Heterogeneity and Aggregate Fluctuations

Minsu Chang*
University of Pennsylvania

Frank Schorfheide
University of Pennsylvania, CEPR, PIER, NBER

Preliminary Version: March 19, 2018

Abstract

We specify a vector autoregressive model that stacks macroeconomic aggregates and cross-sectional distributions to provide semi-structural evidence about the interaction of aggregate and distributional dynamics. The specification of our functional VAR is motivated by a linearization of a reduced-form model in which dynamics of aggregate variables and a function of the lagged cross-sectional distribution of individual-level decisions or states, and the units (households or firms) base their decisions on lagged macroeconomic aggregates and lagged cross-sectional distributions. To make the functional VAR analysis tractable, we approximate the log-densities of the cross-sectional distributions as well as the transition kernels in the functional VARs by sieves. We apply our techniques to study the dynamics of technology shocks, per capita GDP, employment, and the earnings distribution. (JEL C11, C32, C52, E32)

Key words: Econometric Model Evaluation, Earnings Distribution, Functional Vector Autoregressions, Heterogeneous Agent Models, Technology Shocks

* Correspondence: M. Chang, F. Schorfheide: University of Pennsylvania, Department of Economics, Philadelphia, PA 19104. Email: minsuc@sas.upenn.edu (Chang), schorf@ssc.upenn.edu (Schorfheide). We thank Yongsung Chang, Xiaohong Chen, and Dirk Krüger for helpful suggestions. Schorfheide gratefully acknowledges financial support from the National Science Foundation under Grant 1424843.

1 Introduction

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

The goal of this paper is to develop and apply econometric tools that can provide semi-structural evidence about the interaction of aggregate and distributional dynamics. More specifically, we specify a vector autoregressive model that stacks macroeconomic aggregates and cross-sectional distributions. We motivate the specification of our functional VAR by a linearization of a reduced form model in which dynamics of aggregate variables and a function of the lagged cross-sectional distribution of individual-level decisions or states, and the units (households or firms) base their decisions on lagged macroeconomic aggregates and lagged cross-sectional distributions. To make the functional VAR analysis tractable, we approximate the log-densities of the cross-sectional distributions as well as the transition kernels in the functional VARs by sieves. This leads to a finite-dimensional VAR in terms of the aggregate variables as well as the coefficients of the sieve approximations.

The functional VAR analysis is implemented with spline basis functions. Log-splines are a popular tool in the statistics literature to approximate densities non-parametrically. In a first step, for each time period we estimate the coefficients of the log-spline density approximation. We then condition on these estimates and estimate a Bayesian VAR on the macroeconomic variables and the spline coefficients. Based on the estimated VAR, we can construct impulse responses of cross-sectional distributions to aggregate shocks as well as responses of aggregate variables to shocks that primarily move the cross-sectional distribution. We can use historical decompositions to track the relative importance of aggregate and distributional shocks over time. We illustrate our toolkit in a simulation study and then apply it to study the interaction of total factor productivity, per capita GDP, employment, and the cross-sectional distribution of earnings.

Our paper is related to several strands of literatures. First, there is a large body of research that builds on the seminal paper by Krusell and Smith (1998), henceforth KS. KS combine a neoclassical stochastic growth model with a heterogeneous agent economy in which households face uninsurable idiosyncratic income risk. The equilibrium in the KS model has exactly the features described above: households' decisions depend on the aggregate technology shock as well as the cross-sectional distribution of skills and wealth. In turn, the entirety of the household-level decisions determine the cross-sectional distribution. One of the key findings was that in the stationary stochastic equilibrium, the behavior of the macroeconomic aggregates can be almost perfectly described using only the mean of the wealth distribution. However, in extensions of the benchmark model, this approximate aggregation result is no longer true and these models exhibit a richer interaction between aggregate and distributional dynamics.

Chang, Kim, and Schorfheide (2013) consider a HA model with indivisible labor supply and use an RA model to approximate the aggregate dynamics implied by the HA model. They find that once a labor supply shock is included in the RA model, the RA model does quite well approximating the aggregate dynamics of the HA model. This suggests that modeling the dynamics of cross-sections is not of first-order importance if the analysis focuses on macroeconomic aggregates. However, there is an important caveat: preference and technology parameter estimates of the RA model are not invariant to policy changes and the bias in the RA models policy predictions is large compared to predictive intervals that reflect parameter uncertainty.

Both our functional VAR analysis as well as the solution of models with heterogeneous agents requires a parsimonious representation of cross-sectional distributions. For instance, Krusell and Smith (1998) represent the wealth distribution by its mean. Reiter (2010) uses a discretization of the wealth distribution and a Markov transition matrix to characterize movements in the wealth distribution. The transition probabilities are functions of the aggregate states. Algan, Allais, and Den Haan (2008) and Winberry (2017) use moments to characterize the distribution of productivity and capital. From these moments, it is then possible to recover a density with a class of distributions that belong to the exponential family.[1] Childers (2015) considers linearization methods for models with function-valued states, using a wavelet representation. These approximations can be interpreted as sieves with different types of basis functions. Our analysis uses spline basis functions to approximate log densities, which dates back to Kooperberg and Stone (1990).

[1]This approach is somewhat related to the density estimator discussed in Efron and Tibshirani (1996).

Finally, there is an extensive literature on the statistical analysis of functional time series. General treatments are provided in the books by Bosq (2000), Ramsey and Silverman (2005), and Horvath and Kokoszka (2012). Chang, Kim, and Park (2016) use a functional time series process to capture the evolution of densities. They are focusing on non-stationary directions in the functional time series.

The remainder of this paper is organized as follows. In Section 2 we outline a basic reduced-form model in which the cross-sectional distribution matters for unit-level and aggregate dynamics. We then linearize this model and show that it can lead us to a restricted functional VAR. In Section 3 we remove some of the restrictions to make the empirical specification more flexible. Implementation details such as the choice of basis functions and the specification of the coefficient VAR are discussed in Section 4. A numerical illustration is provided in Section 5 and Section 6 contains the empirical application. Finally, Section 7 concludes. Detailed derivations and additional results are relegated to the Online Appendix.

2 A Basic Model

We begin with a model in which the evolution of an individual-level variable x_{it} depends on the cross-sectional distribution of x_{it-1} as well as a lagged aggregate variable Z_{t-1}. Likewise, the evolution of Z_t depends on its lag and on the cross-sectional distribution of x_{it-1}. Let $p_t(x)$ denote the cross-sectional distribution of x_{it}. Then

$$Z_t = \psi Z_{t-1} + \int \Psi(x) p_{t-1}(x) dx + \sigma \eta_t \tag{1}$$

$$x_{it} = \phi x_{it-1} + \int \Phi(x) p_{t-1}(x) dx + \gamma Z_{t-1} + \epsilon_{it}. \tag{2}$$

Based on the x_{it} equation, we can obtain the cross-sectional distribution of x_{it}. Let $f_\epsilon(\cdot)$ denote the density of ϵ_{it}, then

$$p_t(x) = \int f_\epsilon \left(x - \phi \tilde{x} - \int \Phi(x) p_{t-1}(x) dx - \gamma Z_{t-1} \right) p_{t-1}(\tilde{x}) d\tilde{x}. \tag{3}$$

We will make the following additional assumptions

$$\eta_t \sim N(0,1), \quad \epsilon_{it} \sim N(0, \sigma_\epsilon^2), \quad \Psi(x) = \psi_p x, \quad \Phi(x) = \phi_p x.$$

Direct Solution of Model. Suppose that $\sigma = 0$ and that there exists no aggregate uncertainty. It can be verified that a time-invariant solution to the model described in (1)

to (2) is given by

$$Z_t = Z_* = 0, \quad p_t(x) = p_*(x) \sim N\big(\mu_{x,*}, \sigma^2_{x,*}\big), \quad \mu_{x,*} = 0, \quad \sigma^2_{x,*} = \frac{\sigma^2_\epsilon}{1-\phi^2}. \tag{4}$$

We will now allow for aggregate fluctuations, i.e., $\sigma > 0$, and assume that the x_{it} process was initialized with draws from $p_*(x)$. Notice that (2) preserves the normality of the x_{it}'s over time. Moreover, the cross-sectional variance of x_{it} does not change: $\sigma^2_{x,t} = \sigma^2_{x,*}$. Thus, the evolution of $p_t(x)$ can be characterized by the evolution of its mean. Let $\mu_{x,t}$ denote the cross-sectional mean of the x_{it}s in period t. This mean is a function of the sequence of aggregate shocks $\{Z_t\}$. Taking expectations of both sides of (2) yields

$$\mu_{x,t} = (\phi + \phi_p)\mu_{x,t-1} + \gamma Z_{t-1}, \tag{5}$$

which leads to

$$p_t(x) \sim N\big(\mu_{x,t}, \sigma^2_x\big). \tag{6}$$

Because the cross-sectional variance of $x_{i,t}$ is time-invariant, we drop t and $*$, respectively, from the subscript.

We proceed by representing the log-density of $p_t(x)$ through a sieve. Because we are dealing with normal distributions, we will use a polynomial basis and the sieve approximation will be exact. Let

$$\zeta_0(x) = 1, \quad \zeta_1(x) = x, \quad \zeta_2(x) = x^2 \tag{7}$$

and write

$$\ell_t(x) = \ln p_t(x) = \sum_{k=0}^{2} \alpha_{k,t}\zeta_k(x). \tag{8}$$

For the normal distribution derived previously, the coefficients for the sieve approximation are

$$\alpha_{0,t} = -\frac{1}{2}\ln(2\pi\sigma^2_x) - \frac{1}{2\sigma^2_x}\mu^2_{x,t}, \quad \alpha_{1,t} = \frac{1}{\sigma^2_x}\mu_{x,t}, \quad \alpha_{2,t} = -\frac{1}{2\sigma^2_x}.$$

The evolution of the coefficients is given by

$$\begin{aligned} \alpha_{0,t} &= -\frac{1}{2}\ln(2\pi\sigma^2_x) - \frac{1}{2\sigma^2_x}\left((\phi+\phi_p)\sigma^2_x\alpha_{1,t-1} + \gamma Z_{t-1}\right)^2 \\ \alpha_{1,t} &= (\phi+\phi_p)\alpha_{1,t-1} + \frac{\gamma}{\sigma^2_x}Z_{t-1} \\ \alpha_{2,t} &= -\frac{1}{2\sigma^2_x}. \end{aligned} \tag{9}$$

Notice that (i) the coefficient $\alpha_{0,t}$ makes sure that the density is normalized to one. (ii) The law of motion for $\alpha_{1,t}$ is linear and depends on the aggregate state Z_{t-1}. (iii) The model can

be interpreted as a functional autoregressive model in which $p_t(x)$ and Z_t evolve jointly over time.

Perturbation Solution of the Model. We can write the system that comprises (1) and (3) as

$$\mathbb{E}_t\big[F\big(Z_{t+1}, p_{t+1}(\cdot), Z_t, p_t(\cdot), \sigma\big)\big] = 0. \tag{10}$$

Here σ is the perturbation parameter. By setting $\sigma = 0$ we can shut down aggregate uncertainty and construct a steady state that satisfies

$$F\big(Z_*, p_*(\cdot), Z_*, p_*(\cdot), 0\big) = 0. \tag{11}$$

This steady state reflects only idiosyncratic uncertainty and its objects are a value Z_* for the aggregate variable and a distribution $p_*(x)$ for the unit-specific variables. We previously derived the steady state and its values are given in (4).

The steady state $p_*(x)$ satisfies the functional equation

$$p_*(x) = \int f_\epsilon\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) d\tilde{x}. \tag{12}$$

Although we know the solution to this functional equation, we will now verify it by direct calculation. This calculation will become useful later on, when we linearize the functional equation. Define

$$\omega_\epsilon = \frac{1}{\sigma_\epsilon^2}, \quad \omega_x = \frac{1}{\sigma_x^2} = \frac{1-\phi^2}{\sigma_\epsilon^2}.$$

Then

$$\begin{aligned}
&\int f_\epsilon\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) d\tilde{x} \\
&= (2\pi)^{-1}(\sigma_\epsilon^2\sigma_x^2)^{-1/2} \int \exp\left\{-\frac{1}{2}\left[\omega_\epsilon x^2 - 2\omega_\epsilon\phi x\tilde{x} + \omega_\epsilon\phi^2\tilde{x}^2 + \omega_x\tilde{x}^2 - 2\omega_x\mu_x\tilde{x} + \omega_x\mu_x^2\right]\right\} d\tilde{x} \\
&= (2\pi)^{-1}(\sigma_\epsilon^2\sigma_x^2)^{-1/2} \int \exp\left\{-\frac{1}{2}\left[(\omega_\epsilon\phi^2 + \omega_x)\tilde{x}^2 - 2(\omega_\epsilon\phi x + \omega_x\mu_x)\tilde{x} + \omega_\epsilon x^2 + \omega_x\mu_x^2\right]\right\} d\tilde{x}
\end{aligned}$$

Now define

$$\begin{aligned}
\sigma_{\tilde{x}}^2 &= \frac{1}{\omega_\epsilon\phi^2 + \omega_x} = \frac{\sigma_\epsilon^2}{\phi^2 + 1 - \phi^2} = \sigma_\epsilon^2 \\
\mu_{\tilde{x}} &= \sigma_{\tilde{x}}^2(\omega_\epsilon\phi x + \omega_x\mu_x) = \phi x + (1-\phi^2)\mu_x = \phi x.
\end{aligned} \tag{13}$$

Here we used $\mu_x = 0$. Thus,

$$\begin{aligned}
&\int f_\epsilon\left(x-\phi\tilde{x}\right)p_*(\tilde{x})d\tilde{x} \qquad (14)\\
&= (2\pi)^{-1}(\sigma_\epsilon^2\sigma_x^2)^{-1/2}\int\exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_\epsilon^2}(\tilde{x}-\mu_{\tilde{x}})^2-\frac{1}{\sigma_\epsilon^2}\mu_{\tilde{x}}^2+\omega_\epsilon x^2\right]\right\}d\tilde{x}\\
&= (2\pi\sigma_x^2)^{-1/2}\exp\left\{-\frac{1}{2}\left[-\frac{1}{\sigma_\epsilon^2}\mu_{\tilde{x}}^2+\omega_\epsilon x^2\right]\right\}\int(2\pi\sigma_\epsilon^2)^{-1/2}\exp\left\{-\frac{1}{2\sigma_\epsilon^2}(\tilde{x}-\mu_{\tilde{x}})^2\right\}d\tilde{x}\\
&= (2\pi\sigma_x^2)^{-1/2}\exp\left\{-\frac{1}{2}\left[-\frac{1}{\sigma_\epsilon^2}\phi^2x^2+\frac{1}{\sigma_\epsilon^2}x^2\right]\right\}\\
&= (2\pi\sigma_x^2)^{-1/2}\exp\left\{-\frac{1}{2\sigma_x^2}x^2\right\}=p_*(x),
\end{aligned}$$

as desired.

