Firm heterogeneity and aggregate fluctuations: A functional VAR model for multidimensional distributions

Massimiliano Marcellino
Bocconi University, Baffi Centre, CEPR

Bank of Italy, 15 September 2025

Introduction

- The study of household and firm **heterogeneity** plays a crucial role in understanding **macroeconomic fluctuations**.
- Heterogeneous household/firm models (e.g. HANK models), often imply that **heterogeneity** offers additional insight for the analysis of aggregate fluctuations.

  But:

  - Semi-structural models of aggregate fluctuations typically neglect micro-level heterogeneity.

- **Functional VAR models** emerged as a new tool for semi-structural analysis of aggregate fluctuations while accounting for micro-level heterogeneity. Chang Chen Schorfheide (2024)
- In practice, FunVARs can be used for modeling **dynamic interactions** between **distributions** and traditional **macroeconomic time series**.

Why functional VARs?

- Analyze the propagation of **aggregate shocks**, while controlling for the distribution of **micro level heterogeneity**.
- Analyze the effect of macroeconomic shocks on the **micro-level distributions**.
- **Validate** the predictions from **heterogeneous agent models** using micro and aggregate macro data.

But also

- **Sharpen identification** of **aggregate shocks**, through additional identifying restrictions.
- **Nowcast distributions** observed at lower frequency exploiting macro time series observed at higher frequency.

Structure of the presentation

1. We start presenting a functional VAR model for a univariate distribution (household income/consumption) that we use to study the distributional effects of uncertainty shocks. This part introduces the theoretical model and the estimation.
2. Next we present a functional VAR model to handle multivariate distributions and we use it to analyze firm heterogeneity and aggregate fluctuations in the US.
3. Finally, if time permits, we briefly discuss nowcasting distributions with a functional MIDAS model.

The Distributional Effects of Economic Uncertainty

Florian Huber   Massimiliano Marcellino   Tommaso Tornese

[1]University of Salzburg

[2]Bocconi University, Baffi Centre, CEPR

[3]Università Cattolica del Sacro Cuore

Motivation

- Uncertainty shocks among drivers of the business cycle (e.g. Castelnuovo, 2019; Fernandez-Villaverde and Guerron-Quintana, 2020);
- **Limited research on the distributional implications;**
- Modeling distributional dynamics using standard methods is problematic:
  - Percentiles in standard VARs percentiles crossing;
  - Moments in standard VARs number of moments is indefinite;

Related literature

- Use of Functional-VAR to model joint dynamics of aggregate variables and income distribution (e.g. Chang M. et al., 2024).
- Econometric methods for functional data:
  - Bayesian (Kowal et al., 2017; Chang M. et al., 2024);
  - Frequentist (Chang et al., 2016; Hu and Park, 2016);
- Recent empirical applications of functional models:
  - Inoue and Rossi (2018): monetary policy as functional schocks;
  - Meeks and Monti (2023): Phillips curve with heterogeneous beliefs;
  - Chang et al. (2022): effects of shocks on heterogeneous inflation expectations;
  - Bjornland et al.(2023): effects of oil shocks on the distribution of stock returns;
  - Chang and Schorfheide (2024): effects of monetary policy on earnings/consumption distribution.

Contribution

- How to treat distributions for Functional Data Analysis (FDA): different transformations have different pro and cons;
- How to summarize the density through Functional-PCA (FPCA, Ramsay and Silverman, 1997): advantages over alternative methods (e.g. splines, Chang et al., 2024);
- What are the effects of uncertainty shocks on income/consumption distribution;
- Robustness of the results to different modeling strategies;
- Estimation of the effects through Functional Local Projections.

Preview of the results

- Show through simulations that:
  - FPCA on $p_t(\cdot)$ provides best approximations, but produces inadmissible distribution responses to shocks (i.e. densities with negative regions);
  - FPCA on $\log(p_t(\cdot))$ ensures non-negativity (not unit-integration) of distributions, but provides worst approximations;
  - FPCA on Log Quantile Density (LQD, Petersen and Muller, 2016) ensures non-negativity and unit-integration of distributions, and provides accurate approximations;
- Propagation of uncertainty shocks in two phases (Carriero et al, 2024):
  - Short run: Unemployment increases; Investments are reduced; Share of employed with low relative income decreases and share of low-consumption households increases $\longmapsto$ Decrease inequality among employees, but higher consumption inequality;
  - Longer horizon: Unemployment is reabsorbed; Labor productivity decreases; Mass of low-income workers increases $\longmapsto$ Increased inequality among employees, but consumption inequality is reabsorbed.

The model

- Assume that income/consumption observations are $\xi_{it} \sim iid\ \ p_t$, $\xi_{it} \in \Xi$;
- Define $f_t(\xi) = g(p_t(\xi)) - \bar{g}$ to be some de-meaned transformation of the distribution (i.e. $\bar{g} = \frac{1}{T}\sum_{t=1}^{T} g(p_t(\xi))$);
- Specify y_t to be a vector of macro/financial aggregate variables.
- The F-VAR(p) is (see e.g. Inoue and Rossi, 2021; Chang et al., 2024):

$$y_t = c_y + \sum_{l=1}^{p} B_{l,yy} y_{t-l} + \sum_{l=1}^{p} \int B_{l,yf}\left(\acute{\xi}\right) f_{t-l}\left(\acute{\xi}\right) d\acute{\xi} + u_{y,t}$$

$$f_t(\xi) = c_f(\xi) + \sum_{l=1}^{p} B_{l,fy}(\xi)\, y_{t-l} + \sum_{l=1}^{p} \int B_{l,ff}\left(\xi, \acute{\xi}\right) f_{t-l}\left(\acute{\xi}\right) d\widetilde{\xi} + u_{f,t}(\xi)$$

Functional VAR (F-VAR)

- Approximate the functions by terminating the Karhunen-Loeve at some truncation point, K:
$$f_t(\xi) \approx \sum_{k=1}^{K} \zeta_k(\xi) * \alpha_{k,t} = \boldsymbol{\zeta}'(\xi)\,\boldsymbol{\alpha}_t; \;\; u_{f,t}(\xi) \approx \sum_{k=1}^{K} \zeta_k(\xi) * \widetilde{u}_{k,t} = \boldsymbol{\zeta}'(\xi)\,\widetilde{\boldsymbol{u}}_t;$$
where $\zeta(\xi)$ is a $K \times 1$ vector of coefficients, and $\boldsymbol{\alpha}_t$ and $\widetilde{\boldsymbol{u}}_t$ are $K \times 1$ random vectors.
- The F-VAR becomes a standard finite-dimensional Factor Augmented VAR:
$$\begin{bmatrix} y_t \\ \alpha_t \end{bmatrix} = \begin{bmatrix} c_y \\ \widetilde{\boldsymbol{c}}_f \end{bmatrix} + \sum_{l=1}^{p} \begin{bmatrix} B_{l,yy} & B_{l,yf}\,C_\alpha \\ B_{l,fy} & B_{l,ff}\,C_\alpha \end{bmatrix} \begin{bmatrix} y_{t-l} \\ \alpha_{t-l} \end{bmatrix} + \begin{bmatrix} u_{y,t} \\ \widetilde{u}_{f,t} \end{bmatrix},$$
where $C_\alpha \equiv \int \boldsymbol{\delta}\left(\acute{\xi}\right) \boldsymbol{\zeta}'\left(\acute{\xi}\right) d\acute{\xi}$, and $u_t = \left[u'_{y,t}, \widetilde{u}'_t\right]'$ has zero mean and variance Ω.
- Inference can now be performed applying conventional techniques;
- It can be given a structural interpretation based on identifying assumptions;

Empirical Approach

- We follow a three-step approach:

1. Estimate the distribution of interest for every t from a sample of draws;
2. Transform the distributions and approximate the resulting function through FPCA;
3. Jointly model the FPCs and a set of random variables with a (Bayesian) VAR.

FDA on Distributions

LQD transfomation

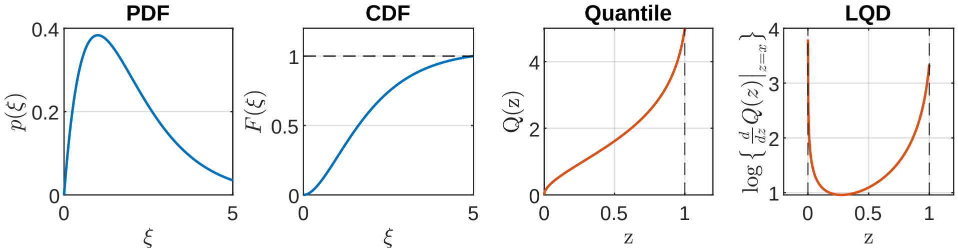
Performing FDA on distributions poses unique challenges:

- Consider the Log Quantile Density (LQD) associated with $p_t(\xi)$ (Petersen and Muller, 2016);

$$g\left(p_t\left(\cdot\right)\right) = \log\left\{\frac{d}{dz}Q_t\left(z\right)\bigg|_{z=x}\right\} = -\log\left\{p_t\left(Q_t\left(x\right)\right)\right\}$$

where $Q(z) = F^{-1}(z)$ is the quantile function (inverse cdf), and $x \in [0,1]$.

