NT}B_1(K,\lambda) + \max_{(K,\lambda)\in\mathcal{K}_N\times\Omega}\frac{1}{NT}B_2(K,\lambda) \\ &\lesssim \frac{1}{NT}\left(\sqrt{\frac{\overline{K}_N^3}{N}} + \frac{T}{2N}\overline{K}_N\right). \quad \blacksquare\end{aligned} \tag{A.37}$$

Lemma 2 *The likelihood increments for the Kalman filter and the VAR approximation take the following form*

$$\begin{aligned}L_*(D_{1:T}^N|\theta,K) &: D_t^N|(D_{1:t-1}^N,\theta,K) \sim \mathcal{N}\big(\mu_t^*(\theta),\Sigma^*(\theta)\big) \\ L_{KF}(D_{1:T}^N|\theta,K) &: D_t^N|(D_{1:t-1}^N,\theta,K) \sim \mathcal{N}\big(\mu_t^{KF}(\theta),\Sigma_t^{KF}(\theta)\big),\end{aligned} \tag{A.38}$$

where

$$\mu_t^{KF}(\theta) = \mu_t^*(\theta) + \frac{1}{N}\Delta_t^\mu(\theta), \quad \Sigma_t^{KF}(\theta) = \Sigma^*(\theta) + \frac{1}{N}\Delta_t^\Sigma(\theta).$$

Hence

$$\sup_{\theta\in\Theta(K),t}|F_t(\theta)| = O(K).$$

Proof of Lemma 2. For the KF likelihood it can be verified that uniformly over time t,

$$P_{t-1|t-1}(\alpha\alpha) = O(1/N), \quad \alpha_{t-1|t-1} - \hat{\alpha}_{t-1} = O(1/N).$$

Define $\bar{P}_{t-1|t-1} = NP_{t-1|t-1}(\alpha\alpha)$ to make the dependence on the cross-sectional sample size clear. Let

$$\left[\begin{array}{c} Y_t \\ \alpha_t \end{array}\right] \Bigg| (Y_{1:t-1}, X_{1:t-1}) \sim \mathcal{N}\left(\left[\begin{array}{c} Y_{t|t-1} \\ \alpha_{t|t-1} \end{array}\right], \left[\begin{array}{cc} P_{t|t-1}(yy) & P_{t|t-1}(y\alpha) \\ P_{t|t-1}(\alpha y) & P_{t|t-1}(\alpha\alpha) \end{array}\right]\right), \tag{A.39}$$

where

$$\begin{aligned}
Y_{t|t-1} &= \Phi_{yy}Y_{t-1} + \Phi_{y\alpha}\hat{\alpha}_{t-1} + \frac{1}{N}\Phi_{y\alpha}N(\alpha_{t-1|t-1} - \hat{\alpha}_{t-1}) \\
\alpha_{t|t-1} &= \Phi_{\alpha y}Y_{t-1} + \Phi_{\alpha\alpha}\hat{\alpha}_{t-1} + \frac{1}{N}\Phi_{\alpha\alpha}N(\alpha_{t-1|t-1} - \hat{\alpha}_{t-1}) \\
P_{t|t-1}(yy) &= \Sigma_{yy} + \frac{1}{N}\Phi_{y\alpha}\bar{P}_{t-1|t-1}(\alpha\alpha)\Phi'_{y\alpha} \\
P_{t|t-1}(\alpha y) &= \Sigma_{\alpha y} + \frac{1}{N}\Phi_{\alpha\alpha}\bar{P}_{t-1|t-1}(\alpha\alpha)\Phi'_{y\alpha} \\
P_{t|t-1}(\alpha\alpha) &= \Sigma_{\alpha\alpha} + \frac{1}{N}\Phi_{\alpha\alpha}\bar{P}_{t-1|t-1}(\alpha\alpha)\Phi'_{\alpha\alpha}.
\end{aligned}$$

We can now factorize the joint distribution of (Y_t, α_t) into a marginal distribution of Y_t and a conditional distribution of $\alpha_t|Y_t$:

$$Y_t|(Y_{1:t-1}, X_{1:t-1}) \sim \mathcal{N}\big(Y_{t|t-1}, P_{t|t-1}(yy)\big) \tag{A.40}$$

$$\alpha_t|(Y_{1:t}, X_{1:t-1}) \sim \mathcal{N}\big(\alpha_{t|y,t-1}, P_{t|y,t-1}(\alpha\alpha)\big), \tag{A.41}$$

where

$$\begin{aligned}
\alpha_{t|y,t-1} &= \alpha_{t|t-1} + P_{t|t-1}(\alpha y)[P_{t|t-1}(yy)]^{-1}(Y_t - Y_{t|t-1}) \\
P_{t|y,t-1}(\alpha\alpha) &= P_{t|t-1}(\alpha\alpha) - P_{t|t-1}(\alpha y)[P_{t|t-1}(yy)]^{-1}P_{t|t-1}(y\alpha).
\end{aligned}$$

Finally, the distribution of $\hat{\alpha}_t$ conditional on Y_t and $t-1$ information is

$$\hat{\alpha}_t|(Y_{1:t}, X_{1:t-1}) \sim \mathcal{N}(\hat{\alpha}_{t|y,t-1}, F_{t|y,t-1}), \tag{A.42}$$

where

$$\begin{aligned}
\hat{\alpha}_{t|y,t-1} &= \alpha_{t|y,t-1} \\
F_{t|y,t-1} &= P_{t|y,t-1}(\alpha\alpha) + \frac{1}{N}\hat{V}_t.
\end{aligned}$$

Thus, the time-t increment of the KF likelihood $L_{KF}(D^N_{1:T}|\theta, K)$ is given by

$$Y_t|(Y_{1:t-1}, X_{1:t-1}) \sim \mathcal{N}\big(Y_{t|t-1}, P_{t|t-1}(yy)\big), \quad \hat{\alpha}_t|(Y_{1:t}, X_{1:t-1}) \sim \mathcal{N}(\hat{\alpha}_{t|y,t-1}, F_{t|y,t-1}). \tag{A.43}$$

We now turn to $L_*(D_{1:T}^N|\theta,K)$ which ignores the uncertainty about the latent state α_t and equates it with $\hat{\alpha}_t$, setting $P_{t|t}(\alpha\alpha)=0$. Denoting the resulting moments with a $*$ superscript, we can define

$$\begin{aligned}
Y_{t|t-1}^* &= \Phi_{yy}Y_{t-1}+\Phi_{y\alpha}\hat{\alpha}_{t-1}\\
\alpha_{t|t-1}^* &= \Phi_{\alpha y}Y_{t-1}+\Phi_{\alpha\alpha}\hat{\alpha}_{t-1}\\
P_{t|t-1}^*(yy) &= \Sigma_{yy}\\
P_{t|t-1}^*(\alpha y) &= \Sigma_{\alpha y}\\
P_{t|t-1}^*(\alpha\alpha) &= \Sigma_{\alpha\alpha}\\
\alpha_{t|y,t-1}^* &= \alpha_{t|t-1}^*+\Sigma_{\alpha y}\Sigma_{yy}^{-1}(Y_t-Y_{t|t-1}^*)\\
P_{t|y,t-1}^*(\alpha\alpha) &= \Sigma_{\alpha\alpha}-\Sigma_{\alpha y}\Sigma_{yy}^{-1}\Sigma_{y\alpha}\\
\hat{\alpha}_{t|y,t-1}^* &= \Phi_{\alpha y}Y_{t-1}+\Phi_{\alpha\alpha}\hat{\alpha}_{t-1}+\Sigma_{\alpha y}\Sigma_{yy}^{-1}\big(Y_t-[\Phi_{yy}Y_{t-1}+\Phi_{y\alpha}\hat{\alpha}_{t-1}]\big)\\
F_{t|y,t-1}^* &= \Sigma_{\alpha\alpha}-\Sigma_{\alpha y}\Sigma_{yy}^{-1}\Sigma_{y\alpha}.
\end{aligned}$$

It can be shown that the $*$ moments differ from the KF moments by a term that is $O(1/N)$ uniformly in t, i.e.,

$$\begin{aligned}
Y_{t|t-1} &= Y_{t|t-1}^*+O(1/N)\\
P_{t|t-1}(yy) &= P_{t|t-1}^*(yy)+O(1/N)\\
\hat{\alpha}_{t|y,t-1} &= \hat{\alpha}_{t|y,t-1}^*+O(1/N)\\
F_{t|y,t-1} &= F_{t|y,t-1}^*+O(1/N).
\end{aligned}$$

By the definition of $F_t(\theta)$ in the main text and the results above, we obtain uniformly over θ,t,K

$$\begin{aligned}
F_t(\theta) &= \operatorname{tr}\left([\Sigma^*(\theta)]^{-1}\Delta_t^{\Sigma}(\theta)\right)+2\big(D_t^N-\mu_t^*(\theta)\big)'\big(\Sigma^*(\theta)\big)^{-1}\Delta_t^{\mu}(\theta)\\
&\quad+\big(D_t^N-\mu_t^*(\theta)\big)'\delta_t^{\Sigma}(\theta)\big(D_t^N-\mu_t^*(\theta)\big)+o(1)
\end{aligned}\tag{A.44}$$

with

$$\delta_t^{\Sigma}(\theta)=-[\Sigma^*(\theta)]^{-1}\Delta_t^{\Sigma}(\theta)[\Sigma^*(\theta)]^{-1}.$$

Thus $\sup_{\theta\in\Theta(K),t}|F_t(\theta)|=O(K)$. ■

A.7 Hyperparameter Determination

The subsequent statements about hyperparameter selection based on the exact log MDD $\ln p(D_{1:T}^N|K,\omega)$ and the approximate MDD $\ln p_*(D_{1:T}^N|K,\omega)$ require an assumption about the

data generating process (DGP). We follow the steps outlined in Section A.1 to specify the DGP which is represented by the probability measure $\mathbb{P}^o$. Associated with this probability measure are pseudo-optimal parameters (Φ_o, Σ_o) which we stack in the vector $\theta_o^{(K)}$.

In finite samples, it is beneficial to use a Bayes estimator for the VAR that shrinks the distance between the MLE and a prior mean, which we describe in detail in the paper. To capture the benefit of shrinkage in an asymptotic analysis in which $T \longrightarrow \infty$, we need to assume that the prior mean is not "too far" from the true parameter $\theta_o^{(K)}$. We do so in a drifting coefficients framework in which the prior variance of θ shrinks at rate $1/T$ and the prior mean $\underline{\theta}_T^{(K)}$ drifts so that it stays within a $1/\sqrt{T}$ radius of $\theta_o^{(K)}$. This setup ensures that the prior probability assigned to neighborhoods around $\theta_o^{(K)}$ does not vanish asymptotically. Recall from Section A.1 the definition of the set of integers for the cross-sectional sieve dimension:

$$\mathcal{K}_N \equiv \left\{ K \in \mathbb{N} : \underline{K} \leq K \leq \overline{K},\ \underline{K} = \underline{c} \times [N \ln N]^{\frac{1}{1+2s}},\ \overline{K} = \overline{c} \times \frac{N^{1/3}}{(\ln N)^2},\ 0 < \underline{c}, \overline{c} < \infty \right\}.$$

Let Ω be a grid for the localized hyperparameter ω. To simplify the subsequent calculations, we will make the following additional assumptions:

Assumption 2 *For every $K \in \mathcal{K}_N$ the following conditions are satisfied:*
(i) the log likelihood function is exactly quadratic and takes the form

$$\ln L_*(D_{1:T}^N|\theta, K) = \ln L_*(D_{1:T}^N|\hat{\theta}_*^{(K)}, K) - \frac{T}{2}(\theta - \hat{\theta}_*^{(K)})'\hat{\mathcal{V}}_*^{-1}(\theta - \hat{\theta}_*^{(K)}).$$

(ii) The prior distribution is Normal and takes the form of a g-prior; see Zellner (1983):

$$\theta|K, \omega \sim \mathcal{N}\left(\underline{\theta}_T^{(K)}, \frac{\omega}{T}\hat{\mathcal{V}}_*\right), \quad \textit{i.e.,} \quad \underline{\mathcal{V}} = \hat{\mathcal{V}}_*,$$

where the hyperparameter $\omega \in \Omega$ and the prior density is denoted by $f_T(\theta|\omega)$.
(iii) Under the DGP $\mathbb{P}^o$: $T\big(\hat{\theta}_*^{(K)} - \underline{\theta}_T^{(K)}\big)'\hat{\mathcal{V}}_*^{-1}\big(\hat{\theta}_*^{(K)} - \underline{\theta}_T^{(K)}\big) = O_p(K^2)$.

Assumption 2 would be satisfied for Φ in a VAR, but not for Σ in finite samples. However, it could be asymptotically justified. By assuming that it holds for any (T, K) we avoid having to keep track of the remainder term. The Gaussian prior for θ specified in (ii) ensures that the posterior will also be Gaussian. The use of the g-prior (originally called reference informative prior (RIP)) will lead to a convenient MDD formula from which one can derive an analytical formula for the optimal hyperparameter choice. Finally, Assumption 2 is a

high-level assumption about the convergence rate of the MLE that can be derived from low-level assumptions on the DGP. Underlying the assumption is that the MLE $\hat{\theta}_*^{(K)}$ is consistent for the pseudo-optimal $\theta_o^{(K)}$ and that the prior mean drifts toward $\theta_o^{(K)}$ (see above).

Now define

$$(K_*, \omega_*) = \text{argmax}_{(K,\omega)\in\mathcal{K}_N\times\Omega} \ln p_*(D_{1:T}^N|K,\omega), \tag{A.45}$$
$$(\widehat{K}, \widehat{\omega}) = \text{argmax}_{(K,\omega)\in\mathcal{K}_N\times\Omega} \ln p(D_{1:T}^N|K,\omega), \tag{A.46}$$

and for all $K \in \mathcal{K}_N$,

$$\omega_*^o(K) = \text{argmax}_{\omega\in\Omega} \ln p_*(D_{1:T}^N|K,\omega), \tag{A.47}$$
$$\widehat{\omega}^o(K) = \text{argmax}_{\omega\in\Omega} \ln p(D_{1:T}^N|K,\omega). \tag{A.48}$$

Theorem 3 *Suppose Assumptions 1 and 2 are satisfied and* $T\bar{K}_N/N = o(1)$. *Let* $\overline{K}_N = \bar{c} \times \frac{N^{1/3}}{(\ln N)^2}$ *be an integer. Let* $T\overline{K}_N/N = o(1)$ *and* $\overline{K}_N/T = o(1)$. *Then, for any* $\varepsilon > 0$, *for all* $K \in \mathcal{K}_N$,

$$\mathbb{P}^o\Big\{ |\omega_*^o(K) - \widehat{\omega}^o(K)| < \varepsilon \Big\} \longrightarrow 1$$

as $(T, N) \longrightarrow \infty$.