We will now linearize the function $F\big(Z_{t+1},p_{t+1}(\cdot),Z_t,p_t(\cdot),\sigma\big)$ around $(Z_*,p_*,Z_*,p_*,0)$. However, rather than expanding the function in terms of the densities $p(x)$, we will make an expansion in terms of log densities, because, as in the direct solution of the model, we will represent the log densities by a sieve. Let

$$p_t(x)=\exp\{\ell_t(x)\},\quad p_*(x)=\exp\{\ell_*(x)\}.$$

We provide a more detailed discussion of linearizations in function spaces in the Appendix. Define

$$\tilde{Z}_t=Z_t-Z_*,\quad \tilde{\ell}_t(x)=\ell_t(x)-\ell_*(x).$$

We continue to use the sieve in (8) to represent $\ell_t(x)$. Because the sieve is linear in the $\alpha_{k,t}$ coefficients, we can generate $\tilde{\ell}_t(x)$ and $\ell_*(x)$ using coefficients $\tilde{\alpha}_{k,t}$ and $\alpha_{k,*}$, respectively, that satisfy

$$\tilde{\alpha}_{k,t}=\alpha_{k,t}-\alpha_{k,*}.$$

Using the linearization approach, we can now approximate the law of motion for Z_t as

$$\begin{aligned}
\tilde{Z}_t &= \psi\tilde{Z}_{t-1}+\int\psi_p x p_*(x)\tilde{l}_{t-1}(x)dx+\sigma\eta_t \qquad (15)\\
&= \psi\tilde{Z}_{t-1}+\sum_{k=0}^{2}\alpha_{k,t-1}\psi_p\int x^{k+1}p_*(x)dx+\sigma\eta_t.
\end{aligned}$$

Note that according to this approximation the conditional mean of $\tilde{Z}_t$ is linear in $\tilde{Z}_{t-1}$ and the $\alpha_{k,t-1}$'s. Because $p_*(x)\sim N(0,\sigma_x^2)$, the first and third moments are zero and the second moment is σ_x^2. Thus,

$$\tilde{Z}_t=\psi\tilde{Z}_{t-1}+\psi_p\sigma_x^2\alpha_{1,t-1}+\sigma\eta_t. \qquad (16)$$

Because $\mu_{x,t-1} = \sigma_x^2 \alpha_{1,t-1}$ this approximation result is exact and consistent with the direct calculation.

We now linearize (3) to obtain an approximate law of motion for $\tilde{\ell}_t(x)$. Using the fact that $Z_* = 0$ and $\int \Phi(x) p_*(x) dx = 0$ we obtain

$$\begin{aligned} p_*(x)\tilde{l}_t(x) &= -\gamma \left[\int f_\epsilon^{(1)}\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) d\tilde{x} \right] \tilde{Z}_{t-1} \\ &\quad - \left[\int f_\epsilon^{(1)}\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) d\tilde{x} \right] \int \phi_p \tilde{x} p_*(\tilde{x}) \tilde{\ell}_{t-1}(\tilde{x}) d\tilde{x} \\ &\quad + \int f_\epsilon\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) \tilde{\ell}_{t-1}(\tilde{x}) d\tilde{x}. \end{aligned} \tag{17}$$

We proceed by evaluating the three terms that appear on the r.h.s. of (17). First,

$$\begin{aligned} &\int f_\epsilon^{(1)}\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) d\tilde{x} \\ &= -\frac{1}{\sigma_\epsilon^2} \int (x - \phi\tilde{x}) f_\epsilon\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) d\tilde{x} \\ &= -\frac{1}{\sigma_\epsilon^2} x \int f_\epsilon\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) d\tilde{x} + \frac{\phi}{\sigma_\epsilon^2} \int \tilde{x} f_\epsilon\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) d\tilde{x} \\ &= -\frac{1}{\sigma_\epsilon^2} x p_*(x) + \frac{\phi}{\sigma_\epsilon^2} \mu_{\tilde{x}} p_*(x) \\ &= -\frac{x}{\sigma_x^2} p_*(x). \end{aligned}$$

The first equality follows from differentiating the Normal pdf. The third equality can be obtained by adapting the calculations in (14). The last equality utilizes the definition of $\mu_{\tilde{x}}$ in (13). Second, using the same manipulations as in (15) we obtain that

$$\phi_p \int \tilde{x} p_*(\tilde{x}) \tilde{\ell}_{t-1}(x)(\tilde{x}) d\tilde{x} = \phi_p \sigma_x^2 \alpha_{1,t-1}.$$

Third, we have

$$\begin{aligned} &\int f_\epsilon\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) \tilde{\ell}_{t-1}(\tilde{x}) d\tilde{x} \\ &= \sum_{k=0}^{K} \alpha_{k,t-1} \int \zeta_k(\tilde{x}) f_\epsilon\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) d\tilde{x} \\ &= \sum_{k=0}^{K} \alpha_{k,t-1} \int \tilde{x}^k f_\epsilon\left(x - \phi\tilde{x}\right) p_*(\tilde{x}) d\tilde{x} \\ &= \left[\tilde{\alpha}_{0,t-1} + \tilde{\alpha}_{1,t-1}\mu_{\tilde{x}} + \tilde{\alpha}_{2,t-1}(\mu_{\tilde{x}}^2 + \sigma_{\tilde{x}}^2)\right] p_*(x) \\ &= \left[\tilde{\alpha}_{0,t-1} + \tilde{\alpha}_{1,t-1}\phi x + \tilde{\alpha}_{2,t-1}(\phi^2 x^2 + \sigma_\epsilon^2)\right] p_*(x). \end{aligned}$$

The last two equalities follow from (14) and (13). Plugging these expressions into (17) yields the following law of motion

$$\begin{aligned} &\tilde{\alpha}_{0,t} + \tilde{\alpha}_{1,t}x + \tilde{\alpha}_{2,t}x^2 \\ &= \frac{1}{\sigma_x^2}\gamma\tilde{Z}_{t-1}x + \phi_p\tilde{\alpha}_{1,t-1}x + \tilde{\alpha}_{0,t-1} + \tilde{\alpha}_{1,t-1}\phi x + \tilde{\alpha}_{2,t-1}(\phi^2x^2 + \sigma_\epsilon^2). \end{aligned} \tag{18}$$

We deduce that the laws of motion for $\tilde{\alpha}_{k,t}$ are given by

$$\begin{aligned} \tilde{\alpha}_{0,t} &= \tilde{\alpha}_{0,t-1} + \tilde{\alpha}_{2,t-1}\sigma_\epsilon^2 \\ \tilde{\alpha}_{1,t} &= (\phi + \phi_p)\tilde{\alpha}_{1,t-1} + \frac{\gamma}{\sigma_x^2}\tilde{Z}_{t-1} \\ \tilde{\alpha}_{2,t} &= \tilde{\alpha}_{2,t-1}\phi^2. \end{aligned} \tag{19}$$

The law of motion for $\tilde{\alpha}_{1,t}$ matches that obtained from the direct calculation. The law of motion for $\tilde{\alpha}_{2,t}$ is consistent with $\tilde{\alpha}_{2,t} = 0$, which means that $\alpha_{2,t} = \alpha_{2,*}$ is constant. The law of motion for $\alpha_{0,t}$ needs to be adjusted to make sure the densities integrate to one in every period.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote the underlying probability space for the aggregate uncertainty. The following objects in the above analysis are non-stochastic: the steady state values

$$Z_*,\ p_*(x),\ \ell_*(x),\ \mu_{x,*}, \sigma_x^2, \alpha_{k,*},$$

and the basis functions $\zeta_k(x)$. The following objects are stochastic (which we now highlight through the argument ω):

$$Z_t(\omega),\ \tilde{Z}_t(\omega),\ p_t(x;\omega),\ \ell_t(x;\omega),\ \tilde{\ell}_t(x;\omega),\ \mu_{x,t}(\omega),\ \alpha_{k,t}(\omega),\ \tilde{\alpha}_{k,t}(\omega).$$

3 A More Flexible Specification

At a more abstract level, the model sketched in Section 2 takes the form of a first-order functional vector autoregression, which we write as

$$\begin{aligned} \tilde{Z}_t(\omega) &= b_{zz}\tilde{Z}_{t-1}(\omega) + \int B_{z\ell}(\tilde{x})p_*(\tilde{x})\tilde{\ell}_{t-1}(\tilde{x};\omega)d\tilde{x} + \sigma\eta_t(\omega) \\ p_*(x)\tilde{\ell}_t(x;\omega) &= B_{\ell z}(x)\tilde{Z}_{t-1}(\omega) + \int B_{\ell\ell}(x,\tilde{x})p_*(\tilde{x})\tilde{\ell}_{t-1}(\tilde{x})d\tilde{x}. \end{aligned} \tag{20}$$

For now, we assume that both Z_t and x are scalar. The objects to be estimated in this model are the coefficient b_{zz}, the functions $B_{z\ell}(x)$ and $B_{\ell z}(x)$, and the transition kernel $B_{\ell\ell}(x,\tilde{x})$.

The first equation in (20) resembles (15) and the second equation mimics (17). This model is singular, in the sense that the transition equation for the density is deterministic.

The function $B_{z\ell}(x)$ determines which features of the cross-sectional distribution feed back into the aggregate variable $\tilde{Z}_t(\omega)$. In Section 2 we made it proportional to x so that it would capture the mean of the cross-sectional distribution. One could easily generate it to capture higher-order moments or tail probabilities, e.g.,

$$B_{z\ell}(x) = b_{z\ell,1}x + b_{z\ell,2}x^2 + b_{z\ell,3}x^3 \quad \text{or} \quad B_{z\ell}(x) = b_{z\ell,1}\mathbb{I}\{x \leq \underline{x}\} + b_{z\ell,2}\mathbb{I}\{x \geq \bar{x}\},$$

where $\mathbb{I}\{x \leq a\}$ is the indicator function that is one if $x \leq a$ and is zero otherwise.

The function $B_{\ell x}(x)$ captures how the aggregate shock $\tilde{Z}_{t-1}(\omega)$ alters the cross-sectional distribution. In the basic model in Section 2 we assumed that $\hat{Z}_{t-1}$ shifts the mean of all cross-sectional units by $\gamma\hat{Z}_{t-1}$, which led to

$$B_{\ell z}(x) = -\gamma \int f_\epsilon^{(1)}(x - \phi(\tilde{x}))p_*(\tilde{x})d\tilde{x} = \frac{\gamma}{\sigma_x^2}xp_*(x).$$

Alternatively, suppose that the effect of the aggregate shock on the mean of x_{it} is a function of x_{it-1}, e.g., $\gamma x_{it-1}Z_{t-1}$, then we obtain

$$B_{\ell z}(x) = -\gamma \int \tilde{x} f_\epsilon^{(1)}(x - \phi(\tilde{x}))p_*(\tilde{x})d\tilde{x} = \gamma\left(\frac{1}{\sigma_\epsilon^2}x - \frac{\phi}{\sigma_\epsilon^2}(\phi^2x^2 + \sigma_\epsilon^2)\right)p_*(x).$$

Thus, $B_{\ell z}(x)$ is a quadratic function of x.

Finally, consider the transition kernel $B_{\ell\ell}(x, \tilde{x})$. In the basic model, it is given by

$$B_{\ell\ell}(x, \tilde{x}) = -\phi_p\left[\int f_\epsilon^{(1)}\left(x - \phi\bar{x}\right)p_*(\bar{x})d\bar{x}\right]\tilde{x} + f_\epsilon\left(x - \phi\tilde{x}\right) = \frac{\phi}{\sigma_x^2}xp_*(x) + p_N(x|\phi\tilde{x}, \sigma_\epsilon^2).$$

Here the first term captures the effect of letting $p_{t-1}(x)$ deviate from the steady state $p_*(x)$ on the conditional mean of the x_{it}'s and the second term captures the effect of perturbing $p_{t-1}(x)$ on the convolution of $f_\epsilon(\cdot)$ and $p_{t-1}(\tilde{x})$. To make the model more flexible, we can generalize one or both of these terms; but the second term should remain a transition density.

3.1 Approximating Log Densities

Recall that we represented the densities as follows:

$$p_t(x) = \exp\{\ell_t(x)\}, \quad p_*(x) = \exp\{\ell_*(x)\}, \quad \tilde{\ell}_t(x) = \ell_t(x) - \ell_*(x). \tag{21}$$

We will continue to approximate log density functions by sieves:

$$\ell_*(x) = \sum_{k=0}^{K} \alpha_{k,*}\zeta_k(x), \quad \tilde{\ell}_t(x) = \sum_{k=0}^{K} \tilde{\alpha}_{k,t}\zeta_k(x), \quad \ell_t(x) = \sum_{k=0}^{K} \alpha_{k,*}\zeta_k(x), \tag{22}$$

which leads to

$$\tilde{\ell}_t(x) = \sum_{k=0}^{K} (\alpha_{k,t} - \alpha_{k,*})\zeta_k(x) = \sum_{k=0}^{K} \tilde{\alpha}_{k,t}\zeta_k(x). \tag{23}$$

We define the $K+1$ vector $\tilde{\alpha}_k$ that stacks the basis function coefficients $\tilde{\alpha}_{k,t}$.

We assume that $\zeta_0(x) = 1$. This means that the coefficients $\alpha_{0,*}$ and $\alpha_{0,t}$ can be interpreted as normalization constants:

$$\alpha_{0,j} = -\ln \int \exp\left\{\sum_{k=1}^{K} \alpha_{k,j}\zeta_k(x)\right\} dx \quad \text{for} \quad j \in \{*, t\}.$$

In what follows, we will develop an autoregressive law of motion for $\tilde{\ell}_t(x)$.

3.2 Approximating the Coefficient Functions

In order to operationalize the evaluation of (20) we will also use sieves to approximate the functions $B_{z\ell}(x)$, $B_{\ell z}(x)$, and the transition matrix $B_{\ell\ell}(x, \tilde{x})$. We will turn (20) into a vector autoregressive law of motion for $(\tilde{Z}_t, \tilde{\alpha}_t)$.

First, suppose that we use the following sieve approximation for $B_{z\ell}(x)$:

$$B_{z\ell}(x) = \sum_{j=0}^{J} b_{z\ell,j}\xi_j(x). \tag{24}$$

Then,

$$\begin{aligned} \int B_{z\ell}(x)p_*(x)\tilde{\ell}_{t-1}(x;\omega)dx &= \sum_{j=0}^{J} \int b_{z\ell,j}\xi_j(x)p_*(\tilde{x})\tilde{\alpha}_{k,t-1}(\omega)\zeta_k(x)dx \\ &= \sum_{k=0}^{K} \left[\sum_{j=0}^{J} b_{z\ell,j}\left(\int \xi_j(x)\zeta_k(x)p_*(x)dx\right)\right] \tilde{\alpha}_{k,t-1}(\omega). \end{aligned} \tag{25}$$

Recall that we previously defined the $(K+1)\times 1$ vector $\tilde{\alpha}_{t-1}(\omega)$ with elements $\tilde{\alpha}_{k,t-1}$. Now also define the $1\times(J+1)$ vector $b_{z\ell}$ with elements $b_{z\ell,j}$ and the $(J+1)\times(K+1)$ matrix $C_{z\ell}$ with elements $c_{z\ell,jk} = \int \xi_j(x)\zeta_k(x)p_*(x)dx$. This leads to the following compact expression:

$$\int B_{z\ell}(x)p_*(x)\tilde{\ell}_{t-1}(x;\omega)dx = b_{z\ell}C_{z\ell}\tilde{\alpha}_{t-1}(\omega), \tag{26}$$

where the elements of C_{zl} are known and $b_{z\ell}$ is a vector of coefficients that needs to be estimated.