LQD transformation



FDA on Distributions

FPCA on the LQD transformation

- Use Tsay's (2016) FPCA approach:
  - Let X denote a $T \times N$ matrix with $(t, i)^{th}$ element $x_{t,i} = f_t(\xi_i) = g(\hat{p}_t(\xi_i)) - \bar{g}(\xi)$, $t = 1, ..., T$, $i = 1, ..., N$.
  - Truncated Singular Value Decomposition (SVD):

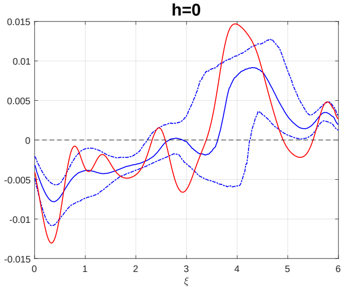
$$X = SVD' + E;$$

  - The principal components D will serve as functional basis $\zeta(\cdot)$;
  - The scores VS_t' (S_t: t-th row of S), will serve as factors $\boldsymbol{\alpha}_t$;

Functional IRFs

- **Functional IRF** let us analyze **changes in the distribution** of interest w.r.t its steady state after a shock occurs.
- Given the finite dimensional representation of the Functional-VAR, the IRFs of $\boldsymbol{\alpha_t}$ at horizon h after a shock $\varepsilon_j = d$, $IRF_\alpha (h, \varepsilon_j = d)$, can be easily computed;
- These need to be mapped back to IRFs for the distribution of interest, $IRF_p (h, \varepsilon_j = d)$, by:
  1. Computing the model-implied steady-state distribution:
  $$p_{ss}(\cdot) = g^{-1}\left(\zeta(\cdot)' \boldsymbol{\alpha}_{ss} + \bar{g}\right)$$
  2. Computing the expected distribution h periods after the shock:
  $$p_{ss+h}(\cdot)\,|_{\varepsilon_j=d} = g^{-1}\left(\zeta(\cdot)' \left(\boldsymbol{\alpha}_{ss} + IRF_\alpha (h, \varepsilon_j = d)\right) + \bar{g}\right)$$
  3. Computing the difference between the two:
  $$IRF_p (h, \varepsilon_j = d) = p_{ss+h}(\cdot)\,|_{\varepsilon_j=d} - p_{ss}(\cdot)$$

Interpretation of F-IRFs



- The figure shows the difference between $p_{ss+h}(\cdot)\,|_{\varepsilon_j = std(\varepsilon_j)}$ and $p_{ss}(\cdot)$.
  - The horizontal axis shows the support Ξ;
  - The vertical axis mesures the difference between the two densities;
  - h is the horizon of the response.
  - In this example, the red is the true FIRF, in blue the estimated one with credible sets.

Simulated Data: Krusell and Smith (1998) DGP

- Simulated data borrowed from Chang M. et al. (2022);
- $T = 160$ artificial observations from the SVAR(1) resulting from the log-linearized solution of the Krusell and Smith (1998) model. Observe:
  - Productivity level, the capital stock, the employment level ($n_v = 3$);
  - Centered moments of the distribution of assets among the employed;
- A sample of $N = 9230$ is drawn from the asset distribution;
- The impact matrix implied by the model is lower triangular, the first structural shock is a productivity shock.

Krusell and Smith (1998) DGP: F-IRFs

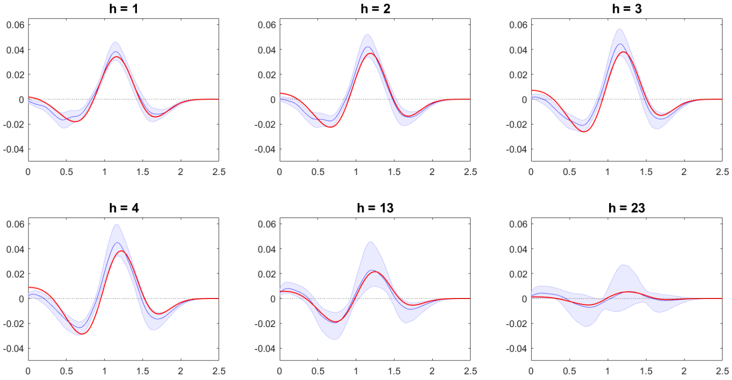


Figure: Red lines: true responses of $p_t(\cdot)$ to one standard deviation technology shocks. Solid blue lines: posterior median responses, dashed blue lines: 90% credible bands. The number of FPC is selected as the smallest one for which 90% of variance is explained. (Timing convetion is different than the one in Chang et al.(2024): here shock happens at $t = 0$, there at $t = 2$)

Uncertainty shocks: the model

- Augment the VAR model analyzed by Jurado et al. (JLN, 2015) by including alternately income/consumption distributions as $f_t(\cdot)$;
- Earnings-to-GDP data constructed by Chang M et al. (2022) based on the Current Population Survey (CPS);
- Consumption data from thr Consumption Expenditure Survey;
- Convert the monthly SVAR of JLN in a quarterly F-SVAR model and focus on the period 1989:Q1 - 2017:Q3 for income, and 1990:Q2 - 2016:Q4 for consumption;
- The $n_v = 11$ endogenous variables included in the model are: (i) real GDP, (ii) real PCE, (iii) GDP deflator, (iv) real wages, (v) real investments, (vi) labor productivity, (vii) unemployment rate, (viii) Federal Funds Rate, (ix) S&P500 index, (x) M2 growth rate, and (xi) JLN's macro-uncertainty measure;
- Assume $K = 7$ (different Ks do not affect results for $K > 2$);
- The macro-uncertainty shock is identified by ordering the uncertainty measure last among the aggregate variables in a Cholesky identification scheme.

Uncertainty shocks: IRFs

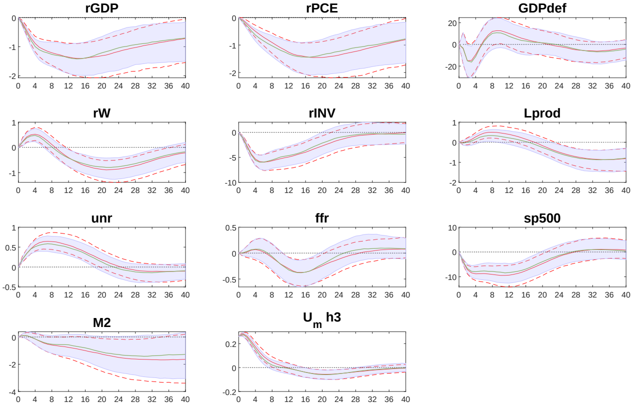
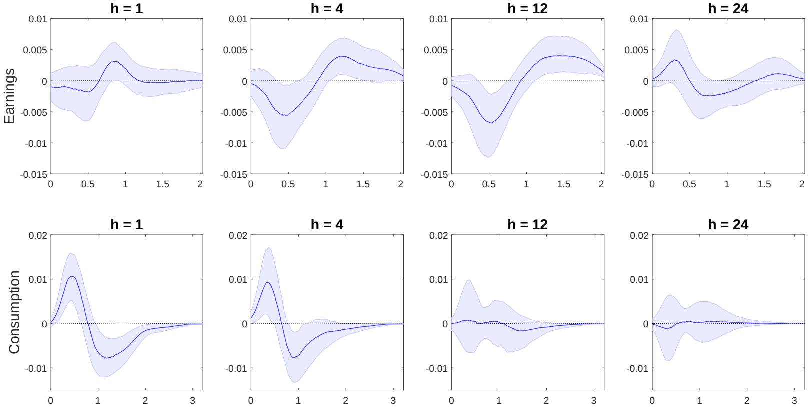


Figure: IRFs to an uncertainty shock implied by: a SVAR (red), F-SVAR (blue), 68% credible bands (dashed lines and shaded areas).

Uncertainty shocks: F-IRFs

Response of distributions to an uncertainty shock:



- The horizontal axis measures the earnings/consumption relative to the per-capita level;
- h denotes the horizon in quarters.

Uncertainty shocks: Gini IRF

The F-IRFs of income distribution can be mapped to the IRF of the Gini coefficient:

- Earnings inequality decreases in the short run, but it increases at longer horizons. The degree of consumption inequality increases significantly in the short run.

Uncertainty shocks: comments

- Aggregate effects:
  - IRFs predicted by the F-SVAR are similar to those generated by the standard SVAR.
- Distributional effects:
  - Propagation in two phases:
    - ⋆ In the short run (up to 3 years): while unemployment increases, the share of workers with low relative income decreases, and the mass of people employed receiving income above the average increases;
    - ⋆ In the longer run: while unemployment is reabsorbed, the share of occupied with low relative income increases to the detriment of the middle-income class (probably due to the decrease of labor productivity triggered by the decrease in investments experienced at short horizons).

