Measuring the Effects of Aggregate Shocks on Unit-Level Outcomes and Their Distribution

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

[1]University of Bonn

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

October 15, 2024

Two Questions

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

Q2: How does the x of a particular cross-sectional unit respond to an aggregate shock?

Potential Empirical Strategies

Q1: Effect of an aggregate shock on cross-sectional distribution	Q2: Response of cross-sectional unit to aggregate shock
Functional VAR	**Panel model to track individuals (Panel LP)**
Repeated cross sections	Panel data
Chang, Chen, Schorfheide (2024); Chang, Schorfheide (2024); Ettmeier (2023), Lenza, Savoia (2024), (...)	Holm, Paul, Tischbirek (2021); Almuzara and Sancibrian (2023); Amberg, Jansson, Klein, Rogantini Picco (2022); Andersen, Johannesen, Jorgesen, Peydro (2021)

→ **What we do:** Compare modelling the distributional response directly (**Q1**) vs. computing the unit-level responses and converting them into distributional response (**Q2**).

→ **Note:** in existing literature, unit-level modeling is mostly based on panel local projections (LPs).

Today's overview / contributions

Methodological contribution:

- Replace the functional part of the VAR by income dynamics for cross-sectional units ($\Rightarrow$ csuVAR). LPs can be viewed as an approximation.
- csuVAR delivers unit-level responses that can be aggregated to distributional response.

Empirical contributions (this talk):

- Use administrative data set from Germany on labor earnings that has panel structure.
- Estimate fVAR and csuVAR and compare their implied earnings distribution to productivity shock.

Empirical contributions (ongoing research):

- Use panel information to shed light into explanations for distributional shifts
- Explore role of unit-level heterogeneity for distributional responses. How do certain groups respond?
- Quantify relative importance of GE vs. PE effects / "missing intercept".

Related literatures

- **Functional (VAR) Approaches:** used widely outside of economics; recent applications in economics; user-specified basis functions vs. functional principal components analysis.

- **Income dynamics literature:** modeling of unit-level income processes; emphasis is on separating permanent from transitory income components.

- **Panel models with aggregate shocks:** often focus on nonparametric identification; joint modeling of unit-level and aggregate dynamics; panel local projections.

- **Estimating heterogeneous coefficients in panel models:** typically correlated RE setup; cross section provides information about CRE distribution; use as prior for unit-level time series regressions; empirical Bayes vs. full Bayes implementations.

Stylized example: interaction between macro and micro dynamics

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

$$A_{yy} Y_t = B_{yy} Y_{t-1} + \int B_{yl}(\tilde{x})[\ln p_{t-1}^x(\tilde{x})] d\tilde{x} + D_y^{1/2} \epsilon_{y,t}, \quad \epsilon_{y,t} \overset{iid}{\sim} p_\epsilon(\epsilon). \qquad (1)$$

- **Cross-sectional unit dynamics:**

$$x_{it} = \gamma_{i0} + \gamma_{iY} Y_t + \rho x_{it-1} + u_{it}, \quad u_{it} \overset{iid}{\sim} p_u(u), \quad (\gamma_{i0}, \gamma_{iY}) \overset{iid}{\sim} p_\gamma(\gamma_0, \gamma_Y). \qquad (2)$$

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

$$p_t^x(x) = \int \int p_u(x - \gamma_{i0} - \gamma_{iY} Y_t - \rho \bar{x}) p_\gamma(\gamma_0, \gamma_Y) p_{t-1}^x(\bar{x}) d(\gamma_0, \gamma_Y) d\bar{x}. \qquad (3)$$

- **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:

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

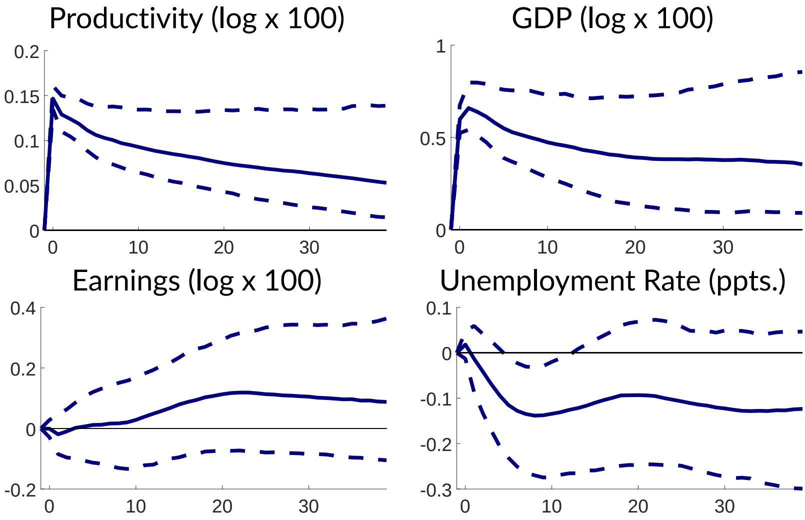
- Unemployment rate

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

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.

