Distributional Effects of Aggregate Shocks: Functional vs. Panel Approaches

Stephanie Ettmeier[1] Chi Hyun Kim[1] Frank Schorfheide[2]

[1]University of Bonn

[2]University of Pennsylvania, CEPR, NBER, PIER

June 27, 2024

Questions and empirical strategies in the existing literature

Q1: What is the effect of an aggregate shock on the cross-sectional distribution of x?

Q2: How does the x of particular households/firms respond to an aggregate shock?

Questions and empirical strategies in the existing literature

Q1: Effect of an aggregate shock on cross-sectional distribution	**Q2: Response of cross-sectional unit to aggregate shock**
Functional VAR Repeated cross sections Chang, Chen, Schorfheide (2024); Chang, Schorfheide (2024); Ettmeier (2023), Lenza, Savoia (2024)	**Panel model to track individuals** Panel data Holm, Paul, Tischbirek (2021); Almuzara and Sancibrian (2023); Amberg, Jansson, Klein, Rogantini Picco (2022); Andersen, Johannesen, Jorgesen, Peydro (2021)
VAR with inequality stats Repeated cross sections Coibion, Gorodnichenko, Kueng, Silvia (2017); Furceri, Loungani, Zdzienicka (2018); Guerello (2018)	**Pseudo panel to track groups** Repeated cross sections Anderson, Inoue, Rossi (2016); Cloyne, Ferreira, Surico (2020); Mitman, Broer, Kramer (2022)
Indirect calculation: multiply income component or asset share by aggregate IRF One cross section McKay, Wolf (2023); Lenza, Slacalek (2023); Del Canto, Grigsby, Qian, Walsh (2023)	

Overview / Contribution

→ Administrative data set from Germany that has panel structure but can also be used as repeated cross sections

- **Empirical contribution 1:** estimate a functional VAR (fVAR) to measure the response of the earnings distribution to a productivity shock for Germany

- **Methodological contribution:** replace the functional part of the VAR by cross-sectional-unit-level income dynamics equation (csuVAR)

- **Empirical contribution 2:** compare the csuVAR responses to the fVAR results
  - Panel information can be used to shed more insights into explanations for the distributional shifts
  - How do certain groups of individuals react?

Stylized example: interaction between macro- and microdynamics

- Aggregate variable y_t, cross-sectional variable x_{it} with density $p_t^x(x)$
- Macro dynamics:

$$y_t = B_{yy} y_{t-1} + \int B_{yl}(\tilde{x}) [\ln p_{t-1}^x(\tilde{x})] d\tilde{x} + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} p_\epsilon(\epsilon) \tag{1}$$

- Individual-level dynamics:

$$x_{it} = \lambda_{i1} y_t + \lambda_{i2} y_{t-1} + \phi_{xx} x_{it-1} + \eta_{it}, \quad \eta_{it} \overset{iid}{\sim} p_\eta(\eta), \quad (\lambda_{i1}, \lambda_{i2}) \overset{iid}{\sim} p_\lambda(\lambda_1, \lambda_2) \tag{2}$$

- Density (functional) dynamics (can be linearized):

$$p_t^x(x) = \int \int p_\eta(x - \lambda_1 y_t - \lambda_2 y_{t-1} - \phi_{xx}\bar{x}) p_\lambda(\lambda_1, \lambda_2) p_{t-1}^x(\bar{x}) d(\lambda_1, \lambda_2) d\bar{x} \tag{3}$$

- **fVAR approach:** estimate (1) and of (3) linearized wrt. $\ell_t(x) = \ln p_t^x(x)$
- **csuVAR approach:** estimate (1) and (2)

DATA SET

German administrative data: SIAB

- Panel data set containing a 2% sample of all individuals ever registered in the social security system
- Covers $\approx$ 80% of German labor force: excludes self-employed and civil servants
- Data on daily earnings, together with working days per spell
- Earnings are top-coded at the social security contribution ceiling
- Sample selection: 1992:Q1 - 2019:Q4