Further developments: Functional LP

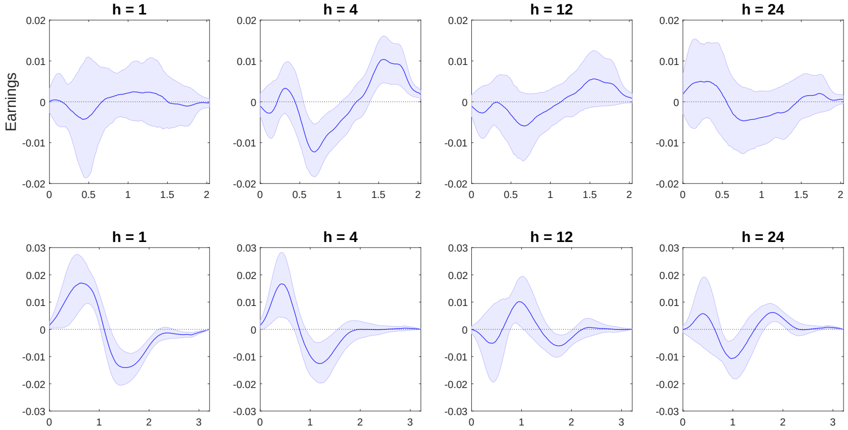
- F-IRFs can be also estimated by Local Projections;
  1. Estimate responses of α_t: $IR_\alpha(t,h,d_i) = E[\alpha_{t+h} \mid \varepsilon_t = d_i, \Im_t] - E[\alpha_{t+h} \mid \varepsilon_t = 0, \Im_t]$;
  2. Compute the Functional IRFs through the mapping: $IR_f(\xi,t,h,d_i) = \zeta'(\xi) \times IR_\alpha(t,h,d_i)$;

  Suppose $[y_{1t}, y_{2t}, \ldots, y_{n_v-1,t}]$ is predetermined w.r.t. $[U_t, \alpha_t]'$.
  The joint response of α to an impulse in U can be estimated through the multivariate regression:

$$\alpha_{t+h} = a^h + B_1^h y_{t\setminus U} + \beta_1^h U_t + \sum_{l=1}^{p} B_{l+1}^h \left[\ y'_{t-l}, \quad \alpha'_{t-l} \ \right]' + e_{h,t}$$

  where $IR_\alpha(t,h,d_i = [1,0,\ldots,0]') = \beta_1^h$.
- Can be estimated by OLS with (system-wide) HAR standard errors;
- Results obtained by Functional LP are equivalent.

Uncertainty Shocks: F-IRFs



Firm heterogeneity and aggregate fluctuations: a functional VAR model for multidimensional distributions

Massimiliano Marcellino   Andrea Renzetti   Tommaso Tornese

[1]Bocconi University, Baffi Centre, CEPR

[2]Bocconi University

[3]Università Cattolica del Sacro Cuore

This paper

- Introduce a **Functional VAR** model for modelling **firm heterogeneity** arising in **more than one dimension**.

- Approximate the multivariate distribution of firm's level characteristics using dimension reduction techniques for **tensor data objects**:
  - **Principal component analysis on unfolded tensors**.
  - **Multilinear principal component analysis**.
  - **CP decomposition**.

- Combining both **firm-level data** and **aggregate macroeconomic data**:
  - Analyze the micro and macro propagation of **TFP shocks** and **monetary policy shocks** in the US economy.
  - **Identification** of **cross-sectional uncertainty shocks** and estimation of their effects on the US economy.

Related literature

- Semi structural multivariate time series models which leverage both micro and macro data for analyzing aggregate fluctuations.

  Chang Chen Schorfheide (2024); Chang Schorfheide (2024); Huber Marcellino Tornese (2024); Huang (2024)

  $\rightarrow$ This paper: Modelling **heterogeneity** arising in **more than one dimension**.

- Aggregate fluctuations and **firm heterogeneity**.
  Lenza Savoia (2024)

Why modeling the joint distribution?

Functional IRF of the marginal distributions of x_1 and x_2 after the shock

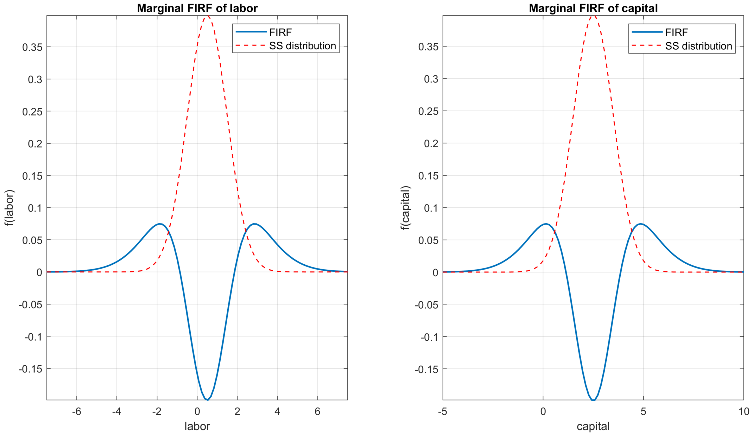


Figure: The figure shows the functional IRFs of the marginal labor and capital distributions after a shock. In blue we report the change in the probability mass with respect to the steady state distribution, reported in red.

Why modeling the joint distribution?

FIRF of the **joint distribution** of (x_1, x_2) after the shock

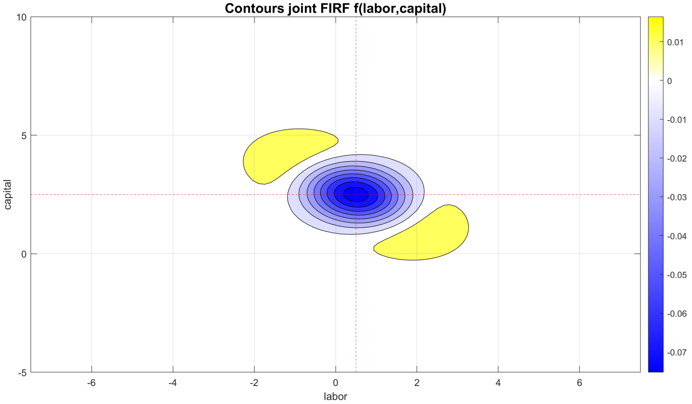


Figure: The figure reports the contours from a bivariate functional IRFs, compatible with the changes in the marginal distributions. In blue we report negative changes in the probability mass of the joint distribution w.r.t the steady state distribution, while in yellow we report positive changes.

Why modeling the joint distribution?

FIRF of the **joint distribution** of (x_1, x_2) after the shock

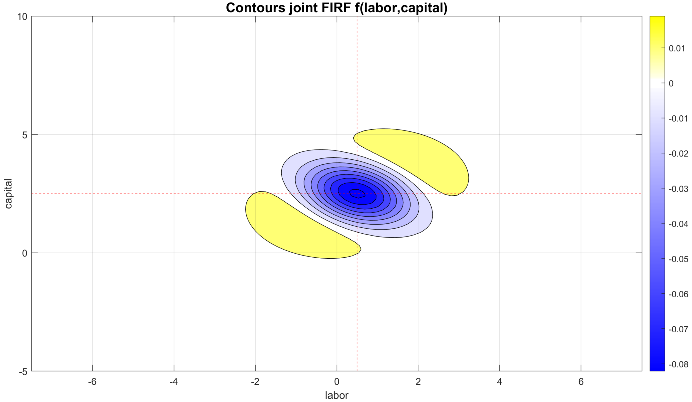


Figure: The figure reports the contours from another bivariate functional IRFs, equally compatible with the changes in the marginal distributions. In blue we report negative changes in the probability mass of the joint distribution w.r.t the steady state distribution, while in yellow we report positive changes.

Why modeling the joint distribution? / 2

- Identifying macroeconomic shocks—especially when these shocks are naturally defined as exogenous shifts in specific features of a joint distribution.
- For example, **cross-sectional uncertainty shocks** can be interpreted as an exogenous shift in the cross-sectional dispersion of firm-level productivity.
- In terms of the observables, these shocks manifest as exogenous changes in the cross-sectional dispersion of firms' output, conditional on their capital and labor inputs. For example considering:

$$output_{it} = f(k_{it}, l_{it}) + \sigma_t \varepsilon_{it} \tag{1}$$

  CSU shocks manifest as $\uparrow \sigma_t$.