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



Empirical fVAR Analysis

fVAR specification

- Each t: econometrician observes Y_t and a sample of N *iid* draws x_{it} from $p_t(x)$
- Measurement equation for micro data:

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

- State-transition equations:

$$\begin{aligned} A_{yy} Y_t &= B_{yy} Y_{t-1} + \boldsymbol{B}_{yl}[\ell_{t-1}] + D_y^{1/2} \epsilon_{y,t} \\ A_{ly}(x) Y_t + \boldsymbol{A}_{ll}[\ell_t](x) &= B_{ly}(x) Y_{t-1} + \boldsymbol{B}_{ll}[\ell_{t-1}](x) + \boldsymbol{D}_l^{1/2}[\epsilon_{l,t}](x), \\ &\text{e.g. } \boldsymbol{B}_{ll}[\ell_{t-1}](x) = \int B_{ll}(x,\tilde{x})\ell_{t-1}(\tilde{x})d\tilde{x}. \end{aligned}$$

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

$$\ell_t(x) \approx \ell_t^{(K)}(x) = \sum_{k=1}^{K} \alpha_{k,t}\zeta_k(x).$$

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

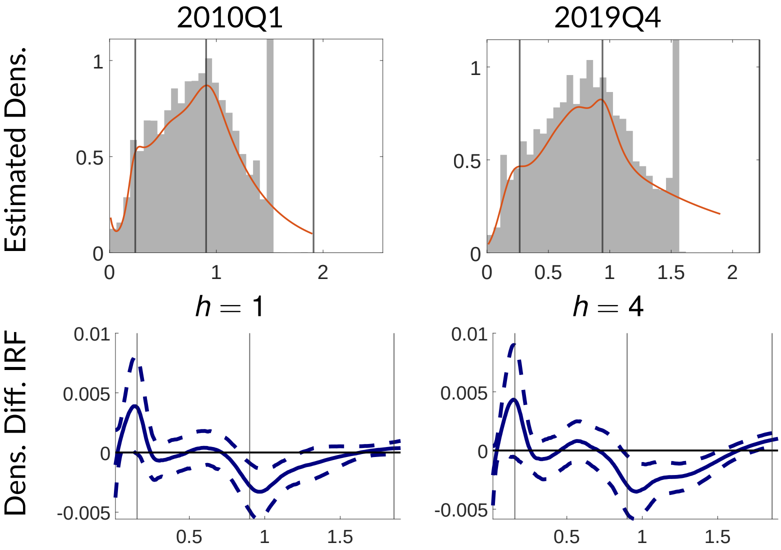
fVAR - cookbook recipe

Let's say you have some macro series (labor productivity, GDP, unemployment,...) and cross-sectional data on individual level labor earnings for each quarter $t = 1, \ldots, T$.

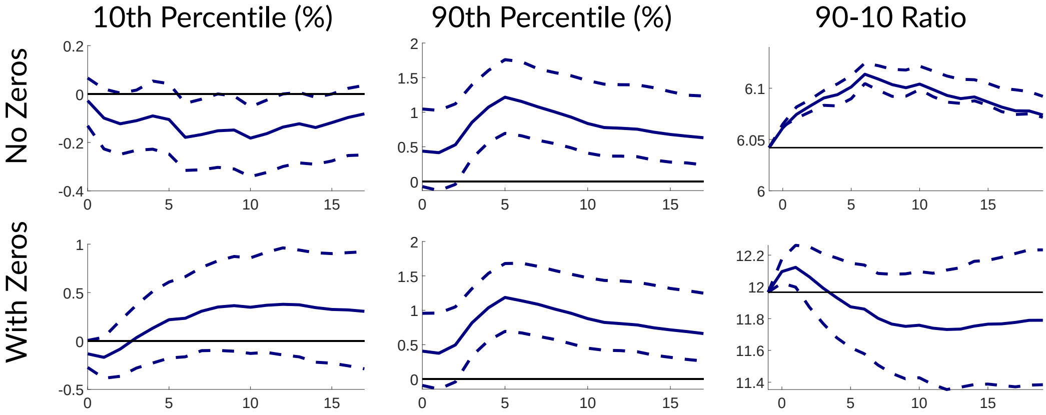
$\Rightarrow$ Convenient and easy to use three-step approach:

1. For each period t separately, estimate a log-spline density for the cross sectional observations X_t. This yields spline coefficients $\hat{\alpha}_t$.
2. Then either estimate a VAR in $\hat{\alpha}_t$ and macro variables Y_t, ignoring the estimation error associated with $\hat{\alpha}_t$, or estimate a linear state-space model.
3. Use estimated model for $(Y_t, \hat{\alpha}_t)$ to compute IRFs of aggregate variables *and* cross-sectional densities.