Micro-level observables

- Employment status: $s_{it} \in \{1(E, \text{employed}), 2(U, \text{unemployed}), 3(O, \text{out of sample})\}$ by moving units in O state when they drop out of sample

- Observe labor earnings when working: $\tilde{x}_{it}\mathbb{I}\{s_{it} = 1\}$. $\rightarrow$ Average:

$$\bar{x}_t = \frac{\sum_{i=1}^{N} \tilde{x}_{it}\mathbb{I}\{s_{it} = 1\}}{\sum_{i=1}^{N} \mathbb{I}\{s_{it} = 1\}}. \tag{4}$$

- Standardization + inverse hyperbolic sine transformation of observed earnings to remove trend and capture spatial correlation due to aggregate shocks:

$$x_{it} = f(\tilde{x}_{it}/\bar{x}_t). \tag{5}$$

- Unemployment rate

$$UR_t = \frac{\sum_{i=1}^{N} \mathbb{I}\{s_{it} = 2\}}{\sum_{i=1}^{N} \mathbb{I}\{s_{it} = 1\} + \sum_{i=1}^{N} \mathbb{I}\{s_{it} = 2\}} \tag{6}$$

Aggregate data and shocks

- **Macro variables:**
  - Log labor productivity, measured as total GDP/total hours worked (Federal Statistical Office Germany)
  - Log real GDP per capita (Federal Statistical Office Germany)
  - Log average earnings (SIAB) $\ln \bar{x}_t$ from above
  - Unemployment rate *or* EE EO UU UO OE OO transition probabilities (SIAB)
- **Recursive shock identification:** shock to labor productivity which is ordered first
- **Note:** analysis can be done with monetary or fiscal shocks, but we wanted to maximize variation generated by shock

FUNCTIONAL VAR

fVAR for empirical analysis

- Each t: econometrician observes Y_t and a sample of N *iid* draws x_{it} from $p_t(x)$.
- State-transition equations:

$$\begin{aligned} Y_t &= B_{y0} + B_{yy} Y_{t-1} + \boldsymbol{B}_{yl}[\ell_{t-1}] + u_{y,t} \\ \ell_t(x) &= B_{l0}(x) + B_{ly}(x) Y_{t-1} + \boldsymbol{B}_{ll}[\ell_{t-1}](x) + u_{l,t}(x), \\ & \text{e.g. } \boldsymbol{B}_{ll}[\ell_{t-1}](x) = \int B_{ll}(x,\tilde{x})\ell_{t-1}(\tilde{x})d\tilde{x} \end{aligned}$$

- Measurement equation for micro data:

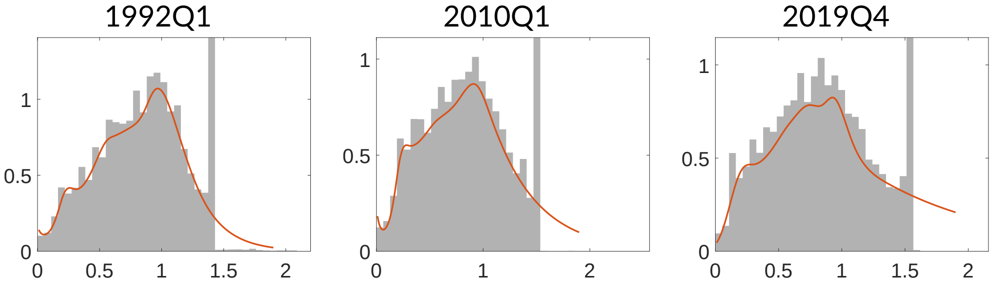
$$x_{it} \sim p_t(x) = \frac{\exp\{\ell_t(x)\}}{\int \exp\{\ell_t(x)\}dx}.$$

- Use a sieve approximation for $\ell_t(x)$ (and operators) to obtain K-dim model:

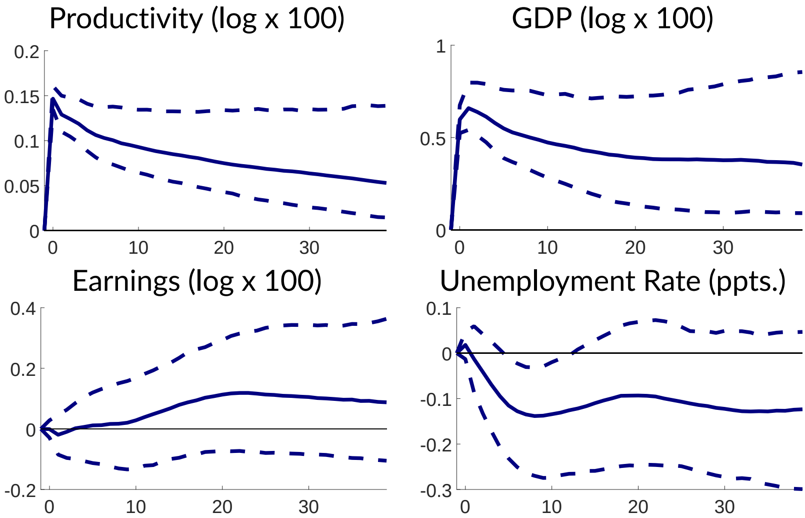
$$\ell_t(x) \approx \ell_t^{(K)}(x) = \sum_{k=1}^{K} \alpha_{k,t}\zeta_k(x) = \zeta'(x)\alpha_t.$$

- **Bayesian estimation:** see Chang, Chen, and Schorfheide (JPE, 2024).

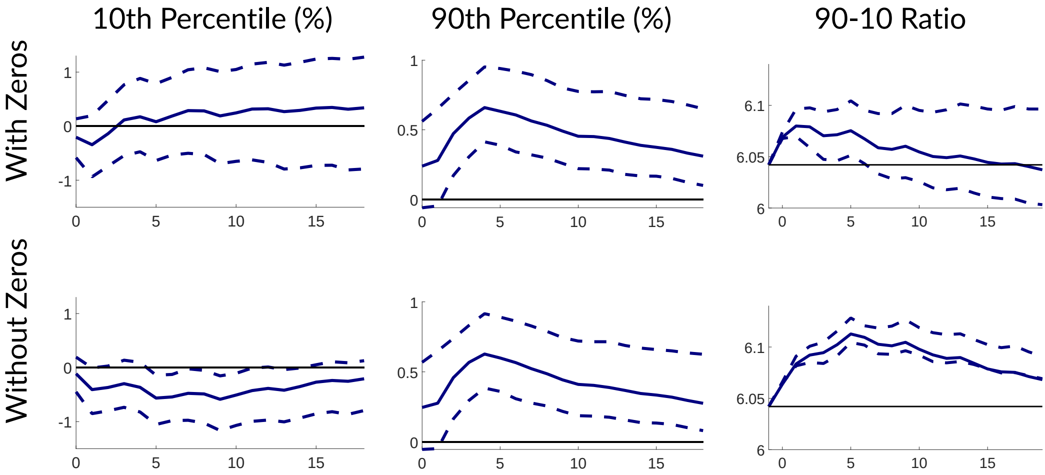
Estimated densities for three time periods



IRFs of the aggregate variables (1 std dev productivity shock)



IRF of percentiles and inequality statistics



Next steps

- Keep the aggregate VAR part but replace the functional dynamics in previous model with unit level x_{it} dynamics
- Transition probabilities for employment status (E, U, O) replace unemployment rate in vector of aggregate variables
- Generate IRFs at unit level and then aggregate into distributional response