Second, in order to preserve the basis for the approximation of $\tilde{l}_t(x;\omega)$ in the forward iteration of (20) we express

$$B_{\ell z}(x) = \sum_{k=0}^{K} b_{\ell z,k}\zeta_k(x)p_*(x). \tag{27}$$

Let $b_{\ell z}$ be the $(K+1)\times 1$ vector with elements $b_{\ell z,k}$ that need to be estimated.

Finally, we use the following sieve approximation for $B_{\ell\ell}(x,\tilde{x})$:

$$B_{\ell\ell}(x,\tilde{x}) = \sum_{i=0}^{K}\sum_{j=0}^{J} b_{\ell\ell,ij}\zeta_i(x)p_*(x)\xi_j(\tilde{x}). \tag{28}$$

Then,

$$\begin{aligned}\int B_{\ell\ell}(x,\tilde{x})p_*(\tilde{x})d\tilde{x} &= \sum_{i=0}^{K}\sum_{j=0}^{J}\sum_{k=0}^{K}\int b_{\ell\ell,ij}\zeta_i(x)p_*(x)\xi_j(\tilde{x})\tilde{\alpha}_{k,t-1}(\omega)\zeta_k(\tilde{x})p_*(\tilde{x})d\tilde{x} \qquad (29)\\ &= \sum_{i=0}^{K}\sum_{j=0}^{J} b_{\ell\ell,ij}\left(\sum_{k=0}^{K}\left(\int \xi_j(\tilde{x})\zeta_k(\tilde{x})p_*(\tilde{x})d\tilde{x}\right)\tilde{\alpha}_{k,t-1}(\omega)\right)\zeta_i(x)p_*(x)\end{aligned}$$

Now define the constants $c_{\ell\ell,jk} = \int \xi_j(\tilde{x})\zeta_k(\tilde{x})p_*(\tilde{x})d\tilde{x}$ and the $(J+1)\times(K+1)$ matrix $C_{\ell\ell}$ with elements $c_{\ell\ell,jk}$. Using the previous definition of $\tilde{\alpha}_{t-1}(\omega)$, we can now write

$$\sum_{k=0}^{K}\left(\int \xi_j(\tilde{x})\zeta_k(\tilde{x})p_*(\tilde{x})d\tilde{x}\right)\tilde{\alpha}_{k,t-1}(\omega) = C_{\ell\ell,j\cdot}\tilde{\alpha}_{t-1}(\omega)$$

where $C_{\ell\ell,j\cdot}$ denotes the j'th row of the matrix $C_{\ell\ell,j\cdot}$, then

$$\begin{aligned}\int B_{\ell\ell}(x,\tilde{x})p_*(\tilde{x})d\tilde{x} &= \left(\sum_{i=0}^{K}\left[\sum_{j=0}^{J} b_{\ell\ell,ij}C_{\ell\ell,j\cdot}\tilde{\alpha}_{t-1}(\omega)\right]\zeta_i(x)\right)p_*(x) \qquad (30)\\ &= \left(\sum_{i=0}^{K}\tilde{\alpha}_{i,t}\zeta_i(x)\right)p_*(x).\end{aligned}$$

Overall, we have obtained a vector autoregressive law of motion for $[\tilde{Z}_t,\tilde{\alpha}_t]$:

$$\begin{bmatrix}\tilde{Z}_t(\omega)\\ \tilde{\alpha}_t(\omega)\end{bmatrix} = \begin{bmatrix} b_{zz} & b_{z\ell}C_{z\ell}\\ b_{\ell z} & B_{\ell\ell}C_{\ell\ell}\end{bmatrix}\begin{bmatrix}\tilde{Z}_{t-1}(\omega)\\ \tilde{\alpha}_{t-1}(\omega)\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix}\sigma\eta_t(\omega), \tag{31}$$

where $B_{\ell\ell}$ is the $(K+1)\times(J+1)$ matrix of coefficients $b_{\ell\ell,ij}$. We can reparameterize the system in terms of $b^*_{zl} = b_{zl}C_{zl}$ and $B^*_{ll} = B_{ll}C_{ll}$, though we might have to impose some rank reductions.

4 Implementation Details

We now provide some of the implementation details. The choice of basis functions is described in Section 4.1 and the specification of the VAR for the basis function coefficients is discussed in Section 4.2.

4.1 Basis Functions

A convenient basis for the log density is a spline of degree $m = 3$. This connects the analysis to log-spline density estimation; see Kooperberg and Stone (1990). A spline is a piecewise polynomial functions with knots x_s, $s = 1, \ldots, S$:

$$\begin{aligned}\mathrm{Spl}(m, S) &= \sum_{k=0}^{m} a_k \left(x^k \mathbb{I}\{x \le x_S\} + x_S^k \mathbb{I}\{x > x_S\}\right) \\ &\quad + \sum_{s=1}^{S-1} b_s \left(\left[\max\{x - x_s,\, 0\}\right]^m \mathbb{I}\{x \le x_S\} + (x_S - x_s)^m \mathbb{I}\{x > x_S\}\right) \\ &\quad + \sum_{k=1}^{m} c_k \left[\max\{x - x_S,\, 0\}\right]^m.\end{aligned}$$

Kooperberg and Stone (1990) suggest to make the function linear and upward sloping on the interval $(-\infty, x_1)$ and linear and downward sloping on the interval $[x_S, \infty]$. Thus, for $m = 3$ this would lead to $a_2 = a_1 = 0$ and $c_2 = c_3 = 0$. This leads to tails of a Laplace density, which are a bit thicker than Gaussian tails. Translating this specification into our $\zeta_j(x)$, $j = 0, \ldots, K = S + 1$ notation, we obtain:

$$\begin{aligned}\zeta_0(x) &= 1 \\ \zeta_1(x) &= (x\mathbb{I}\{x \le x_S\} + x_S \mathbb{I}\{x > x_S\}) \\ \zeta_2(x) &= \left(\left[\max\{x - x_1,\, 0\}\right]^3 \mathbb{I}\{x \le x_S\} + (x_S - x_1)^3 \mathbb{I}\{x > x_S\}\right) \\ &\vdots \\ \zeta_S(x) &= \left(\left[\max\{x - x_{S-1},\, 0\}\right]^3 \mathbb{I}\{x \le x_S\} + (x_S - x_{S-1})^3 \mathbb{I}\{x > x_S\}\right) \\ \zeta_{S+1}(x) &= \max\{x - x_S,\, 0\}.\end{aligned} \tag{32}$$

4.2 Coefficient VAR

Starting point for the functional VAR analysis is the sequence $\{\hat{\alpha}_t\}_{t=1}^{T+H}$. In our experience, the α_t's exhibit some collinearity, which we will remove in a preliminary transformation step

using principal-component analysis (PCA). Let $\bar{\alpha} = \frac{1}{T+H}\sum_{t=1}^{T+H}\hat{\alpha}_t$. The goal is to express the demeaned vector of spline coefficients by a lower-dimensional $\tilde{K} \leq K+1$ vector $\tilde{\alpha}_t$ of compressed coefficients or factors:

$$\hat{\alpha}_t = \bar{\alpha} + \Lambda'\tilde{\alpha}_t + \text{residual}, \tag{33}$$

where Λ' is a $(K+1)\times\tilde{K}$ matrix of loadings. Consider the following eigenvalue decomposition of the the sample covariance matrix of the α_t's:

$$\hat{\mathbb{V}}[\alpha_t] = \frac{1}{T+h}\sum_{t=1}^{T+H}(\alpha_t - \bar{\alpha})(\alpha_t - \bar{\alpha})' = V'\Xi V, \tag{34}$$

where V is a matrix of eigenvectors and Ξ is a diagonal matrix of eigenvectors. We define the $(T+H)\times\tilde{K}$ matrix $\tilde{\alpha}$ with rows $\tilde{\alpha}_t'$ as the matrix of eigenvectors associated with eigenvalues that exceed $1E-10$ in absolute value. The loadings can be computed as

$$\hat{\Lambda} = (\tilde{\alpha}'\tilde{\alpha})^{-1}\tilde{\alpha}'(\hat{\alpha} - \iota_{T+H}\bar{\alpha}'), \tag{35}$$

where $\hat{\alpha}$ is the $(T+H)\times(K+1)$ matrix with rows $\hat{\alpha}_t'$ and ι_{T+H} is a $(T+H)\times 1$ vector of ones. Using a cut-off value for the eigenvalues of $1E-10$ these estimated loadings generate essentially a perfect fit. We will subsequently use a VAR to model the joint evolution of $(\tilde{Z}_t, \tilde{\alpha}_t)$. Using the matrix $\hat{\Lambda}$ we convert the $\tilde{\alpha}_t$'s back into $\hat{\alpha}_t$'s.

We proceed by estimating a VAR for $y_t = [\tilde{Z}_t, \tilde{\alpha}_t']'$. Let $w_t = [y_{t-1}', \ldots, y_{t-p}']'$, let Y be a matrix with rows y_t' and W be a matrix with rows w_t'. Using this notation we can write the VAR as

$$y_t = D_1 y_{t-1} + \ldots + D_p y_{t-p} + u_t, \quad u_t \sim N(0, \Sigma_u), \quad Y = WD + U, \tag{36}$$

where U is a matrix with rows $u_t's$ and $D = [D_1, \ldots, D_p]'$. We combine the Gaussian VAR with an improper prior of the form

$$p(D, \Sigma_u) \propto |\Sigma_u|^{-1/2}. \tag{37}$$

We use the estimated VAR in (36) to describe the joint evolution of $(\tilde{Z}_t, \tilde{\alpha}_t)$.

VAR-based forecasts and impulse responses of $\tilde{\alpha}_t$ can be converted back into forecasts and impulse responses of $\hat{\alpha}_t$ using (33) and replacing Λ by $\hat{\Lambda}$ obtained in (35). In a last step, we re-compute the coefficients $\hat{\alpha}_{0,t}$ so that the corresponding densities integrate to one. By combining the coefficients $\hat{\alpha}_{k,t}$ with the basis functions $\zeta_k(x)$ we can recover the densities $\hat{p}_t(x)$.

5 Simulation Experiment

We now consider a data generating process (DGP) in which $p_t(x)$ is a mixture of two normal distributions with time-varying means, stochastic volatility, and time-varying weights. Let $p_N(x|\mu, \sigma^2)$ be the probability density function (pdf) associated with a $N(\mu, \sigma^2)$ random variable. Then

$$p_t(x) = \lambda_t p_N\big(x|\mu_{1t}, \sigma_{1t}^2\big) + (1 - \lambda_t) p_N\big(x|\mu_{2t}, \sigma_{2t}^2\big), \tag{38}$$

where

$$\begin{aligned} \mu_{it} &= (1 - \rho_{i\mu})\mu_i + \rho_{i\mu}\mu_{it-1} + \gamma_{i\mu}\tilde{Z}_t + \omega_{i\mu}\epsilon_{it} \\ \ln \sigma_{it} &= (1 - \rho_{i\sigma}) \ln \sigma_i + \rho_{i\sigma} \ln \sigma_{it-1} + \gamma_{i\sigma}\tilde{Z}_t + \omega_{i\sigma}\eta_{it} \\ \lambda_t &= \Phi_N(\kappa_t), \quad \kappa_t = \rho_\kappa \kappa_{t-1} + \gamma_\kappa \tilde{Z}_t + \omega_\kappa \eta_{\kappa t}, \end{aligned} \tag{39}$$

for $i = 1, 2$. Here the innovations ϵ_{it}, η_{it}, and $\eta_{\kappa t}$ are $iidN(0, 1)$. Moreover,

$$\tilde{Z}_t = \rho_z \tilde{Z}_{t-1} + \omega_z \epsilon_{z,t}. \tag{40}$$

Thus, the density $p_t(x)$ is a mixture of two normal distributions. The means and log volatilities of the mixture components follow linear Gaussian AR(1) processes. These processes are shifted by the aggregate variable $\mathcal{Z}_t$, which also evolves according to an AR(1) process. Finally, the mixture component probability λ_t varies over time. We use the cdf of a standard Normal random variable, denoted by $\Phi_N(\cdot)$, to map the real-valued process κ_t into the unit interval.

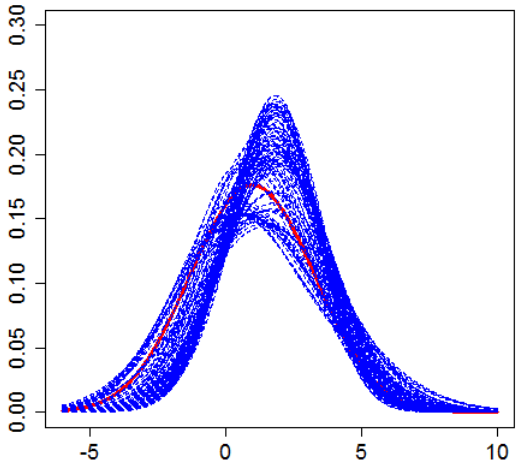
The parameterization of the DGP for the simulation experiment is summarized in Table 1. Initializing the autoregressive processes at their steady states, we are simulating the model for $T_0 + T + H$ periods. We discard the first T_0 observations, will treat the subsequent T observations as the estimation sample, and use the remaining H observations for forecast exercises. We set $T_0 = 40$, $T = 80$, and $H = 12$. To illustrate the variation of the densities $p_t(x)$ over time, we plot the steady state density as well as the time t densities in the right panel of Table 1.