FunVAR

Modeling the dynamic interaction between micro level heterogeneity and the aggregates

We consider the following Functional VAR model

$$\boldsymbol{y_t} = \boldsymbol{c_y} + \sum_{s=1}^{p} \boldsymbol{B}_{l,yy}\boldsymbol{y_{t-s}} + \sum_{s=1}^{p} \int \boldsymbol{B_{s,yl}}(\boldsymbol{x}) l_{t-s}(\boldsymbol{x}) d\boldsymbol{x} + \boldsymbol{u_{y,t}} \tag{2}$$

$$l_t(\boldsymbol{x}) = c_l(\boldsymbol{x}) + \sum_{s=1}^{p} \boldsymbol{B}_{s,ly}(\boldsymbol{x})\boldsymbol{y}_{t-s} + \sum_{s=1}^{P} \int B_{ll}(\boldsymbol{x}, \boldsymbol{x}') l_{t-s}(\boldsymbol{x}') d\boldsymbol{x}' + u_{l,t}(\boldsymbol{x}) \tag{3}$$

where $\boldsymbol{y}_t$ are the macroeconomic aggregates and $\boldsymbol{x}$ are the firm-level characteristics (it is a **vector**), $l(.)$ is the Centered Log Ratio transformation of the **multivariate** distribution function. For example for $\boldsymbol{x} = [x_1, x_2]$ with x_1 capital and x_2 labor we have:

$$l_t(\boldsymbol{x}) := \text{CLR}(f(x_1, x_2)) = \log(f(x_1, x_2)) - \frac{1}{|\Omega|} \int_{\Omega} \log(f(x_1, x_2))\, dx_1 dx_2 \tag{4}$$

see (Petersen et. al (2021)) for details on this transformation.

Challenges of approximating multidimensional distributions

- **Curse of Dimensionality:** Multidimensionality of $\boldsymbol{x}$ requires a much larger number of basis functions compared to the unidimensional case. (e.g. splines)
- **Increased Computational Costs**.

We resort to:

- PCA on the vectorized tensor.
- Multilinear principal component analysis.
- CP decomposition.

.

Approximation of the multidimensional distribution

- Let's consider $\boldsymbol{x} = [x_1, x_2]$ with x_1 capital and x_2 labor. We assume that we observe:

$$l_t(\boldsymbol{x})^{\text{obs}} = l_t(\boldsymbol{x}) + \varepsilon_t \tag{5}$$

on a bi-dimensional grid, that is $\{l_t(x_{1,i}, x_{2,j})^{\text{obs}} : i = 1, \dots, N_1,\ j = 1, \dots, N_2\}$

where

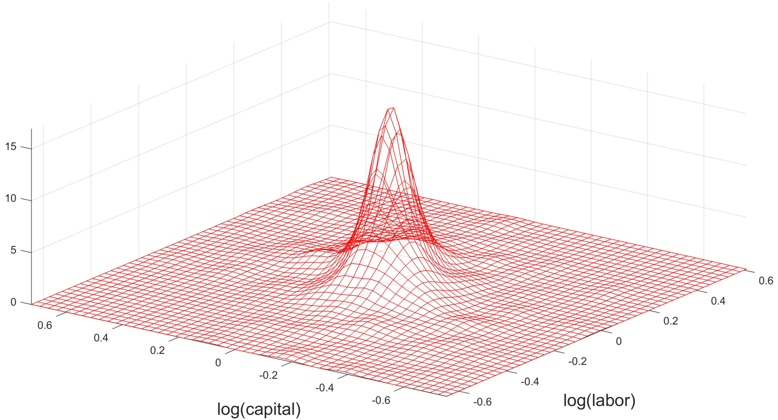
$$l(\boldsymbol{x})^{\text{obs}} = CLR(\hat{f}(\boldsymbol{x}))$$

and

$$\hat{f}(x_1, x_2) = \frac{1}{N_1 N_2 h_1 h_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} K\left(\frac{x_1 - x_{1,i}}{h_1}, \frac{x_2 - x_{2,j}}{h_2}\right)$$

is a multivariate kernel probability density estimate.

Firms' joint capital and labor distribution



Approximation of the multidimensional distribution

We assume that the true CLR transformed distribution can be expanded with a finite number of basis functions, that is:

$$l_t(x_1, x_2) = \sum_{j=1}^{K} \beta_{j,t} h_j(x_1, x_2) \tag{6}$$

This assumption lets us write the **functional VAR** as a **factor augmented VAR**.

$$\begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{\beta}_t \end{bmatrix} = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \begin{bmatrix} \boldsymbol{y}_{t-1} \\ \boldsymbol{\beta}_{t-1} \end{bmatrix} + \ldots + \boldsymbol{\Phi}_p \begin{bmatrix} \boldsymbol{y}_{t-p} \\ \boldsymbol{\beta}_{t-p} \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_{y,t} \\ \tilde{\boldsymbol{u}}_{q,t} \end{bmatrix} \tag{7}$$

$$\underbrace{vec(\boldsymbol{L}_t)}_{N_1 N_2 \times 1} = \underbrace{\boldsymbol{H}}_{N_1 N_2 \times K} \underbrace{\beta_t}_{K \times 1} + \underbrace{\varepsilon_t}_{N_1 N_2 \times 1} \tag{8}$$

Estimation

We estimate the Factor augmented VAR in two steps:

- Step 1: Estimate the basis functions, i.e the loadings $\boldsymbol{H}$.
  - PCA on vectorized tensor.
  - Multilinear principal component analysis.
  - CP decomposition.

- Step 2: Estimate the factor augmented VAR model.

$$\begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{\beta}_t \end{bmatrix} = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \begin{bmatrix} \boldsymbol{y}_{t-1} \\ \boldsymbol{\beta}_{t-1} \end{bmatrix} + \ldots + \boldsymbol{\Phi}_p \begin{bmatrix} \boldsymbol{y}_{t-p} \\ \boldsymbol{\beta}_{t-p} \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_{y,t} \\ \tilde{\boldsymbol{u}}_{q,t} \end{bmatrix} \tag{9}$$

$\boldsymbol{\beta}_{1:T}$ can be either estimated in step 1 and treated as observed in step 2 or treated as latent factors and jointly estimated in step 2 (e.g. Doz Giannone Riechlin (2011)).

Unfolding and principal component analysis

- Unfolding, consists in the vectorization of $\boldsymbol{L}_t$.
- Principal component analysis is performed on the unfolded data and seeks to minimize

$$\min_{\boldsymbol{H},\boldsymbol{\beta}_t} \frac{1}{T} \sum_{t=1}^{T} \left\| vec(\boldsymbol{L}_t) - \boldsymbol{H}\boldsymbol{\beta}_t \right\|^2 \tag{10}$$

- Pros: Simple, need to determine the number of principal components K.
- Cons: In small sample can be difficult to precisely estimate all the $N_1 N_2 K$ parameters in $\boldsymbol{H}$.

Multilinear principal component analysis

- Multilinear principal component analysis assumes:

$$l_t(x_1, x_2) = \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} \beta_{ij,t} h_i(x_1) h_j(x_2) = (\boldsymbol{h}(x_1) \otimes \boldsymbol{h}(x_2))' \boldsymbol{\beta}_t \tag{11}$$

- In terms of the observable we have:

$$\boldsymbol{L}_t = \boldsymbol{H}_2 \boldsymbol{B_t} \boldsymbol{H}_1' + \boldsymbol{E_t} \tag{12}$$

and seek to find $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ that minimize (Hung et. al (2012))

$$\min_{\boldsymbol{H}_1, \boldsymbol{H}_2} \frac{1}{T} \sum_{t=1}^{T} \left\| \boldsymbol{L}_t - \boldsymbol{H}_2 \boldsymbol{B}_t \boldsymbol{H}_1' \right\|^2 \tag{13}$$

- Pros: Lower number of parameters to estimate.
- Cons: Conditional on a choice of K_1 and K_2. Can be restrictive.

CP decomposition

- CP analysis assumes:

$$I(x_1, x_2) = \sum_{k=1}^{K} \beta_{t,k} h_k^{(1)}(x_1) h_k^{(2)}(x_2) \ . \tag{14}$$

- We estimate the loadings by Alternating Least Squares (ALS) (Carroll and Chang 1970; Harshman 1970). An alternative is iterated PCA Babii, Ghysels, and Pan (2024).

- Pros: Lower number of parameters to estimate w.r.t PCA. Only K to determine w.r.t multilinear PCA.
- Cons: Less flexible when different modes of the data exhibit varying complexity or correlation structures

Factors augmented VAR model

- We rely on **cross validation** for choosing between PCA on the vectorized tensor, multilinear principal component analysis and CP decomposition.
- We perform Bayesian inference on parameters of the factor augmented VAR model

$$\begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{\beta}_t \end{bmatrix} = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \begin{bmatrix} \boldsymbol{y}_{t-1} \\ \boldsymbol{\beta}_{t-1} \end{bmatrix} + \ldots + \boldsymbol{\Phi}_p \begin{bmatrix} \boldsymbol{y}_{t-p} \\ \boldsymbol{\beta}_{t-p} \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_{y,t} \\ \tilde{\boldsymbol{u}}_{q,t} \end{bmatrix} \qquad \begin{bmatrix} \boldsymbol{u}_{y,t} \\ \tilde{\boldsymbol{u}}_{q,t} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right) \quad (15)$$

- As the factor augmented VAR can become high dimensional, we leverage Bayesian shrinkage exploiting the asymmetric conjugate prior by Chan (2022).
- Semi-structural analysis regards $\boldsymbol{u}_t$ as as a linear combination of the structural shocks ε_t.

