Replication Report for Chang, Chen, and Schorfheide (2024)

Zhiheng You

1 Introduction

I replicate the empirical analysis section of the recent Journal of Political Economy (JPE) paper titled "Heterogeneity and Aggregate Fluctuations" by Minsu Chang, Xiaohong Chen, and Frank Schorfheide (hereafter referred to as CCS). CCS develop a state space model that can provide semi-structural evidence on the interaction between aggregate and distributional dynamics at business cycle frequencies. They then fit their model to data on aggregate variables and cross-sectional labor earnings, and uncover three key empirical findings: first, Granger-causal relationship from the cross-sectional income distribution to the aggregate variables is weak; second, an expansionary TFP shock decreases earnings inequality in our sample because it raises earnings at the bottom of the earnings distribution; third, the responses of aggregate output to distributional shocks are insignificant. In this report, I extend the original analysis by providing additional computational details on the estimation procedure—particularly those omitted from the main paper and its appendix—and presenting supplementary results that complement the findings reported in CCS.

2 Functional State Space Model

To make this report self-contained, I begin by presenting the functional state space model proposed by CCS. I extend their original notation to allow for multiple VAR lags, as in the implementation I work with multiple lags and use a data-dependent method to choose the lag periods.

Transition equation: For an $n_y \times 1$ vector of macroeconomic aggregates Y_t and a cross-sectional log densities $\ell_t(x) = \ln p_t(x)$, we decompose them into a deterministic component $(Y_*, \ell_*(x))$ and fluctuations around it:

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

The deviations from the deterministic component $(Y_t, \ell_t(x))$ evolve jointly according to the following linear functional VAR:

$$\tilde{Y}_t = \sum_{h=1}^{p} B_{yy}^{(h)} \tilde{Y}_{t-h} + \sum_{h=1}^{p} \int B_{yl}^{(h)}(\tilde{x}) \tilde{\ell}_{t-h}(\tilde{x}) d\tilde{x} + u_{y,t},$$
$$\tilde{\ell}_t(x) = \sum_{h=1}^{p} B_{ly}^{(h)}(x) \tilde{Y}_{t-h} + \sum_{h=1}^{p} \int B_{ll}^{(h)}(x, \tilde{x}) \tilde{\ell}_{t-h}(\tilde{x}) d\tilde{x} + u_{l,t}(x).$$

We represent the density, kernels, and functional innovation using K-dimensional basis functions[1]:

$$\ell_t(x) = \zeta'(x)\alpha_t,$$
$$B_{ll}^{(h)}(x, \tilde{x}) = \zeta'(x) B_{ll}^{(h)} \xi(\tilde{x}), \quad B_{yl}^{(h)}(x) = B_{yl}^{(h)} \xi(\tilde{x}),$$
$$B_{ly}^{(h)}(x) = \zeta'(x) B_{ly}^{(h)}, \quad u_{l,t}(x) = \zeta'(x) u_{\alpha,t},$$

where $\zeta(x)$ and $\xi(x)$ are two $K \times 1$ vectors of basis functions and $u_{\alpha,t}$ is a $K \times 1$ vector of innovations. Then the VAR system can be rewritten as

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

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

Measurement equation: In every period $t = 1, \ldots, T$ an econometrician observes aggregates Y_t and a sample of N_t i.i.d. draws $x_{it}, i = 1, \ldots, N_t$ from cross-sectional density $p_t(x)$. Denote $\mathcal{L}(\alpha_t \mid X_t) = \frac{1}{N_t} \sum_{i=1}^{N_t} \ell_t(x_{it})$. Then the measurement equation for the time t cross-sectional observations $X_t = [x_{1t}, \ldots, x_{N_t,t}]'$ can be expressed as

$$p(X_t \mid \alpha_t) = \exp\{N_t \mathcal{L}(\alpha_t \mid X_t)\}, \quad \mathcal{L}(\alpha_t \mid X_t) = \bar{\zeta}'(X_t)\alpha_t - \varphi(\alpha_t),$$
$$\bar{\zeta}(X_t) = \frac{1}{N_t} \sum_{i=1}^{N_t} \zeta(x_{it}), \quad \varphi(\alpha_t) = \ln \int \exp\{\zeta'(x)\alpha_t\} dx.$$

[1]By the standard argument for sieve estimators, the functional can be approximated as a linear combination of basis functions, with the approximation error approaching zero as K increases. Here I omit the (K) superscript to compress the notation.

3 Estimation of Aggregate VAR

Let Y_t be a $n \times 1$ ($n = 3$) vector of TFP growth, GDP growth, and unemployment rate, and $W_t = Y_t - Y_*$. Then the aggregate VAR model can be written as: from $t = 1, \ldots, T$

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

In matrix form,

$$W = \tilde{Z}\Phi + U,$$

where $W = [W_1, \ldots, W_T]'$, $\tilde{Z} = [\tilde{Z}_1, \ldots, \tilde{Z}_T]'$ with rows $\tilde{Z}_t' = \left[W_{t-1}', \ldots, W_{t-p}'\right]$, and $\Phi = [\Phi_1, \ldots, \Phi_p]'$.

Prior: To perform Bayesian estimation on the aggregate VAR model, we impose a Minnesota-like prior on the reduced-form parameters Φ and Σ:

$$\begin{gathered}
\Sigma \sim IW(\underline{\nu}, \underline{S}), \quad \underline{S} = \operatorname{diag}\left(\underline{s}_1^2, \ldots, \underline{s}_n^2\right), \\
\operatorname{vec}(\Phi) \mid \Sigma \sim \mathcal{N}(0, \Sigma \otimes \underline{V_\phi}), \\
np \times np \text{ matrix } \underline{V_\phi} \text{ diagonal with } \left[\underline{V_\phi}\right]_{ll} = \frac{1}{\lambda_1 \lceil \frac{l}{n} \rceil^2}.
\end{gathered}$$

We set $\underline{\nu} = 2$ and $\underline{s}_i^2 = [\Sigma^{ols}]_{ii}$, where $\Sigma^{ols} = W'W - W'\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'W$.

Choice of hyperparameters: We maximize the marginal data density (MDD) to select hyperparameters λ_1 and p. The log MDD can be derived analytically as

$$\begin{aligned}
\ln p(W) = & -\frac{nT}{2}\ln(2\pi) - \frac{n\underline{\nu}}{2}\ln 2 - \ln\Gamma_n\left(\tfrac{\underline{\nu}}{2}\right) + \ln\Gamma_n\left(\tfrac{\underline{\nu}+T}{2}\right) \\
& - \tfrac{n}{2}\ln\left|\underline{V_\phi}\right| - \tfrac{n}{2}\ln\left|\tilde{Z}'\tilde{Z} + \underline{V_\phi}^{-1}\right| - \frac{\underline{\nu}+T}{2}\ln\left|\underline{S} + \widehat{E}'\widehat{E}\right| + \tfrac{\underline{\nu}}{2}\ln\left|\underline{S}\right|,
\end{aligned}$$

where $\Gamma_n()$ is the multivariate gamma function, $\widehat{E} = W - \tilde{Z}\widehat{\Phi}$, and $\widehat{\Phi} = \left(\tilde{Z}'\tilde{Z} + \underline{V_\phi}^{-1}\right)^{-1}\tilde{Z}'W$. We pick the grid for λ_1 to be $\{\exp(-10+k), k = 0, 1, 2, \ldots, 30\}$, and that for p to be $\{1, 2, 3, 4\}$. The optimal hyperparameters are $\hat{\lambda}_1 = e^3$ and $\hat{p} = 4$.