5.1 Log-Spline Approximation

For each time period t we generate $N = 2,000$ cross-sectional draws from the sequence of densities $p_t(x)$ because in practice, researchers do not observe sequences of densities. Instead,

Table 1: Parameterization of DGP and Cross-sectional Densities

Para	Comp. 1	Comp. 2
μ_i	0.0	2.0
$\rho_{i\mu}$	.90	.80
$\gamma_{i\mu}$	.20	0.0
$\omega_{i\mu}$	.05	.05
σ_i	2.0	2.0
$\rho_{i\sigma}$	.95	.95
$\gamma_{i\sigma}$	.01	.01
$\omega_{i\sigma}$	.05	.05
ρ_κ	.90	
γ_κ	-.05	
ω_κ	.10	
ρ_z	.95	



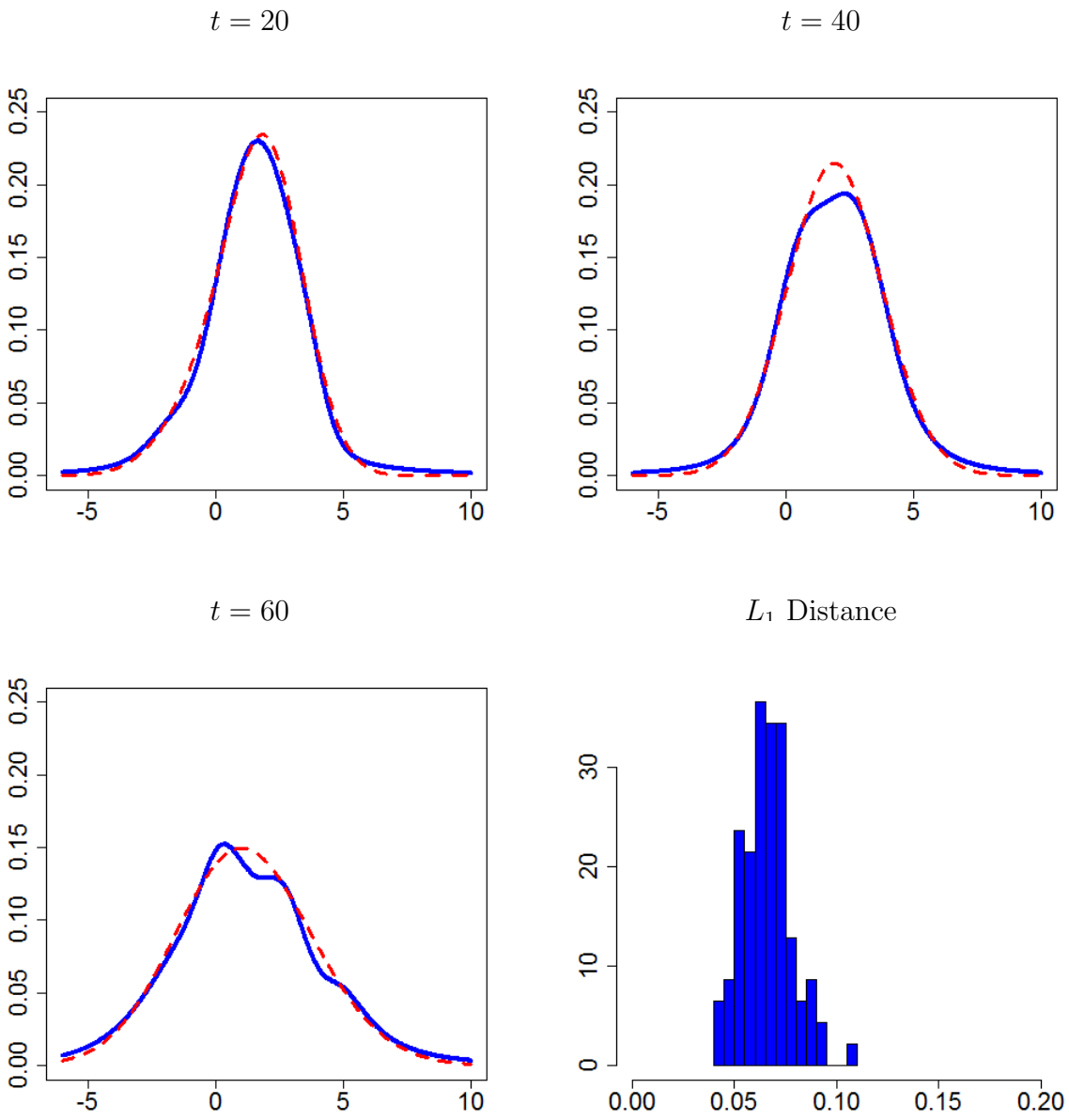
Notes: We set $\omega_z = \sqrt{1-\rho_z^2}$ such that $\mathbb{V}[\tilde{Z}] = 1$. The figure on the right depicts the steady state density (solid red) as well as the simulated densities $p_t(x)$ (dotted blue).

they observe cross-sectional draws or panels obtained from these densities. Thus, in order to operationalize the functional VAR analysis, we need to map the cross-sectional draws into a density estimate. To do so, we construct log-spline density estimation, previously discussed in Section 4.1.

The density estimation is implemented using the R package *logspline*. The key difficulty in log-spline density estimation is the placement of the knots and the determination of the number of knots. While the model determination is largely automated in the R package, we face the additional difficulty that the knot placement, which determines the basis functions in (32), has to be identical in all time periods, because all the time-variation is absorbed in the law-of-motion of the basis function coefficients. As the density shifts over time, there may be temporarily too few observations in the tails to execute the log-spline density estimation.

We pool the observations x_{it} across time periods and compute the 1%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, and 99% quantiles. We place knots at these quantiles as well as close to the minimum and the maximum of our x_{it} draws. In order to be able to estimate log-spline approximations for each time period t using these knots, we generate $N_* = 100$ artificial

Figure 1: Cross-Sectional Density and Its Log-Spline Estimate



Notes: The first three panels compare the true densities $p_t(x)$ (red, dashed) to their log-spline estimates $\hat{p}_t(x)$ (blue, solid). The last panel depicts a histogram of the L_1 distance between true and estimated densities across all time periods.

equally-spaced observations, spanning the interval between the minimum and the maximum of the actual x_{it} observations. Based on the samples of actual and artificial observations, the log-spline estimation procedure converges for all time periods t.

A comparison of the estimated density with the original density for three specific time

periods is provided in Figure 1. For $t = 20$ the estimated and the true density are almost identical, whereas in other periods there is a small discrepancy. For $t = 60$, for instance, the estimated density is a bit "bumpier" than the original densities, but it still captures the location and scale of the true cross-sectional distribution.

5.2 Results from an Estimated Functional VAR

We now conduct a functional VAR analysis conditional on the basis function coefficient vectors $\{\hat{\alpha}_t\}_{t=1}^{T+H}$ obtained from the log-spline density estimation. We use the VAR representation and the estimation method described in Section 4.2. We set the number of lags to $p = 1$. In slight abuse of notation we orthogonalize the VAR innovations u_t in (36) as follows:

$$u_t = \Sigma_{tr}\Omega\epsilon_t, \quad \epsilon_t \sim N(0, I). \tag{41}$$

Here Σ_{tr} is the lower-triangular Cholesky factor of Σ_u and Ω is an orthonormal matrix with the property $\Omega\Omega' = I$.

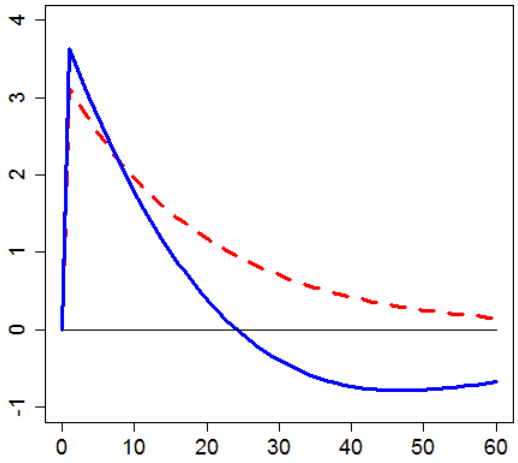
Impulse Response Function to a Technology Innovation. We now generate an impulse response function to an innovation $\epsilon_{z,t}$. According to the DGP, shocks to the density $p_t(x)$ do not affect $\tilde{Z}_t$. Thus, to identify the innovation $\epsilon_{z,t}$ we let $\tilde{Z}_t$ be the first element of the vector y_t in the VAR and set $\Omega = I$. Let $e_{(1)} = [1, 0, \ldots, 0]'$, such that the impact of a one-standard-deviation innovation to $\tilde{Z}_t$ is given by

$$\frac{\partial}{\partial \epsilon_{z,t}} y_t = \Sigma_{tr} e_{(1)}. \tag{42}$$

For now, we fix the D_1 matrix (recall that $p = 1$) and Σ_{tr} at their posterior means and then iterating (36) forward. The size of the shock is 10 standard deviations, i.e., $\bar{\epsilon}_z = 10$. The response of $\tilde{Z}_{t+h}$, $h = 0, \ldots, H$ is depicted in Figure 2. According to the DGP, $\tilde{Z}_{t+h}$ is an AR(1) process with $\rho_z = 0.95$. Thus, the IRF decays monotonically to zero. When estimating the functional VAR, we do not restrict the coefficients on $\tilde{\alpha}_{t-1}$ in the $\tilde{Z}_t$ equation to be equal to zero. In turn, the estimated VAR implies that after 25 period $\tilde{Z}_t$ becomes negative and then subsequently slowly reverts back to zero.

The response of the densities is depicted in Figure 3. The right panels of the figure shows the responses computed from the functional VAR (computed at the posterior mean estimates), and the left panels depict responses from the DGP. The DGP responses are computed by initializing the system (38) to (40) in the steady state and then subjecting it to

Figure 2: Impulse Response of $\tilde{Z}_t$



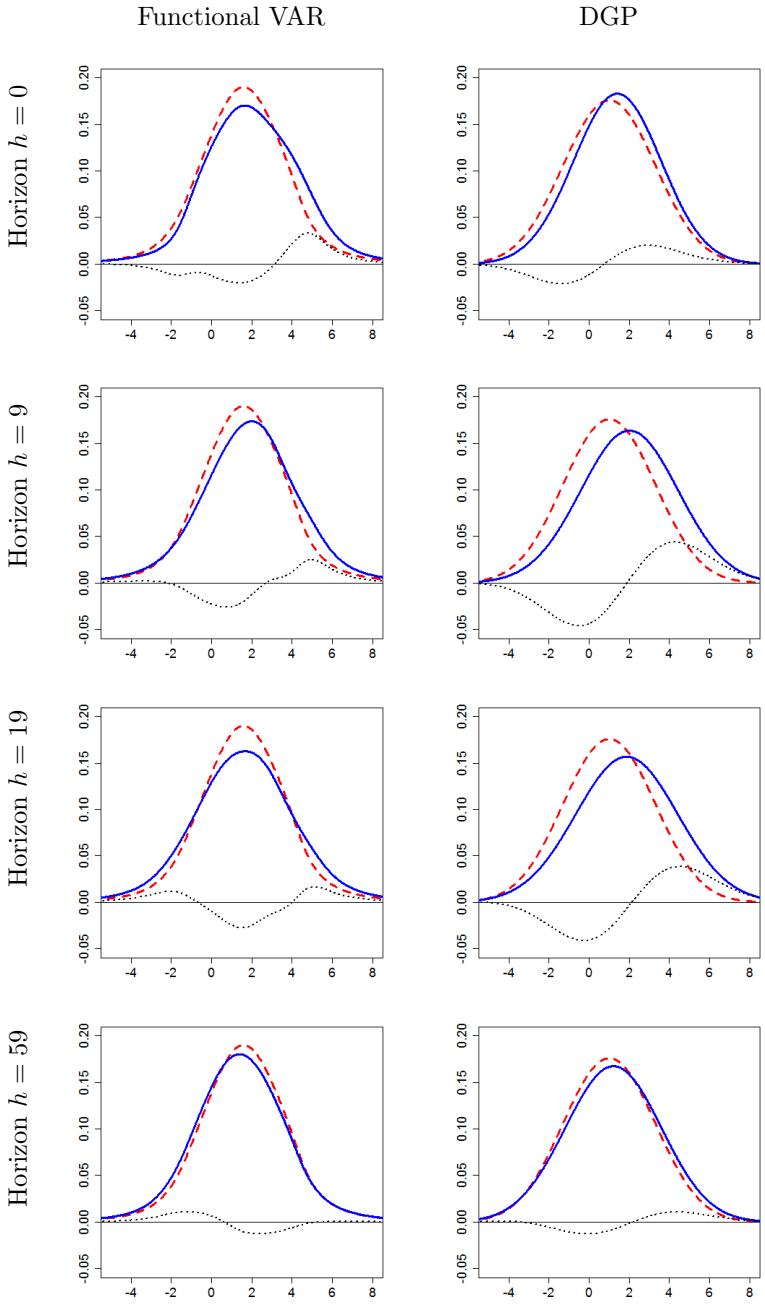
Notes: This figure depicts the IRF of $\tilde{Z}_t$ to a 10-standard-deviations innovation, computed from the DGP (red, dashed) and the estimated functional VAR (blue, solid). The functional VAR IRFs are computed at the posterior mean of the parameters.

the shock $\epsilon_{z,T+1} = \bar{\epsilon}_z$, setting all other innovations equal to zero. The dashed red densities depict the baselines. Note that the baseline densities are slightly different for the functional VAR and the DGP. For the functional VAR they are obtained by setting $\alpha_t = \bar{\alpha}$, whereas for the DGP they are obtained by setting the processes μ_{it}, σ_{it}, and λ_t to their steady state values.

According to the DGP, the main effect of $\tilde{Z}_t$ is that it raises the mean of the first mixture component ($\gamma_{i\mu} = 0.2$). This tends to shift the $p_t(x)$ distribution to the right relative to the baseline distribution, which is visible in the top panels of Figure 3. The impact effects ($h = 1$) in the functional VAR and the DGP are fairly similar. In the DGP, the shift of probability toward the right continues in subsequent periods; see $h = 9$ and $h = 19$. Eventually, the density reverts back to its steady state. At horizon $h = 19$ the estimated VAR response looks quite different from the DGP response, as probability mass already shifts back toward the left. The resulting density is roughly centered around the same point that the baseline density is centered, but it now has fatter left and right tails. This pattern is broadly consistent with the path of $\tilde{Z}_t$ in the functional VAR (see Figure 2), which decays faster than under the DGP and eventually becomes negative.