CROSS-SECTIONAL UNIT VAR

csuVAR model

- Y_t stacks macro observables: labor productivity, real GDP pc, log average earnings
- Define unobserved transition probabilities: $\Pi_{jk,t} = \mathbb{P}\{s_{it} = j | s_{it-1} = k\}$
- $\mathcal{D}_{1:N,1:T}$ collects the micro observables $(s_{it}, x_{it}\mathbb{I}\{s_{it} = 1\})$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$
- For $p = 1$ we obtain factorization:

$$
\begin{aligned}
& p(Y_{1:T}, \Pi_{1:T}, \mathcal{D}_{1:N,1:T} | \theta) \\
&= \prod_{t=1}^{T} \Bigg(\underbrace{p(Y_t, \Pi_t | Y_{t-1}, \Pi_{t-1}, \ell_{t-1}, \theta) p(\ell_{t-1} | \mathcal{D}_{it-1})}_{\text{VAR part}} \\
&\quad \underbrace{\times\, p(s_{1:N,t} | \Pi_t, \mathcal{D}_{it-1}) \prod_{i=1}^{N} p(x_{it}\mathbb{I}\{s_{it} = 1\} | s_{it}, Y_t, Y_{t-1}, \mathcal{D}_{it-1}, \theta)}_{\text{panel part}} \Bigg)
\end{aligned}
\tag{7}
$$

csuVAR model: four-step estimation

1. $\hat{\Pi}_t$ from transition counts

2. θ_{agg} based on VAR with $\hat{\Upsilon}_t = [Y_t, \hat{\Pi}_t]$ with $\hat{\ell}_{t-1}(x)$ as additional explanatory variable (not yet implemented for current results)

3. θ_{mic} **based on panel model which includes:**

   - $U \mapsto E$ transitions: $p(x_{it}\mathbb{I}\{s_{it} = 1\}|s_{it} = 1, Y_t, s_{it-1} = 2, \theta_{mic});$
   - $O \mapsto E$ transitions: $p(x_{it}\mathbb{I}\{s_{it} = 1\}|s_{it} = 1, Y_t, s_{it-1} = 3, \theta_{mic});$
   - $E \mapsto E$ transitions: $p(x_{it}\mathbb{I}\{s_{it} = 1\}|s_{it} = 1, Y_t, \mathcal{D}_{it-1}, \theta_{mic}).$

4. Obtain IRFs of cross-sectional units + aggregate to densities ▸ IRF computation

csuVAR Model: $E \mapsto E$ Transitions

- (Simplified) earnings process (Z_t is a linear function of (Y_t, Y_{t-1}))

$$x_{it} = \rho x_{it-1} + \alpha_i + \beta_i' Z_t + \eta_{it}, \quad \eta_{it} \sim \mathcal{N}(0, \sigma_i^2). \tag{8}$$

- (Parametric) correlated random effects assumption:

$$(\alpha_i, \beta_i, \sigma_i^2) \overset{iid}{\sim} p(\alpha, \beta | x_{i0}, \sigma^2, \xi) p(\sigma^2 | \xi) \tag{9}$$

$$p(\alpha, \beta | x_{i0}, \sigma^2, \hat{\xi}) \equiv \mathcal{N}\left(\begin{bmatrix} 0.08 + 0.11 x_{i0} \\ -.016 \\ .004 \end{bmatrix}, \sigma^2 \begin{bmatrix} 0.21 & & \\ 0 & \frac{6E-6}{\lambda} & \\ 0 & -\frac{7E-7}{\lambda} & \frac{2E-7}{\lambda} \end{bmatrix} \right), \lambda = .0025$$

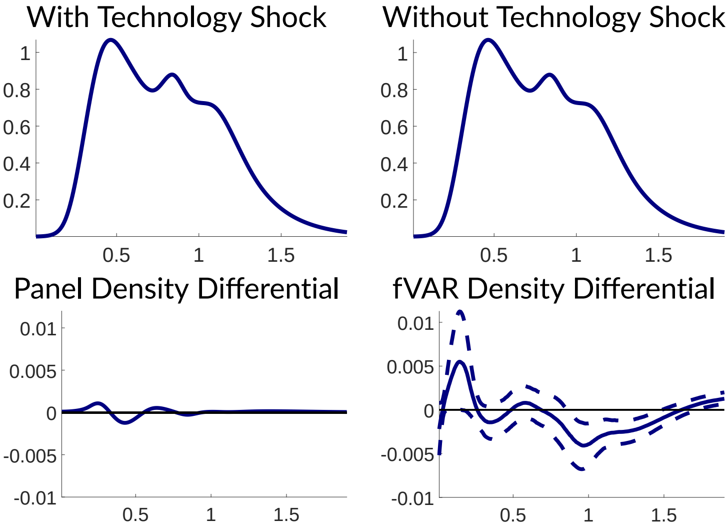
$$p(\sigma^2 | \hat{\xi}) \equiv IG(\underline{\nu} = 5, \underline{s}^2 = 5 \cdot 0.18^2)$$