Posterior: We draw Φ and Σ from the following posterior distributions

$$\Sigma \,\big|\, W \;\sim\; IW\Big(\underline{\nu}+T,\; \underline{S}+\widehat{E}'\,\widehat{E}\Big),$$
$$\mathrm{vec}(\Phi) \;\big|\; \Sigma,\; W \;\sim\; \mathcal{N}\Big(\mathrm{vec}(\widehat{\Phi}),\; \Sigma \otimes \big(\tilde{Z}'\tilde{Z}+\underline{V_\phi}^{-1}\big)^{-1}\Big).$$

Impulse responses: The IRFs are identified from Cholesky decomposition, where we order TFP growth first, GDP growth second, and the employment rate third. For each draw of Φ and Σ, we compute the corresponding IRF. Figure 1 demonstrates the 80% confidence band of impulse responses.

4 Estimation of the Functional State Space Model

4.1 Overview

The measurement equation can be approximated by

$$\hat{\alpha}_t = \alpha_t + N_t^{-1/2}\eta_t, \quad \eta_t \sim \mathcal{N}\left(0, \hat{V}_t\right),$$

where $\hat{\alpha}_t$ is the maximizer of likelihood function $\mathcal{L}^{(K)}\left(\alpha_t \mid X_t\right)$, and $\hat{V}_t$ is the negative inverse Hessian associated with $\mathcal{L}^{(K)}\left(\alpha_t \mid X_t\right)$ evaluated at $\hat{\alpha}_t$.

Hence, the state space model can be written as

$$\hat{\alpha}_t = \alpha_t + N_t^{-1/2}\eta_t, \quad \eta_t \sim \mathcal{N}\left(0, \hat{V}_t\right),$$
$$\begin{bmatrix} Y_t - Y_* \\ \alpha_t - \alpha_* \end{bmatrix} = \sum_{h=1}^{p} \begin{bmatrix} \Phi_{yy}^{(h)} & \Phi_{y\alpha}^{(h)} \\ \Phi_{\alpha y}^{(h)} & \Phi_{\alpha\alpha}^{(h)} \end{bmatrix} \begin{bmatrix} Y_{t-h} - Y_* \\ \alpha_{t-h} - \alpha_* \end{bmatrix} + \begin{bmatrix} u_{y,t} \\ u_{\alpha,t} \end{bmatrix}, \quad u_t \sim \mathcal{N}(0, \Sigma).$$

CCS propose a two-step procedure: in the first step, one computes the sequence of MLEs $\hat{\alpha}_t$ from the cross-sectional observations X_t period by period; in the second step, one estimates the a linear state space model where $\hat{\alpha}_t$s and Y_ts are treated as observables and α_ts are latent states.

Regarding the second step, standard Bayesian estimation involves a Gibbs sampler that iterates between: (a) drawing (Φ, Σ) given $(Y_{-p+1:T}, \alpha_{-p+1:T})$, where Φ collects all the VAR coefficients; (b) drawing $\alpha_{-p+1:T}$ given $(Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}, \Phi, \Sigma)$ (simulation smoothing). As

usual, to implement the simulation smoothing, one can first run the Kalman filter to obtain posterior distributions $s_t|Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}, \Phi, \Sigma$ for $t = 1, \ldots, T$, where $s_t = [\alpha_t, \ldots, \alpha_{t-p+1}]'$. Starting from $s_T|Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}, \Phi, \Sigma$, one can then draw blocks backwards $s_{T-p}|s_T$, $s_{T-2p}|s_{T-p}, s_T$, $s_{T-3p}|s_{T-2p}, s_{T-p}, s_T$... (I omit $Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}, \Phi, \Sigma$ in the conditioning set).

The organization of Section 4 is as follows. Section 4.2 discusses how to estimate $\hat{\alpha}_t$s with MLE. Section 4.3 provides details on how to draw (Φ, Σ) given $(Y_{-p+1:T}, \alpha_{-p+1:T})$ in the state-transition equation, as well as how to construct IRFs based on sampled (Φ, Σ). Section 4.4 describes the smoothing simulator that draws $\alpha_{-p+1:T}$ given $(Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}, \Phi, \Sigma)$. Section 4.5 discusses how to select hyperparameters.

4.2 Density Estimation

As in CCS, I work with the density of transformed earnings x instead of that of raw earnings to GDP ratio z. The relationship between x and z is

$$x = g(z) = \ln\left(z + \left(z^2 + 1\right)^{1/2}\right) = \sinh^{-1}(z).$$

Since CPS applies top coding to the earnings data, we need to use the likelihood function with censoring to estimate the α_ts. Define $c_t = \max_{i=1,\ldots,N} x_{it}$, $N_{t,\max} = \sum_{i=1}^{N} \mathbb{I}\{x_{it} = c_t\}$, $\pi_t = \mathbb{P}(x_{it} \geq c_t)$, and $\bar{\zeta}_t(X_t) = \frac{1}{N_t}\sum_{i=1}^{N_t} \zeta(x_{it})\,\mathbb{I}\{x_{it} < c_t\}$. Then the likelihood function for period t observations is

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

The MLE estimator for π_t is $N_{t,\max}/N_t$, while that for α_t is

$$\hat{\alpha}_t = \text{argmax}_{\alpha_t}\,\mathcal{L}^{(K)}(\alpha_t \mid X_t) = \text{argmax}_{\alpha_t}\,\bar{\zeta}_t'(X_t)\,\alpha_t - \frac{N_t - N_{t,\max}}{N_t}\ln\int_0^{c_t}\exp\{\zeta'(x)\alpha_t\}\,dx.$$

A by-product of the estimation is $\hat{V}_t$, the negative inverse Hessian associated with $\mathcal{L}^{(K)}(\alpha_t \mid X_t)$

evaluated at $\hat{\alpha}_t$, whose (k,l)-th element can be calculated as

$$[\hat{V}_t]_{kl} = -\left(\frac{N_t - N_{t,\,\max}}{N_t}\right)\int_0^{c_t}\left(\zeta_k(x) - \int_0^{c_t}\zeta_k(x)\bar{p}\left(x \mid \hat{\alpha}_t\right)dx\right) \times \left(\zeta_l(x) - \int_0^{c_t}\zeta_l(x)\bar{p}\left(x \mid \hat{\alpha}_t\right)dx\right)\bar{p}\left(x \mid \hat{\alpha}_t\right)dx, \tag{2}$$

where

$$\bar{p}\left(x \mid \hat{\alpha}_t\right) = \frac{\exp\left\{\sum_{k=1}^{K}\hat{\alpha}_{k,t}\zeta_k(x)\right\}}{\int_0^{c_t}\exp\left\{\sum_{k=1}^{K}\hat{\alpha}_{k,t}\zeta_k(x)\right\}dx}\mathbb{I}\left\{x < c_t\right\}.$$

For the choice of basis function, I pick the spline function from upper bound $\bar{x}$ to 0 as in CCS:

$$\zeta_K(x) = \max\{\bar{x} - x, 0\}$$
$$\zeta_k(x) = \left[\max\left\{x_k - x, 0\right\}\right]^3, \quad k = K-1, \ldots, 1.$$

The knots $\{x_k\}$ are placed at certain percentiles (see Table 2 in CCS).