Estimated densities and their IRFs



IRF of percentiles + inequality statistics (1 std dev productivity shock)



Empirical csuVAR Analysis

Empirical csuVAR model

Observables and Unobservables:

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

Two Simplifying Assumptions (For Today):

- Lower triangular structure: $p_t^x(x)$ does not affect Y_t contemporaneously.
  $\Rightarrow$ Allows us to estimate the macro part *separately* from the micro part.
- Transition from $s_{it-1} \mapsto s_{it}$ does not depend on unit-level characteristics.

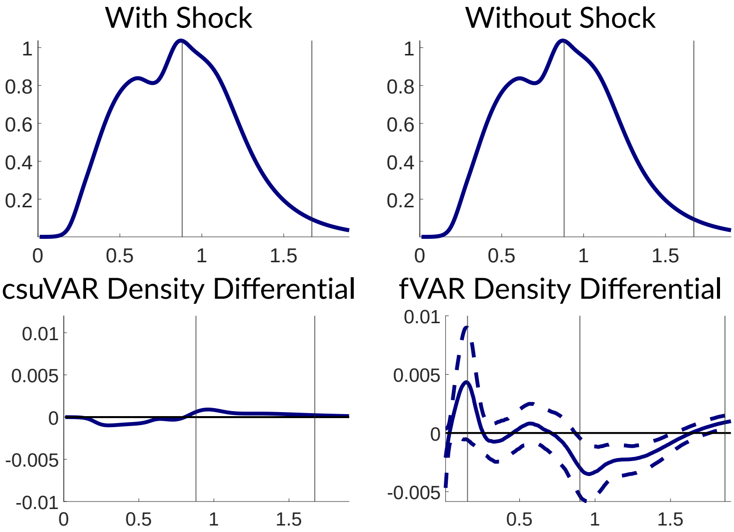
Likelihood Function

$$
\begin{aligned}
& p\left(Y_{1:T}, \Pi_{1:T}, \ell_{1:T}, \mathcal{D}_{1:N,1:T}, \gamma_{1:N}, \sigma^2_{1:N} \middle| Y_0, \Pi_0, \ell_0, \theta_{VAR}, \theta_{csu}, \xi\right) \\
= & \prod_{t=1}^{T} \Bigg\{ \underbrace{p(Y_t, \Pi_t | Y_{t-1}, \Pi_{t-1}, \ell_{t-1}, \theta_{VAR})}_{\text{VAR Part}} \times \underbrace{p(s_{1:N,t} | \Pi_t, \mathcal{D}_{1:N,t-1})}_{\text{Panel Part I}} \\
& \times \underbrace{\left(\prod_{i=1}^{N} \left[p(\gamma_i, \sigma_i^2 | x_{it}, \xi) \right]^{\mathbb{I}\{t=\tau_{i0}\}} \left[p(x_{it} | Y_t, Y_{t-1}, \mathcal{D}_{it-1}, \gamma_i, \sigma_i^2, \theta_{csu}) \right]^{\mathbb{I}\{s_{it}=1\}} \right)}_{\text{Panel Part II}} \\
& \times \underbrace{\left(\prod_{t=1}^{T} \mathbb{I}\{\ell_t = f_t(\cdot)\} \right)}_{\text{Panel Part III}} \Bigg\}.
\end{aligned}
$$

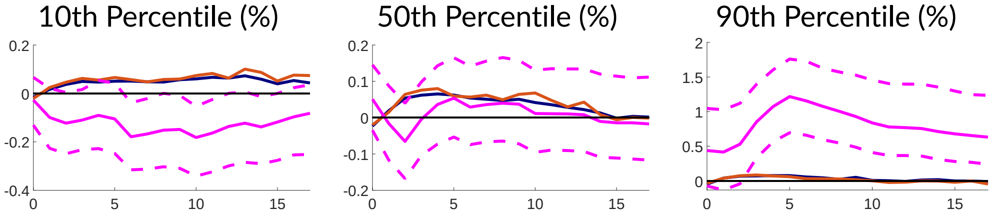
csuVAR Model: Four-step Estimation

1. $\hat{\Pi}_t$ from transition counts; use large sample argument to replace Π_t by $\hat{\Pi}_t$.
2. θ_{agg} based on VAR with $\hat{Y}_t = [Y_t, \hat{\Pi}_t]$.
3. θ_{mic} **based on panel model which includes:**
   - $E \mapsto E$ transitions: $p(x_{it}\mathbb{I}\{s_{it} = 1\}|s_{it} = 1, Y_t, \mathcal{D}_{it-1}, \theta_{mic})$;
   - $U \mapsto E$ transitions: $p(x_{it}\mathbb{I}\{s_{it} = 1\}|s_{it} = 1, Y_t, s_{it-1} = 2, \theta_{mic})$;
   - $O \mapsto E$ transitions: $p(x_{it}\mathbb{I}\{s_{it} = 1\}|s_{it} = 1, Y_t, s_{it-1} = 3, \theta_{mic})$.
4. Obtain IRFs of cross-sectional units + aggregate to log densities.