- **Bayesian estimation...** (in progress). Results are obtained by taking a couple of short cuts.

csuVAR model: some remarks

- Typically panel and VAR models are not integrated. **We do integrate the two components**
- Large sample arguments are required to
  - replace latent $\Pi_{kt,t}$s by estimates $\hat{\Pi}_{jk,t} = N_{jk,t} / \sum_{j=1}^{3} N_{jk,t}$;
  - replace ℓ_{t-1} by $\hat{\ell}_{t-1}(\mathcal{D}_{it-1})$
- Model and parameters are set up so panel and VAR part can be estimated separately
- We assumed that transition probabilities for s_{it} do not depend on unit-level characteristics. Abstract from selection effects for now
- "Missing intercept" (McKay and Wolf; Barnichon and Mesters) implicitly through feedback from lagged cross-sectional distribution into aggregate variables (not yet implemented)

Impulse response of cross-sectional densities, $h = 4$



Remarks

- **csuVAR results are very preliminary**
- Density differentials are qualitatively similar, but quantitatively different
- Distributional effect obtained from panel analysis is more muted. Modeling of income dynamics needs to be refined in various dimensions (negative income in simulations, top coding, etc.)
- Once there is a match between distributional responses, panel information can be used to shed more insights into explanations for the distributional shifts. (How do certain groups of individuals react?)
- How does all of this relate to panel local projections?

Conclusion

- **Use German administrative data to compare two empirical approaches...**

- fVAR modeling:
  - (+) repeated cross-sections suffice
  - (+) unit-level behavior and heterogeneity does not need to be explicitly modeled
  - (−) cannot track behavior of cross-sectional units.

- csuVAR modeling:
  - (+) ability to track unit-level behavior
  - (−) estimation requires panel data
  - (−) challenging to specify unit-level law-of-motion: heterogeneity, non-Gaussianity, nonlinearity

- In practice, researchers are limited by the availability of data sets. Insights from this research may be useful for combining different types of data sets and conducting analyses with mixed-frequency data.

Impulse response of cross-sectional units + densities

Fix an event date T_0, e.g., 2010:Q1. Shock happens in Q2.

For $h = 0, \ldots, H$**:** for baseline "0" trajectories (no productivity shock at $h = 0$) and shocked "s" trajectories (one std dev productivity shock at $h = 0$).

1. Using cross-sectional data from period $t = T_0 + h - 1$ estimate coefficients for cross-sectional log densities $\ell^0_{t+h-1}(\cdot)$ and $\ell^s_{t+h-1}(\cdot)$
2. Iterate estimated aggregate VAR forward to obtain Y^0_{t+h} and Y^s_{t+h} (based on Y^0_{t+h-1}, Y^s_{t+h-1}, ℓ^0_{t+h-1}, and ℓ^s_{t+h-1}).
3. Conditional on Y^0_{t+h} and Y^s_{t+h}, generate $(x^0_{it+h}\mathbb{I}\{s^s_{it+h} = 1\}, s^s_{it+h})$ and $(x^s_{it+h}\mathbb{I}\{s^s_{it+h} = 1\}, s^s_{it+h})$ for $i = 1, \ldots, N$.
4. Estimate log spline densities to obtain $p^0_{t+h}(\cdot)$ and $p^s_{t+h}(\cdot)$.

Back