In terms of the optimization algorithm, I use Newton's method with Jacobian and Hessian computed by automatic differentiation. If error occurs, I then switch to BFGS.

To remove the seasonality in the estimated $\hat{\alpha}_t$s, I first regress $\hat{\alpha}_t$s on four seasonal (quarter) dummies, and then take the residuals as the new $\hat{\alpha}_t$s. This is equivalent to detracting the seasonal mean from the original $\hat{\alpha}_t$s. Then we can define α_* to be the seasonal mean of $\hat{\alpha}_t$, except in the IRF computation where we let α_* to be the all-time mean.

Since the time variation in cross-sectional densities might be concentrated in a lower-dimensional space, we need to compress the coefficients before estimating the state space model. Define $\hat{\tilde{\alpha}}_t = \hat{\alpha}_t - \alpha_*$ and $T \times K$ matrix $\hat{\tilde{\alpha}} = [\hat{\tilde{\alpha}}_1', \ldots, \hat{\tilde{\alpha}}_T']$. We conduct a principal components analysis (PCA) 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 $\hat{a} = \hat{\tilde{\alpha}}\hat{M}$, and $\hat{\Lambda} = \left(\hat{a}'\hat{a}\right)^{-1}\hat{a}'\hat{\tilde{\alpha}}$, where $\hat{a}$ is the $T \times \tilde{K}$ matrix with rows $\hat{a}_t'$.

4.3 VAR Estimation

Prior: Similar to Section 3, we impose a Minnesota-like shrinkage prior on the parameters. However, this prior allows for cross-variable shrinkage, that is, shrinking coefficients on the lags of the other group of variables compared to the lags of their own group. Here,

"group" refers to either aggregate variables Y_ts or the cross-sectional distribution variables α_ts. It is important to note that the Kronecker-product structure of the original prior, $p(\text{vec}(\Phi) \mid \Sigma)$, requires symmetric treatment of own lags and lags of other variables. Consequently, constructing a new prior is necessary to accommodate this asymmetry. CCS adopt the asymmetric conjugate priors proposed by Chan (2022) to address this requirement.

We follow the notation in Section 3 and write the state-transition equation as

$$W_t = \sum_{h=1}^{p} \Phi_h W_{t-h} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma),$$

where $n \times 1$ vector $W_t = [(Y_t - Y_*)', (\alpha_t - \alpha_*)']'$, and Y_*, α_* are computed as the time series average of Y_t, α_t. In matrix form,

$$W = \tilde{Z}\Phi + U,$$

where $W = [W_1, \ldots, W_T]'$, $\tilde{Z} = [\tilde{Z}_1, \ldots, \tilde{Z}_T]'$ with rows $\tilde{Z}_t' = \left[W_{t-1}', \ldots, W_{t-p}'\right]$, and $\Phi = [\Phi_1, \ldots, \Phi_p]'$.

One can first rewrite the VAR in the following quasi-structural form:

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

where D is a diagonal matrix with diagonal elements D_i and A a lower-triangular matrix with ones on the diagonal. Further define $\mathcal{A}_i = [A_{i,1}, \ldots, A_{i,i-1}]$, $\tilde{W}_{<i,t} = -[W_{1,t}, \ldots, W_{i-1,t}]'$, for $i = 2, \ldots, n$. Then, we can write the i th equation as

$$W_{i,t} = \tilde{W}_{<i,t}'\mathcal{A}_i + \tilde{Z}_t' B_{\cdot i} + \epsilon_{i,t}.$$

To simplify the notation further, let $k_i = k + i - 1$, and define the $k_i \times 1$ vectors $Z_{i,t} = \left[\tilde{W}_{<i,t}', \tilde{Z}_t\right]'$ and $\beta_i = [\mathcal{A}_i', B_{\cdot i}']'$. Write the i th 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

$$W_i = Z_i\beta_i + \epsilon_i.$$

CCS impose the following prior on the quasi-structural parameters (β_i, D_i):

$$\beta_i \mid D_i \sim \mathcal{N}\left(\underline{\beta}_i, D_i \underline{V}_i^{\beta}\right), \quad D_i \sim IG\left(\frac{\underline{\nu}+i-n}{2}, \frac{\underline{s}_i^2}{2}\right),$$

where the prior of α_i component of β_i takes the form

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

and the prior of $B_{\cdot i}$ component takes the form

$$B_{\cdot i} \sim \mathcal{N}\left(0, \underline{V}_i^B\right), \quad \left[\underline{V}_i^B\right]_{f(j,h)} = \left[\underline{V}_i^{\phi}\right]_{f(j,h)} + \sum_{l=1}^{i-1}\left(\left[\underline{V}_l^{\phi}\right]_{f(j,h)} + \frac{1}{\underline{s}_l^2}\left[\underline{\phi}_l\right]_{f(j,h)}^2\right),$$

with $f(j,h) = (h-1)n + j$ and $\underline{V}_i^{\phi}$ is a $np \times np$ diagonal matrix with

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

To interpret the hyperparameters, λ_1 controls the overall precision of VAR coefficients $B_{\cdot i}$, λ_2 governs the relative precision for a to Y transmission compared to within-group transmission, and λ_3 captures the relative precision for Y to a transmission.

We set $\lambda_3 = 1$, $\underline{\nu} = 2$, and $\underline{s}_i^2 = [\Sigma^{ols}]_{ii}$, where $\Sigma^{ols} = W'W - W'\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'W$.

Posterior and MDD: Define precision matrices $\underline{P}_i^{\beta} = \left(\underline{V}_i^{\beta}\right)^{-1}$ and $\bar{P}_i^{\beta} = \left(\bar{V}_i^{\beta}\right)^{-1}$. Then the updating equations for the posterior take the form

$$\beta_i \mid (D_i, W_i) \sim \mathcal{N}\left(\bar{\beta}_i, D_i \bar{V}_i^{\beta}\right), \quad D_i \sim IG\left(\bar{\nu}_i, \bar{S}_i\right),$$

where

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

And the MDD of the VAR system [2]can be computed analytically as follows:

$$\begin{aligned}\mathcal{L}(Y,\alpha|\lambda_1,\lambda_2,p) = &-\frac{Tn}{2}\ln(2\pi)\\ &+\sum_{i=1}^{n}\left[\frac{1}{2}\left(\ln\left|\underline{P}_i^{\beta}\right|-\ln\left|\bar{P}_i^{\beta}\right|\right)+\underline{\nu}\ln|\underline{S}_i|-\bar{\nu}\ln\left|\bar{S}_i\right|-\ln\Gamma\left(\underline{\nu}_i\right)+\ln\Gamma\left(\bar{\nu}_i\right)\right].\end{aligned} \tag{3}$$

For each posterior draw of (β_i, D_i), we can back out the reduced form parameter (Φ, Σ) using the following formula:

$$\Phi_h = A^{-1}B_h, \quad \Sigma = A^{-1}DA^{-1},$$

where A and B_h are constructed from $\mathcal{A}_i$s and $B_{.i}$s, respectively.