Proof of Theorem 3. To simplify the notation we will drop (K) superscripts. We first introduce extra notation. Let

$$m_*(D_{1:T}^N|K,\omega) = \int L_*(D_{1:T}^N|\theta,K) f_T(\theta|\omega) d\theta$$
$$L_*(D_{1:T}^N|\theta,K) = \prod_{t=1}^T p_G(Y_t, \alpha_t = \hat{\alpha}_t | Y_{t-1}, \alpha_{t-1} = \hat{\alpha}_{t-1}, \theta, K).$$

Then

$$\ln p_*(D_{1:T}^N|K,\omega) = \sum_{t=1}^T \ln p_{pen}(X_t|\hat{\alpha}_t, K) + \ln m_*(D_{1:T}^N|K,\omega). \tag{A.49}$$

Step 1. We begin by analyzing $\ln p_*(D_{1:T}^N|K,\omega)$. The first term in (A.49), $\sum_{t=1}^T \ln p_{pen}(X_t|\hat{\alpha}_t, K)$, depends on K only and does not depend on ω at all. The second term, $\ln m_*(D_{1:T}^N|K,\omega)$, depends on ω explicitly, and depends on K indirectly through the data D_t^N dimension $(n_y + K)$, and through the dimension of the θ parameter $d \equiv \dim(\theta) = O(K^2)$. For any fixed K, we compute

$$\omega_*^o(K) = \text{argmax}_{\omega\in\Omega} \ln p_*(D_{1:T}^N|K,\omega) = \text{argmax}_{\omega\in\Omega} \ln m_*(D_{1:T}^N|K,\omega). \tag{A.50}$$

The term $m_*(D_{1:T}^N|K,\omega)$ has two components: the likelihood function $L_*(D_{1:T}^N|\theta,K)$, and the Gaussian priors $f_T(\theta|\omega)$. Because we are using a drifting sequence of priors both components will concentrate as T goes to infinity: the likelihood function $L_*(D_{1:T}^N|\theta,K)$ will concentrate around its MLE $\hat{\theta}_*$ and the prior will concentrate around the prior mode $\underline{\theta}_T$. Under Assumption 2(i) and (ii) the posterior density for θ takes the form

$$f(\theta|D_{1:T}^N,\omega) = (2\pi)^{-d/2}\left|\frac{\bar{\mathcal{V}}_*}{T}\right|^{-1/2}\exp\left\{-\frac{T}{2}(\theta-\bar{\theta}_{T*})'[\bar{\mathcal{V}}_*]^{-1}(\theta-\bar{\theta}_{T*})\right\}, \tag{A.51}$$

where

$$\bar{\mathcal{V}}_* = \frac{\omega}{\omega+1}\hat{\mathcal{V}}_*, \quad \bar{\theta}_{T*} = \frac{\omega}{\omega+1}\hat{\theta}_* + \frac{1}{\omega+1}\underline{\theta}_T. \tag{A.52}$$

Rewriting Bayes Theorem, we can express the log MDD as

$$\begin{aligned}\ln m_*(D_{1:T}^N|K,\omega) &= \ln L_*(D_{1:T}^N|\theta,K) + \ln f_T(\theta|\omega) - \ln f(\theta|D_{1:T}^N,\omega) \\ &= \ln L_*(D_{1:T}^N|\hat{\theta}_*,K) - \frac{1}{2}\ln\left|\frac{\omega\underline{\mathcal{V}}}{T}\right| + \frac{1}{2}\ln\left|\frac{\bar{\mathcal{V}}_*}{T}\right| \\ &\quad -\frac{T}{2}\left(\hat{\theta}_*'\hat{\mathcal{V}}_*^{-1}\hat{\theta}_* + \frac{1}{\omega}\underline{\theta}_T'\underline{\mathcal{V}}^{-1}\underline{\theta}_T - \bar{\theta}_{T*}'\bar{\mathcal{V}}_*^{-1}\bar{\theta}_{T*}\right).\end{aligned} \tag{A.53}$$

Using (A.52) we can simplify the log MDD formula as follows:

$$\begin{aligned}&\ln m_*(D_{1:T}|K,\omega) \\ &= \ln L_*(D_{1:T}|\hat{\theta}_*,K) - \frac{T}{2}\frac{1}{1+\omega}(\hat{\theta}_{T*}-\underline{\theta}_T)'\hat{\mathcal{V}}_*^{-1}(\hat{\theta}_{T*}-\underline{\theta}_T) - \frac{d}{2}\ln|1+\omega|,\end{aligned} \tag{A.54}$$

where d is the dimension of θ. It can be verified that the MDD is maximized by

$$\omega_*^o(K) = \max\left\{\frac{T(\hat{\theta}_{T*}-\underline{\theta}_T)'\hat{\mathcal{V}}_*^{-1}(\hat{\theta}_{T*}-\underline{\theta})}{d} - 1,\ 0\right\}. \tag{A.55}$$

Step 2. From Theorem 2 we have that

$$\begin{aligned}&\max_{(K,\omega)\in\mathcal{K}_N\times\Omega}\left|\ln p(D_{1:T}^N|K,\omega) - \ln p_*(D_{1:T}^N|K,\omega)\right| \\ &\lesssim \left(\sqrt{\frac{\overline{K}_N^3}{N}} + \frac{T}{2N}\overline{K}_N\right) \equiv \eta_{NT} = o(1).\end{aligned} \tag{A.56}$$

We want to show that uniformly over $K\in\mathcal{K}_N$, for any $\varepsilon>0$,

$$\mathbb{P}^o\Big\{\,|\omega_*^o(K)-\widehat{\omega}^o(K)|<\varepsilon\Big\}\longrightarrow 1. \tag{A.57}$$

Note that:

$$
\begin{aligned}
&\mathbb{P}^o\left(|\omega_*^o(K)-\widehat{\omega}^o(K)|\geq\varepsilon,\ \forall K\in\mathcal{K}_N\right)\\
&\quad=_{(1)}\ \mathbb{P}^o\left(0<\ln p(D_{1:T}^N|K,\widehat{\omega}^o(K))-\ln p(D_{1:T}^N|K,\omega_*^o(K)),\ \forall K\in\mathcal{K}_N\right)\\
&\quad\leq_{(2)}\ \mathbb{P}^o\Bigg(0<\ln p_*(D_{1:T}^N|K,\widehat{\omega}^o(K))-\ln p_*(D_{1:T}^N|K,\omega_*^o(K))\\
&\qquad\quad+2\max_{\omega\in\Omega}\left|\ln p(D_{1:T}^N|K,\omega)-\ln p_*(D_{1:T}^N|K,\omega)\right|,\ \forall K\in\mathcal{K}_N\Bigg)\\
&\quad=_{(3)}\ \mathbb{P}^o\Bigg(\ln p_*(D_{1:T}^N|K,\omega_*^o(K))-\ln p_*(D_{1:T}^N|K,\widehat{\omega}^o(K))\\
&\qquad<\quad 2\max_{\omega\in\Omega}\left|\ln p(D_{1:T}^N|K,\omega)-\ln p_*(D_{1:T}^N|K,\omega)\right|,\ \forall K\in\mathcal{K}_N\Bigg)\\
&\quad\leq_{(4)}\ \mathbb{P}^o\left(0<\ln p_*(D_{1:T}^N|K,\omega_*^o(K))-\max_{\omega\in\Omega,|\omega^o(K)-\omega|\geq\varepsilon}\ln p_*(D_{1:T}^N|K,\omega)\lesssim 2\eta_{NT},\ \forall K\in\mathcal{K}_N\right).
\end{aligned}
$$

The steps can be justified as follows. If the left-hand side has to hold for all $\epsilon>0$, then (1) follows because the MDD evaluated at its optimizer is greater that the MDD evaluated at the alternative value $\omega_*^o(K)$. (2) We replace the $\ln p$ MDD by the $\ln p_*$ MDD using a bound on the approximation error. (3) re-arranges terms. (4) We use the bound in (A.56) to remove the max expression from (3) and we replace $\ln p_*(D_{1:T}^N|K,\widehat{\omega}^o(K))$ by the maximum over ω. We proceed as follows:

$$
\begin{aligned}
0\ &<\ 2\left(\ln p_*(D_{1:T}^N|K,\omega_*^o(K))-\max_{\omega\in\Omega,|\omega_*^o(K)-\omega|\geq\varepsilon}\ln p_*(D_{1:T}^N|K,\omega)\right)\\
&=\ 2\left(\ln m_*(D_{1:T}^N|K,\omega_*^o(K))-\max_{\omega\in\Omega,|\omega_*^o(K)-\omega|\geq\varepsilon}\ln m_*(D_{1:T}^N|K,\omega)\right)\\
&=\ -\left(\frac{1}{1+\omega_*^o(K)}T(\hat{\theta}_{T*}-\underline{\theta}_T)'\hat{\mathcal{V}}_*^{-1}(\hat{\theta}_{T*}-\underline{\theta}_T)+d\ln|1+\omega_*^o(K)|\right)\\
&\quad-\left\{\max_{\omega\in\Omega,|\omega_*^o(K)-\omega|\geq\varepsilon}-\left(\frac{1}{1+\omega}T(\hat{\theta}_{T*}-\underline{\theta}_T)'\hat{\mathcal{V}}_*^{-1}(\hat{\theta}_{T*}-\underline{\theta}_T)+d\ln|1+\omega|\right)\right\}\\
&=\ \left\{\min_{\omega\in\Omega,|\omega_*^o(K)-\omega|\geq\varepsilon}\left(\frac{1}{1+\omega}T(\hat{\theta}_{T*}-\underline{\theta}_T)'\hat{\mathcal{V}}_*^{-1}(\hat{\theta}_{T*}-\underline{\theta}_T)+d\ln|1+\omega|\right)\right\}\\
&\quad-\left(\frac{1}{1+\omega_*^o(K)}T(\hat{\theta}_{T*}-\underline{\theta}_T)'\hat{\mathcal{V}}_*^{-1}(\hat{\theta}_{T*}-\underline{\theta}_T)+d\ln|1+\omega_*^o(K)|\right)\\
&\equiv\ g_1(N,T,\varepsilon)
\end{aligned}
$$

Now consider

$$
\mathbb{P}^o\big(0<g_1(N,T,\varepsilon)\lesssim 4\eta_{NT},\ \forall K\in\mathcal{K}_N\big). \tag{A.58}
$$

Because $(\hat{\theta}_{T*}-\underline{\theta}_T)'\hat{\mathcal{V}}_*^{-1}(\hat{\theta}_{T*}-\underline{\theta}_T)$ is $O_p(K^2)$ or it is diverging, whereas $\eta_{NT}\longrightarrow 0$, we deduce

that

$$\frac{\eta_{NT}}{g_1(N,T,\varepsilon)} = o_p(1),$$

hence

$$\mathbb{P}^o\left(|\omega_*^o(K) - \widehat{\omega}^o(K)| \geq \varepsilon\right) \longrightarrow 0. \quad \blacksquare$$

B Implementation Details

B.1 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

$$\bar{\zeta}_t(X_t) = \frac{1}{N_t}\sum_{i=1}^{N_t}\zeta(x_{it})\mathbb{I}\{x_{it} < c_t\}. \tag{A.59}$$

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

$$\begin{aligned}\mathcal{L}^{(K)}(\alpha_t, \pi_t|X_t) &= \frac{N_{t,max}}{N_t}\ln\pi + \frac{N_t - N_{t,max}}{N_t}\ln(1-\pi_t) \\ &\quad + \bar{\zeta}_t'(X_t)\alpha_t - \frac{N_t - N_{t,max}}{N_t}\ln\int_0^{c_t}\exp\left\{\zeta'(x)\alpha_t\right\}dx.\end{aligned} \tag{A.60}$$

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

$$\hat{\pi}_t = \text{argmax}_{\pi\in[0,1]}\, \mathcal{L}^{(K)}(\alpha_t, \pi_t|X_t) = N_{t,max}/N_t. \tag{A.61}$$

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

$$\begin{aligned}\hat{\alpha}_t &= \text{argmax}_{\alpha_t}\, \mathcal{L}^{(K)}(\alpha_t, \pi_t|X_t) \\ &= \text{argmax}_{\alpha_t}\, \bar{\zeta}_t'(X_t)\alpha_t - \frac{N_t - N_{t,max}}{N_t} \ln \int_0^{c_t} \exp\left\{\zeta'(x)\alpha_t\right\} dx.\end{aligned} \tag{A.62}$$

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

$$p(x|\alpha_t) = \frac{\exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\}}{\int_0^\infty \exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\} dx}. \tag{A.63}$$

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

$$\begin{aligned}&\mathcal{L}_{kl}^{(2)}(\alpha_t|\pi_t, X_t) \\ &= -\left(\frac{N_t - N_{t,max}}{N_t}\right) \int_0^{c_t} \left(\zeta_k(x) - \int_0^{c_t} \zeta_k(x)\bar{p}(x|\alpha_t)dx\right) \\ &\qquad \times \left(\zeta_l(x) - \int_0^{c_t} \zeta_l(x)\bar{p}(x|\alpha_t)dx\right) \bar{p}(x|\alpha_t)dx.\end{aligned} \tag{A.64}$$

Thus, compared to the standard case in Section A.2, the limits of integration change and there is an additional factor $(N_t - N_{t,max})/N_t$.

B.2 Transformations of the $\hat{\alpha}_t$s

Compression/Standardization. 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 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:

$$(\alpha_t - \alpha_*)' = a_t'\Lambda, \tag{A.65}$$

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

$$\hat{a} = \hat{\tilde{\alpha}}\hat{M}, \quad \hat{\Lambda} = (\hat{a}'\hat{a})^{-1}\hat{a}'\hat{\tilde{\alpha}}, \tag{A.66}$$

where $\hat{a}$ is the $T \times \tilde{K}$ matrix with rows $\hat{a}_t'$. Even if $\tilde{K} = K$ this operation standardizes the basis function coefficients α_t.

We can now replace (11) by[26]

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

To evaluate the MDD formulas such as (19), we replace K by $\tilde{K}$, $\mathcal{L}^{(K)}(\hat{\alpha}_t|X_t)$ by $\mathcal{L}^{(\tilde{K})}(\alpha_* + \hat{\Lambda}'\hat{a}_t|X_t)$, and we change the term $\sum_{t=1}^T \ln|\hat{V}_t|^{1/2}$ to $\sum_{t=1}^T \ln|(\hat{\Lambda}\hat{V}_t^{-1}\hat{\Lambda}')^{-1}|^{1/2}$.The measure-

[26]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 5 and 6.

ment equation (13) is replaced by

$$\hat{a}_t = a_t + N^{-1/2}\eta_t, \quad \eta_t \sim \mathcal{N}\big(0, (\hat{\Lambda}\hat{V}_t^{-1}\hat{\Lambda}')^{-1}\big). \tag{A.67}$$

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.