Simulation from Winberry (2018) heterogeneous firms model

Approximation by principal component on the vectorized tensor

Bivariate ($log(k), log(l)$) distribution from Winberry (2018) RBC heterogeneous firm model

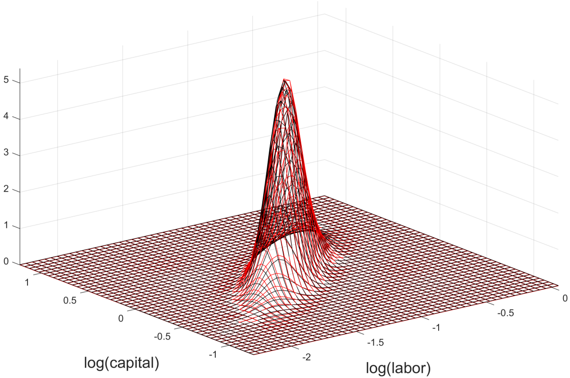


Figure: The figure shows the true (red) and the approximated (black) bivariate log-capital and log-labor distribution from for one sample period in the simulation.

Evidences from simulation: Winberry (2018) RBC heterogeneous firm model

IRFs of aggregate to TFP shock

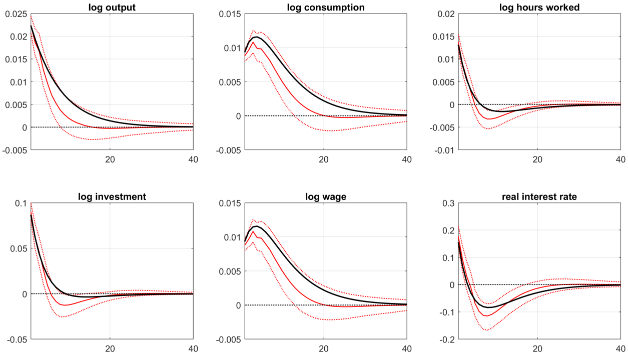


Figure: The figure shows the Impulse Response Functions (IRFs) of the macroeconomic aggregates to a one standard deviation TFP shock. In black we report the IRF of the Winberry (2018) heterogeneous firm model. In red bold line we report the posterior mean estimate from the FunVAR while in dashed red line we report the 5^{th} and 95^{th} credible sets.

Evidences from simulation: Winberry (2018) RBC heterogeneous firm model

Functional IRF of the bivariate (capital, labor) distribution $h = 4$

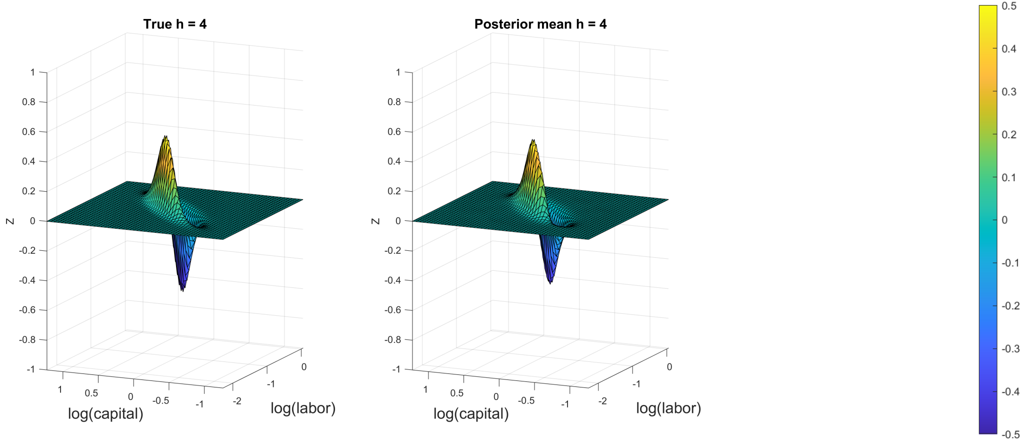


Figure: The figure shows the bivariate FIRF, i.e. the change in the probability mass w.r.t. the steady state following the TFP shock. On the left hand side the true FIRF, on the right hand side the posterior mean estimate from the FunVAR.

Evidences from simulation: Winberry (2018) RBC heterogeneous firm model

Functional IRF of the bivariate (capital, labor) distribution $h = 8$

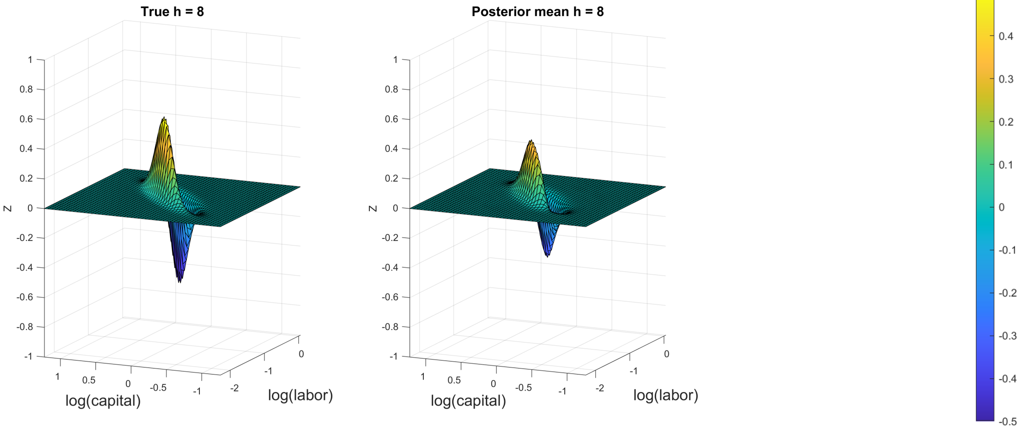


Figure: The figure shows the bivariate FIRF, i.e. the change in the probability mass w.r.t. the steady state following the TFP shock. On the left hand side the true FIRF, on the right hand side the posterior mean estimate from the FunVAR.

Evidences from simulation: Winberry (2018) RBC heterogeneous firm model

Functional IRF of the bivariate (capital, labor) distribution $h = 24$

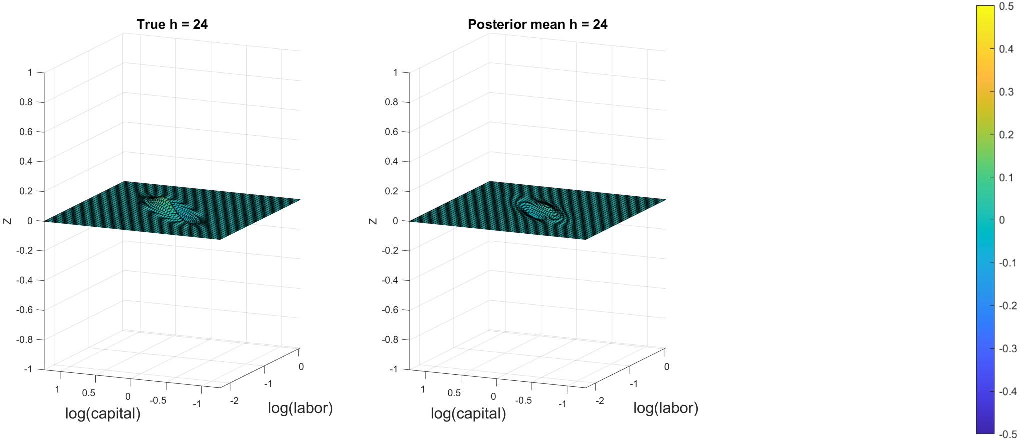


Figure: The figure shows the bivariate FIRF, i.e. the change in the probability mass w.r.t. the steady state following the TFP shock. On the left hand side the true FIRF, on the right hand side the posterior mean estimate from the FunVAR.

Evidences from simulation: Winberry (2018) RBC heterogeneous firm model

Functional IRF of the bivariate (capital, labor) distribution $h = 4$

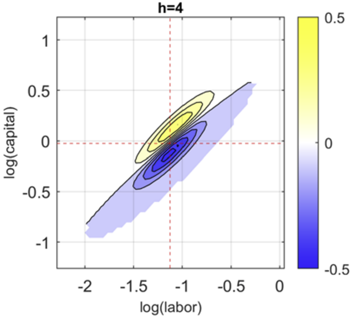


Figure: The figure shows the contours from the true bivariate FIRFs from the heterogeneous firm model. Yellow (blue) for positive (negative) changes in probability mass w.r.t the steady state distribution. Dashed red lines are the steady state mean values in the heterogeneous firm model.