Impulse responses: I compute four types of impulse responses: IRFs of aggregate variables to aggregate shocks, IRFs of cross-sectional density to aggregate shocks, IRFs of inequality measure to aggregate shocks, and IRFs of aggregate variables to distributional shocks. For the first type, the construction is the same as in Section 3. For the second type, I first transform the IRF of a_t (denoted by $\dot{a}_t$) into the new density coefficients α_t (denoted by $\dot{\alpha}_t$): $\dot{\alpha}_t' = \alpha_*' + \dot{a}_t'\hat{\Lambda}$, where $\hat{\Lambda}$ is defined in Section 4.2. Define

$$p^{(K)}\left(x \mid \alpha_t\right) = \frac{\exp\left\{\zeta'(x)\alpha_t\right\}}{\int \exp\left\{\zeta'(\tilde{x})\alpha_t\right\} d\tilde{x}}. \tag{4}$$

Then the IRF of cross-sectional distribution can be obtained as

$$IRF(x) = p^{(K)}\left(x \mid \dot{\alpha}_t\right) - p^{(K)}\left(x \mid \alpha_*\right).$$

Here I emphasize the difference between results computed from density with no pointmass at 0 and those from density with pointmass at 0. The above definition applies to density with no pointmass at 0, i.e., the unemployed are excluded from the distribution. To consider the density with pointmass at 0, we need to assign the unemployed with zero earnings. Specifically, denote the unemployment rate is u_t, then Eq (4) should be modified as

$$p^{(K)}\left(x \mid \alpha_t\right) = u_t \cdot \delta(x) + (1-u_t)\cdot\frac{\exp\left\{\zeta'(x)\alpha_t\right\}}{\int \exp\left\{\zeta'(\tilde{x})\alpha_t\right\} d\tilde{x}}, \tag{5}$$

where $\delta(x)$ is a Dirac distribution at $x = 0$.

[2]This is not the MDD of the functional state space model.

For the third type, I focus on three types of inequality measures: 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 earnings-to-density distribution. I first back out the after-shock earnings-to-density distribution from the transformed distribution using $p_z^{(K)}(z \mid \dot{\alpha}_t) = p_x^{(K)}(g(z) \mid \dot{\alpha}_t)|g'(z)|$, where $g(.)$ is defined in Section 4.2, and $g'(z) = 1/(z^2+1)^{1/2}$. Then I compute the relative measures $\mathcal{M}(\dot{\alpha}_t)$ based on $p_z^{(K)}(z \mid \dot{\alpha}_t)$. The IRFs are defined as $\mathcal{M}(\dot{\alpha}_t) - \mathcal{M}(\alpha_*)$.

For the fourth type, I define the distributional shock as a structural shock (i.e., a linear combination of the reduced-form shock) that maximizes the Gini coefficient of the on-impact (i.e., the period that shock happens) earnings-to-density distribution. To find the linear weights, I randomly draw 500 weights that sum to 1 and pick the optimal q^* that maximizes the Gini coefficient. After finding the optimal weights, the problem is reduced to computing the IRF of aggregate variables to a structural shock.

4.4 Smoothing Simulator

As laid out in Section 4.1, to implement the simulation smoothing, one can first run the Kalman filter and then draw blocks of latent states backwards from $t = T$. In the Appendix Section B.6, the authors provide the exact formula on how standard Kalman filter/simulation smoother can be extended to multiple lags. The basic idea is to first transform the VAR(p) system into its companion form (VAR(1)) and then follow the standard procedure. Therefore, I save the paragraphs here and follow the exact formula stated in their appendix in the implementation.

4.5 Hyperparameter Determination

To determine the hyperparameters $(K, \lambda_1, \lambda_2, p)$, we can maximize the log MDD of the functional state space system. Under the assumptions made in CCS, the log MDD can be approximated by

$$\frac{KT}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln\left|\hat{V}_t\right| + \sum_t N_t \mathcal{L}^{(K)}\left(\hat{\alpha}_t, \pi_t \mid X_t\right) + \mathcal{L}(Y, \hat{\alpha}|\lambda_1, \lambda_2, p), \tag{6}$$

where $\hat{V}_t$ is defined in Eq(2), $\mathcal{L}^{(K)}\left(\hat{\alpha}_t, \pi_t \mid X_t\right)$ is defined in Eq(1), and $\mathcal{L}(Y, \hat{\alpha}|\lambda_1, \lambda_2, p)$ is defined in Eq(3).

Since in practice we replace $\hat{\alpha}$ with PCA-compressed coefficients $\hat{a}$, the measurement equation is now

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

Eq(6) should thus be modified as

$$\frac{\tilde{K}T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln\left|\left(\hat{\Lambda}\hat{V}_t^{-1}\hat{\Lambda}'\right)^{-1}\right| + \sum_t N_t \mathcal{L}^{(\tilde{K})}\left(\alpha_* + \hat{\Lambda}'\hat{a}_t \mid X_t\right) + \mathcal{L}(Y, \hat{a}|\lambda_1, \lambda_2, p). \quad (7)$$

5 Estimation of Inequality VAR

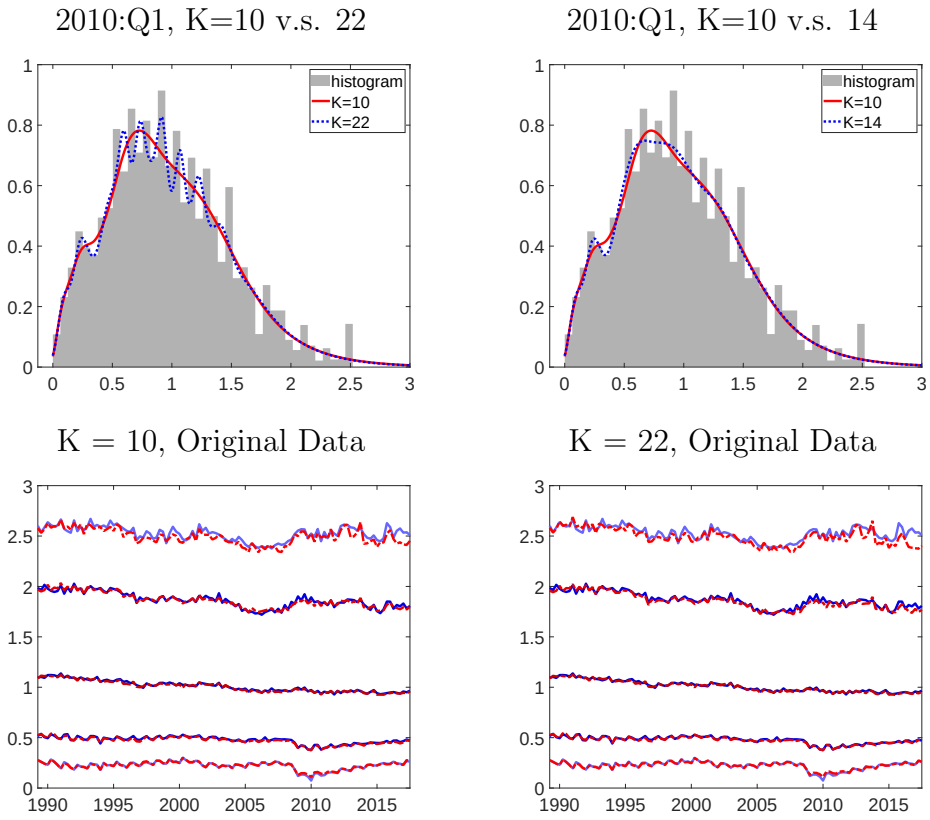
For comparison with the functional state space model, CCS also estimate VAR models with inequality measures included in Y_t. They consider two specifications: Specification 1 combines aggregate variables with the fraction of individuals earning less than GDP per capita and the Gini coefficient computed from cross-sectional observations, while Specification 2 adds the 10th, 20th, 50th, 80th, and 90th percentiles. The estimation procedure follows exactly the same approach as described in Section 3.