B.3 Recovering Cross-Sectional Densities

Based on the estimated state-transition equation 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 prior/posterior draw of a_t from the relevant posterior distribution). First, use (A.65) 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

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

B.4 Construction of Prior Distribution

The specification of the prior for the VAR coefficients follows Chan (2022). We subsequently provide detailed information about the construction of this prior.

Recall from the main text that the state-transition equation is a VAR of the form

$$W_t = \sum_{j=1}^{p} \Phi^{(j)} W_{t-j} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma), \tag{A.68}$$

where $W_t = [(Y_t - Y_*)', (\alpha_t - \alpha_*)']'$. If we define W as the $T \times n$ matrix with rows W_t', Z as the $T \times k$ matrix with rows $\tilde{Z}_t' = [W_{t-1}', \ldots, W_{t-p}']$, and $\Phi = [\Phi_1, \ldots, \Phi_p]'$, then we can write

the model in matrix form as

$$W = \tilde{Z}\Phi + U. \tag{A.69}$$

Reparameterizing the VAR. To generate a prior distribution we rewrite the VAR in (A.68) in quasi-structural form:

$$AW_t = \sum_{h=1}^{p} B_h W_{t-h} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, D), \tag{A.70}$$

where D is a diagonal matrix with diagonal elements D_i and A a lower-triangular matrix with ones on the diagonal. Note that unlike in other parts of the paper D here summarizes innovation variances instead of the "data."

Let $k = np$ and define the $k \times n$ matrix $B = [B_1, \ldots, B_p]'$ such we can write (A.70) as

$$W_t' A' = \tilde{Z}_t' B + \epsilon_t'. \tag{A.71}$$

Subsequently, we use $M_{i.}$ to denote the ith row of matrix M. Likewise $M_{.j}$ is the jth column.

Using the lower-triangular structure of the A matrix, define the $(i-1) \times 1$ vectors

$$\alpha_i = \big[A_{i,1}, \ldots, A_{i,i-1}\big], \quad \tilde{W}_{<i,t} = -\big[W_{1,t}, \ldots, W_{i-1,t}\big]', \quad i = 2, \ldots, n.$$

Then, we can write the ith equation as

$$W_{i,t} = \tilde{W}_{<i,t}' \alpha_i + \tilde{Z}_t' B_{.i} + \epsilon_{i,t}. \tag{A.72}$$

Note that $W_{i,t}$ depends on the contemporaneous variables $W_{1,t}, \ldots, W_{i-1,t}$. But since the system is triangular and A has ones on the diagonal, the determinant of the Jacobian A associated with the change of variables from ϵ_t to W_t is equal to one and the likelihood function has the usual Gaussian form.

To simplify the notation further, let $k_i = k + i - 1$, and define the $k_i \times 1$ vectors $Z_{i,t} = [\tilde{W}_{<i,t}', \tilde{Z}_t]'$ and $\beta_i = [\alpha_i', B_{.i}']'$. Write the ith equation as

$$W_{i,t} = Z_{i,t}' \beta_i + \epsilon_{i,t}.$$

Define W_i to be the $T \times 1$ vector with elements $W_{i,t}$ and Z_i the $T \times k_i$ matrix with rows $Z_{i,t}'$. Then we obtain

$$W_i = Z_i \beta_i + \epsilon_i. \tag{A.73}$$

Overview of Prior. We assume that the parameters are *a priori* independent across equations, i.e.,

$$p(\beta, D) = \prod_{i=1}^{n} p(\beta_i|D_i)p(D_i). \tag{A.74}$$

For each pair (β_i, D_i) we use a Normal-Inverse Gamma (NIG) distribution of the form

$$\beta_i|D_i \sim \mathcal{N}\big(\underline{\beta}_i, D_i\underline{V}_i^{\beta}\big), \quad D_i \sim IG\big(\underline{\nu}_i, \underline{S}_i\big). \tag{A.75}$$

The prior density takes the form

$$\begin{aligned} p(\beta_i, D_i) &= (2\pi)^{-k_i/2}|\underline{V}_i^{\beta}|^{-1/2}\frac{\underline{S}_i^{\underline{\nu}_i}}{\Gamma(\underline{\nu}_i)}D_i^{-(\underline{\nu}_i+1+k_i/2)} \\ &\quad \times \exp\left\{-\frac{1}{D_i}\left[\underline{S}_i + \frac{1}{2}(\beta_i - \underline{\beta}_i)'(\underline{V}_i^{\beta})^{-1}(\beta_i - \underline{\beta}_i)\right]\right\}. \end{aligned}$$

In the remainder of this subsection we discuss the construction of $\underline{\beta}_i$, $\underline{V}_i^{\beta}$, $\underline{\nu}_i$, and $\underline{S}_i$. The prior is obtained by transforming a prior for the reduced-form parameters (Φ, Σ) into a prior for the quasi-structural parameters $(\beta_1, \ldots, \beta_{n_w}, D)$.

Prior for D_i and the α_i component of β_i. We start from a prior for $\Sigma = A^{-1'}DA^{-1}$:

$$\Sigma \sim IW\big(\underline{\nu}, \underline{S}\big), \quad \underline{S} = \text{diag}(\underline{s}_1^2, \ldots, \underline{s}_n^2). \tag{A.76}$$

Chan (2021) shows that this prior implies

$$D_i \sim IG\left(\frac{\underline{\nu} + i - n}{2}, \frac{\underline{s}_i^2}{2}\right), \quad i = 1, \ldots, n. \tag{A.77}$$

Thus, a comparison with (A.75) indicates that we are setting

$$\underline{\nu}_i = \frac{\underline{\nu} + i - n}{2}, \quad \underline{S}_i = \frac{\underline{s}_i^2}{2}. \tag{A.78}$$

Moreover, (A.76) implies that

$$A_{ij}|D_i \sim \mathcal{N}\left(0, \frac{D_i}{\underline{s}_j^2}\right), \quad 1 \leq j < i, \quad i = 2, \ldots, n, \tag{A.79}$$

which determines the prior for the α_i component of β_i.

Prior for the $B_{\cdot i}$ component of β_i. The prior will take the form

$$B_{\cdot i} \sim \mathcal{N}\big(\underline{B}_{\cdot i}, \underline{V}_i^{B}\big), \tag{A.80}$$

where $\underline{V}_i^B$ is assumed to be diagonal.

To specify a prior for $B_{\cdot i}$, we loosely map *a priori* beliefs about $(\alpha_i, \Phi_{\cdot i})$ into beliefs about $B_{\cdot i}$. To simplify the notation a bit, let $\phi_i = \Phi_{\cdot i}$ and suppose that the researcher starts with the belief that

$$\phi_i \sim \mathcal{N}\big(\underline{\phi}_i,\, D_i \underline{V}_i^{\phi}\big) \tag{A.81}$$

with $\underline{\phi}_i = 0$. The prior covariance matrix $\underline{V}_i^{\phi}$ is assumed to be diagonal with elements

$$[\underline{V}_i^{\phi}]_{ll} = \frac{1}{\lambda_1} \begin{cases} \frac{1}{\underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if vars } (i,j) \text{ belong to same block} \\ \frac{1}{\lambda_2 \underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if var } i \text{ belongs to } Y \text{ and } j \text{ belongs to } a \\ \frac{1}{\lambda_3 \underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if var } i \text{ belongs to } a \text{ and } j \text{ belongs to } Y \end{cases}$$

The quasi-structural-form coefficients (B) are related to the reduced form coefficients via

$$[B_h]_{ij} = [\Phi_h]_{ij} + \sum_{l=1}^{i-1} A_{il} [\Phi_h]_{lj}. \tag{A.82}$$

An example of this relationship is given below. Taking expectation of (A.82) and using $\mathbb{E}[A_{ij}] = 0$, we deduce that

$$\mathbb{E}\big[[B_h]_{ij}\big] = \mathbb{E}\big[[\Phi_h]_{ij}\big],$$

which leads to

$$\underline{B}_{\cdot i} = 0. \tag{A.83}$$

We use a prior covariance matrix $\underline{V}_i^B$ that is diagonal. The entries on the diagonal are specified as follows: we first express the variance of a generic element $[B_h]_{ij}$ in terms of the variances of A_{ij} and $[B_h]_{ij}$:

$$\begin{aligned} \mathbb{V}\big[[B_h]_{ij}\big] &= \mathbb{E}\big[\mathbb{V}\big([B_h]_{ij}\big)|A\big] + \mathbb{V}\big[\mathbb{E}\big([B_h]_{ij}\big)|A\big] \\ &= \mathbb{E}\left[\mathbb{V}\big([\Phi_h]_{ij}\big) + \sum_{l=1}^{i-1} A_{il}^2 \mathbb{V}\big([\Phi_h]_{lj}\big)\right] + \mathbb{V}\left[\mathbb{E}([\Phi_h]_{ij}) + \sum_{l=1}^{i-1} A_{il}\mathbb{E}([\Phi_h]_{lj})\right] \\ &= \mathbb{V}\big([\Phi_h]_{ij}\big) + \sum_{l=1}^{i-1} \mathbb{V}(A_{il})\mathbb{V}\big([\Phi_h]_{lj}\big) + \sum_{l=1}^{i-1} \mathbb{V}(A_{il})\big(\mathbb{E}([\Phi_h]_{lj})\big)^2. \end{aligned}$$

We use the index function

$$f(j,h) = (h-1)n + j \tag{A.84}$$

to arrange the $\mathbb{V}\big[[B_h]_{ij}\big]$ terms on the diagonal of the $k \times k$ matrix $\underline{V}_i^B$. Using the definition of the index function, the expressions for $\mathbb{V}[A_{il}]$ and $\mathbb{E}([\Phi_h]_{lj})$ from the Normal distribution

Table A-1: Hyperparameters for VAR Prior

Parameter	Description
$\underline{\nu} = 2$	Degrees of freedom for IG distribution
$\underline{s}_i = \text{StDev}(W_i)$	Shape para for IG; use sample standard dev.
λ_1	Overall precision of prior
λ_2	Relative precision for a to Y transmission
$\lambda_3 = 1$	Relative precision for Y to a transmission
$\lambda_4 = 2$	Decay rate for prior variance on lags

in (A.79), and the expression for $\mathbb{V}\big([\Phi_h]_{lj}\big)$ from the Normal distribution in (A.81), we can write

$$\mathbb{V}\big[[B_h]_{ij}\big] = D_i[\underline{V}_i^{\phi}]_{f(j,h)} + \sum_{l=1}^{i-1} \frac{D_i}{\underline{s}_l^2}\left[D_l[\underline{V}_l^{\phi}]_{f(j,h)} + [\underline{\phi}_l]^2_{f(j,h)}\right].$$

We now replace the variance parameter D_l by the hyperparameter $\underline{s}_l^2$. This ensures that $\underline{V}_i^B$ is not a function of the (unknown) variance parameter D_i and simplifies posterior calculations. Using $\mathbb{V}\big[[B_h]_{ij}\big] = D_i[\underline{V}_i^B]_{f(j,h)}$ we obtain

$$[\underline{V}_i^B]_{f(j,h)} = [\underline{V}_i^{\phi}]_{f(j,h)} + \sum_{l=1}^{i-1}\left([\underline{V}_l^{\phi}]_{f(j,h)} + \frac{1}{\underline{s}_l^2}[\underline{\phi}_l]^2_{f(j,h)}\right). \tag{A.85}$$

Summary. The overall prior takes the form (A.75). The prior for D_i is given by (A.77). The prior for β_i is obtained by combining (A.79) with (A.80), where mean and variance are given in (A.83) and (A.85), respectively. The hyperparameters for the prior are summarized in Table A-1. We set $\underline{s}_i$ equal to the sample standard deviation of W_i.

B.5 Posterior Sampling and MDD

Model and prior are set up so that the coefficients can be estimated equation by equation:

$$p(W,\beta,D) = \prod_{i=1}^{N}\left((2\pi D_i)^{-1/2}\exp\left\{-\frac{1}{2D_i}(W_i - Z_i\beta_i)'(W_i - Z_i\beta_i)\right\}p(\beta_i|D_i)p(D_i).\right) \tag{A.86}$$

Because the prior is conjugate, the posterior stays in the NIG family. It takes the form

$$\beta_i|(D_i,W_i) \sim \mathcal{N}\big(\bar{\beta}_i, D_i\bar{V}_i^{\beta}\big), \quad D_i \sim IG\big(\bar{\nu}_i, \bar{S}_i\big), \tag{A.87}$$

Instead of working with covariance matrices, it is more efficient to work with precision matrices. Define:

$$\underline{P}_i^{\beta} = \big(\underline{V}_i^{\beta}\big)^{-1}, \quad \bar{P}_i^{\beta} = \big(\bar{V}_i^{\beta}\big)^{-1}.$$

The updating equations for the posterior take the form

$$\begin{aligned}
\bar{P}_i^{\beta} &= \underline{P}_i^{\beta} + Z_i' Z_i \\
\bar{\beta}_i &= \big(\bar{P}_i^{\beta}\big)^{-1}\big(\underline{P}_i^{\beta}\underline{\beta}_i + Z_i' W_i\big) \\
\bar{\nu}_i &= \underline{\nu}_i + T/2 \\
\bar{S}_i &= \underline{S}_i + \frac{1}{2}\big(W_i' W_i + \underline{\beta}_i' \underline{P}_i^{\beta} \underline{\beta}_i - \bar{\beta}_i' \bar{P}_i^{\beta} \bar{\beta}_i\big).
\end{aligned}$$

The MDD can be computed analytically as follows:

$$\begin{aligned}
\ln p(W) &= -\frac{Tn}{2}\ln(2\pi) + \sum_{i=1}^{n}\Bigg[\frac{1}{2}\big(\ln|\underline{P}_i^{\beta}| - \ln|\bar{P}_i^{\beta}|\big) \\
&\qquad + \underline{\nu}\ln|\underline{S}_i| - \bar{\nu}\ln|\bar{S}_i| - \ln\Gamma(\underline{\nu}_i) + \ln\Gamma(\bar{\nu}_i)\Bigg].
\end{aligned} \tag{A.88}$$

B.6 Filtering and Simulation Smoothing with Multiple Lags

In this subsection we show how the standard Kalman filter / simulation smoother can be extended to multiple lags. We abstract from the compression/standardization step discussed in Section B.2, use α_t for the basis function coefficients, and ignore the deterministic components α_* and Y_*.