Impulse Response Function to a Distributional Innovation. Innovations to the basis

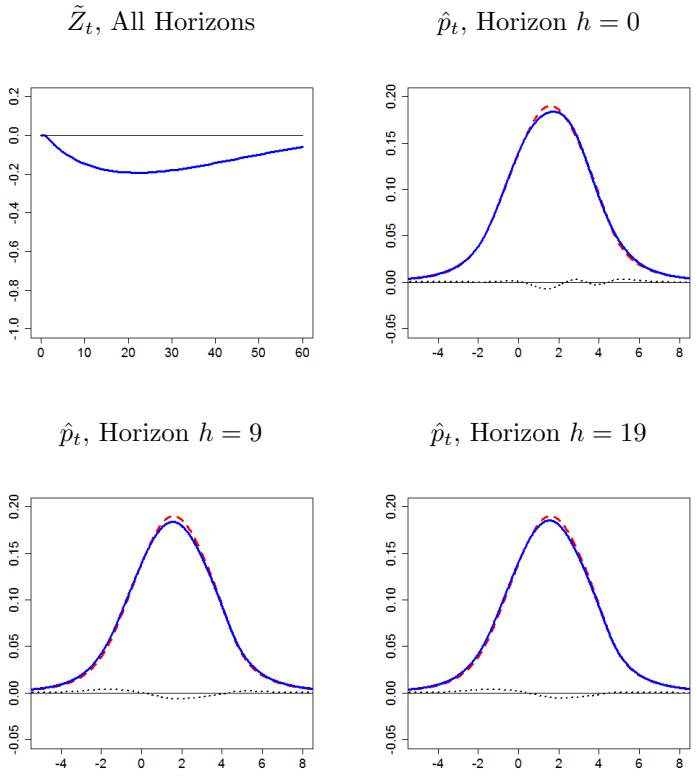
Figure 3: Density Response to Technology Innovation $\epsilon_{z,t}$



Notes: Each panel depicts the baseline density (red, dashed), the density after a 10-standard-deviation innovation to $\tilde{Z}_t$ (blue, solid), and their difference (black, dotted). The functional VAR IRFs are computed at the posterior mean of the parameters.

function coefficients $\tilde{\alpha}_t$ are much more difficult to interpret and unlike in the case of the technology innovation, there are no obvious exclusion restrictions that could be used to identify an innovation that is of particular interest. Instead, building on work by Uhlig

Figure 4: Responses to Distributional Innovation ϵ_t^*



Notes: The top left panel depicts the impulse response function of $\tilde{Z}_t$ to a 1-standard-deviation distributional innovation ϵ_t^*. Each of the remaining panels depicts the baseline density (red, dashed), the density after a 1-standard-deviation distributional innovation ϵ_t^* (blue, solid), and their difference (black, dotted). The functional VAR IRFs are computed at the posterior mean of the parameters.

(2005) and Kurmann and Otrok (2013), we consider the shock that has the largest average effect (in terms of contribution to the variance) over a period of $\bar{h} = 8$ periods on the aggregate variable $\tilde{Z}_t$. In the Online Appendix we describe how we construct a unit-length vector q^*, such that the impulse response function can be generated by propagating the forward the impact effect

$$\frac{\partial}{\partial \epsilon_t^*} y_t = \Sigma_{tr} q^*. \tag{43}$$

The results are depicted in Figure 12. The distributional shock with the maximal effect on the fluctuations of $\tilde{Z}_t$ hat a hump-shaped effect on $\tilde{Z}_t$. For about 20 periods the aggregate series is falling and then slowly reverts to steady state. Recall that according to the DGP the unconditional variance of $\tilde{Z}_t$ is equal to one and $\tilde{Z}_t$ is independent of the distributional shocks. Thus, the non-zero response is either due to estimation noise or an inaccurate linearization. The effect of the distributional shock on $\hat{p}_t(x)$ is small. Initially there is a

transfer of probability mass from the region around to median into the right tail of the distribution. Subsequently, the distribution swings back and some mass is transferred into the left tail. Eventually, the distribution reverts back to its steady state.

Historical Decomposition. Figure 5 provides information about the importance of the technology shock innovation for fluctuations in the cross-sectional distribution. In the left panel we plot the actual path of the coefficient $\tilde{\alpha}_{3,t}$ as well as two counterfactual paths which are generated by either shutting down the technology innovation or by shutting down all distributional innovations.

To generate the counterfactual series, recall that given estimates of D and Σ for the VAR in (36) as well as the identification assumption in (41) we can compute structural residuals

$$\hat{\epsilon}_t = \hat{\Sigma}_{tr}^{-1}(y_t - \hat{D}_1 y_{t-1} - \ldots - \hat{D}_p y_{t-p}) \tag{44}$$

and then generate counterfactual series y_t^* by iterating the VAR difference equation forward (starting from $(y_{-p+1}, \ldots, y_0)$:

$$y_t^* = \hat{D}_1 y_{t-1}^* + \ldots + \hat{D}_p y_{t-p}^* + \hat{\Sigma}_{tr} M_\epsilon \hat{\epsilon}_t, \tag{45}$$

where M_ϵ is a diagonal matrix with ones and zeros on the diagonal that selects elements of the $\hat{\epsilon}_t$ vector. The counterfactual paths depicted in the left panel of the figure are obtained by either shutting down the technology innovation, or by shutting down the distributional innovations. The counterfactual paths also reflect the effect of the initial condition $(y_{-p+1}, \ldots, y_0)$. The figure indicates that the path of $\tilde{\alpha}_{2,t}$ is mostly explained by the non-technology innovations.

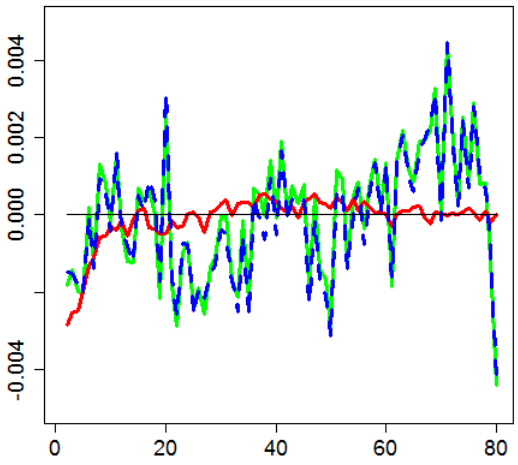
To generate the right panel of Figure 5, we first convert the counterfactual paths $\tilde{\alpha}_t^*$ back into densities. Second, we compute a sequence of densities $\hat{p}_t^0$ by shutting down all shocks. The $\hat{p}_t^0$ densities capture the effect of the initial values $y_{1-p}, \ldots, y_0$. Third, we compute the L_1 distance between the actual density $\hat{p}_t(x)$ and the no-shocks densities $\hat{p}_t^0(x)$, denoted by $\Delta_{L_1}(\hat{p}_t, \hat{p}_t^0)$, and between the counterfactual densities $\hat{p}_t^*(x)$ and $\hat{p}_t^0(x)$, denoted by $\Delta_{L_1}(\hat{p}_t^*, \hat{p}_t^0)$. Finally, we define our measure of the importance of a particular innovation $\epsilon_{i,t}$. Suppose that $\hat{p}_t^*(x)$ is generated by shutting down the sequence of innovations $\epsilon_{i,t}$. Then,

$$IMP_t^*(\epsilon_i) = 1 - \frac{\Delta_{L_1}(\hat{p}_t^*, \hat{p}_t^0)}{\Delta_{L_1}(\hat{p}_t, \hat{p}_t^0)} \tag{46}$$

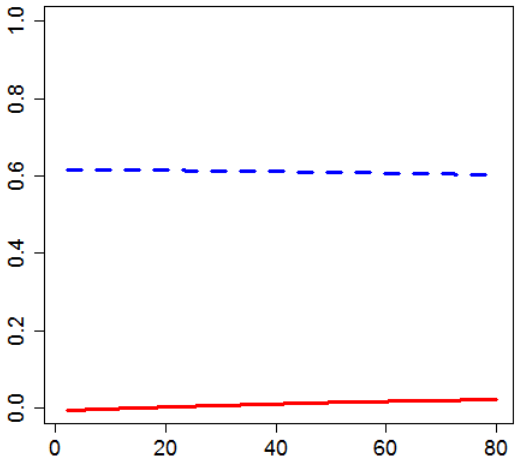
can be viewed as a measure of the importance of the omitted innovation. If omitting the innovation has essentially no effect on $\tilde{\alpha}_t^*$, then $\hat{p}_t^*(x) \approx \hat{p}_t(x)$ and $IMP_t^*(\epsilon_i) \approx 0$. If the

Figure 5: Importance of Technology Shock for Density Fluctuations

$\tilde{\alpha}_{3,t}$ Coefficient Decomposition



L_1 Distance Decomposition



Notes: Left panel: actual path (green, solid) of $\tilde{\alpha}_{2,t}$, counterfactual path without technology innovations only (blue, dashed), and counterfactual path without non-technology (red, solid) innovations. Right panel: importance measure (46) for technology innovations (red, solid) and non-technology innovations (blue, dashed).

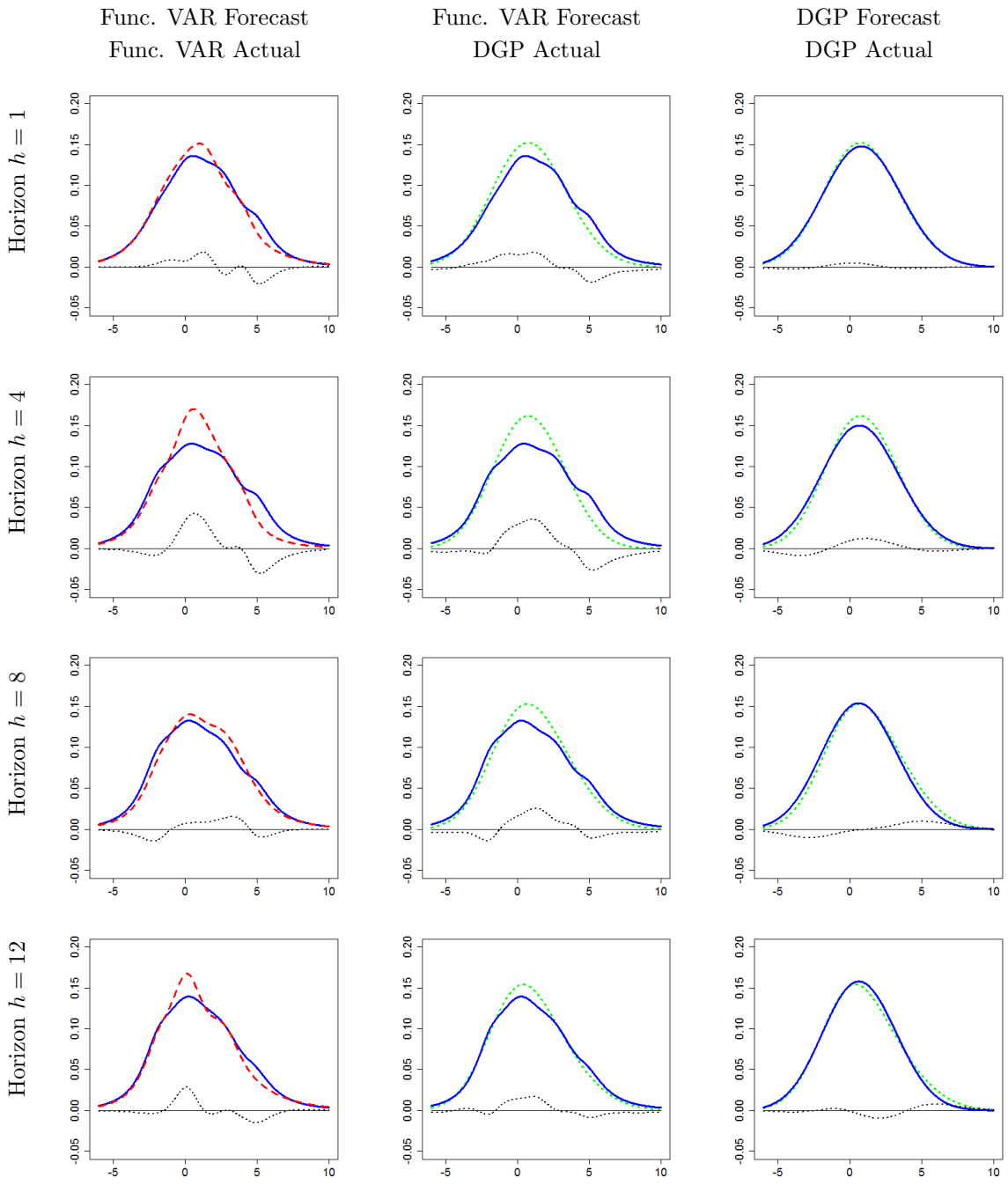
shock explains most of the fluctuations in $\tilde{\alpha}_t$, then $\Delta_{L_1}(\hat{p}_t^*, \hat{p}_t^0)$ is close to zero and $IMP_t^*(\epsilon_i)$ is approximately one.

Because the mapping from the $\tilde{\alpha}_t$, $\tilde{\alpha}_t^0$, and $\tilde{\alpha}_t^*$ coefficients into the L_1 distances between the densities $\hat{p}_t$, $\hat{p}_t^0$, and $\hat{p}_t^*$ is nonlinear, the specific magnitudes (other than the extremes of zero and one) are difficult to interpret. Shutting down the technology innovations leaves $\hat{p}_t^*$ essentially unchanged and has a negligible effect on the L_1 distance to the no-shock density $\hat{p}_t^0$. Shutting down the distributional innovations reduces the L_1 distance to $\hat{p}_t^0$ by 40%. Over time, the importance measure is fairly stable.

Forecasting. We now generate point forecasts of $p_{T+h}(x)$. To obtain forecasts from the functional VAR, we condition on the posterior mean estimates of D and iterate (36) forward conditional on y_T, while setting innovations dated $t > T$ equal to zero. The point forecasts from the DGP are generated by simply iterating Equations (38) to (40) forward, while setting innovations dated $t > T$ equal to zero. A comparison of forecast and actual densities is provided in Figure 6.