6 Results

6.1 Density Estimation

Figure 1 presents the fitted densities and percentiles of the earnings-to-GDP distribution for different Ks. CCS compare the fitted densities for $K = 10$ and $K = 22$ (see Figure 8 in their paper), and argue that $K = 10$ is more suitable than $K = 22$ as the latter captures undesirable jaggedness of the histograms caused by the granular reporting design of the survey data—a pattern I successfully replicate in the upper left panel. However,they do not explain why they specifically pick $K = 10$. As a robustness check, I compare the fitted densities for $K = 10$ and $K = 14$ in the upper right panel.The two densities appear nearly identical, except around 0.4, where the destiny for $K = 10$ is smoother, while the density for $K = 14$ captures the jaggedness. This comparison supports CCS's choice of $K = 10$ for presenting results. Moreover, the estimated percentiles under both specifications closely match the sample percentiles, except for some small inaccuracies in the 90th percentile toward the end of the sample. This finding is also consistent with the results reported by CCS.

Figure 1: Fitted Densities and Percentiles of Earnings/GDP Distribution



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

6.2 Hyperparameter Determination

Table 1 reports the optimal hyperparameter choices for different values of K (Columns 1-5). Overall, my results align with Table 4 in CCS, except for discrepancies in the optimal value of λ_2. Specifically, the values of $\hat{\lambda}_2$ match at $K = 14$ and $K = 22$ but differ for other values of K. OOne possible explanation is that the Granger causality in the cross-sectional distribution is weak, making it difficult to detect how changes in λ_2 influence the MDD. As a result, the optimal estimate $\hat{\lambda}_2$ can vary significantly with even minor changes in implementation.

As in CCS, I also compare the optimal hyperparameter selection with the case where λ_2 is fixed at 1 (Columns 6-8). The resulting change in MDD is much smaller than the changes observed across different K values. This is because λ_2 primarily affects the model's fit in the time series dimension rather than the cross-sectional dimension.

Following CCS, I adopt $K = 10$ for the subsequent analysis. The corresponding optimal hyperparameters are $\hat{\lambda}_1 = 54.6$, $\hat{\lambda}_2 = 4.9E + 8$, and $\hat{p} = 1$. The values of $\hat{\lambda}_1$ and $\hat{p}$ are consistent with CCS, but $\hat{\lambda}_2$ differs from the original value of 1096.6 reported in their paper. This discrepancy in $\hat{\lambda}_2$ is also one reason why my subsequent replication results may slightly differ from those of CCS.

Table 1: 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	20.1	4	0	54.6	4	0
6	20.1	4.9E+8	1	8559	54.6	1	8552
8	54.6	4.9E+8	1	9387	54.6	1	9376
10	54.6	4.9E+8	1	9392	148.4	1	9381
14	148.4	4.9E+8	1	9724	403.4	1	9715
18	403.4	4.9E+8	1	9779	403.4	1	9774
22	403.4	7.4	1	12810	403.4	1	12803

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 $\hat{p}(K)$.

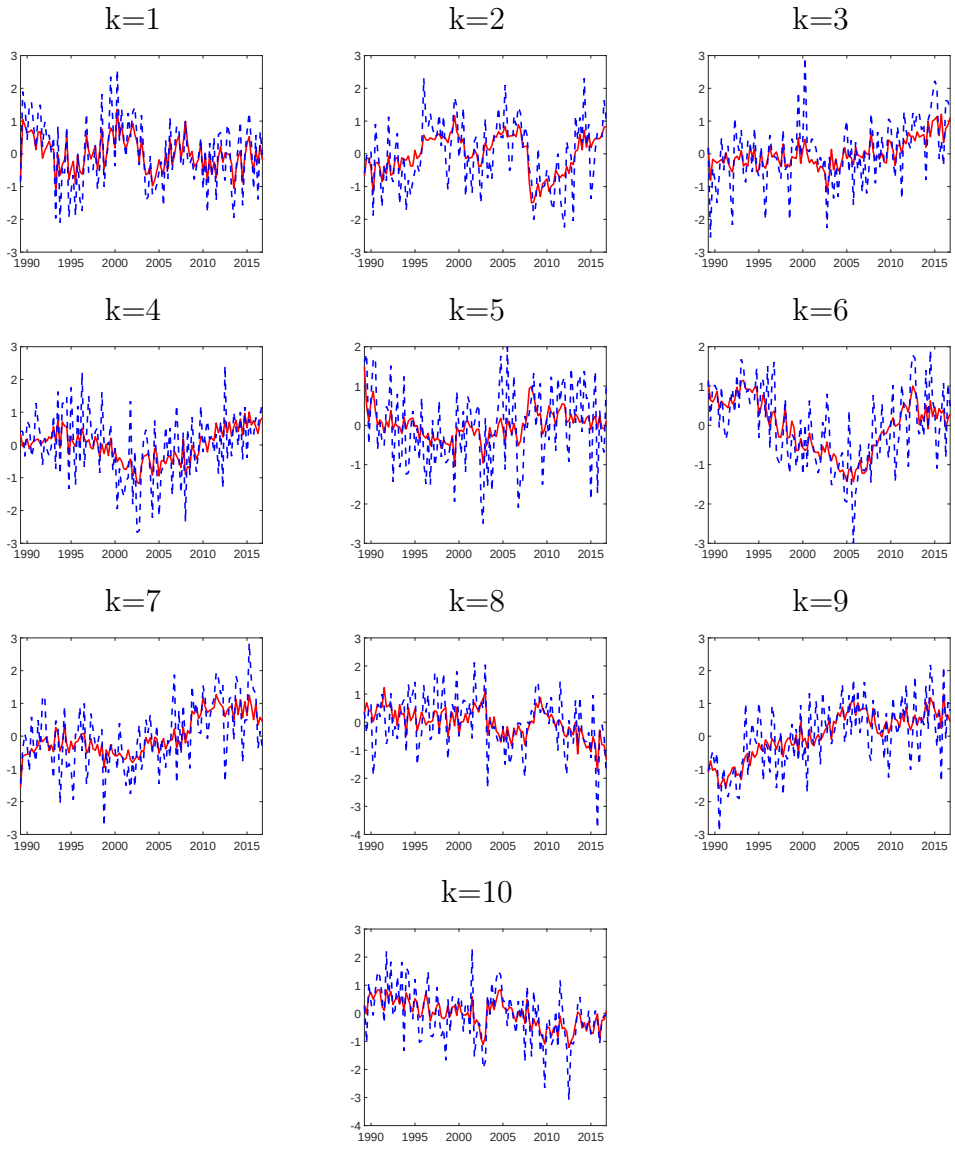
6.3 Simulation Smoothing

To illustrate the simulation smoother, Figure 2 overlays the estimated $\hat{a}_{k,t}$ versus the smoothed $a_{k,t}$s generated as output of the smoother. As intended, the smoothed series are inherently smoother than the original series. Recall that the transformation from $\alpha_{k,t}$ to $a_{k,t}$ demeans and orthogonalizes the series. As a result, different series capture fluctuations around zero at varying frequencies.