B.6.1 Companion Form

To derive the updating equations for the filter and simulation smoother we express the state-transition equation in companion form. We illustrate the companion form notation for $p = 2$. The generalization is straightforward. We define

$$W_t' = [Y_t', Y_{t-1}', \alpha_t', \alpha_{t-1}'] \tag{A.89}$$

and partition W_t' into

$$W_t' = [Z_t', s_t'], \quad Z_t' = [Y_t', Y_{t-1}'], \quad s_t = [\alpha_t, \alpha_{t-1}]'.$$

The companion-form law of motion for w_t can be written as

$$\begin{bmatrix} Y_t \\ Y_{t-1} \\ \alpha_t \\ \alpha_{t-1} \end{bmatrix} = \begin{bmatrix} \Phi_{1,yy} & \Phi_{2,yy} & \Phi_{1,y\alpha} & \Phi_{2,y\alpha} \\ I_y & 0 & 0 & 0 \\ \Phi_{1,\alpha y} & \Phi_{2,\alpha y} & \Phi_{1,\alpha\alpha} & \Phi_{2,\alpha\alpha} \\ 0 & 0 & I_\alpha & 0 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ \alpha_{t-1} \\ \alpha_{t-2} \end{bmatrix} + \begin{bmatrix} I_y & 0 \\ 0 & 0 \\ 0 & I_\alpha \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{y,t} \\ u_{\alpha,t} \end{bmatrix}, \tag{A.90}$$

or, more compactly, as

$$W_t = \Psi W_{t-1} + M u_t. \tag{A.91}$$

We further partition the selection matrix M into

$$M = \begin{bmatrix} M_y & M_\alpha \end{bmatrix}$$

such that

$$M_y' W_t = Y_t, \quad M_\alpha' W_t = \alpha_t.$$

In some instances, we need to extract the subvectors Z_t and s_t from W_t, which is done by the selection matrices

$$\Xi_z' = \begin{bmatrix} I_z & 0 \end{bmatrix}, \quad Z_t = \Xi_z' W_t, \quad \Xi_s' = \begin{bmatrix} 0 & I_s \end{bmatrix}, \quad s_t = \Xi_s' W_t.$$

For the forward iterations of the Kalman filter, it is useful to separate s_t and Z_t. To that end, define

$$\Psi_{ss} = \Xi_s' \Psi \Xi_s = \begin{bmatrix} \Phi_{1,\alpha\alpha} & \Phi_{2,\alpha\alpha} \\ I_\alpha & 0 \end{bmatrix}, \quad \Psi_{sz} = \Xi_s' \Psi \Xi_z = \begin{bmatrix} \Phi_{1,\alpha y} & \Phi_{2,\alpha y} \\ 0 & 0 \end{bmatrix}.$$

Moreover, we define

$$M_{s\alpha} = \Xi_s' M_\alpha = \begin{bmatrix} I_\alpha \\ 0 \end{bmatrix}, \quad \Phi_{yz} = \begin{bmatrix} \Phi_{1,yy} & \Phi_{2,yy} \end{bmatrix}, \quad \Phi_{ys} = \begin{bmatrix} \Phi_{1,y\alpha} & \Phi_{2,y\alpha} \end{bmatrix}.$$

Using this notation, the measurement equation can be written as

$$\hat{\alpha}_t = M_{s\alpha}' s_t + \eta_t, \quad \eta_t \sim N(0, V_t). \tag{A.92}$$

B.6.2 Forward Filtering

The forward filtering iterations are obtained from a modified Kalman filter that recognizes that y_t is directly observable. Thus, only s_t is a latent state variable.

Recursive Assumption. We start from the assumption that

$$s_{t-1}|(Y_{1:t-1}, \hat{\alpha}_{1:t-1}) \sim \mathcal{N}\big(s_{t-1|t-1}, P_{t-1|t-1}(ss)\big). \tag{A.93}$$

We use the notation $P_{t-1|t-1}(ss)$ to indicate that one can define a larger matrix, conforming with w_t, which is partitioned as follows:

$$P_{t-1|t-1} = \begin{bmatrix} P_{t-1|t-1}(zz) & P_{t-1|t-1}(zs) \\ P_{t-1|t-1}(zs) & P_{t-1|t-1}(ss) \end{bmatrix},$$

with the understanding that

$$P_{t-1|t-1}(zz) = 0, \quad P_{t-1|t-1}(zs) = 0, \quad P_{t-1|t-1}(sz) = 0.$$

Initialization. Just in a regular VAR analysis, we will condition the likelihood function on observations to initialize the lags associated with the determination of (Z_1, s_1). Note that the initial lags of Y_t are directly observed under suitable definition of the sample period and hence we know Z_0. To initialize the latent state we set

$$s_{0|0} = \begin{bmatrix} \hat{\alpha}_0 \\ \vdots \\ \hat{\alpha}_{-p+1} \end{bmatrix}, \quad \mathrm{P}_{0|0}(ss) = \mathrm{diag}\big[V_0, \cdots, V_{-p+1}\big].$$

Thus, we assume that

$$s_0|(Y_{-p+1:0}, \hat{\alpha}_{-p+1:0}) \sim \mathcal{N}\big(s_{0|0}, P_{0|0}\big).$$

For the smoother below, it will become important to properly account for the initialization because we will also need draws of $s_0 = \alpha_{-p+1:0}$ to set up the Gibbs sampler.

Forecasting. We begin by forecasting (Y_t, s_t) jointly, using (A.90). Let

$$\begin{bmatrix} Y_t \\ s_t \end{bmatrix} \Big| (Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) \sim \mathcal{N}\left(\begin{bmatrix} y_{t|t-1} \\ s_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1}(yy) & P_{t|t-1}(ys) \\ P_{t|t-1}(sy) & P_{t|t-1}(ss) \end{bmatrix} \right), \tag{A.94}$$

where

$$\begin{aligned} Y_{t|t-1} &= \Phi_{yz} Z_{t-1} + \Phi_{ys} s_{t-1|t-1} \\ s_{t|t-1} &= \Psi_{sz} Z_{t-1} + \Psi_{ss} s_{t-1|t-1} \\ P_{t|t-1}(yy) &= \Phi_{ys} P_{t-1|t-1}(ss) \Phi'_{ys} + \Sigma_{yy} \\ P_{t|t-1}(sy) &= \Psi_{ss} P_{t-1|t-1}(ss) \Phi'_{ys} + M_{s\alpha} \Sigma_{\alpha y} \\ P_{t|t-1}(ss) &= \Psi_{ss} P_{t-1|t-1}(ss) \Psi'_{ss} + M_{s\alpha} \Sigma_{\alpha\alpha} M'_{s\alpha}. \end{aligned}$$

We can now factorize the joint distribution of (Y_t, s_t) into a marginal distribution of Y_t and a conditional distribution of $s_t|Y_t$:

$$\begin{aligned} Y_t|(Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) &\sim \mathcal{N}\big(Y_{t|t-1}, P_{t|t-1}(yy)\big) \\ s_t|(Y_t, Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) &\sim \mathcal{N}\big(s_{t|y,t-1}, P_{t|y,t-1}(ss)\big), \end{aligned} \tag{A.95}$$

where

$$\begin{aligned} s_{t|y,t-1} &= s_{t|t-1} + P_{t|t-1}(sy)[P_{t|t-1}(yy)]^{-1}(Y_t - Y_{t|t-1}) \\ P_{t|y,t-1}(ss) &= P_{t|t-1}(ss) - P_{t|t-1}(sy)[P_{t|t-1}(yy)]^{-1} P_{t|t-1}(ys). \end{aligned}$$

Finally, the distribution of $\hat{\alpha}_t$ conditional on y_t and $t-1$ information is

$$\hat{\alpha}_t|(Y_t, Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) \sim N(\hat{\alpha}_{t|y,t-1}, F_{t|y,t-1}), \tag{A.96}$$

where

$$\begin{aligned} \hat{\alpha}_{t|y,t-1} &= M'_{s\alpha} s_{t|y,t-1} \\ F_{t|y,t-1} &= M'_{s\alpha} P_{t|y,t-1}(ss) M_{s\alpha} + V_t. \end{aligned}$$

Updating. The updating step in the Kalman filter is done conditional on Y_t. Generically,

$$\begin{aligned} & p(s_t|\hat{\alpha}_t, Y_t, Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) \\ & \propto \; p(\hat{\alpha}_t|s_t, Y_t, Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) p(s_t|y_t, Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}). \end{aligned} \tag{A.97}$$

We can use the standard updating formulas, substituting in the means and variances that are conditional on Y_t:

$$\begin{aligned} s_{t|t} &= s_{t|y,t-1} + P_{t|y,t-1}(ss) M_{s\alpha} F^{-1}_{t|y,t-1}(\hat{\alpha}_t - \hat{\alpha}_{t|y,t-1}) \\ P_{t|t}(ss) &= P_{t|y,t-1}(ss) - P_{t|y,t-1}(ss) M_{s\alpha} F^{-1}_{t|y,t-1} M'_{s\alpha} P_{t|y,t-1}(ss). \end{aligned} \tag{A.98}$$

B.6.3 Simulation Smoothing

The goal is to generate draws from the distribution of $\alpha_{-p+1:T}$ given $(Y_{-p+1:T}, \hat{\alpha}_{-p+1:T})$. Note that the filter in its final step generates the distribution $s_T|(Y_{-p+1:T}, \hat{\alpha}_{-p+1:T})$. Because of the companion form, a draw s_T^i determines $\alpha_T^i, \ldots, \alpha_{T-p+1}^i$. Thus, the distribution $s_{T-1}|(s_T, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T})$ is degenerate. In the implementation of the smoother, we will draw blocks as follows:

$$s_T, \quad s_{T-p}|s_T, \quad s_{T-2p}|(s_{T-p}, s_T), \quad s_{T-3p}|(s_{T-2p}, s_{T-p}, s_T), \quad \ldots.$$

Because $s_{T-jp} = [\alpha'_{T-jp}, \ldots, \alpha'_{T-(j+1)p+1}]'$, this approach generates the entire sequence $\alpha_{-p+1:T}$. However, special care needs to be given to $\alpha_{-p+1:0}$ and the fact that T may not be divisible by p.

Preliminaries. Assuming for now that T is a multiple of p and that $T = bp$, where b is the number of blocks, the simulation smoother relies on the factorization

$$p(Y_{-p+1:T}, \alpha_{-p+1:T}|Y_{-p+1:0}, \alpha_{-p+1:0}) = \prod_{j=1}^{b} p(Y_{(j-1)p+1:jp}, \alpha_{(j-1)p+1:jp}|Y_{-p+1:(j-1)p}, \alpha_{-p+1:(j-1)p}),$$

where, because of the VAR(p) structure of the state-transition equation,

$$\begin{aligned} &p(Y_{(j-1)p+1:jp}, \alpha_{(j-1)p+1:jp}|Y_{-p+1:(j-1)p}, \alpha_{-p+1:(j-1)p}) \\ &\quad = \quad p(Y_{(j-1)p+1:jp}, \alpha_{(j-1)p+1:jp}|Y_{(j-2)p+1:(j-1)p}, \alpha_{(j-2)p+1:(j-1)p}). \end{aligned}$$

The conditional distribution on the right-hand side, can be obtained by iterating the companion form (A.91) p periods forward:

$$W_{t+p} = \Psi^p W_t + M \sum_{h=0}^{p-1} \Psi^h u_{t+p-h}.$$

Thus,

$$W_{t+p}|W_t \sim \mathcal{N}\left(\Psi^p W_t, \sum_{h=0}^{p-1} \Psi^h M \Sigma M' \Psi^{h'}\right). \tag{A.99}$$

Now, set $t = (j-1)p$ and note that

$$W_{(j-1)p} = [Y_{(j-1)p}, \alpha_{(j-1)p}, \ldots, Y_{(j-2)p+1}, \alpha_{(j-2)p+1}],$$

as required.

Generic Smoother. We first modify the generic derivation of the smoother to account for the presence of both $Y_{-p+1:T}$ and $\hat{\alpha}_{-p+1:T}$ in the conditioning set and the block sampling. As before, it is convenient for now to assume that $T = bp$ and the time index t shifts in steps of p periods, that is, $t = (j-1)p$, where $j = 0, \ldots, b$. Consider the following factorization

$$\begin{aligned}
&p(\alpha_{t-p+1:t}, \alpha_{t+1:T}, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) \\
&\quad = \int p(\alpha_{-p+1:T}, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) d\alpha_{-p+1:t-p} \\
&\quad = \int p(\alpha_{-p+1:t}, Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) \\
&\qquad \times \left(\prod_{\tau=t+1}^{T} p(Y_\tau | Y_{\tau-p:\tau-1}, \alpha_{\tau-p:\tau-1}) p(\alpha_\tau | Y_\tau, Y_{\tau-p:\tau-1}, \alpha_{\tau-p:\tau-1}) p(\hat{\alpha}_\tau | \alpha_\tau) \right) d\alpha_{-p+1:t-p} \\
&\quad = p(\alpha_{t-p+1:t}, Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) \\
&\qquad \times \prod_{\tau=t+1}^{t+p} p(Y_\tau | Y_{\tau-p:\tau-1}, \alpha_{\tau-p:\tau-1}) p(\alpha_\tau | Y_\tau, Y_{\tau-p:\tau-1}, \alpha_{\tau-p:\tau-1}) p(\hat{\alpha}_\tau | \alpha_\tau) \\
&\qquad \times \text{terms without } \alpha_{t-p+1:t}.
\end{aligned}$$

Maintaining the assumption that $T = bp$ and $t = (j-1)p$, we can deduce

$$\begin{aligned}
&p(s_t | s_{t+p}, s_{t+2p}, \ldots, s_T, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) \\
&\quad = p(\alpha_{t-p+1:t} | \alpha_{t+1:T}, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) \\
&\quad \propto p(\alpha_{t-p+1:t}, \alpha_{t+1:T}, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) \\
&\quad \propto p(\alpha_{t-p+1:t}, Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) p(Y_{t+1:t+p}, \alpha_{t+1:t+p} | Y_{t-p+1:t}, \alpha_{t-p+1:t}) \\
&\quad = p(s_t, Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) p(Z_{t+p}, s_{t+p} | Z_t, s_t).
\end{aligned} \tag{A.100}$$