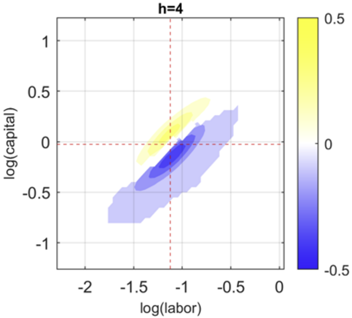


Figure: The figure shows the posterior mean estimate from the FunVAR only if the $15^{th} - 85^{th}$ credible set does not contain zero.Yellow (blue) for positive (negative) changes in probability mass w.r.t the steady state distribution. Dashed red lines are the steady state mean values.

Evidences from simulation: Winberry (2018) RBC heterogeneous firm model

Functional IRF of the bivariate (capital, labor) distribution $h = 8$

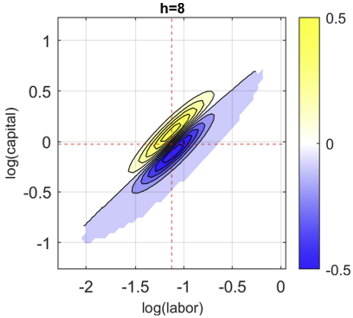


Figure: The figure shows the contours from the true bivariate FIRFs from the heterogeneous firm model. Yellow (blue) for positive (negative) changes in probability mass w.r.t the steady state distribution. Dashed red lines are the steady state mean values in the heterogeneous firm model.

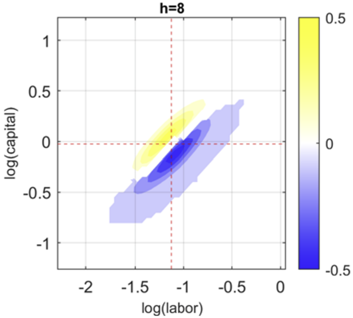


Figure: The figure shows the posterior mean estimate from the FunVAR only if the $15^{th} - 85^{th}$ credible set does not contain zero.Yellow (blue) for positive (negative) changes in probability mass w.r.t the steady state distribution. Dashed red lines are the steady state mean values .

Evidences from simulation: Winberry (2018) RBC heterogeneous firm model

Functional IRF of the bivariate (capital, labor) distribution $h = 24$

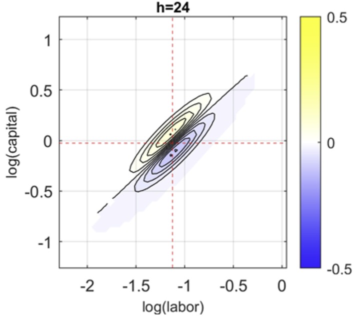


Figure: The figure shows the contours from the true bivariate FIRFs from the heterogeneous firm model. Yellow (blue) for positive (negative) changes in probability mass w.r.t the steady state distribution. Dashed red lines are the steady state mean values in the heterogeneous firm model.

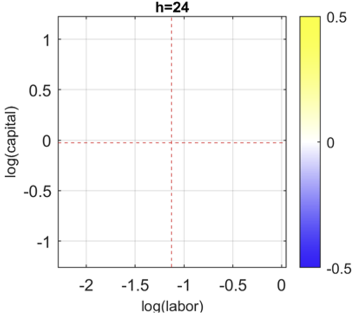


Figure: The figure shows the posterior mean estimate from the FunVAR only if the $15^{th} - 85^{th}$ credible set does not contain zero.Yellow (blue) for positive (negative) changes in probability mass w.r.t the steady state distribution. Dashed red lines are the steady state mean values.

Empirical analysis and pre-view of the results

- We use the model for studying the effects of TFP shocks and monetary policy shocks on the **joint capital and labor distribution** of US firms.
- We find that **TFP shocks** have a persistent effect on the joint labor-capital distribution, with the effects peaking between one and two years after the shock. Results on TFP shocks
- We find that **monetary policy shocks** have a less persistent contractionary effect on the joint labor-capital distribution. Results on MP shocks
- We use the model for identifying **cross-sectional uncertainty shocks** and estimate their effects on the US macroeconomy.
- We find **cross-sectional uncertainty shocks** to be **contractionary**.

Cross-sectional uncertainty shocks

Cross sectional uncertainty shocks

- A significant number of structural models highlight the importance of cross-sectional uncertainty as a key driver of macroeconomic fluctuations. (Bloom 2009; Christiano et al. 2014; Bloom et al. 2018; Arellano et al. 2019)
- In terms of the observables, these shocks manifest as exogenous changes in the cross-sectional dispersion of firms' output, conditional on their capital and labor inputs.
- We assume that the relationship between the FVAR residuals and the cross-sectional uncertainty shocks ε_t^{CSU}, is as follows:

$$\boldsymbol{u}_t = \boldsymbol{\Sigma_{tr}} \boldsymbol{q} \varepsilon_t^{CSU} \tag{16}$$

- We identitify CSU shocks by finding the orthogonal rotation $\boldsymbol{q}$ that maximizes the change in the variance of the output distribution conditional on capital and labor endowment on impact.

Data and estimation sample

Data:

- **Aggregate data:** real GDP, consumption, investment, employment and real wages.
- **Firm level data:** Compustat data on employment (l), capital (k), and sales (y).

Estimation sample: 1984Q4 - 2019Q4

Cross sectional uncertainty shocks

Impulse response function of the aggregate variables to a CSU shock

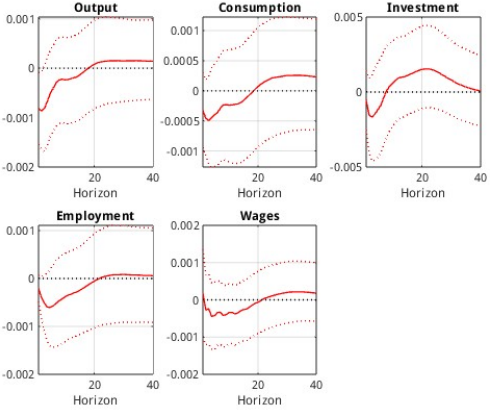


Figure: The figure shows the impulse response function to a cross sectional uncertainty shock.

Nowcasting distributions: a functional MIDAS model

Massimiliano Marcellino   Andrea Renzetti   Tommaso Tornese

[1]Bocconi University, Baffi Centre, CEPR

[2]Bocconi University

[3]Università Cattolica del Sacro Cuore

Introduction

- **Micro-level distributions** of household and firm characteristics originating from surveys are typically **released infrequently**, creating a **lag** in economic insights.
- Monitoring and anticipating real-time changes in distributions necessitates seeking timely signals from variables and indicators available before the official data release.
- However, very limited research has been devoted to **nowcasting and forecasting distributions** using higher frequency indicators.

This paper

- Introduce a **functional MIDAS model** to leverage high-frequency information for forecasting and nowcasting distributions observed at a lower frequency.
- Approximate the functional MIDAS model with a finite-dimensional SUR-MIDAS regression model.
- To identify the relevant predictors, we develop a **Group Lasso with spike-and-slab** prior for the **SUR MIDAS regression model**.
- We use the model to nowcast the **yearly** U.S. households' income distribution from the March CPS-ASEC leveraging **quarterly** macroeconomic and financial indicators.
  - Exploiting high frequency macroeconomic and financial indicators enhances forecast accuracy for the entire target distribution and for key features of the distribution that signal changes in **inequality**.

Nowcasting distributions in a mixed frequency framework

- We aim to forecast/nowcasts a distribution of interest (e.g. the household income distribution) observed at a lower frequency $f_t(x)$.
- We aim to exploit a potentially high dimensional vector of regressors observed at a higher frequency (e.g macro/financial indicators) $\boldsymbol{x}_t^{(m)}$.
- We construct a model that links $f_t(x)$ to its own lags $L(l)f_t(x)$ and the lags of the high frequency indicators $L(l/m)\boldsymbol{x}_{\boldsymbol{t}}^{(m)}$.

The model

Functional MIDAS regression model

We consider the following functional MIDAS regression model:

$$q_{t+h}(\tau) = c_q(\tau) + \underbrace{\sum_{l=1}^{p_q} \int_0^1 B_{qq,l}(\tau, \tau') L(l) q_t(\tau') d\tau'}_{\text{lags of } q_t(\tau)} + \underbrace{\sum_{l=0}^{p_x - 1} \boldsymbol{B}_{qx,l}(\tau) L(l/m) \boldsymbol{x}_t^{(m)}}_{\text{lags of } \boldsymbol{x}_t^{(m)}} + u_{q,t+h}(\tau), \tag{17}$$

where $\boldsymbol{x}_t^{(m)}$ is the $n_x \times 1$ vector of high-frequency macroeconomic indicators, $\boldsymbol{q}_{t+h}(\tau)$ is the LQD transformation of the household income probability density function, $L(.)$ is the lag operator, such that $L(1/m)\boldsymbol{x}_t = \boldsymbol{x}_{t-1/m}^{(m)}$ and $L(1) q_t(\tau) = q_{t-1}(\tau)$.