6.4 Aggregate VAR v.s. Functional State Space Model

Figure 3 compares the impulse responses of aggregate variables to aggregate shocks generated by the aggregate VAR and the functional model. To facilitate a clearer comparison of IRFs, I report 66% confidence bands instead of the 80% confidence bands used in CCS. Overall, my results match the shape and magnitude of the impulse responses reported in Figure 9 of CCS. Specifically, a TFP growth shock leads to a permanent increase in the level of TFP. GDP also rises permanently, and employment responds positively, consistent with

Figure 2: Estimated $\hat{a}_{k,t}$ versus Smoothed $a_{k,t}$

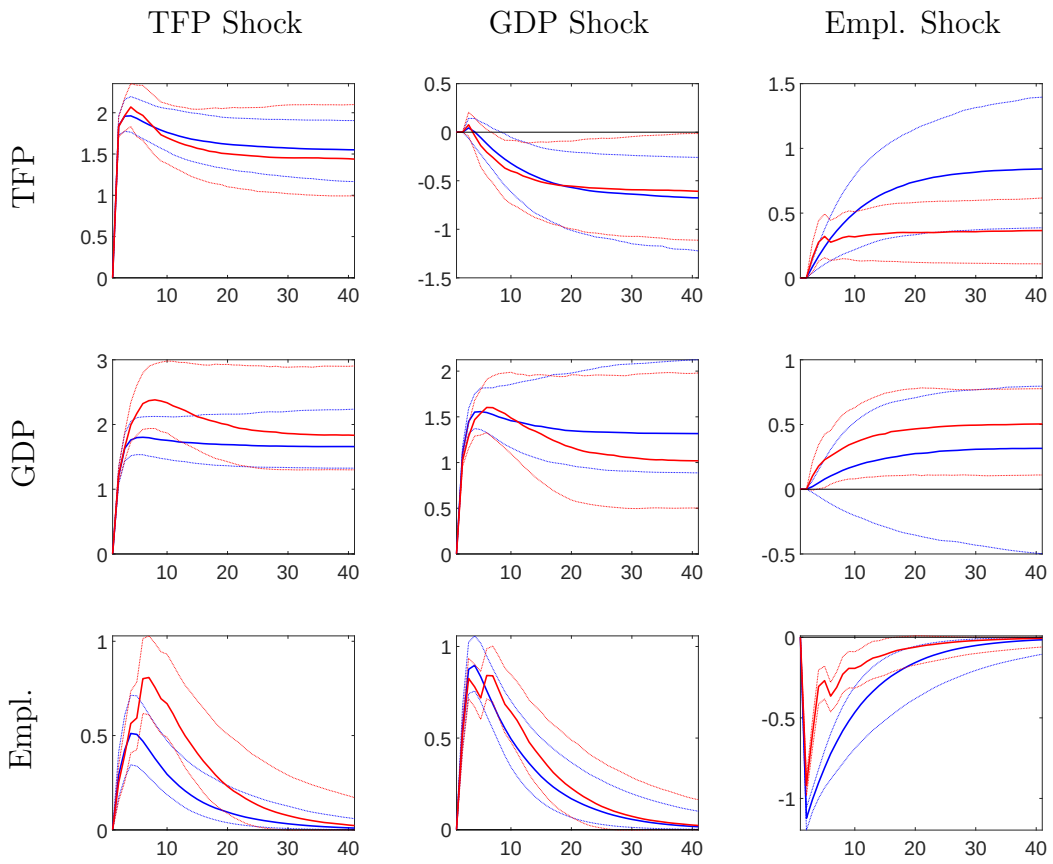


Notes: The red solid lines correspond to the smoothed $a_{k,t}$ series, whereas the blue dashed lines represent the estimated series $\hat{a}_{k,t}$.

predictions from the RBC model. A GDP growth shock results in a permanent increase in GDP, a temporary employment boom, and a long-run decline in measured TFP. Similarly, an employment shock causes a drop in the employment rate while increasing TFP and GDP with a one-period lag.

However, one major difference between my results and those in CCS is that, the IRFs produced by the two models in Figure 3 exhibit significant quantitative differences, particularly the responses to the employment shock. Specifically, the responses of TFP and employment to the employment shock are larger in the functional model than in the aggregate VAR, while the permanent positive response of GDP is significant in the aggregate VAR but insignificant in the functional model. This discrepancy arises because, in my implementation, the MDD criterion selects $\hat{p} = 4$ for the aggregate VAR but $\hat{p} = 1$ for the functional model. This difference in model specification leads to varying IRFs. In contrast, CCS select a lag length of 1 for both models, and due to the weak Granger causality of the cross-sectional distribution, the causal relationships among aggregate variables remain largely unchanged when distributional information is incorporated.