The first proportionality follows from Bayes Theorem. The second proportionality follows from dropping the factor $p(Y_{t+p+1:T}, \alpha_{t+p+1:T} | Y_{t+1:t+p}, \alpha_{t+1:t+p})$ because it does not depend on $\alpha_{t-p+1:t}$. The last equality uses $s_t = [\alpha'_{t-p+1}, \ldots, \alpha'_t]'$ and $Z_t = [Y'_{t-p+1}, \ldots, Y'_t]'$. Because

$$p(s_t, Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) = p(s_t | Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) p(Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}),$$

it follows that

$$p(s_t | s_{t+p}, s_{t+2p}, \ldots, s_T, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) \propto p(s_t | Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) p(Z_{t+p}, s_{t+p} | Z_t, s_t). \tag{A.101}$$

Smoothing Formulas for the Linear Gaussian Model. We previously established that

$$s_t | (Y_{1:t}, \hat{\alpha}_{1:t}) \sim \mathcal{N}\big(s_{t|t}, P_{t|t}(ss)\big).$$

As previously shown in (A.99), iterating the companion form forward for p periods yields

$$(Z_{t+p}, s_{t+p})|(Z_t, s_t) \sim \mathcal{N}\left(\Psi^p W_t, \sum_{h=0}^{p-1} \Psi^h M \Sigma M' \Psi^{h'}\right). \tag{A.102}$$

Let

$$W_{t|t} = [Z_t', s_{t|t}']'$$

and define

$$P_{t+p|t}(ww) = \Psi^p \Xi_s P_{t|t}(ss) \Xi_s' \Psi^p + \sum_{h=0}^{p-1} \Psi^h M \Sigma M' \Psi^{h'}. \tag{A.103}$$

The joint distribution of (s_t, W_{t+p}) is given by

$$\left[\begin{array}{c} s_t \\ W_{t+p} \end{array}\right] \Big| (\cdot) \sim \mathcal{N}\left(\left[\begin{array}{c} s_{t|t} \\ \Psi^p W_{t|t} \end{array}\right], \left[\begin{array}{cc} P_{t|t}(ss) & P_{t|t}(ss)\Xi_s'\Psi^{p'} \\ \Psi^p \Xi_s P_{t|t}(ss) & P_{t+p|t}(ww) \end{array}\right]\right). \tag{A.104}$$

Then, we sample s_t from

$$s_t \sim \mathcal{N}\big(s^i_{t|t+p}, P_{t|t+p}(ss)\big), \tag{A.105}$$

where

$$s_{t|t+p} = s_{t|t} + P_{t|t}(ss)\Xi_s'\Psi^{p'} P^{-1}_{t+p|t}(ww)\left(\left[\begin{array}{c} Z_{t+p} \\ s_{t+p} \end{array}\right] - \Psi^p W_{t|t}\right) \tag{A.106}$$

$$P_{t|t+p}(ss) = P_{t|t}(ss) - P_{t|t}(ss)\Xi_s'\Psi^{p'} P^{-1}_{t+p|t}(ww)\Psi^p \Xi_s P_{t|t}(ss). \tag{A.107}$$

T is not a multiple of p. Consider the following example. Suppose that $T = 10$ and $p = 3$. In this case we can use the formulas to generate

$$\begin{aligned} &p(s_{10}|Y_{-2:10}, \hat{\alpha}_{-2:10}), \quad p(s_7|s_{10}, Y_{-2:10}, \hat{\alpha}_{-2:10}), \\ &p(s_4|s_7, s_{10}, Y_{-2:10}, \hat{\alpha}_{-2:10}), \quad p(s_1|s_4, s_7, s_{10}, Y_{-2:10}, \hat{\alpha}_{-2:10}). \end{aligned}$$

The last step gives us

$$p(\alpha_{-1:1}|\alpha_{2:10}, Y_{-2:10}, \hat{\alpha}_{-2:10}),$$

but we still lack

$$p(\alpha_{-2}|\alpha_{-1:10}, Y_{-2:10}, \hat{\alpha}_{-2:10}).$$

In order to generate the draws needed to initialize the vector autoregressive law of motion of the state transition equation, we take the following short-cut: (i) discard $\alpha_{-1:0}$, (ii) redraw s_0 from (A.105) by evaluating the formulas $s_{t|t+p}$ and $P_{t|t+p}(ss)$ for $t = 0$.

C 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

$$x'(\epsilon, x; s), \quad R(s), \quad W(s), \quad \mu'(s). \tag{A.108}$$

We will subsequently construct approximations to these functions.

C.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 $\mathcal{A}$,

$$\int \mathbb{I}\{x \in \mathcal{A}\} d\mu'_\epsilon = \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \int \mathbb{I}\{x'(\tilde{\epsilon}, x; s) \in \mathcal{A}\} d\mu_{\tilde{\epsilon}}. \tag{A.109}$$

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:

$$\widehat{m}'_\epsilon = \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \left(\int_{x > \underline{x}} \mathbb{I}\{x'(\tilde{\epsilon}, x; s) = \underline{x}\} d\mu_{\tilde{\epsilon}} + \mathbb{I}\{x'(\tilde{\epsilon}, \underline{x}; s) = \underline{x}\} \widehat{m}_{\tilde{\epsilon}} \right). \tag{A.110}$$

We 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

$$\begin{aligned} \widehat{m}'_\epsilon \;=\; & \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \Bigg[(1 - \widehat{m}_{\tilde{\epsilon}}) \int \left(\int \mathbb{I}\{\eta_i \leq \underline{x} - x'(\tilde{\epsilon}, x; s)\} p_\eta(\eta) d\eta \right) p_{\tilde{\epsilon}}(x) dx \\ & + \widehat{m}_{\tilde{\epsilon}} \int \mathbb{I}\{\eta_i \leq \underline{x} - x'(\tilde{\epsilon}, \underline{x}; s)\} p_\eta(\eta) d\eta \Bigg], \end{aligned} \tag{A.111}$$

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. η is an idiosyncratic deviation from the decision rule $x'(\cdot)$ that ensures the distribution of $x|(x > \underline{x})$.

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

$$\begin{aligned} p'_\epsilon(x) &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg[(1-\widehat{m}_{\tilde{\epsilon}})\frac{\int p_\eta(x - x'(\tilde{\epsilon},\tilde{x};s))\mathbb{I}\{x > \underline{x}\}p_{\tilde{\epsilon}}(\tilde{x})d\tilde{x}}{\int\int p_\eta(x - x'(\tilde{\epsilon},\tilde{x};s))\mathbb{I}\{x > \underline{x}\}p_{\tilde{\epsilon}}(\tilde{x})d\tilde{x}dx} \\ &\quad + \widehat{m}_{\tilde{\epsilon}}\frac{p_\eta(x - x'(\tilde{\epsilon},\underline{x};s))\mathbb{I}\{x > \underline{x}\}}{\int p_\eta(x - x'(\tilde{\epsilon},\underline{x};s))\mathbb{I}\{x > \underline{x}\}dx}\Bigg], \end{aligned} \tag{A.112}$$

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.111) and (A.112) define a law of motion for the cross-sectional density $q_\epsilon(x)$.

C.2 Firms and Households

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

$$z' = \rho_z z + \sigma_z \omega'. \tag{A.113}$$

The capital stock has to equal the net asset holdings:

$$K(s) = \sum_\epsilon \pi(\epsilon)\left[(1-\widehat{m}_\epsilon)\int x p_\epsilon(x)dx + \widehat{m}_\epsilon \underline{x}\right]. \tag{A.114}$$

Profit maximization of the representative firm implies

$$\begin{aligned} R(s) &= \alpha e^z K^{\alpha-1}(s)L^{1-\alpha} - \delta \\ W(s) &= (1-\alpha)e^z K^\alpha(s)L^{-\alpha}\,. \end{aligned} \tag{A.115}$$

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

$$\psi(\epsilon, x; z, \widehat{m}_\epsilon, p_\epsilon) = \beta\mathbb{E}\left[\big(1 + R(z', \widehat{m}'_\epsilon, p'_\epsilon)\big)c\big(\epsilon', x'; z', \widehat{m}'_\epsilon, p'_\epsilon\big)^{-\sigma}\middle|\epsilon, x; z, \widehat{m}_\epsilon, p_\epsilon\right]. \tag{A.116}$$

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

$$x'_*(\epsilon, x; s) = W(s)\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1 + R(s)\big)x - \psi^{-1/\sigma}(\epsilon, x; s). \tag{A.117}$$

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

$$x'(\epsilon, x; s) = \max\ \big\{\underline{x},\ x'_*(\epsilon, x; s)\big\}. \tag{A.118}$$

Once the asset holdings are determined, consumption is given by

$$c(\epsilon, x; s) = W(s)\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1 + R(s)\big)x - x'(\epsilon, x; s). \tag{A.119}$$

C.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:

$$p^{(K)}_{t,\epsilon}(x) = \exp\left\{\gamma_{t,\epsilon,0} + \gamma_{t,\epsilon,1}(x - m_{t,\epsilon,1}) + \sum_{k=2}^{K} \gamma_{t,\epsilon,k}\left[(x - m_{t,\epsilon,1})^k - m_{t,\epsilon,k}\right]\right\}. \tag{A.120}$$

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:

$$\begin{aligned} m_{t,\epsilon,1} &= \int x p^{(K)}_{t,\epsilon}(x)dx \\ m_{t,\epsilon,k} &= \int (x - m_{t,\epsilon,1})^k p^{(K)}_{t,\epsilon}(x)dx. \end{aligned} \tag{A.121}$$

We also require that the approximate density integrates to one:

$$\gamma_{t,\epsilon,0} = -\ln \int \exp\left\{\gamma_{t,\epsilon,1}(x - m_{t,\epsilon,1}) + \sum_{k=2}^{K} \gamma_{t,\epsilon,k}\left[(x - m_{t,\epsilon,1})^k - m_{t,\epsilon,k}\right]\right\} dx. \tag{A.122}$$

Conditional on the moments $m_{t,\epsilon,k}$, we can use (A.121) and (A.122) 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,

$$\begin{aligned}\widehat{m}_{t+1,\epsilon} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg(\int \mathbb{I}\{x_t'(\tilde{\epsilon},x)=\underline{x}\}(1-\widehat{m}_{t,\tilde{\epsilon}})p_{t,\tilde{\epsilon}}^{(K)}(x)dx \\ &\quad +\mathbb{I}\{x_t'(\tilde{\epsilon},x=\underline{x})=\underline{x}\}\widehat{m}_{t,\tilde{\epsilon}}\Bigg) \end{aligned} \tag{A.123}$$

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

$$\begin{aligned} m_{t+1,\epsilon,1} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg(\int \mathbb{I}\{x_t'(\tilde{\epsilon},x)>\underline{x}\}x_t'(\tilde{\epsilon},x)(1-\widehat{m}_{t,\tilde{\epsilon}})p_{t,\tilde{\epsilon}}^{(K)}(x)dx \\ &\quad +\mathbb{I}\{x_t'(\tilde{\epsilon},\underline{x})>\underline{x}\}x_t'(\tilde{\epsilon},\underline{x})\widehat{m}_{t,\tilde{\epsilon}}\Bigg) \\ m_{t+1,\epsilon,k} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg(\int \mathbb{I}\{x_t'(\tilde{\epsilon},x)>\underline{x}\}[x_t'(\tilde{\epsilon},x)-m_{t+1,\epsilon,1}]^k(1-\widehat{m}_{t,\tilde{\epsilon}})p_{t,\tilde{\epsilon}}^{(K)}(x)dx \\ &\quad +\mathbb{I}\{x_t'(\tilde{\epsilon},\underline{x})>\underline{x}\}[x_t'(\tilde{\epsilon},\underline{x})-m_{t+1,\epsilon,1}]^k\widehat{m}_{t,\tilde{\epsilon}}\Bigg). \end{aligned} \tag{A.124}$$

Conditional on the decision rule $x_t'(\epsilon,x)$ and the initial density approximation $p_{t,\tilde{\epsilon}}^{(K)}(x)$, Equations (A.123) and (A.124) define a law of motion for $m_{t,\epsilon,k}$ and $\widehat{m}_{t,\epsilon}^k$. Combined with (A.120) and (A.121) one obtains a transition equation for $p_{t,\tilde{\epsilon}}^{(K)}(x)$.

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

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

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

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

Thus, for instance, we can define the quadrature approximation

$$\begin{aligned} m^Q_{t+1,\epsilon,1} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg(\sum_{j=1}^{J} \mathbb{I}\{x'_t(\tilde{\epsilon},x_j) > \underline{x}\} x'_t(\tilde{\epsilon},x_j)(1-\widehat{m}_{t,\tilde{\epsilon}})\omega_j p^{(K)}_{t,\tilde{\epsilon}}(x_j) \\ &\quad + \mathbb{I}\{x'_t(\tilde{\epsilon},\underline{x}) > \underline{x}\} x'_t(\tilde{\epsilon},\underline{x})\widehat{m}_{t,\tilde{\epsilon}} \Bigg). \end{aligned} \tag{A.125}$$

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.120) 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:

$$K_t = \sum_{\epsilon} \pi(\epsilon) \left[(1-\widehat{m}_{t,\epsilon}) m_{t,\epsilon,1} + \widehat{m}_{t,\epsilon}\underline{x} \right]. \tag{A.126}$$

In turn, the factor prices can be written as

$$R_t = \alpha e^{z_t} K_t^{\alpha-1} L^{1-\alpha} - \delta \tag{A.127}$$

$$W_t = (1-\alpha) e^{z_t} K_t^{\alpha} L^{-\alpha}. \tag{A.128}$$

Aggregate total factor productivity evolves according to

$$z_t = \rho_z z_{t-1} + \sigma_z \omega_t. \tag{A.129}$$

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

$$\psi^{(L)}_t(\epsilon, x) = \exp\left\{ \sum_{l=1}^{L} \theta_{t,\epsilon,l} T_l(\xi(x)) \right\}, \tag{A.130}$$

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

$$\begin{aligned} x'_{*,t}(\epsilon, x) &= W_t\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1+R_t\big)x - \big[\psi_t^{(L)}(\epsilon, x)\big]^{-1/\sigma} \\ x'_t(\epsilon, x) &= \max\ \big\{\underline{x},\, x'_{*,t}(\epsilon, x)\big\} \\ c_t(\epsilon, x) &= W_t\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1+R_t\big)x - x'_t(\epsilon, x). \end{aligned} \tag{A.131}$$

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:

$$\exp\left\{\sum_{i=1}^{L}\theta_{t,\epsilon,i}T_i(\xi(x_l))\right\} = \beta\sum_{\tilde{\epsilon}}\pi(\tilde{\epsilon}|\epsilon)\mathbb{E}_t\big[(1+R_{t+1})c_{t+1}^{-\sigma}\big(\tilde{\epsilon}, x'_t(\tilde{\epsilon}, x_l)\big)\big], \quad l = 1, \ldots, L. \tag{A.132}$$