Nowcasting exercise

- We consider the household nominal income distribution from the March Current Population Survey Annual Social and Economic Supplements **(CPS-ASEC)** given by the variable `hhincome`.
- We de-trend the household-level nominal income by dividing by 2/3 of nominal GDP per-capita and then compute the inverse hyperbolic sine function transformation as in Chang Schorfheide (2024).
- The sample period is 1968-2023 and the evaluation sample is 1998-2023.

Variable	Mnemonic	Transformation
Real Gross Domestic Product	GDPC1	%Δ
Real Government Receipts	FGRECPTx	%Δ
Federal Government: Current Expenditures	FGEXPND	%Δ
Real Personal Consumption Expenditures	PCECC96	%Δ
Real Disposable Personal Income	DPIC96	%Δ
Real Exports of Goods	EXPGSC1	%Δ
Real Imports of Goods & Services	IMPGSC1	%Δ
All Employees: Total nonfarm	PAYEMS	%Δ
Civilian Labor Force Participation Rate	CIVPART	%Δ
Civilian Unemployment Rate	UNRATE	%Δ
Number of Civilians Unemployed for 27 Weeks and Over	UEMP27OV	%Δ
Average Weekly Hours Of Production And Nonsupervisory Employees	AWHNONAG	%Δ
Average (Mean) Duration of Unemployment	UEMPMEAN	Δ
Housing Starts	HOUST	%Δ
Real Average Hourly Earnings of Production and Nonsupervisory Employees: Total	AHETPIx	%Δ
Real Average Hourly Earnings of Production and Nonsupervisory Employees Construction	CES2000000008x	%Δ
Real Average Hourly Earnings of Production and Nonsupervisory Employees Manufacturing	CES3000000008x	%Δ
Consumer Price Index all items	CPIAUCSL	%Δ
S&P 500	SP_500	%Δ
Effective Federal Funds Rate	FEDFUNDS	Δ
10-Year Treasury Constant Maturity Rate (Percent)	GS10	Δ
Cons. Expectations	UMCSENTx	level

Nowcasting the household income distribution in the US

Forecast accuracy from 1998 to 2023

$h = 0$	Avg KL	Avg HD	Avg QS20	Avg QS50	Avg QS80
RIDGE-SUR-MIDAS	0.44	0.85	1.06	0.92	0.96
BLSS-SUR MIDAS	0.45	0.82	0.76	0.78	0.86

Table: The metrics are expressed in ratio with respect to the VAR approximation of the Functional VAR model with Minnesota type of prior that does not exploit high frequency macroeconomic indicators.

$h = \frac{1}{4}$ $h = \frac{2}{4}$ $h = \frac{3}{4}$

Nowcasting the household income distribution in the US

Forecast accuracy from 1998 to 2023

$h = 0$	RMSE Mean	RMSE Variance	RMSE Skewness	RMSE Kurtosis	RMSE IR	RMSE GINI	RMSE CV
RIDGE-SUR-MIDAS	0.53	0.90	0.49	0.47	0.83	0.33	0.52
BLSS-SUR MIDAS	0.50	0.86	0.55	0.50	0.93	0.38	0.66

Table: The metrics are expressed in ratio with respect to the VAR approximation of the Functional VAR model with Minnesota type of prior that does not exploit high frequency macroeconomic indicators.

$h = \frac{1}{4}$ $h = \frac{2}{4}$ $h = \frac{3}{4}$

Thank you for your attention.

Appendix

Appendix: Firm heterogeneity and aggregate fluctuations: a functional VAR model for multidimensional distributions

Massimiliano Marcellino Andrea Renzetti Tommaso Tornese

Distributional effects of TFP shocks in the US

TFP shocks: the model

Data:

- **Aggregate data:** Fernald (2014) TFP Series, real GDP, real consumption expenditure, real fixed investment, real interest rate, real compensation, employment.
- **Firm level data:** Compustat data on employment (`emp`) and capital (`ppent`).

Identification strategy:

- **Internal instrument** using Fernald (2014) TFP Series.

Estimation sample: 1984Q4 - 2019Q4

Back

TFP shocks

Effect on the aggregate variables

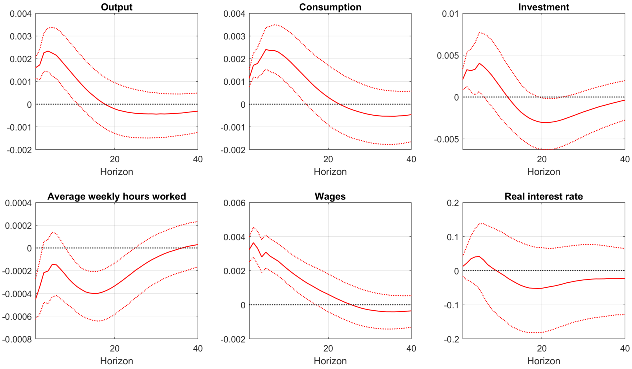


Figure: The figure shows the Impulse Response Functions (IRFs) of the macroeconomic aggregates to a one standard deviation TFP shock. In red bold line we report the posterior mean estimate from the FunVAR while in dashed red line we report the 15^{th} and 85^{th} credible sets. Back

TFP shocks

Effect on the moments of the firm-level joint capital labor distribution

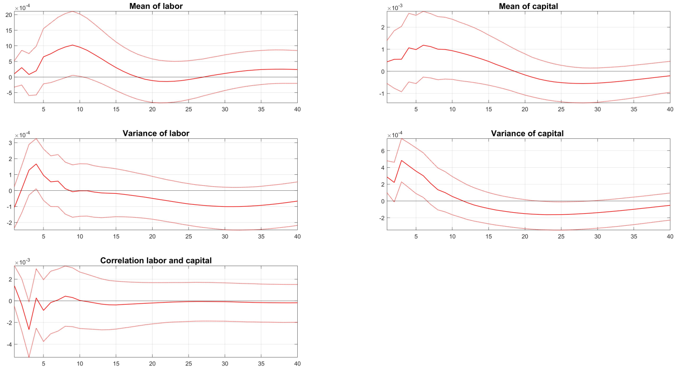
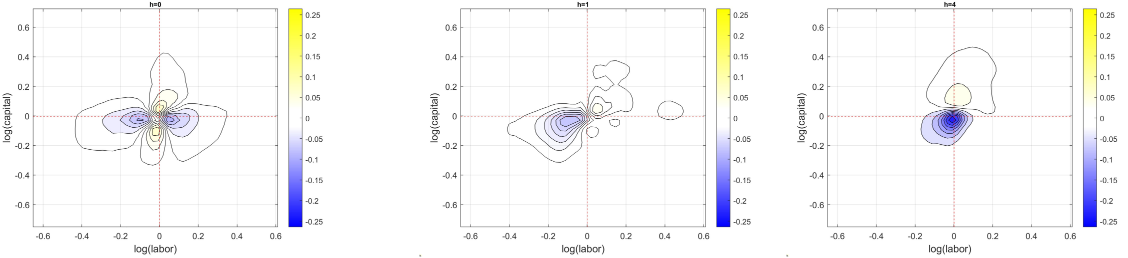


Figure: The figure shows the Impulse Response Functions (IRFs) of the moments of the bivariate capital labor distribution to a one standard deviation TFP shock. In red bold line we report the posterior mean estimate from the FunVAR while in dashed red line we report the 15th and 85th credible sets. Back

TFP shocks

Effect on the firm-level joint capital labor distribution

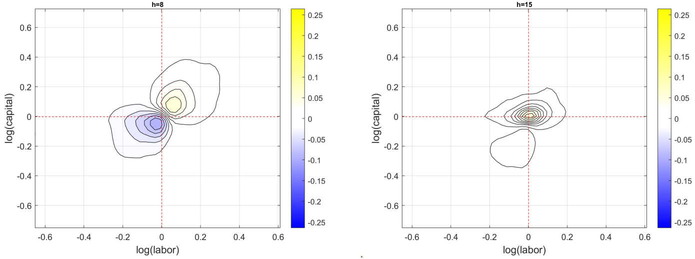


$h = 0$ $h = 1$ $h = 4$

Figure: Contours of the posterior mean estimate of the bivariate labor and capital distribution FIRF following an aggregate TFP shock.The dashed line reports the posterior mean steady state values.The dashed line reports the posterior mean steady state values. Back

TFP shocks

Effect on the firm-level joint capital labor distribution



$h = 8$ $h = 15$

Figure: Contours of the posterior mean estimate of the bivariate labor and capital distribution FIRF following an aggregate TFP shock.The dashed line reports the posterior mean steady state values. Back

Distributional effects of monetary policy shocks in the US

Monetary policy shocks: the model

Data:

- **Aggregate data:** industrial production, the unemployment rate, commodity price index, CPI index, the excess bond premium (Gilchrist and Zakrajšek 2012) and the two-year Treasury yield.
- **Firm level data:** Compustat data on employment (`emp`) and capital (`ppent`).