Figure 3: Responses of Aggregate Variables to Aggregate Shocks

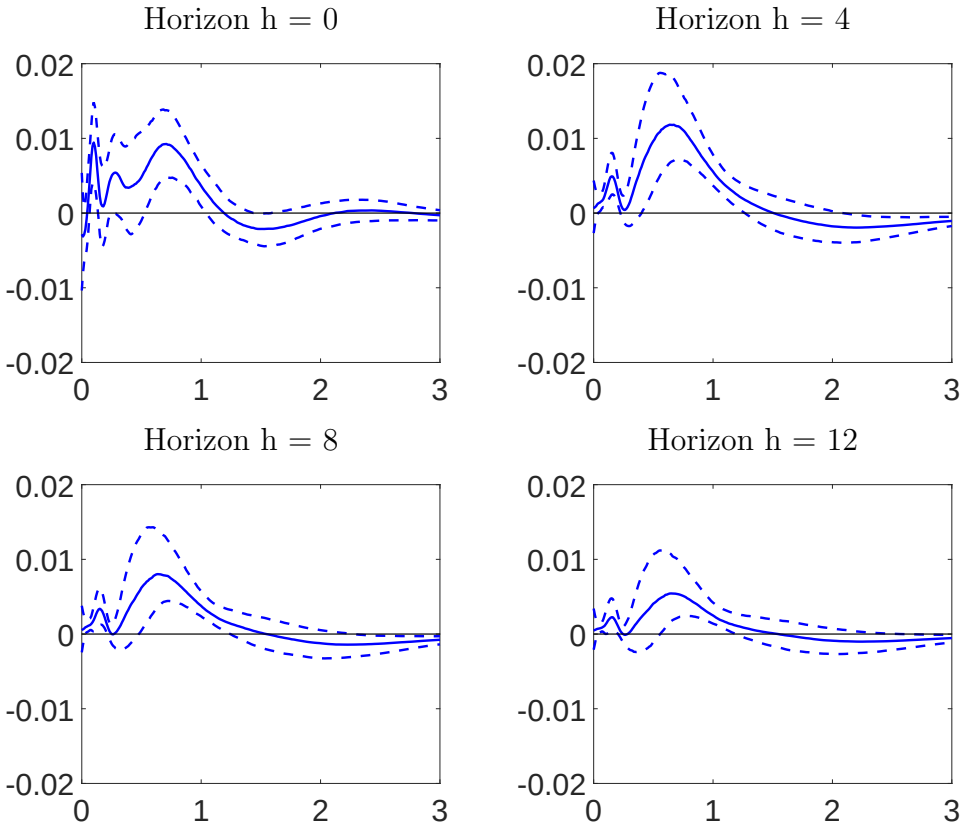


Notes: 66% confidence band of IRFs for three-standard deviation aggregate shocks. Panels depict responses of the log level of TFP and GDP, scaled by 100, and responses of the employment rate in percent. Solid blue responses are based on functional VAR(1) with $K = 10, \hat{\lambda}_1 = 54.6, \hat{\lambda}_2 = 4.9E+08$; dashed red responses are based on aggregate VAR(4) with $\hat{\lambda}_1 = 20.1$.

6.5 Distributional Responses to a TFP shock

Figure 4 plots the the response of the density of transformed earnings to a three-standard deviation TFP innovation. Here the density does not include a pointmass at 0. There are several distinct features. First, the area under the density differential function appears positive, indicating that the employment rate increases in response to the TFP shock. Second, most of the additional probability mass is concentrated between 0.5 and 1.0 across all four horizons shown in the figure. Third, the distributional responses gradually diminish as the horizon h increases. These patterns are consistent with the results reported in Figure 10 of CCS.

Figure 4: Earnings Density Differential (Transformed Data) Response to a TFP Shock



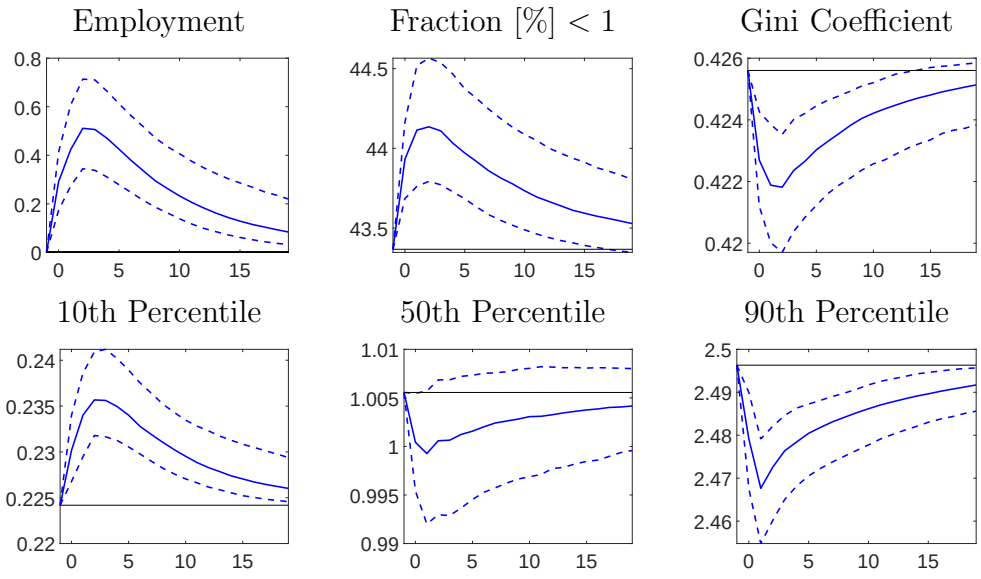
Notes: 66% confidence band of IRFs to a 3-standard-deviations shock to TFP for $K = 10$. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. We depict differences between the shocked and the steady state cross-sectional density at various horizons.

Figures 5 and 6 display the responses of inequality measures to a 3-standard-deviation TFP shock, with and without assigning zero earnings to unemployed individuals, respectively.

My results match Figure 11 in CCS in that we both find TFP shock affects inequality through moving unemployed individuals to the low-income status. Specifically, the 10th percentile rises (posterior median) when the point mass at zero is included but decreases when conditioned only on employed individuals. The responses of the 50th and 90th percentile are not sensitive to the extensive margin.

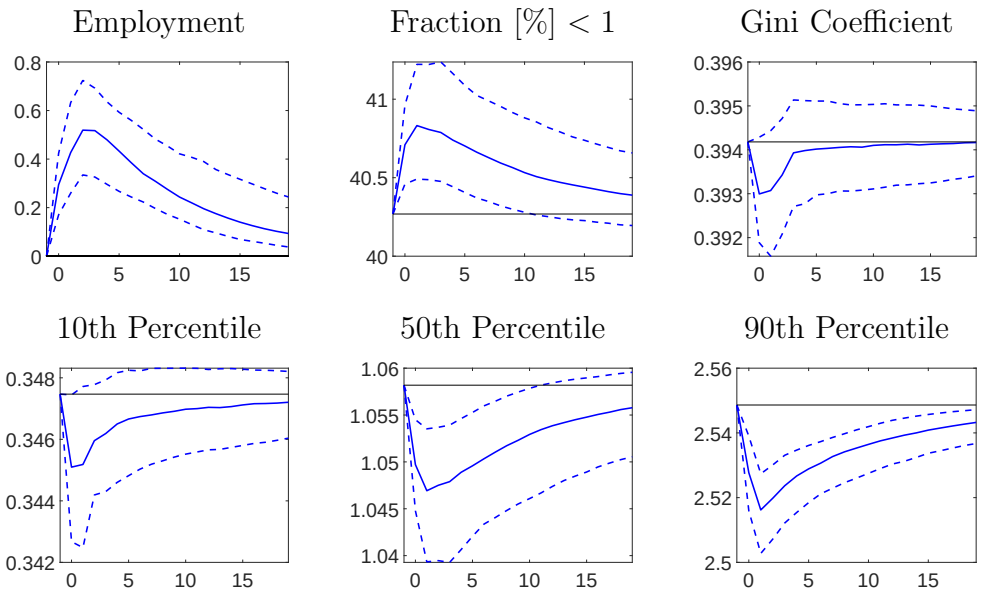
To further support this argument, I compare the IRFs of the Gini coefficient with and without including the zero pointmass. When unemployed individuals are included in the distribution, the 66% confidence band of the Gini coefficient IRF remains below zero for nearly all periods, indicating a consistent negative effect on inequality. However, when conditioned only on employed individuals, the 66% confidence band spans both positive and negative values across all periods. This suggests that the effect of a TFP shock on inequality becomes ambiguous when the extensive margin is not considered.