C.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.120), define

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

so that we can write

$$p_{t,\epsilon,j} = f_p\big(x_j; m_{t,\epsilon,1}, \ldots, m_{t,\epsilon,K}, \gamma_{t,\epsilon,1}, \ldots, \gamma_{t,\epsilon,K}\big), \quad j = 1, \ldots, J, \tag{A.133}$$

where the function $f_p(\cdot)$ is given by (A.120). 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

$$\begin{aligned} m_{t,\epsilon,1} &= \sum_{j=1}^{J} x_j\omega_j f_p\big(x_j; m_{t,\epsilon,1}, \ldots, m_{t,\epsilon,K}, \gamma_{t,\epsilon,1}, \ldots, \gamma_{t,\epsilon,K}\big) \\ m_{t,\epsilon,k} &= \sum_{j=1}^{J}(x_j - m_{t,\epsilon,1})^k\omega_j f_p\big(x_j; m_{t,\epsilon,1}, \ldots, m_{t,\epsilon,K}, \gamma_{t,\epsilon,1}, \ldots, \gamma_{t,\epsilon,K}\big), \quad k = 2, \ldots, K. \end{aligned} \tag{A.134}$$

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

$$\begin{aligned}
m_{t+1,\epsilon,1} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg(\sum_{j=1}^{J} x'_{t,\tilde{\epsilon},j}(1-\widehat{m}_{t,\tilde{\epsilon}})\omega_j p_{t,\tilde{\epsilon},j} + x'_{t,\tilde{\epsilon},1}\widehat{m}_{t,\tilde{\epsilon}} \Bigg) \qquad \text{(A.135)}\\
m_{t+1,\epsilon,k} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg(\sum_{j=1}^{J} [x'_{t,\tilde{\epsilon},j} - m_{t+1,\epsilon,1}]^k(1-\widehat{m}_{t,\tilde{\epsilon}})\omega_j p_{t,\tilde{\epsilon},j} + [x'_{t,\tilde{\epsilon},1} - m_{t+1,\epsilon,1}]^k\widehat{m}_{t,\tilde{\epsilon}} \Bigg),\\
&\qquad k = 2,\ldots,K\\
\widehat{m}_{t+1,\epsilon} &= \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon)\Bigg(\sum_{j=1}^{J} (1-\widehat{m}_{t,\tilde{\epsilon}})\omega_j p_{t,\tilde{\epsilon},j} + \mathbb{I}\{x'_{t,\tilde{\epsilon},1} = 0\}\widehat{m}_{t,\tilde{\epsilon}} \Bigg).
\end{aligned}$$

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

$$\begin{aligned}
K_t &= \sum_{\epsilon} \pi(\epsilon)(1-\widehat{m}_{t,\epsilon})m_{t,\epsilon,1} \qquad \text{(A.136)}\\
R_t &= \alpha e^{z_t} K_t^{\alpha-1} L^{1-\alpha} - \delta\\
W_t &= (1-\alpha)e^{z_t} K_t^{\alpha} L^{-\alpha}\\
z_t &= \rho_z z_{t-1} + \sigma_z \epsilon_{z,t}.
\end{aligned}$$

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:

$$\begin{aligned}
x'_{*,t}(\epsilon, x) &= f_{x'_*}(\epsilon, x; W_t, r_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L}) \qquad \text{(A.137)}\\
&= W_t\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1+R_t\big)x - \big[\psi_t^{(L)}(\epsilon, x; \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L})\big]^{-1/\sigma}\\
x'_t(\epsilon, x) &= f_{x'}(\epsilon, x; W_t, R_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L})\\
&= \max\ \big\{0,\ f_{x'_*}(\epsilon, x; W_t, R_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L})\big\}\\
c_t(\epsilon, x) &= f_c(\epsilon, x; W_t, R_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L})\\
&= W_t\big((1-\tau)\epsilon + b(1-\epsilon)\big) + \big(1+R_t\big)x - f_{x'}(\epsilon, x; W_t, R_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L}).
\end{aligned}$$

Thus,

$$x'_{t,\epsilon,j} = f_{x'}(\epsilon, x_j; W_t, R_t, \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L}), \quad j = 1,\ldots,J. \qquad \text{(A.138)}$$

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

$$\begin{aligned}&\psi_t^{(L)}(\epsilon, x_l; \theta_{t,\epsilon,1}, \ldots, \theta_{t,\epsilon,L}) \qquad\qquad (A.139)\\ &= \beta \sum_{\tilde{\epsilon}} \pi(\tilde{\epsilon}|\epsilon) \mathbb{E}_t\big[(1+R_{t+1}) f_c^{-\sigma}(\tilde{\epsilon}, x_l; W_{t+1}, R_{t+1}, \theta_{t+1,\epsilon,1}, \ldots, \theta_{t+1,\epsilon,L})\big], \quad l = 1, \ldots, L.\end{aligned}$$

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.133) delivers 2J equations that determine $\{p_{t,\epsilon,j}\}$; (A.134) delivers 2K equations that implicitly determine $\{\gamma_{t,\epsilon,k}\}$; (A.135) generates $2(K+1)$ equations that determine the evolution of $\{m_{t,\epsilon,k}\}$ and $\{\hat{m}_{t,\epsilon}\}$; (A.136) comprises 4 equations that determine the aggregate variables K_t, R_t, W_t, and z_t; (A.138) delivers 2J equations that determine $\{x'_{t,\epsilon,j}\}$; and, finally, (A.139) generates 2L equations that determine $\{\theta_{t,\epsilon,l}\}$. Thus, the system contains as many equations as variables.

C.5 Steady State and Local Dynamics

The model can now be solved by finding the steady state of the system defined by Equations (A.133) to (A.139), 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.

D A Second Simulation Experiment

In a second simulation experiment we generate data from a stylized heterogeneous agent New Keynesian (HANK) model which is based on Auclert and Rognlie (2020).

Model. Households choose consumption c_{it} to maximize

$$\max_{c_{it}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t (u(c_{it}) - v(N_t)) \right] \tag{A.140}$$

subject to the budget constraint Budget constraint

$$c_{it} + a_{it} \leq (1 + r_t) a_{it-1} + y_{it}, \qquad a_{it} \geq \underline{a},$$

where a_{it} are asset holdings and y_{it} is an idiosyncratic income process. The interest rate r_t is exogenous and evolves according to

$$r_t = \rho_r r_{t-1} + \epsilon_{r,t}, \quad \epsilon_{r,t} \sim N(0, \sigma_r^2). \tag{A.141}$$

The interest rate innovations are interpreted as monetary policy shocks. It is assumed that there is price stickiness in the background that allows the central bank to control the real interest rate.[27]

Household-level income evolves according to

$$y_{it} = Y_t \cdot \frac{e_{it}^{1+\zeta \log(Y_t)}}{\mathbb{E}\left[e_{it}^{1+\zeta \log(Y_t)} \right]} \tag{A.142}$$

where e_{it} is idiosyncratic state that follows an AR(1) process

$$\ln e_{it} = \rho_e \log e_{it-1} + u_{it}, \quad u_{it} \sim N(0, \sigma_e^2). \tag{A.143}$$

Note that $sd(\ln y_{it}) = (1 + \zeta \ln Y_t) sd(\log s_{it})$. The cyclical properties of inequality and income risk depend on the parameter ζ. When $\zeta > 0$ income risk is procyclical, it is countercyclical if $\zeta < 0$, and acyclical when $\zeta = 0$. Market clearing requires

$$C_t = \int c_{it} di = Y_t = N_t, \qquad A_t = \int a_{it} di = 0. \tag{A.144}$$

[27] Unless prices are fully rigid, the central bank has control over the expected real rate, not the realized real rate. The change in the expected real rate induces a re-evaluation of date zero assets or debts, which will in general affect demand. This effect is absent from the model we consider.

Table A-2: Calibration of HANK Model

Parameter		Perfect Aggregation	Imperfect Aggregation
ρ_e	Persistence of idiosyncratic shock	0.92	
σ_e	Stdev. of idiosyncratic shock	0.7	
ρ_r	Persistence of r	0.7	
σ_r	Stdev. of r	0.01	
ζ	Cyclical income risk	0.0	-0.5
$\underline{a}$	Minimum asset level	-0.000001	-1.0

Notes: Under "Perfect Aggregation" the IRFs to an expansionary monetary policy shock from the HANK model coincide with the IRFs from the representative agent (RA) model. Under "Imperfect Aggregation" the IRFs from the HANK model and the RA model are different.

The calibration of the model economy is summarized in Table A-2. We consider two parameterizations. Under the "Perfect Aggregation" calibration income risk is acyclical and the borrowing constraint $\underline{a}$ is close to zero. Conditional on an exogeneous interest rate sequence r_t, aggregate output Y_t is identical to the output in a representative agent New Keynesian (RANK) economy, obtained by setting $\sigma_\epsilon = 0$, that faces the identical interest rate series. Under the "Imperfect Aggregation" calibration, income risk is countercyclical and aggregate outputs in the HANK and RANK economies differ.

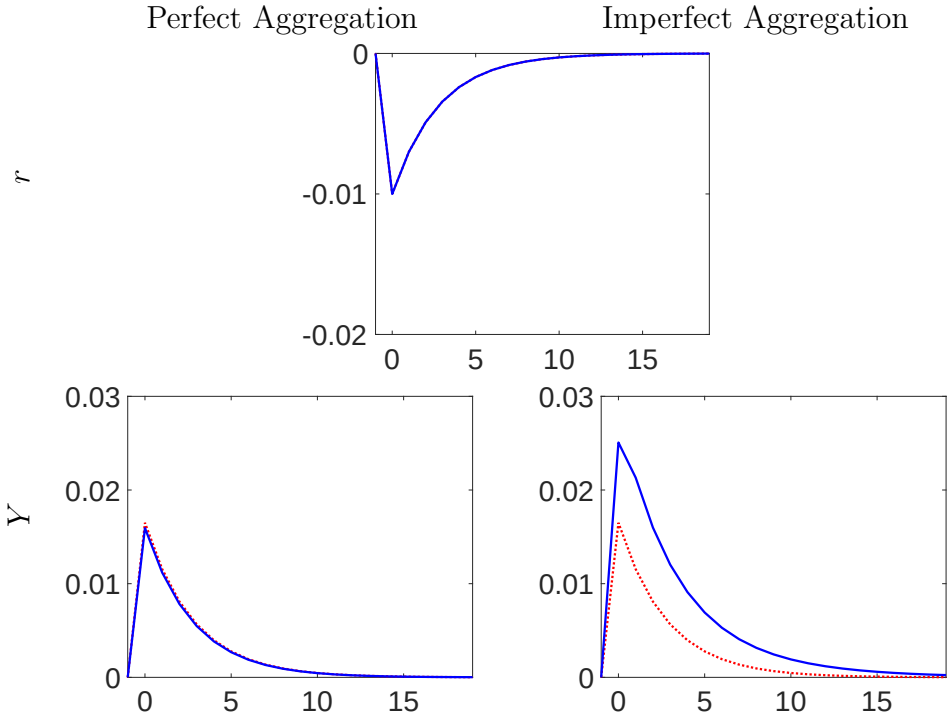
We solve the stylized HANK model $(r_t, Y_t, \{y_{it}\}_{i=1}^N)$ using the codes on the sequence-space Jacobian method provided by the authors.[28] Unlike the Winberry (2018) solution, the sequence-space Jacobian method does not deliver a VAR representation. Instead, it generates a moving average representation (impulse response function).

In Figure A-1 we show model-implied impulse responses of aggregate output to a monetary policy (r_t) shock. The figure compares responses obtained from the HANK model with responses from a RANK model. Under the "Perfect Aggregation" calibration the HANK and RA IRFs are identical. Under "Imperfect Aggregation" the countercyclical income risk leads to an amplification and output responds more strongly in the HANK model.

Model Estimation and IRFs. To simulate aggregate observations from the HANK model, we generate an innovation sequence $\{\epsilon_{r,t}\}_{t=-T_*}^T$ and multiply it by the moving average coefficients depicted in Figure A-1 to construct the aggregate variables r_t and Y_t. Conditional on Y_t, cross-sectional observations can be generated by simulating the idiosyncratic income process in (A.142) and (A.143). In the KS model simulations our cross-sectional observa-

[28]The code is from https://github.com/shade-econ/sequence-jacobian, and https://github.com/shade-econ/nber-workshop-2022. The code simulates $r_t = r_{t-1}^{\text{ex ante}}$ and we have to shift the simulated interest rate by one period before we can estimate our vector autoregressive models.