In the first column of Figure 6 we compare the functional VAR forecast to the log-spline approximation $\hat{p}_{T+h}(x)$. The panels in the second column overlay the funcational

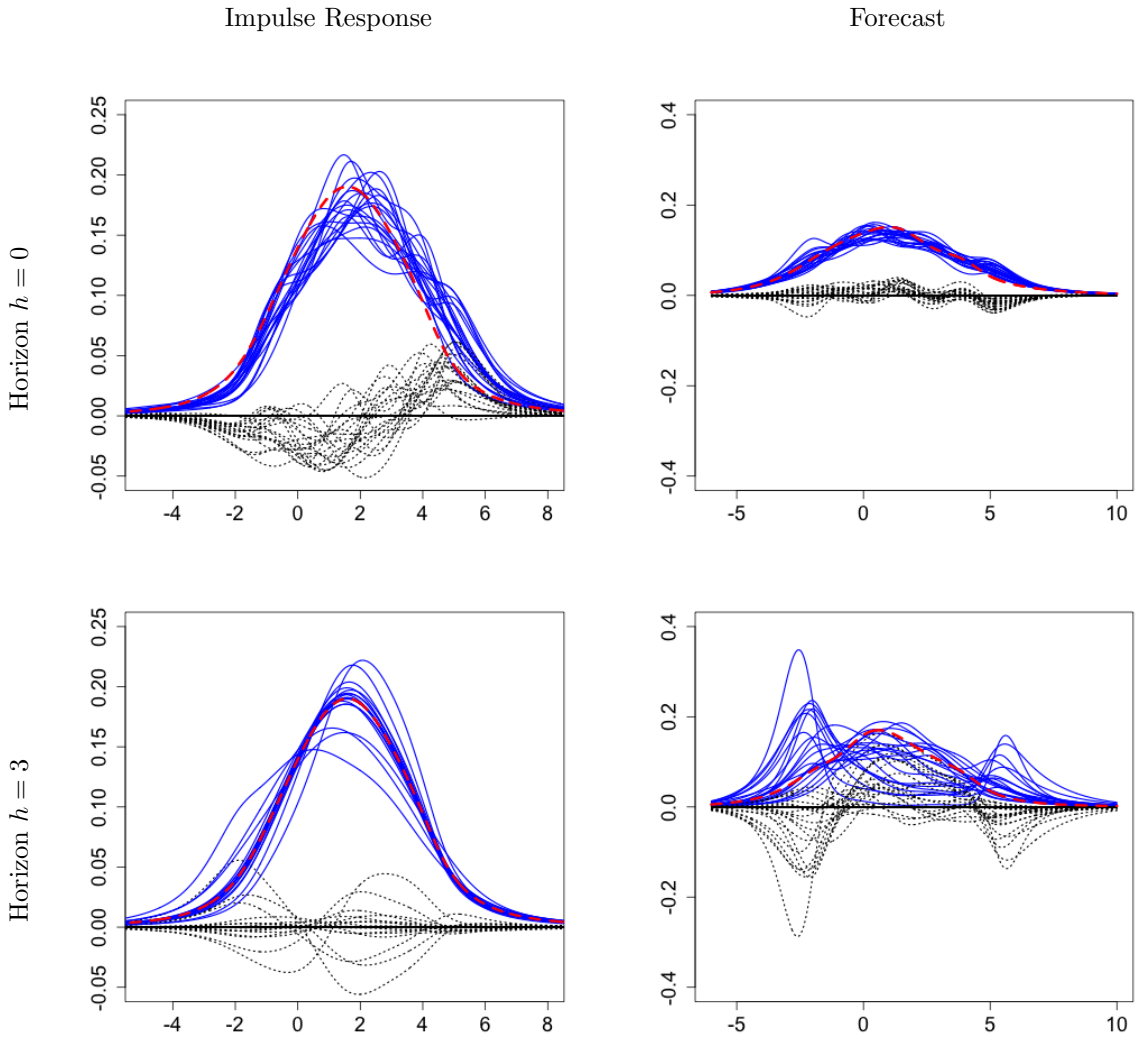
Figure 6: Multi-Step Forecasts



Notes: Each panel depicts the point forecast of $p_{T+h}(x)$ (blue, solid) as well an actual density $\hat{p}_{T+h}$ (red, dashed) or p_{T+h} (green, dotted). The forecast error (actual minus predicted) is indicated by the black dotted line.

VAR forecast with the true $p_{T+h}(x)$ density. Finally, the last column compares a forecast generated from the DGP to the true $p_{T+h}(x)$. The black dotted lines in each panel indicate

Figure 7: Uncertainty Associated with IRFs and Forecasts



Notes: Impulse Responses: each panel depicts the baseline density (red, dashed), draws from the posterior distribution of densities after a 10-standard-deviation innovation to $\tilde{Z}_t$ (blue hairlines), and their difference (black dotted hairlines). Forecasts: each panel depicts draws $\hat{p}_{T+h}(x)$ from the posterior distribution if the shocked distributions as well as the the actual density $\hat{p}_{T+h}$. The distribution of forecast errors (actual minus predicted) is indicated by the black dotted hairlines.

the pointwise density forecast errors. Regardless of whether we evaluate the functional VAR forecast based on $\hat{p}_{T+h}(x)$ or $p_{T+h}(x)$, the shape of the forecast errors is fairly similar. The errors associated with forecasts that are generated from the DGP instead of the functional VAR approximation are generally smaller.

Capturing Uncertainty. The impulse responses in Figure 3 and the forecasts in Figure 6 do not capture parameter and shock uncertainty. While computing statistics that reflect (some of) this uncertainty is fairly straightforward, displaying this uncertainty in a plot is

less straightforward. In Figure 7 we use hairline plots. We condition on the estimates of the sequence of densities $\hat{p}_t(x)$ as well as the mean $\bar{\alpha}$ of the $\hat{\alpha}_t$ sequence. The hairlines are generated by taking subsequences of draws from the posterior distribution of (D, Σ_u), simulating innovations u_t, computing simulated values of y_t, which then can be converted into densities.[2]

6 Empirical Analysis

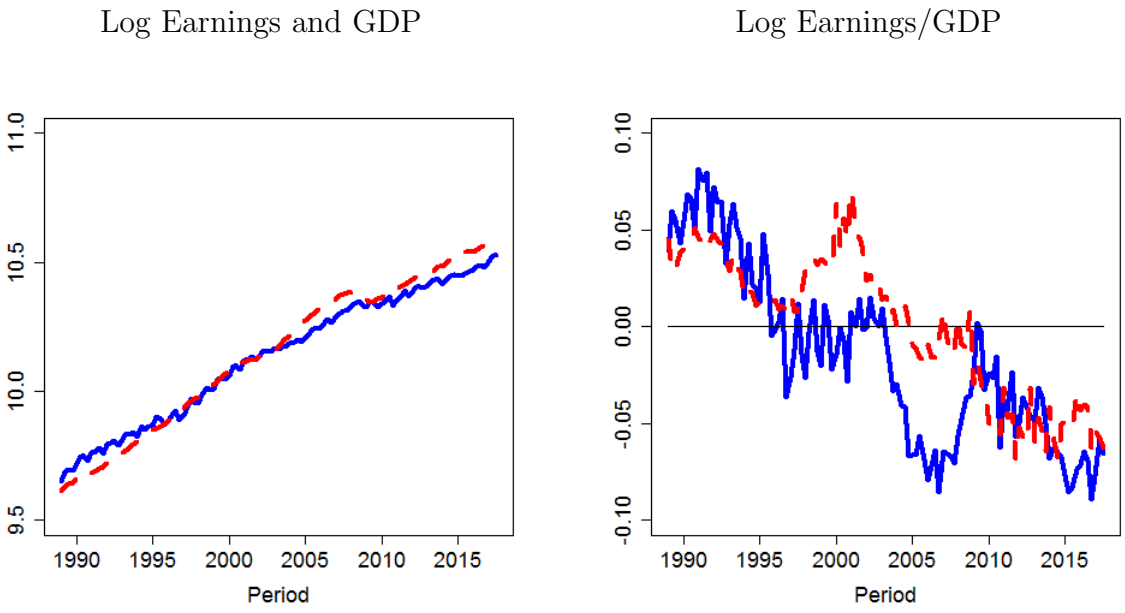
Data. We download quarterly observations on the GDP deflator *GDPDEF*, real per-capita GDP (*A939RX0Q048SBEA*), and the unemployment rate (UNRATE) from FRED. We construct nominal per capita GDP by taking the product of real per capita GDP and the GDP deflator. We downloaded quarterly total factor productivity (TFP) data from Fernald (2012) and use the TFP series *dtfp*, which consists of annualized quarter-on-quarter growth rates. Finally, we obtained the labor share for the nonfarm-business sector from the Bureau of Labor Statistics.

Cross-sectional data on log earnings and employment are obtained from the monthly Current Population Survey (CPS) through the website of the National Bureau of Economic Research (NBER). Our measure of weekly earnings is PRERNWA and we use PREXPLF ("Experienced Labor Force Employment") to construct an employment indicator (zero is "unemployed" and one is "employed"). We drop individuals for which the employment indicator is not available and we recode individuals with non-zero earnings as employed. To obtain a CPS-based unemployment rate, we compute the fraction of observations coded as "unemployed." To obtain the distribution of weekly earnings, we set earnings to zero for units that are coded as "unemployed." Moreover, we drop individuals from the calculation of the earnings distribution which are coded as "employed," but weekly earnings are missing. Finally, we scale weekly earnings to annual earnings by multiplying with 52.

In the left panel of Figure 8 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

[2]For the IRFs we are taking the following shortcut: given a draw (D, Σ_u), we compute the linear IRF of $\tilde{\alpha}_t$, meaning we set all shocks except for $\epsilon_{z,T+1}$ to zero. We then add $\bar{\alpha}$ and convert the resulting coefficients into a density. The baseline density is simply obtained by converting $\bar{\alpha}$ into a density. Alternatively, we could follow the more tedious approach of simulating all shocks when constructing the difference between baseline densities and densities obtained by setting $\epsilon_{z,T+1} = \bar{\epsilon}_z$.

Figure 8: Earnings and GDP



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

of 2/3 to account for the labor share.[3] 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. 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.

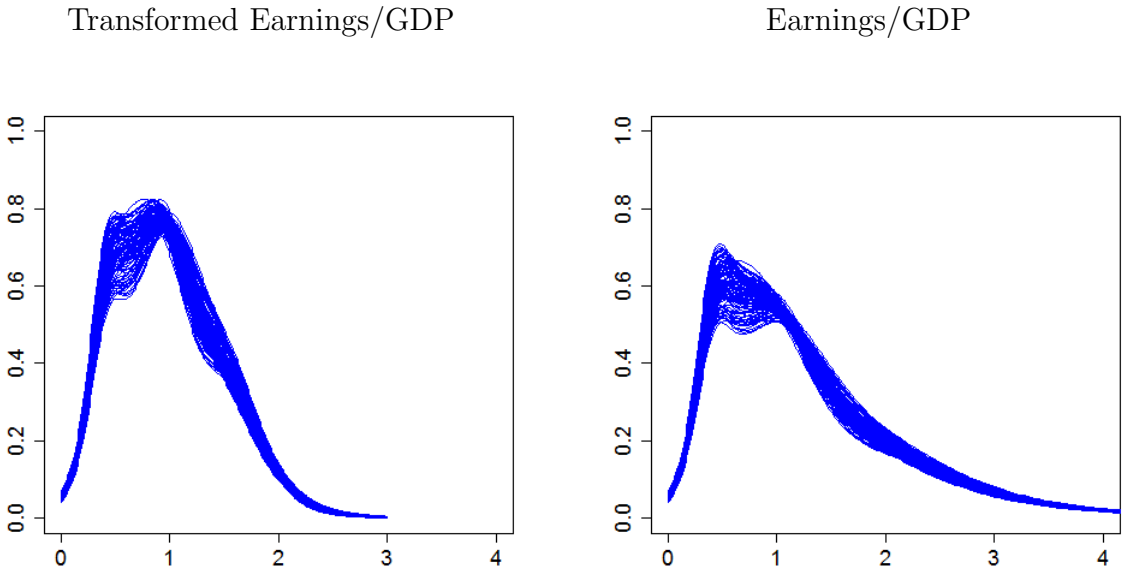
In the remainder of this paper we simply standardize individual-level earnings by per-capita GDP without any labor share adjustments. Rather than taking a logarithmic transformation of the earnings data, we apply the inverse hyperbolic sine transformation, which is given by

$$g(x|\theta) = \frac{\ln(\theta x + (\theta^2 x^2 + 1)^{1/2})}{\theta} = \frac{\sinh^{-1}(\theta x)}{\theta}, \quad x = \frac{\text{Earnings}}{(2/3) \cdot \text{per-capita GDP}}. \tag{47}$$

The function is plotted in the Online Appendix. We set $\theta = 1$. For small values of x the function is approximately equal to x and for large values of x it is equal to $\log(x) + \log(2)$.

[3]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 9: Estimated Log Earnings Distributions



Notes: Each hairline corresponds to the estimated density of earnings for a particular quarter t, where t ranges from 1989:Q1 to 2017:Q3. Transformation is inverse hyperbolic sine transformation in (47).

This transformation avoids the thorny issue of applying a log transformation to earnings that are close to zero.

Log-Spline Estimation. The empirical analysis proceeds in a similar way as the analysis of the simulated data in Sections 5.1 and 5.2. Based on the earnings data for each quarter we estimate a cross-sectional density for the transformed earnings-to-GDP ratio; see (47). The knots for the log-spline density estimation are placed at the 0.01, 0.1, 0.25, 0.5, 0.75, 0.9, and 0.98 quantiles of the distribution of the pooled (across-time periods) transformed earnings data. In each period, we are normalizing the cross-sectional density by the fraction of individuals who reported to be employed. In Figure 9 we overlay the log-spline estimates of the cross-sectional densities. The left panel shows the density of the transformed earnings whereas the right panel shows the densities of the original earnings-to-GDP ratio which is obtained by a change-of-variables.

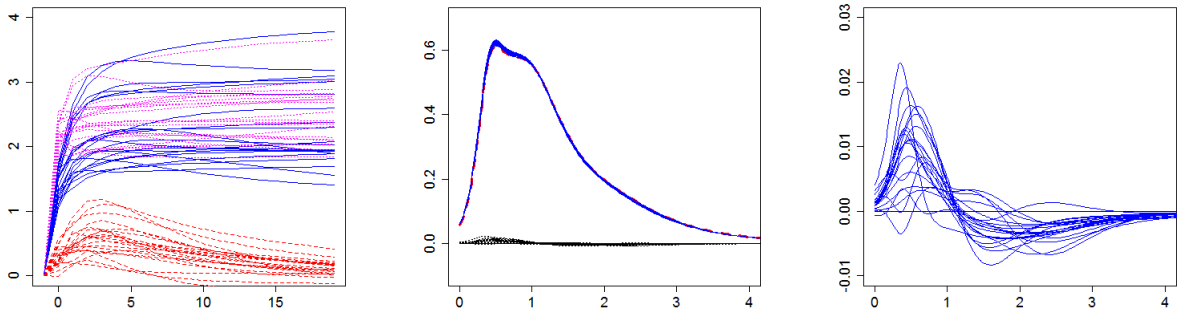
Functional VAR Estimation. We now estimate a functional VAR(1) that takes the form of (36). The vector $\tilde{Z}_t$ is composed of demeaned TFP growth, per-capita real GDP growth, and the unemployment rate computed from the CPS data. The 9×1 vector $\hat{\alpha}_t$ in (33) comprises the vector of coefficients of the log-spline approximation. However, we drop the

Figure 10: Impulse Responses to a TFP Shock

TFP, GDP, Empl. Rate [%]
Horizons $h = 0, \ldots, 19$

Earnings/GDP Distr.
Horizon $h = 4$

Distr. Response
Horizon $h = 4$



Notes: Responses to a 3-standard-deviations shock to TFP. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. Each hairline corresponds to a draw from the posterior distribution. Left panel: TFP (magenta, dotted), GDP (blue, solid) and Employment Rate (red, dashed). Center panel: steady state earnings/GDP density (red dashed), shocked density (blue solid), and difference (black). Right panel: difference between shocked and steady state density.

coefficient that corresponds to the constant term, because the unemployment rate in the $\tilde{Z}_t$ vector encodes the normalization information for the earnings density, which leads to an 8×1 vector. The vector vector $\tilde{\alpha}_t$ that enters the definition of y_t is obtained by removing co-linear elements from $\hat{\alpha}_t$ based on (33) and is of dimension 6×1. Subsequent results about the dynamics of the cross-sectional densities are obtained by converting $\tilde{\alpha}_t$ vectors back into $\hat{\alpha}_t$ vectors, re-normalizing the densities so that they integrate to the employment rate, and then applying the change-of-variable formula to obtain a density for the earnings-to-GDP ratio. Our estimation sample, after computing growth rates for TFP and GDP growth, ranges from 1989:Q2 to 2017:Q3.