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



Percentile responses from fVAR and csuVAR (no zeros)

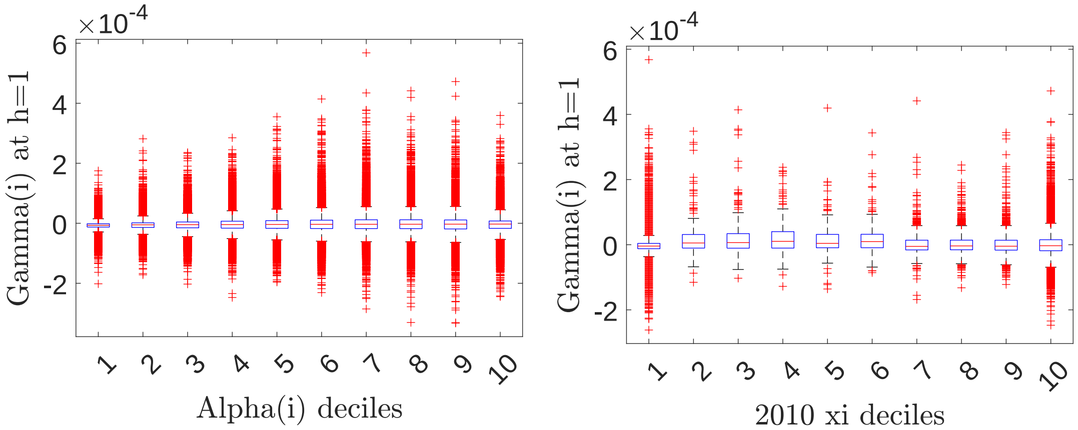


Heterogeneity: Income vs. Agg. Shock Sensitivity

$$x_{it} = \rho x_{it-1} + \alpha_i + \underbrace{\beta_{1i}\text{prod growth}_t + \beta_{2i}\text{GDP growth}}_{\gamma_{i,t}} + u_{it}, \quad u_{it} \sim \mathcal{N}(0, \sigma_i^2). \tag{7}$$

Long-run income deciles

Base period income deciles



Mobility in Response to Aggregate Shock

	1	2	3	4	5	6	7	8	9	10	U	O
1		5	4	7	13	27	10	10	4	6	1	25
2	34		1	2	4	2	2	0	1	0	0	6
3	0	45		1	1	2	0	0	0	0	1	0
4	3	0	56		0	1	2	0	0	0	1	0
5	2	0	0	60		0	0	1	1	0	0	1
6	1	2	0	0	54		4	1	0	0	1	1
7	5	1	0	2	0	23		0	0	0	0	0
8	2	1	0	0	0	0	22		0	0	0	0
9	1	0	0	2	0	1	0	10		1	1	1
10	3	0	0	0	0	0	0	0	11		0	0
U	6	0	4	4	10	10	9	9	13	22		29
O	76	19	6	6	4	19	3	15	8	6	97	

t+1

	1	2	3	4	5	6	7	8	9	10	U	O
1		3	6	10	8	22	15	8	8	4	1	21
2	39		1	6	8	4	4	2	1	0	0	4
3	1	63		0	1	1	1	1	0	0	0	1
4	3	2	69		0	1	0	1	0	0	2	2
5	2	2	0	74		0	0	0	2	0	0	1
6	0	0	0	0	68		0	0	0	0	1	1
7	5	1	0	0	0	41		1	1	0	0	0
8	1	1	1	0	0	0	16		1	1	0	2
9	3	0	1	0	0	0	0	7		5	2	1
10	3	0	0	0	0	0	1	0	4		0	0
U	3	1	2	3	4	11	20	12	12	7		38
O	65	15	7	6	11	9	11	9	9	10	97	

t+4

Approx 200,000 obs. in each decile.

csuVAR model: outlook

- **csuVAR results are preliminary...**
  $\Rightarrow$ Full Bayesian estimation to be implemented.
  $\Rightarrow$ Modeling of income dynamics needs to be refined in various dimensions (negative income in simulations, top coding, etc.)

- Explore unit-level heterogeneity and explain distributional shifts.

  $\rightarrow$ What are characteristics of the individuals that respond strongly to the labor productivity shock?

- Quantify relative importance of GE vs. PE effects

  $\rightarrow$ "Missing intercept" (McKay and Wolf; Barnichon and Mesters) implicitly through feedback from lagged cross-sectional distribution into aggregate variables.

Conclusion

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.