Figure A-1: Y's IRFs to an Expansionary Monetary Policy Shock (RA vs. HANK)



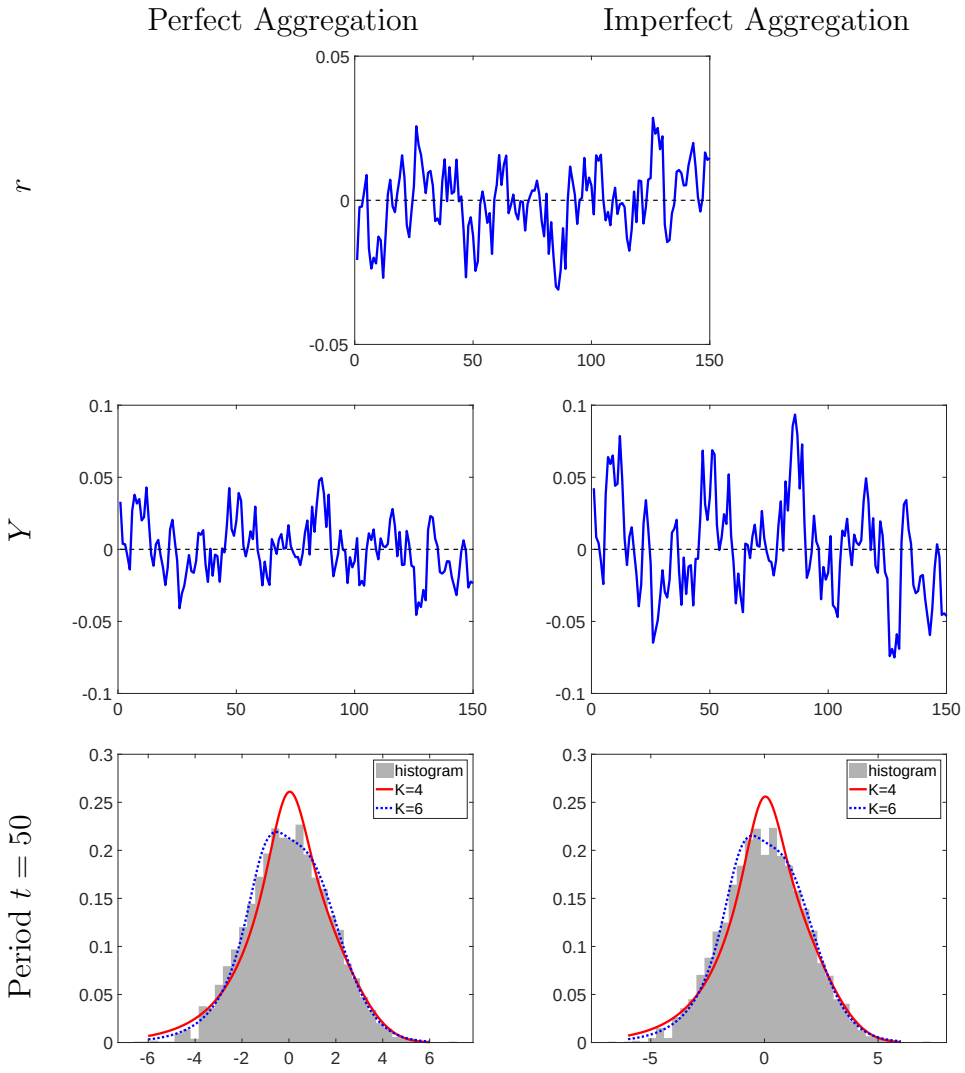
Notes: Solid blue: HANK IRF. Solid red: RANK IRF. Because r_t is exogenous, the r_t responses for HANK and RANK model are identical.

tions were repeated cross-sections. For the HANK model we generate a panel data set by simulating N time series $\ln e_{it}$, $t = 1, \ldots, T$ and then converting them into y_{it}s using (A.142). We generate a sample of size $N = 3,000$ and $T = 150$ and estimate the functional model and an aggregate VAR.[29] Even though our likelihood function was derived under the repeated cross-section assumption, in the HANK simulation the estimation of the functional model also works well based on panel data.

The panels in the first and second row of Figure A-2 show the simulated paths of r_t and Y_t under perfect and imperfect aggregation. We use identical paths for the exogeneous interest rate r_t under the two calibrations. The most notable difference in the simulated time series is that under imperfect aggregation output Y_t is more volatile than under perfect aggregation. This is consistent with the impulse responses plotted in Figure A-1. The panels in the last row of Figure A-2 show histograms of the household-level income data for period $t = 50$. For $K = 6$ the estimated density fits the cross-sectional data well.

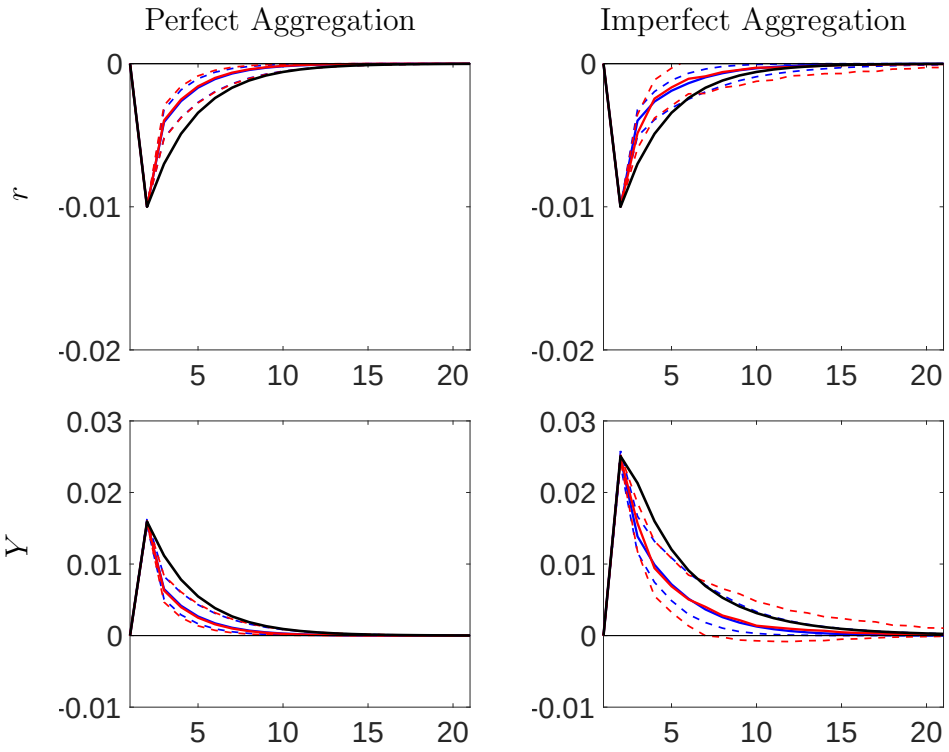
[29]Under perfect aggregation the series r_t and Y_t are perfectly correlated. To facilitate the estimation, we add a small amount of noise to the r_t series which breaks the perfect correlation without distorting the VAR estimates.

Figure A-2: Simulated Data



Estimated IRFs obtained from our functional model and an aggregate VAR in (r_t, Y_t) are plotted in Figure A-3. Under perfect aggregation, the estimated IRFs from the functional model and the aggregate model are essentially identical. Under imperfect aggregation there is a small discrepancy among the credible bands, but the posterior medians are on top of each other. Looking at Figure A-1, the imperfect aggregation in this model mainly creates an amplification of the output response, which does not translate into a large discrepancy among the estimated IRFs. We conclude that a match between IRFs from the functional and the aggregate model does not necessarily imply perfect aggregation.

Figure A-3: IRFs Comparison



Notes: Functional model IRFs: blue. Aggregate Model IRFs: red. Model specifications are selected to maximize MDD.

E More About the Empirical Analysis

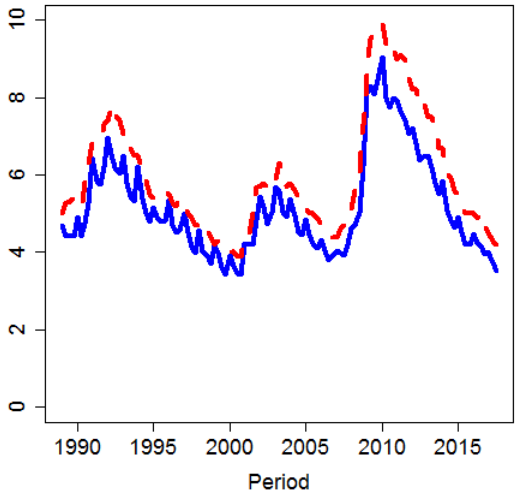
E.1 Data Construction

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) very closely; see the Figure A-4. 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.

In the left panel of Figure A-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

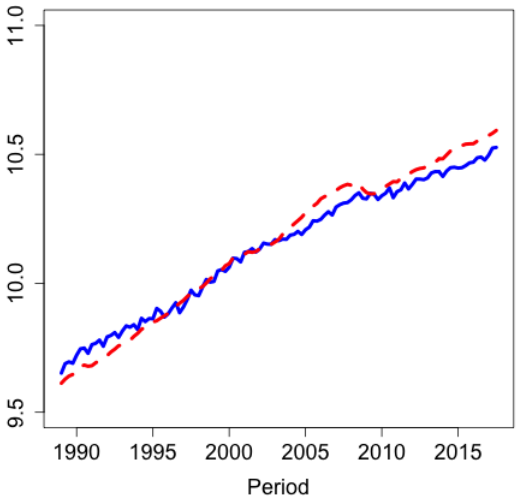
Figure A-4: CPS Unemployment



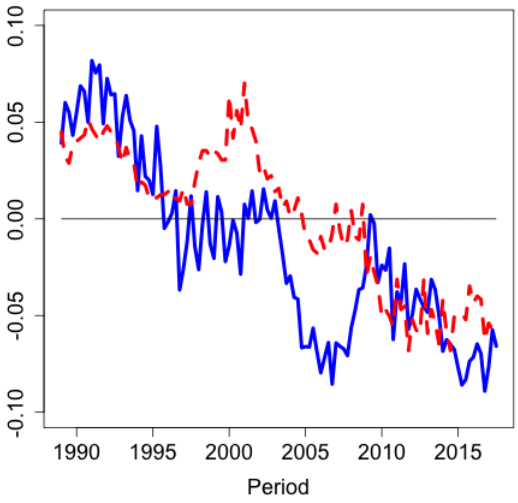
Notes: CPS unemployment rate (blue, solid) and aggregate unemployment rate (red, dashed).

Figure A-5: Earnings and GDP

Log Earnings and GDP



Log Earnings/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.

of 2/3 to account for the labor share.[30] After this re-scaling the mean of log earnings and log 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.

E.2 Data Transformations

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

$$x = g(z|\theta) = \frac{\ln(\theta z + (\theta^2 z^2 + 1)^{1/2})}{\theta} = \frac{\sinh^{-1}(\theta z)}{\theta}. \tag{A.145}$$

The transformation is plotted in the center panel Figure A-6 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

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

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

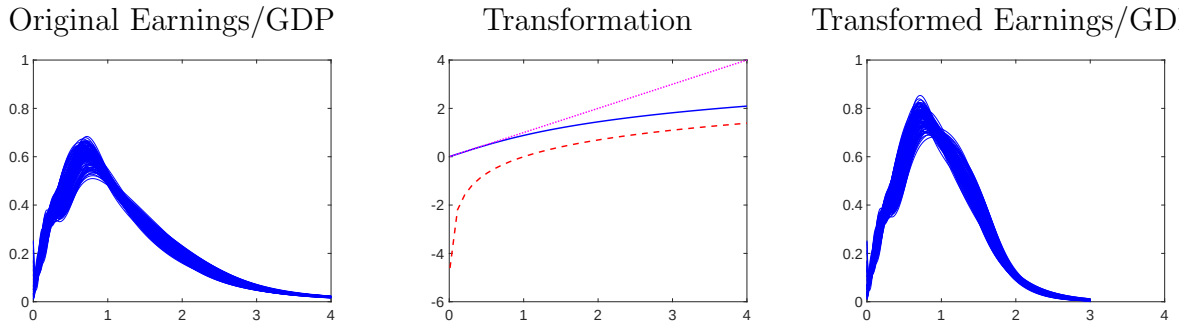
$$p_z(z) = p_x(g(z))|g'(z)|,$$

where

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

[30] Nominal per-capita GDP is obtained by multiplying real per-capita GDP by the GDP deflator (*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 A-6: 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.

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

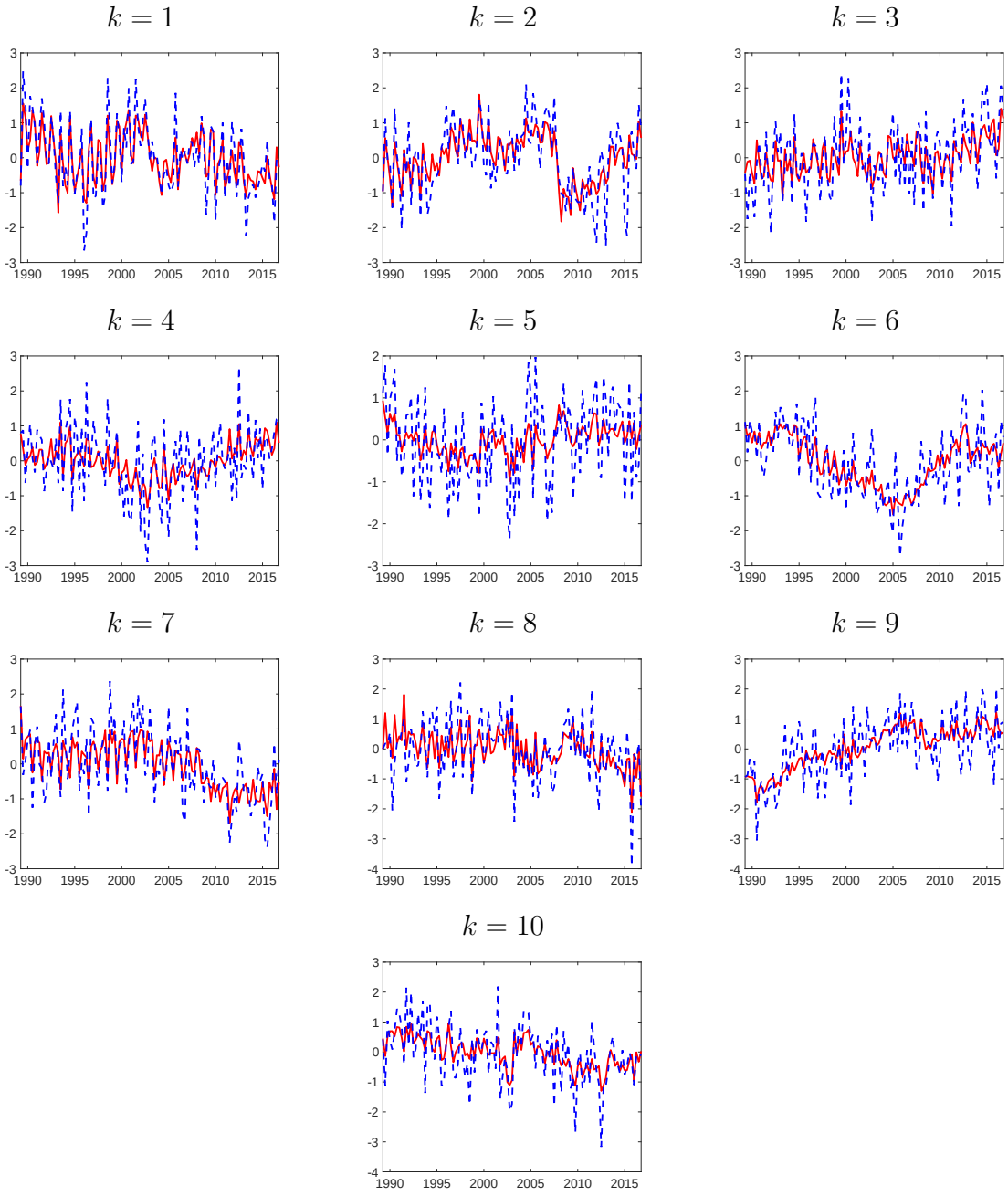
$$p_z(z_j) = p_x(x_j)\big|g'(g^{-1}(x_j))\big|,$$

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-6 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-7: 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 $\hat{a}_{k,t}$.

E.3 Estimated versus Smoothed Coefficients

Figure A-7 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.