Impulse Responses to a TFP Shock. We order TFP growth first, GDP growth second, the CPS-based unemployment third, and the income distribution last. We assume that shocks to GDP growth and the income distribution do not affect measured TFP contemporaneously and use a simple Cholesky factorization to identify the TFP growth innovation. Figure 10 shows impulse responses to a three-standard deviation TFP shock. We take a subsample of 20 draws from the posterior distribution of the VAR parameters and compute IRFs for each of these parameter draws. According to the left panel, in response to this shock the level of TFP rises in between 180 and 300 basis points (the figure depicts the response of log TFP, scaled by 100) in the log-run, whereas log per-capita GDP rises in between 150 and

300 basis points. Some of the TFP responses exhibit a slight hump-shaped pattern whereas most of the GDP responses are monotonic (recall, the responses are based on a functional VAR(1)). We also show responses of the employment rate. Here a number of one means that the employment rate increases by 100 basis points (or equivalently, the unemployment falls by one percentage point).

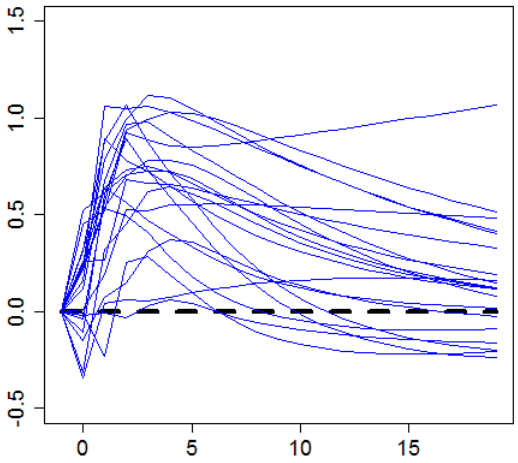
The center panel shows the estimated steady state density of the income distribution (obtained from the mean coefficients $\bar{\alpha}$) as well as the response of this density to the TFP shock after $h = 4$ quarters. Visually, the two densities are difficult to distinguish. We also overlay the difference between the steady state and the shocked density and then plot this difference separately in the right panel of Figure 10. For the center and right panels, a one on the x-axis refers to an individual whose earnings are equal to GDP per capita (adjusted for a labor share of 2/3). Maybe surprisingly, there seems to be a shift of probability mass in the distribution of the log earnings-to-GDP ratio to the left. Now consider a hypothetical individual who shifts from 1.25 to 0.75 in this distribution in response to a TFP shock. Prior to the shock, this individual's earnings were 25% above the per-capita GDP (adjusted for a labor share of 2/3) benchmark, whereas after the shock it dropped to being 25% below the benchmark. Meanwhile the level of per-capita GDP rises by about 2.4%, making the individual significantly worse off. However, dynamics of the distribution of log earnings are mean reverting and the effect dies out fairly quickly.

To assess the overall effect of a technology shock on the earnings distribution we plot the response of the probability mass assigned to individuals with an earnings-to-per-capita-GDP ratio less than one in the left panel of Figure 11. This response is positive but fairly small – less than 1% for most hairlines and around 0.5% on average. The positive response is consistent with a model in which individuals are heterogeneous with respect to their skills and during expansions more low-skilled individuals enter the labor force. The right panel of the figure depicts the response of the Gini coefficient. Most of the hairlines indicate a slight decline of the Gini coefficient (decrease in inequality) after the shock, but the effect is quantitatively small.

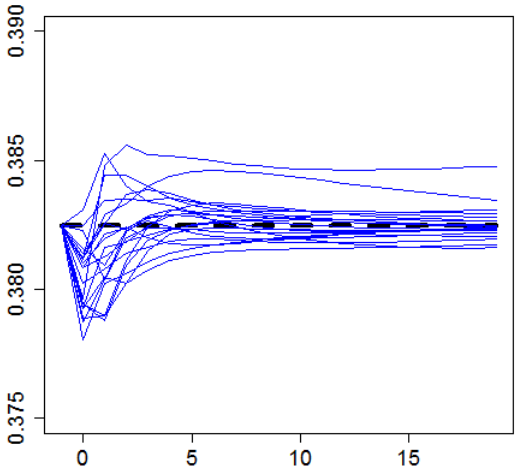
Impulse Responses to a Distributional Shock. Figure 12 shows responses to a three-standard-deviation distributional shock that maximizes the average effect (in terms of contribution to the variance) over $\bar{h}$ periods on aggregate TFP and GDP growth. We normalize the sign of the distributional shock such that it decreases the density of individuals earning around 25% of GDP per capita. The top panels in Figure 12 show responses of the earnings distribution. Panels (1,1) and (1,2) focus on the effect at impact ($h = 0$), whereas Panel (1,3)

Figure 11: Impulse Responses to a TFP Shock – Continued

Fraction of Individuals Earning Less than per-capita GDP [%]



Gini Coefficient



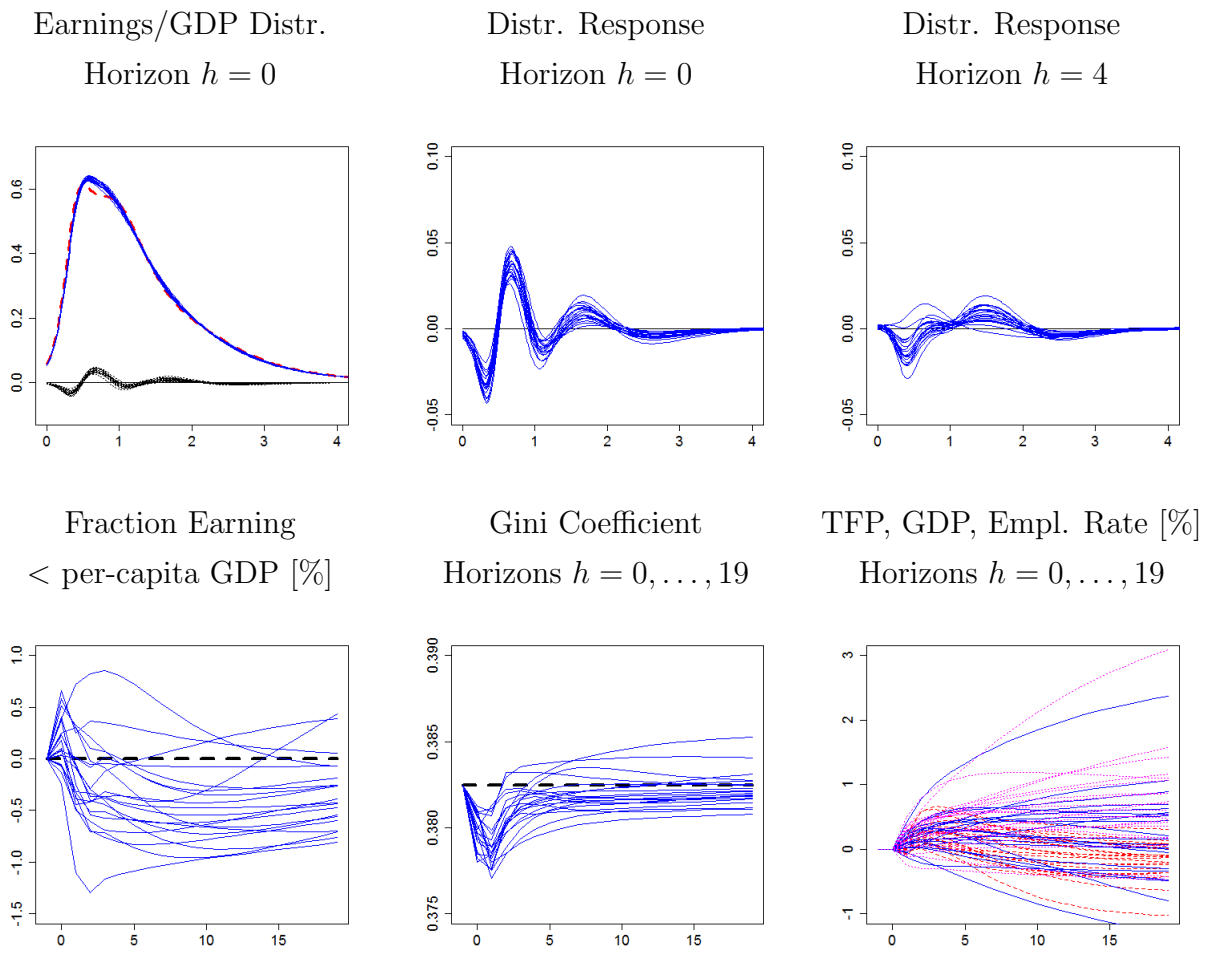
Notes: Responses to a 3-standard-deviations shock to TFP. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. Each hairline corresponds to a draw from the posterior distribution.

shows the response after one year ($h = 4$). Most of the distributional response affects individuals earning less than GDP per capita. Upon impact, the fraction of individuals earning less than 50% decreases whereas the mass of individuals earning between 50% and 100% of per capita GDP increases. More than 50% of the impact effect has vanished after one year. The Gini coefficient drops in response to the shock but quickly reverts back to its steady state. Finally, for most draws the responses of the levels of TFP, output, and employment are positive. However, there is considerable uncertainty about the long-run effects.

7 Conclusion

We developed a vector autoregressive model that stacks macroeconomic aggregates and and cross-sectional distributions to provide semi-structural evidence about the interaction of aggregate and distributional dynamics. We applied the model to examine the effect of a technology shock on the earnings distribution and the effect of a shock that moves the earnings distribution on macroeconomic aggregates. The technique developed in this paper

Figure 12: Responses to a Distributional Innovation ϵ_t^* (Average of TFP/GDP growth)



Notes: Responses to a 3-standard-deviation distributional innovation ϵ_t^*. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. Each hairline corresponds to a draw from the posterior distribution. Panel (1,1): steady state log earnings density (red dashed), shocked density (blue solid), and difference (black). Panels (1,2) and (1,3): difference between shocked and steady state density. Panel (2,3): TFP (magenta, dotted), GDP (blue, solid) and Employment Rate (red, dashed).

should be useful more broadly for the evaluation of heterogeneous agent macro models.

References

Algan, Y., O. Allais, and W. Den Haan (2008): "Solving Heterogeneous-Agent Models with Parameterized Cross-Sectional Distributions," *Journal of Economic Dynamics and Control*, 32(3), 875–908.

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

Carriero, A., G. Kapetanios, and M. Marcellino (2016): "Structural Analysis with Multivariate Autoregressive Index Models," *Journal of Econometrics*, 192, 332–348.

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

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.

Childers, D. (2015): "On the SoluSolu and Application of Rational Expectations Models with Function-Valued States," *Manuscript, Carnegie Mellon University*.

Efron, B., and R. Tibshirani (1996): "Using Specially Designed Exponential Families for Density Estimation," *Annals of Statistics*, 24(6), 2431–2461.

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

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

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

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

Kurmann, A., and C. Otrok (2013): "News Shocks and the Slope of the Term Structure of Interest Rates," *American Economic Review*, 103, 2612–2632.

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

Reinsel, G. (1983): "Some Results on Multivariate Autoregressive Index Models," *Biometrika*, 70(1), 145–156.

Reiter, M. (2010): "Approximate and Almost-Exact Aggregation in Dynamic Stochastic Heterogeneous-Agent Models," *IHS Working Paper, Economics Series*, (258).

Uhlig, H. (2005): "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381–419.

Winberry, T. (2017): "A Toolbox for Solving and Estimating Heterogeneous Agent Macro Models," *Working Paper, Chicago Booth*.

Online Appendix: Does Heterogeneity Matter for Aggregate Fluctuations?

Minsu Chang and Frank Schorfheide

A Derivations

A.1 Models for Functional Data

Some basic terminology from the literature on models for functional data; see Horvath and Kokoszka (2012). The *fully functional* model takes the form

$$Y_t(s) = \int \Psi(s,u)X_t(u)du + \epsilon_t(s).$$

Here both the dependent variable Y_t and explanatory variable X_t are functions. The *scalar response* model takes the form

$$Z_t = \int \Psi(s)X_t(s)ds + \epsilon_t.$$

Here the dependent variable Z_t is a scalar and the explanatory variable is a function. The *functional response* model takes the form

$$Y_t(s) = \Psi(s)Z_t + \epsilon_t.$$

In this model the dependent variable is a function and the explanatory variable is a scalar.

A.2 Functional Derivatives

Recall the definition of a derivative for a function $f(x)$:

$$f^{(1)}(x) = \lim_{h \longrightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

We can rewrite this relationship as

$$\lim_{h \longrightarrow 0} \frac{\|f(x+h) - f(x) - f^{(1)}(x)h\|}{|h|} = 0.$$

Now consider the linearization of the operator $F\big(p(\cdot)\big)$. Let $h(\cdot)$ be a functional discrepancy: $p(\cdot) = p_*(\cdot) + h(\cdot)$. Then, the Frechet derivative (if it exists) has the property

$$\lim_{\|h(\cdot)\|_1 \longrightarrow 0} \frac{\big\|F(p_* + h) - F(p_*) - DF[h]\big\|_2}{\|h\|_1} = 0.$$

This has to be true uniformly. The first-order Taylor approximation of $F(\cdot)$ then takes the form

$$F(p_* + h) = F(p_*) + DF[h] + o\big(\|h\|_1\big).$$

Here $DF[h]$ is a linear operator (in the same way $f^{(1)}(x)h$ is linear in h).