Figure 5: Responses of Inequality Measures to a TFP Shock: Pointmass at 0



Notes: 66% confidence band of IRFs to a 3-standard-deviations shock to TFP for $K = 10$. Including pointmass at 0. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$.

Figure 6: Responses of Inequality Measures to a TFP Shock: No Pointmass

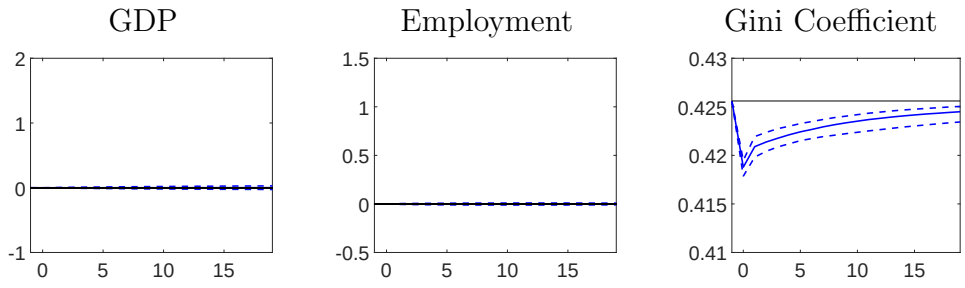


Notes: 66% confidence band of IRFs to a 3-standard-deviations shock to TFP for $K = 10$. No pointmass at 0. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$.

6.6 Aggregate Responses to a Distributional Shock

Figure 7 plots the responses of aggregate variables to a distributional shock. We can see that a pure distributional shock that decreases the Gini coefficient has no effect on aggregate GDP and employment. This conclusion is sharp as the 66% confidence bands are almost focused at zero.

Figure 7: Responses of Aggregate Variables to a Distributional Shock



Notes: 66% confidence band of IRFs to a 1-standard-deviation distributional shock for $K = 10$.

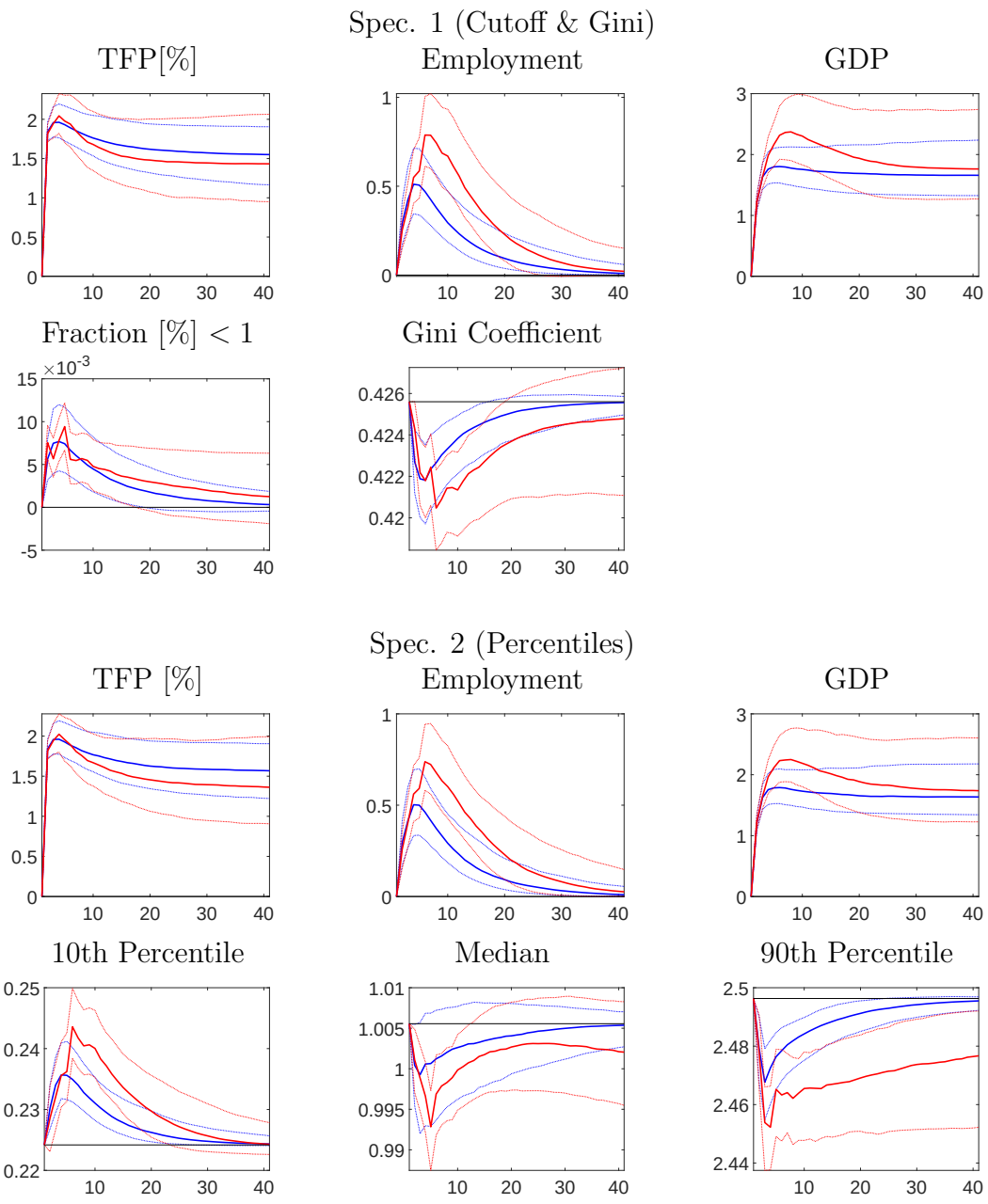
6.7 Functional State Space Model v.s. Inequality VAR

Figure 8 compares the impulse responses to a TFP shock generated by the functional model and VARs with inequality statistics. The replication results closely match Figure 12 in CCS. Additionally, I compute and plot the impulse responses of employment and GDP to provide a more comprehensive analysis. Three distinct features emerge from my results. First, the two inequality VAR specifications produce nearly identical IRFs for aggregate variables: a TFP growth shock causes a permanent increase in both TFP and GDP levels and stimulates employment for a relatively long period. Second, the TFP responses from the functional model and the inequality VARs are aligned. However, at the posterior median, the long-run TFP response in the functional model is slightly lower than that in the inequality VARs. Third, the confidence band generated by the functional approach is much narrower. This is consistent with the formal argument in CCS, which proves that the functional model enables sharper inference by incorporating the full distributional information.

7 Summary

In this project, I replicate the empirical analysis conducted in CCS. My results are generally consistent with their original findings. The functional VAR proves to be a powerful framework for analyzing causal relationships, particularly in contexts where heterogeneity and distributional information are crucial. In the future, I hope to apply this method to other important and interesting empirical settings.

Figure 8: Functional State Space Model v.s. Inequality VAR



Notes: 66% confidence band of IRFs to a 3-standard-deviations shock to TFP. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. 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.