E.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 $h = 1, \ldots, \bar{h}$:

$$q_\alpha^* = \text{argmax } e_i' \left[\sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} \Phi_1^j \Sigma_{tr} M q_\alpha q_\alpha' M' \Sigma_{tr}' (\Phi_1^j)' \right] e_i. \tag{A.146}$$

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

$$\begin{aligned} & e_i' \left[\sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} \Phi_1^j \Sigma_{tr} M q_\alpha q_\alpha' M' \Sigma_{tr}' \Phi_1^{j'} \right] e_i \qquad (A.147) \\ = & \sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} \text{tr} \left[(e_i e_i')(\Phi_1^j \Sigma_{tr} M)(q_\alpha q_\alpha')(M' \Sigma_{tr}' \Phi_1^{j'}) \right] \\ = & \sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} \text{tr} \left[(q_\alpha q_\alpha')(M' \Sigma_{tr}' \Phi_1^{j'})(e_i e_i')(\Phi_1^j \Sigma_{tr} M) \right] \\ = & q_\alpha' \left[\sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} (M' \Sigma_{tr}' \Phi_1^{j'})(e_i e_i')(\Phi_1^j \Sigma_{tr} M) \right] q_\alpha \\ = & q_\alpha' S q_\alpha. \end{aligned}$$

The optimization problem can therefore be expressed as Lagrangian

$$\mathcal{L} = q_\alpha' S q_\alpha - \lambda (q_\alpha' q_\alpha - 1), \tag{A.148}$$

which leads to the first-order condition

$$Sq_\alpha = \lambda q_\alpha. \tag{A.149}$$

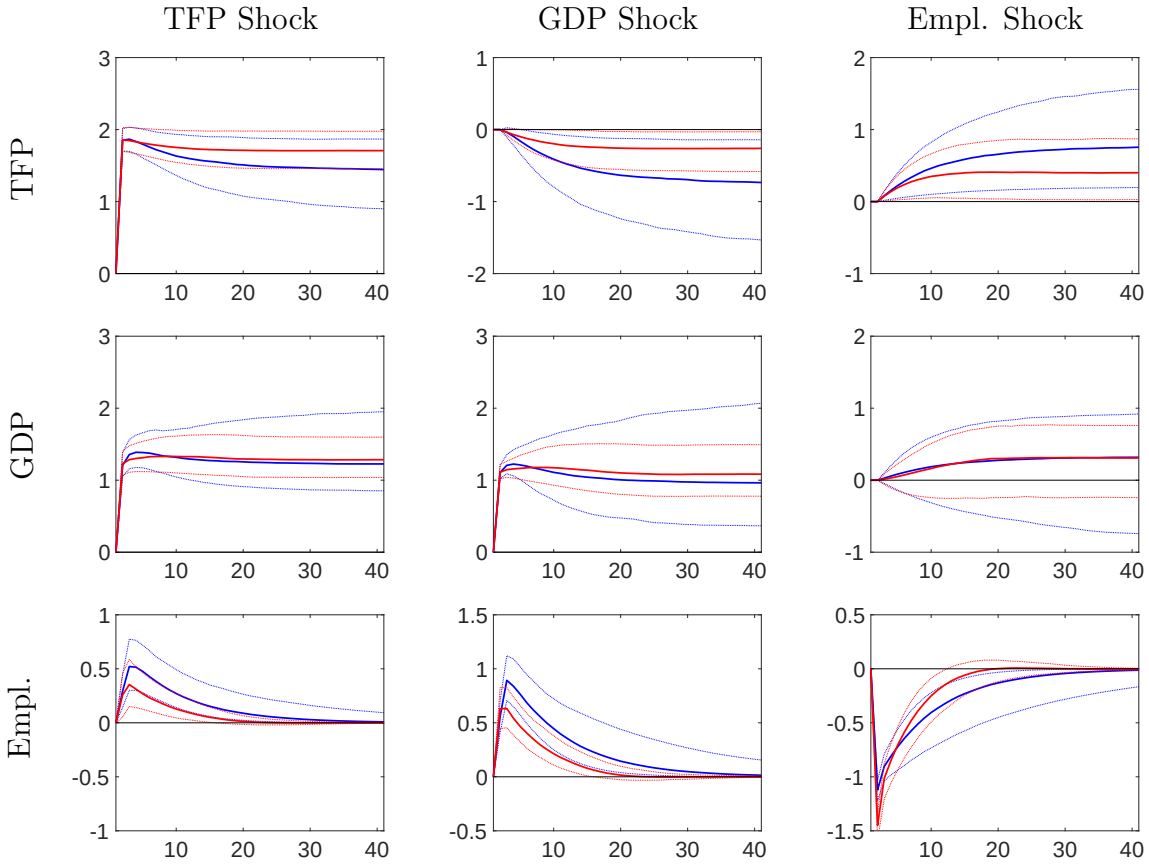
At the first-order condition, we obtain that $\mathcal{L} = \lambda$. Thus, the solution is obtained by finding the eigenvector associated with the largest eigenvalue of the matrix S.

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

In Figure A-8 we compare bands for the impulse responses of aggregate variables to aggregate shocks for $K = 10$ and $K = 22$.

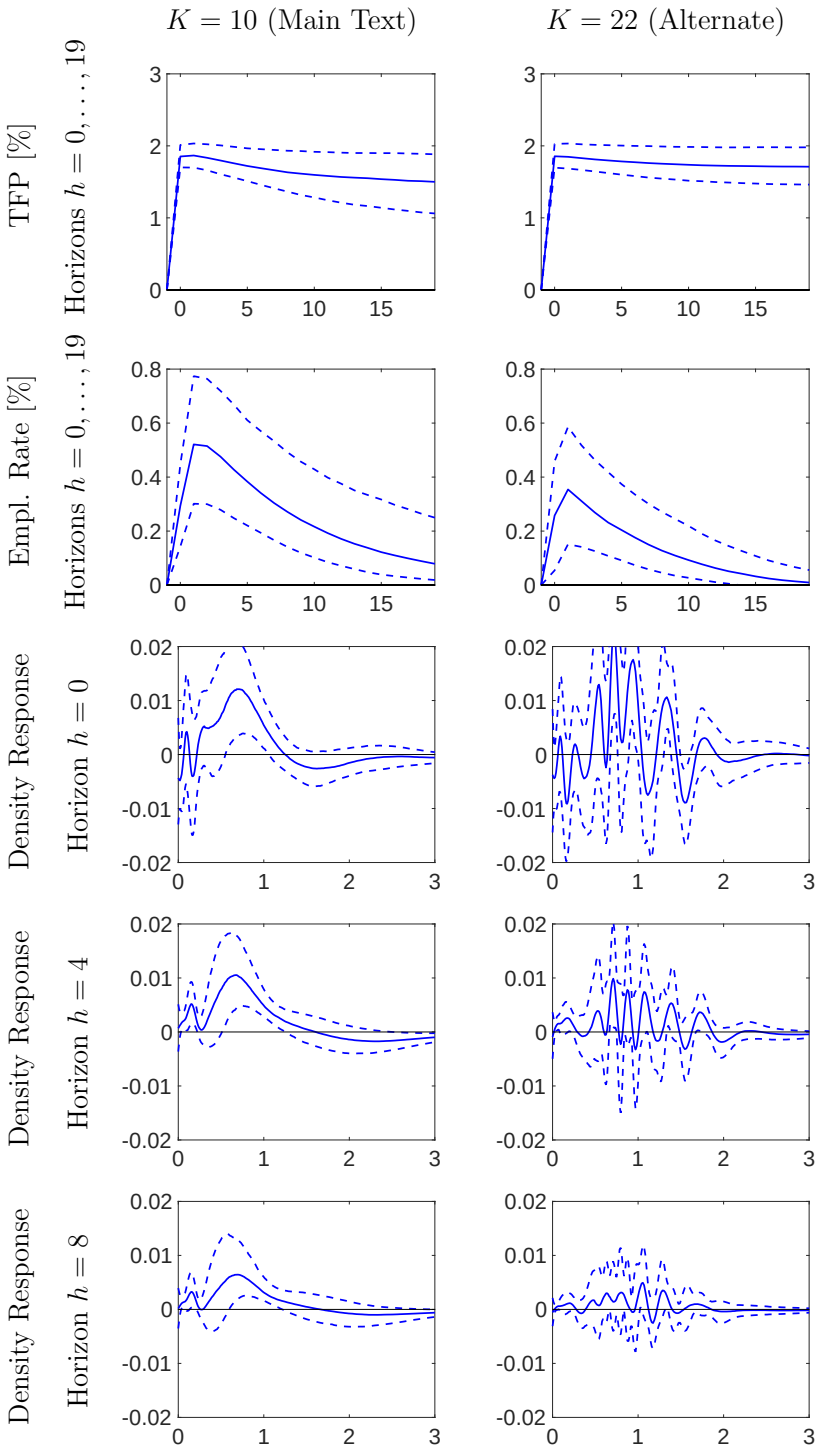
In Figures A-9 and A-10 we compare bands for the impulse responses of the cross-sectional density and inequality measures to aggregate shocks for $K = 10$ and $K = 22$. While the density responses look different, the responses of the inequality measures and percentiles derived from these densities are very similar.

Figure A-8: Responses of Aggregate Variables to Aggregate Shocks: $K = 10$ vs. $K = 22$



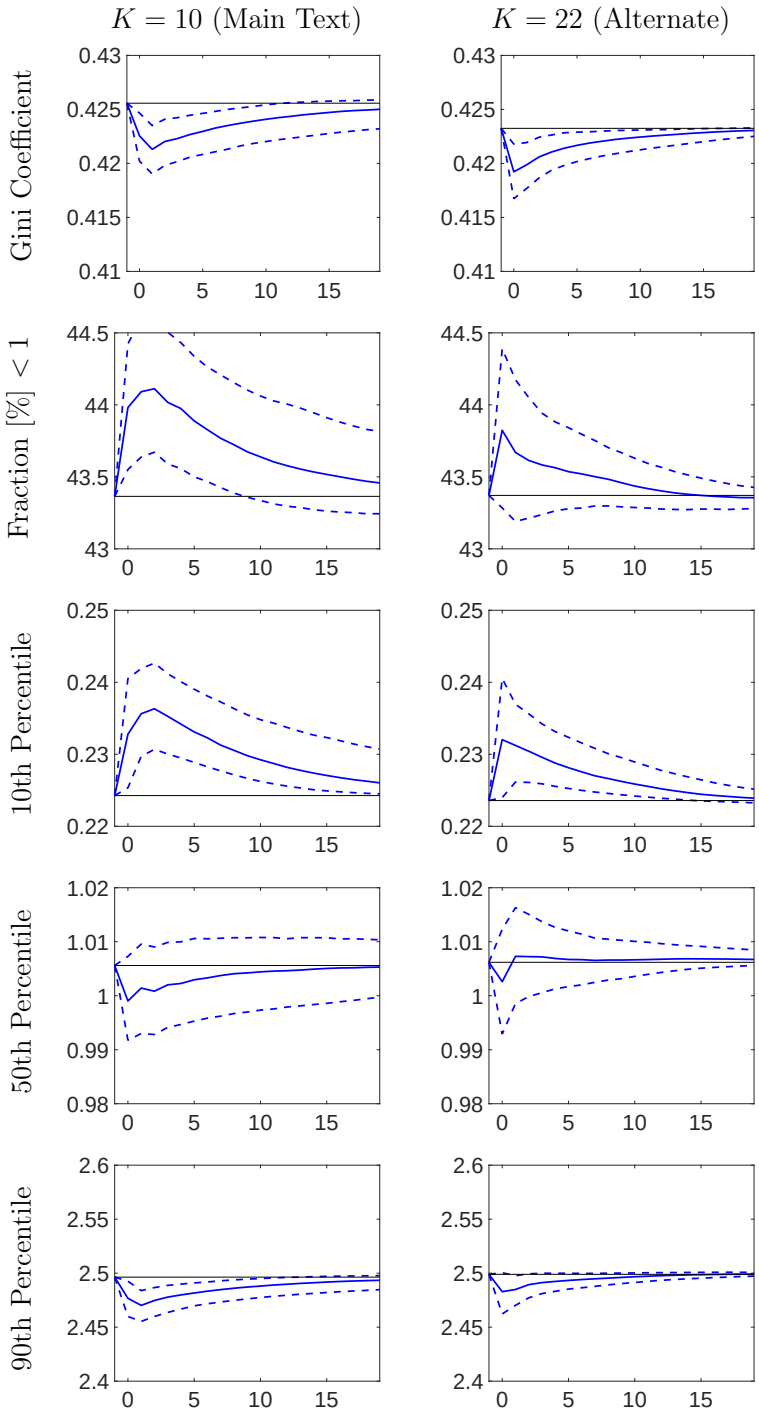
Notes: IRFs for three-standard deviation aggregate shocks (orthogonalized via Cholesky factorization; see (31)). Panels depict responses of the log level of TFP and GDP, scaled by 100, and responses of the employment rate in percent. The bands correspond to pointwise 10th and 90th percentiles of the 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-9: 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.

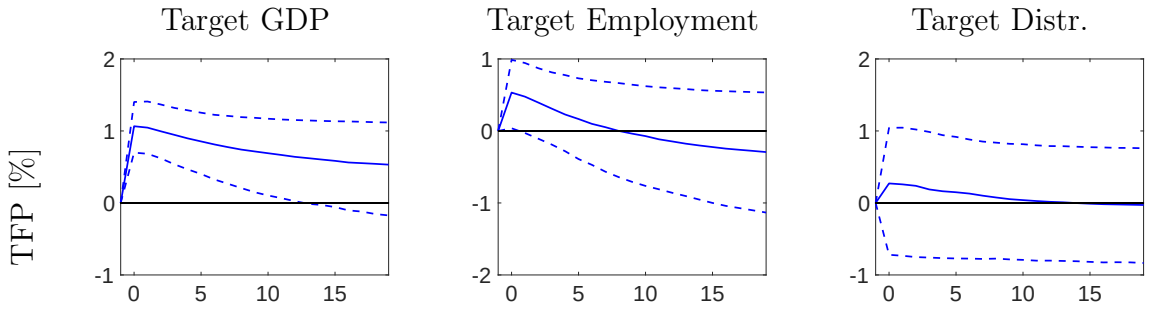
Figure A-10: Inequality Measure (Original Data) Responses to a TFP Shock: $K = 10$ vs. $K = 22$



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.

E.6 Business Cycle Anatomy

Figure A-11: Business Cycle Anatomy – TFP



Notes: 10th, 50th, and 90th percentiles of posterior distributions of IRFs. Sign normalizations: positive response upon impact of "GDP," "Employment," and 10th Percentile.