Identification strategy:

- **Internal instrument** using Bauer and Swanson (2023) orthogonalized MP shocks series.

Estimation sample: 1984Q4 - 2019Q4

Back

Monetary policy shocks

Effect on the aggregate variables

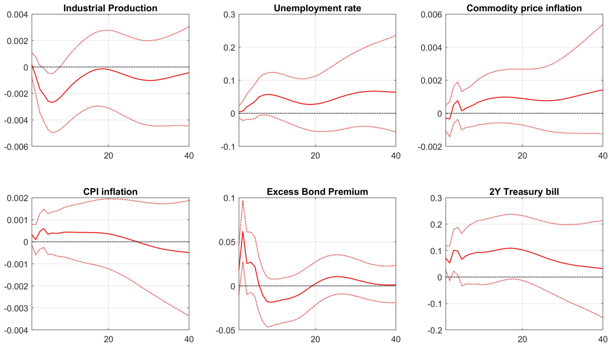


Figure: The figure shows the Impulse Response Functions (IRFs) of the macroeconomic aggregates to a one standard deviation TFP shock. In red bold line we report the posterior mean estimate from the FunVAR while in dashed red line we report the 15^{th} and 85^{th} credible sets. Back

Monetary policy shocks

Effect on the moments of the firm-level joint capital labor distribution

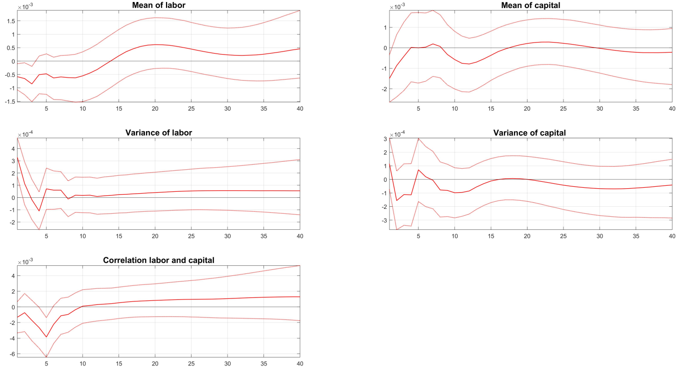
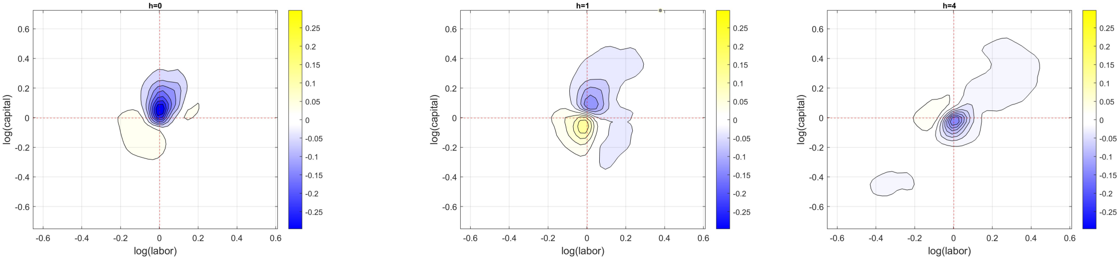


Figure: The figure shows the Impulse Response Functions (IRFs) of the moments of the bivariate capital labor distribution to a one standard deviation TFP shock. In red bold line we report the posterior mean estimate from the FunVAR while in dashed red line we report the 15^{th} and 85^{th} credible sets. Back

Monetary policy shocks

Effect on the firm-level joint capital labor distribution

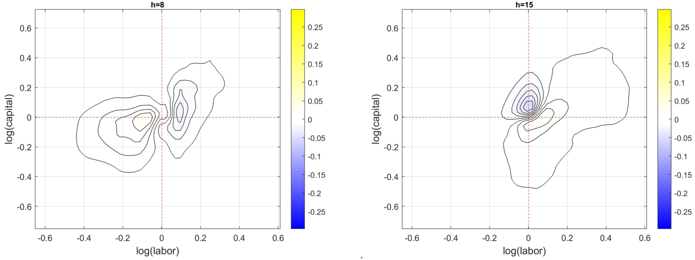


$h = 0$ $h = 1$ $h = 4$

Figure: Contours of the posterior mean estimate of the bivariate labor and capital distribution FIRF following an aggregate monetary policy shock.The dashed line reports the posterior mean steady state values.The dashed line reports the posterior mean steady state values. Back

Monetary policy shocks

Effect on the firm-level joint capital labor distribution



$h = 8$ $\qquad$ $h = 15$

Figure: Contours of the posterior mean estimate of the bivariate labor and capital distribution FIRF following an aggregate monetary policy shock.The dashed line reports the posterior mean steady state values. Back

Appendix: Nowcasting distributions: a functional MIDAS model.

Massimiliano Marcellino   Andrea Renzetti   Tommaso Tornese

Evaluation criteria

- For $h = \{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}\}$ we transform $q_{t+h}(\tau)$ into $f_{t+h}(x)$.
- We compute our nowcasts considering the posterior mean estimates of $f_{t+h}(x)$.
- We evaluate the nowcasts of the household income distribution comparing:
  - Average Kullback–Leibler (KL) divergence and Hellinger Distance (HD) from a kernel density estimate of the income distribution.
  - Average quantile scores (QS) for selected quantiles of the income distribution.
  - Root mean squared error (RMSE) of the following moments and features of the income distribution: mean, variance, kurtosis, Interquartile Range (IR), Gini index and the Coefficient of Variation (CV).

Nowcasting the household income distribution in the US

Forecast accuracy from 1998 to 2023

$h = \frac{1}{4}$	Avg KL	Avg HD	Avg QS20	Avg QS50	Avg QS80
RIDGE-SUR-MIDAS	0.43	0.85	1.07	0.92	0.95
BLSS-SUR MIDAS	0.36	0.85	0.78	0.79	0.81

Table: The metrics are expressed in ratio with respect to the VAR approximation of the Functional VAR model with Minnesota type of prior.

Back

Nowcasting the household income distribution in the US

Forecast accuracy from 1998 to 2023

$h = \frac{2}{4}$	Avg KL	Avg HD	Avg QS20	Avg QS50	Avg QS80
RIDGE-SUR-MIDAS	0.47	0.86	1.11	0.95	0.98
BLSS-SUR MIDAS	0.53	0.85	0.70	0.69	0.74

Table: The metrics are expressed in ratio with respect to the VAR approximation of the Functional VAR model with Minnesota type of prior.

Back

Nowcasting the household income distribution in the US

Forecast accuracy from 1998 to 2023

$h = \frac{3}{4}$	Avg KL	Avg HD	Avg QS20	Avg QS50	Avg QS80
RIDGE-SUR-MIDAS	0.47	0.87	1.13	0.97	1.01
BLSS-SUR MIDAS	0.37	0.84	0.64	0.68	0.78

Table: The metrics are expressed in ratio with respect to the VAR approximation of the Functional VAR model with Minnesota type of prior.

Back

Nowcasting the household income distribution in the US

Forecast accuracy from 1998 to 2023

$h = \frac{1}{4}$	RMSE Mean	RMSE Variance	RMSE Skewness	RMSE Kurtosis	RMSE IR	RMSE GINI	RMSE CV
RIDGE-SUR-MIDAS	0.53	0.89	0.49	0.47	0.83	0.34	0.52
BLSS-SUR MIDAS	0.51	0.90	0.45	0.45	0.89	0.28	0.61

Table: The metrics are expressed in ratio with respect to the VAR approximation of the Functional VAR model with Minnesota type of prior.

Back

Nowcasting the household income distribution in the US

Forecast accuracy from 1998 to 2023

$h = \frac{2}{4}$	RMSE Mean	RMSE Variance	RMSE Skewness	RMSE Kurtosis	RMSE IR	RMSE GINI	RMSE CV
RIDGE-SUR-MIDAS	0.55	0.89	0.52	0.49	0.83	0.35	0.55
BLSS-SUR MIDAS	0.53	0.86	0.61	0.61	0.94	0.45	0.68

Table: The metrics are expressed in ratio with respect to the VAR approximation of the Functional VAR model with Minnesota type of prior.

Back

Nowcasting the household income distribution in the US

Forecast accuracy from 1998 to 2023

$h = \frac{3}{4}$	RMSE Mean	RMSE Variance	RMSE Skewness	RMSE Kurtosis	RMSE IR	RMSE GINI	RMSE CV
RIDGE-SUR-MIDAS	0.57	0.91	0.52	0.49	0.82	0.35	0.55
BLSS-SUR MIDAS	0.50	0.90	0.45	0.44	0.99	0.34	0.58

Table: The metrics are expressed in ratio with respect to the VAR approximation of the Functional VAR model with Minnesota type of prior.

Back

Thanks for your attention!