Example 1: Suppose that $F(p) = \int \Phi(x)p(x)dx$. Then we can write

$$F(p_* + h) = \int \Phi(x)[p_*(x) + h(x)]dx = \int \Phi(x)p_*(x)dx + \int \Phi(x)h(x)dx = F(p_*) + DF(h),$$

where

$$DF(h) = \int \Phi(x)h(x)dx.$$

Here $F(\cdot)$ was a linear operator so that there is no remainder term in the expansion. □

Example 2: Now consider the functional $F(p) = g\left(\int \Phi(x)p(x)dx\right)$. Here we can directly write

$$F(p_* + h) = g\left(\int \Phi(x)p_*(x)dx + \int \Phi(x)h(x)dx\right)$$

A standard first-order Taylor expansion around $\int \Phi(x)p_*(x)dx$ leads to

$$\begin{aligned} F(p_* + h) &= g\left(\int \Phi(x)p_*(x)dx\right) + g^{(1)}\left(\int \Phi(x)p_*(x)dx\right)\int \Phi(x)h(x)dx \\ &\quad + o\left(\left|\int \Phi(x)h(x)dx\right|\right) \\ &= F(p_*) + g^{(1)}\left(\int \Phi(x)p_*(x)dx\right)\int \Phi(x)h(x)dx + o\left(\left|\int \Phi(x)h(x)dx\right|\right) \end{aligned}$$

This suggests that

$$DF[h] = g^{(1)}\left(\int \Phi(x)p_*(x)dx\right)\int \Phi(x)h(x)dx$$

Suppose that the remainder term in the Taylor approximation can be bounded by

$$|\mathcal{R}| \leq M\left(\int \Phi(x)h(x)dx\right)^2.$$

Then

$$\begin{aligned}\frac{\left\|F(p_*+h)-F(p_*)-DF[h]\right\|_2}{\|h\|_1} &= \frac{|\mathcal{R}|}{\|h\|_1} \\ &\leq M\|h\|_1\left(\frac{\left|\int\Phi(x)h(x)dx\right|}{\|h\|_1}\right)^2 \\ &\leq M\|h\|_1\left(\frac{\left|\int|\Phi(x)|\cdot|h(x)|dx\right|}{\int|h(x)|dx}\right)^2\end{aligned}$$

Suppose, for instance, that $|\Phi(x)|$ can be uniformly bounded, then the desired convergence follows and we have established that $DF[h]$ is indeed a Frechet derivative. □

Example 3. Now consider $F(p)=\int g(x-\phi\tilde{x})p(\tilde{x})d\tilde{x}$, which means that $F(\cdot)$ is an operator. Then,

$$F(p_*+h)=\int g(x-\phi\tilde{x})p_*(\tilde{x})d\tilde{x}+\int g(x-\phi\tilde{x})h(\tilde{x})d\tilde{x}.$$

In this case

$$DF[h]=\int g(x-\phi\tilde{x})h(\tilde{x})d\tilde{x}$$

is still a linear operator.

Example 4. Now consider $F(p)=p(x)\int g(x-\phi\tilde{x})p_*(\tilde{x})d\tilde{x}$. Then,

$$\begin{aligned}F(p_*+h) &= [p_*(x)+h(x)]\int g(x-\phi\tilde{x})p_*(\tilde{x})d\tilde{x}+[p_*(x)+h(x)]\int g(x-\phi\tilde{x})h(\tilde{x})d\tilde{x} \\ &= F(p_*)+h(x)\int g(x-\phi\tilde{x})p_*(\tilde{x})d\tilde{x}+p_*(x)\int g(x-\phi\tilde{x})h(\tilde{x})d\tilde{x} \\ &\quad +h(x)\int g(x-\phi\tilde{x})h(\tilde{x})d\tilde{x}.\end{aligned}$$

In this case

$$DF[h]=h(x)\int g(x-\phi\tilde{x})p_*(\tilde{x})d\tilde{x}+p_*(x)\int g(x-\phi\tilde{x})h(\tilde{x})d\tilde{x}.$$

The remainder term is given by

$$\frac{\left\|F(p_*+h)-F(p_*)-DF[h]\right\|_2}{\|h\|_1} = \frac{\left\|h(x)\int g(x-\phi\tilde{x})h(\tilde{x})d\tilde{x}\right\|_2}{\|h(x)\|_1}.$$

B Logspline Density Estimation

Suppose that we have an *iid* sample from a distribution with a log density $\ell(x;\alpha)$. The parameter vector α can be estimated by maximum likelihood. The log likelihood function, imposing that the density has to normalize to one, is given by

$$\sum_{i=1}^{N}\ell(x_i;\alpha)=\sum_{k=1}^{K}\alpha_k\left(\sum_{i=1}^{N}\zeta_k(x_i)\right)-N\ln\left(\int\exp\left\{\sum_{k=1}^{K}\alpha_k\zeta_k(x)\right\}dx\right). \tag{A.1}$$

Using the convention that $\zeta_0(x)=1$, α_0 is a normalization constant, which is determined by

$$\alpha_0=-\ln\left(\int\exp\left\{\sum_{k=1}^{K}\alpha_k\zeta_k(x)\right\}dx\right). \tag{A.2}$$

Taking the derivative with respect to α_k yields

$$\frac{\partial}{\partial\alpha_k}\sum_{i=1}^{N}\ell(x_i;\alpha)=\sum_{i=1}^{N}\zeta_k(x_i)-N\int\zeta_k(x)\exp\left\{\sum_{k=0}^{K}\alpha_k\zeta_k(x)\right\}dx. \tag{A.3}$$

The score process is given by

$$\mathcal{S}(\alpha|x_{1:N})=\sum_{i=1}^{N}\left[\frac{\partial}{\partial\alpha_1}\ell(x_i;\alpha),\ldots\frac{\partial}{\partial\alpha_K}\ell(x_i;\alpha)\right]'.$$

The Hessian $\mathcal{H}(\alpha|x_{1:N})$ is composed of the elements

$$\begin{aligned}\frac{\partial^2}{\partial\alpha_k\alpha_l}\sum_{i=1}^{N}\ell(x_i;\alpha) &= -N\int\zeta_k(x)\zeta_l(x)\exp\left\{\sum_{k=0}^{K}\alpha_k\zeta_k(x)\right\}dx\\ &\quad +N\int\zeta_k(x)\exp\left\{\sum_{k=0}^{K}\alpha_k\zeta_k(x)\right\}dx\int\zeta_l(x)\exp\left\{\sum_{k=0}^{K}\alpha_k\zeta_k(x)\right\}dx.\end{aligned}$$

C Shock Identification

The following identification strategy is based on Uhlig (2005) and Kurmann and Otrok (2013). Recall that we defined $y_t = [\tilde{Z}_t', \tilde{\alpha}_t']'$ and specified a VAR of the form

$$y_t = D_1 y_{t-1} + \ldots + D_p y_{t-p} + u_t, \quad u_t \sim (0, \Sigma).$$

Here the aggregate variable $\tilde{Z}_t$ could be a vector. Let $n_{\tilde{Z}}$ be the dimension of $\tilde{Z}_t$ and $n_{\tilde{\alpha}}$ the dimension of $\tilde{\alpha}_t$. Suppose that the VAR has a moving average representation of the form

$$y_t = \sum_{j=0}^{\infty} H_j u_{t-j}.$$

We assume that the one-step-ahead innovations are function of underlying structural shocks ϵ_t:

$$u_t = \Sigma_{tr} \Omega \epsilon_t,$$

where Σ_{tr} is the lower-triangular Cholesky factor of Σ and Ω is an orthogonal matrix. We assume that the aggregate shocks are given by the first columns of Σ_{tr}. We are interested in a shock to the distribution that has a maximum impact on the aggregate shocks. So, Ω has a block-triangular structure

$$\Omega = \begin{bmatrix} I_{n_{\tilde{Z}}} & 0 \\ 0 & \Omega_{\tilde{\alpha}} \end{bmatrix}.$$

Here $I_{n_{\tilde{Z}}}$ is the $n_{\tilde{Z}} \times n_{\tilde{Z}}$ identity matrix. We are interested in identifying the first column of $\Omega_{\tilde{\alpha}}$, denoted by the $n_{\tilde{\alpha}} \times 1$ vector $q_{\tilde{\alpha}}$ by maximizing the effect of the corresponding shock on the i'th element of the vector y_t.

Let $n = n_{\tilde{Z}} + n_{\tilde{\alpha}}$ and e_i be the $n \times 1$ vector that has a one in position i and zeros elsewhere such that $y_{i,t} = e_i' y_t$. Moreover, define the matrix $M = [0_{n_{\tilde{Z}} \times n_{\tilde{\alpha}}}, I_{n_{\tilde{Z}}}]'$ such that

$$q = \begin{bmatrix} 0 \\ \\ q_{\tilde{\alpha}} \end{bmatrix} = M q_{\tilde{\alpha}}.$$

Using this notation, we can express the contribution of the first $\tilde{\alpha}_t$ innovation to the h-step ahead forecast error as

$$y_{i,t+h} - \mathbb{E}_t[y_{i,t+h}] = \ldots + e_i' \sum_{j=0}^{h-1} H_j \Sigma_{tr} M q_{\tilde{\alpha}} + \ldots \ . \tag{A.4}$$

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

$$q^*_{\tilde{\alpha}} = \text{argmax } e_i' \left[\sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} H_j \Sigma_{tr} M q_{\tilde{\alpha}} q_{\tilde{\alpha}}' M' \Sigma_{tr}' H_j' \right] e_i. \tag{A.5}$$

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

$$\begin{aligned} & e_i' \left[\sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} H_j \Sigma_{tr} M q_{\tilde{\alpha}} q_{\tilde{\alpha}}' M' \Sigma_{tr}' H_j' \right] e_i \qquad (A.6) \\ = & \sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} \text{tr} \left[(e_i e_i')(H_j \Sigma_{tr} M)(q_{\tilde{\alpha}} q_{\tilde{\alpha}}')(M' \Sigma_{tr}' H_j') \right] \\ = & \sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} \text{tr} \left[(q_{\tilde{\alpha}} q_{\tilde{\alpha}}')(M' \Sigma_{tr}' H_j')(e_i e_i')(H_j \Sigma_{tr} M) \right] \\ = & q_{\tilde{\alpha}}' \left[\sum_{h=1}^{\bar{h}} \sum_{j=0}^{h-1} (M' \Sigma_{tr}' H_j')(e_i e_i')(H_j \Sigma_{tr} M) \right] q_{\tilde{\alpha}} \\ = & q_{\tilde{\alpha}}' S q_{\tilde{\alpha}}. \end{aligned}$$

The optimization problem can therefore be expressed as Lagrangian

$$\mathcal{L} = q_{\tilde{\alpha}}' S q_{\tilde{\alpha}} - \lambda (q_{\tilde{\alpha}}' q_{\tilde{\alpha}} - 1), \tag{A.7}$$

which leads to the first-order condition

$$S q_{\tilde{\alpha}} = \lambda q_{\tilde{\alpha}}. \tag{A.8}$$

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.

D Multivariate Autoregressive Index Model

To describe the law of motion of the basis function coefficients, we are using a multivariate autoregressive index model, which is due to Reinsel (1983). We follow the recent treatment in Carriero, Kapetanios, and Marcellino (2016). The model takes the form

$$y_t = A_0 + \sum_{j=1}^{p} A_j \tilde{B} y_{t-j} + u_t, \quad u_t \sim N(0, \Sigma). \tag{A.9}$$

For expositional purposes, we are using y_t to denote the endogenous variables in this regression model. In the context of the functional autoregressive model y_t corresponds to the vector $\tilde{\alpha}_t$. We assume that y_t, u_t, and A_0 are $n \times 1$ vectors. The matrices A_j are of dimension $n \times r$, where $r \leq n$ and the matrix $\tilde{B}$ is of dimension $r \times n$. We can define the $r \times 1$ vector

$$F_t = \tilde{B} y_t,$$

which highlights that the endogenous variables are driven by r "factors." This specification helps us to cope with the problem that in practice there is a strong collinearity among the elements of the y_t vector. We normalize the factors by imposing that

$$\tilde{B} = [I_r, \, B],$$

where B is an $r \times (n - r)$ matrix. If we define the $rp + 1$ vector

$$x_t'(B) = [y_{t-1}'\tilde{B}', \ldots, y_{t-p}'\tilde{B}', 1]$$

and let

$$A = [A_1, \ldots, A_p, A_0]',$$

then the model can be written in matrix form as

$$Y = X(B)A + U, \tag{A.10}$$

where Y is the $T \times n$ matrix with rows y_t' and $X(B)$ is the $T \times (1 + rp)$ matrix with rows $x_t'(B)$. Note that conditional on B, inference on (A, Σ) is "standard," because the model takes the form of a linear Gaussian multivariate regression model.

The inference on the elements of the B matrix is non-standard. Carriero, Kapetanios, and Marcellino (2016) use a block RWMH step to generate draws from the B elements. But there might be a better solution.

E Additional Details on the Empirical Anaysis

E.1 Data Construction

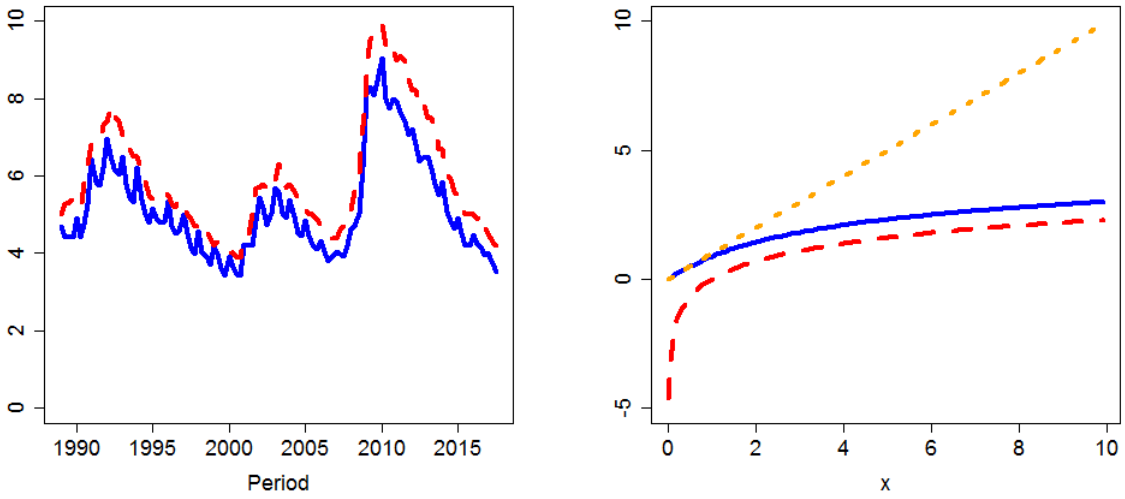
The observations on real per capita GDP, GDP deflator, and the unemployment rate are downloaded from the Federal Reserve Bank of St. Louis' FRED database:
https://fred.stlouisfed.org/.
The TFP series is available from the Federal Reserve Bank of San Francisco:
https://www.frbsf.org/economic-research/indicators-data/total-factor-productivity-tfp/.
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/

E.2 Data Features

The left panel of Figure 13 compares the aggregate unemployment rate to the unemployment rate computed from the CPS data. The levels of the two series are very similar, but the CPS unemployment rate exhibits additional high-frequency flucutations, possibly due to seasonals that have been removed from the aggregate unemployment rate.

Figure 13: CPS Unemployment and Earnings Transformation

Unemployment Rates Income Transformation



Notes: Left panel: CPS unemployment rate (blue, solid) and aggregate unemployment rate (red, dased). Right panel: inverse hyperbolic sine transformation (blue, solid) for $\theta = 1$ given in Eq. (A.11), logarithmic transformation (red, dashed), and 45-degree line (orange, dotted).

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

$$g(x|\theta) = \frac{\ln(\theta x + (\theta^2 x^2 + 1)^{1/2})}{\theta} = \frac{\sinh^{-1}(\theta x)}{\theta}. \tag{A.11}$$

The transformation is plotted in the right panel of Figure 13 for $\theta = 1$. Note that $g(0|\theta) = 0$ and $g^{(1)}(0|\theta) = 1$, that is, for small values of x the transformation is approximately linear. For large values of x the transformation is logarithmic:

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