Federal Reserve Bank of Cleveland Working Paper Series

Incorporating Micro Data into Macro Models Using Pseudo VARs

Gary Koop, Stuart McIntyre, James Mitchell, and Ping Wu

Working Paper No. 26-04

February 2026

Suggested citation: Koop, Gary, Stuart McIntyre, James Mitchell, and Ping Wu. 2026. "Incorporating Micro Data into Macro Models Using Pseudo VARs." Working Paper No. 26-04. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202604.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

Incorporating Micro Data into Macro Models Using Pseudo VARs*

Gary Koop[†], Stuart McIntyre[‡], James Mitchell[§] and Ping Wu[¶]

February 1, 2026

Abstract

This paper develops a method to incorporate micro data, available as repeated cross-sections, into macro VAR models to understand the distributional effects of macroeconomic shocks at business cycle frequencies. The method extends existing functional VAR models by "looking within" the micro distribution to identify the degree to which specific types of micro units are affected by macro shocks. It does so by creating a pseudo-panel from the repeated cross-section and adding these pseudo individuals into the macro VAR. Jointly modeling the micro and macro data leads to a large (pseudo) VAR, and we use Bayesian methods to ensure shrinkage and parsimony. Our application revisits Chang et al. (2024) and compares their functional VAR-based distributional impulse response functions with our proposed pseudo VAR-based ones to identify what types of individuals' earnings are most affected by business-cycle-type shocks. We find that the individuals exhibiting the strongest positive cyclical sensitivity are those in the lower tail of the earnings distribution, particularly men and those without a college education, as well as young workers.

Keywords: Functional VAR; pseudo panel; earnings distribution; business cycle shocks
JEL Codes: C32, C53, E37

*The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland or the Federal Reserve System. We thank Fabio Canova, Alexander Chudik, Kevin Lee, Dirk Krueger, Hashem Pesaran, Barbara Rossi, Ayşegül Şahin, Ron Smith, Martin Weale, and conference and seminar participants at Salzburg University, King's College London, University of Kent, Bank of Canada, Rimini Centre for Economic Analysis 2024, Economic Measurement 2024, Dolomiti Macro Meetings 2025, and CFE 2025.

[†]University of Strathclyde; Economic Statistics Centre of Excellence (ESCoE) (gary.koop@strath.ac.uk)
[‡]University of Strathclyde; ESCoE (s.mcintyre@strath.ac.uk)
[§]Federal Reserve Bank of Cleveland; ESCoE (james.mitchell@clev.frb.org)
[¶]University of Strathclyde; ESCoE (ping.wu@strath.ac.uk)

1 Introduction

Vector autoregressive (VAR) models are the established workhorse empirical tool to analyze macroeconomic variables such as GDP, inflation, and unemployment. When identification restrictions are applied, they can also be used to evaluate the dynamic effects of economic shocks on these variables. Even with such restrictions, structural VAR models remain relatively unconstrained by economic theory, making them valuable as reference points for validating structural macroeconomic models, such as dynamic stochastic general equilibrium (DSGE) models. For example, structural impulse response functions are often compared to, and sometimes aligned with, those derived from empirical VAR models; for example, see Sims (1989), Cogley and Nason (1995), and Chari et al. (2008).

Recent decades have seen growing interest in the distributional effects of economic shocks, spurred, among other factors, by policy discussions about the rise in wage and other heterogeneities in the US and other industrialized nations over the past 40 years (see Acemoglu and Restrepo (2022) and Heathcote et al. (2023)). This interest has been further supported by the availability of richer micro-level data, often derived from large-scale household and firm surveys. In response, structural macroeconomic models with heterogeneous agents have been developed, allowing for differences in income, wealth, and consumption across individuals, households, or firms; for example, see Krusell and Smith (1998), Ahn et al. (2018), Kaplan and Violante (2018), Kaplan et al. (2018), Liu and Plagborg-Møller (2023), and Bayer et al. (2024).

To complement these advances, it is important to develop VARs that account for dynamic heterogeneities in micro-level data and their interactions with macroeconomic variables. Such "micro-macro" VAR models can provide a framework for comparing the predictions of structural heterogeneous-agent models with observed data and provide data-based evidence on how much micro heterogeneity matters for the business cycle and vice versa. They supply basic empirical facts regarding how individual-level dynamics respond to (macro) shocks such as business cycle, monetary policy, or fiscal shocks and allow the researcher to see whether the response to such shocks varies by observable individual-level characteristics. If it does, systematic earnings risk is not homogeneous across individuals and cannot be captured by a time fixed effect, consistent with existing empirical evidence; see, for example, Guvenen et al. (2014).

The key econometric challenge lies in integrating microeconomic data into macroeconomic time series models like the VAR. Several recent papers have taken up this challenge. A simple approach is to include appropriate descriptive statistics that summarize the information in the microeconomic data and include these as variables in a VAR alongside the macroeconomic variables. For instance, Coibion et al. (2017) and Mumtaz and Theophilopoulou (2017) investigate the impact of monetary policy on inequality by including the Gini coefficient or percentiles of the income distribution in a macroeconomic VAR. A drawback of this approach is that, by aggregating micro data in a specific way, important information may be lost.

More recently, *functional* VAR models have been proposed. These use functional time series methods, which involve the dependent variable(s) in a time series model being a function (such as a probability density function) instead of a variable. Chang et al. (2024) develop a functional VAR model that approximates the micro earnings distribution of individuals in the US using a spline and include the sieve coefficients of the spline approximation as additional variables in a macroeconomic VAR.[1] Drawbacks of this approach lie in the fact that it involves this approximation. There is also a question of what basis function to use. For instance, Bjornland et al. (2023), Meeks and Monti (2023), and Huber et al. (2024) use bases other than the spline, and then use functional principal components to summarize the variation in the density. While these functional VAR models let the modeler trace out the effects of aggregate shocks on the entire micro distribution, they do not have the ability to "look within" the distribution to see, within this distribution, specifically which types of micro agents are most (or least) affected by the macro shock of interest.

In this paper, we develop an alternative approach to fill this gap. It is based on a simple idea: Include each individual micro unit as an additional, separate equation in the macroeconomic VAR.[2] In other words, to give an example that anticipates our empirical application, if x_{it} is the earnings of individual i at time t, $\boldsymbol{x}_t$ is a vector of earnings for all N individuals, and $\boldsymbol{y}_t$ is a vector of M conventional macroeconomic variables, then use an $(N + M)$ dimensional VAR with $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ as the dependent variables. Such an approach retains the rich granular information in the microeconomic data, rather than just summarizing it through, for example, the Gini coefficient. Moreover, it does not have to approximate the earnings distribution (say using spline-based methods) as in functional VAR models.

However, our idea likely has two data-driven problems when applying it in practice. The first is that many relevant micro data sets contain information across many thousands of individuals. This would lead to an enormous micro-macro VAR. Beginning with Banbura et al. (2010), who showed the superior forecast performance of a large Bayesian VAR involving over a hundred dependent variables, researchers have been successfully working with larger and larger VARs. In more recent years, papers such as Feldkircher et al. (2022) have found that Bayesian VARs forecast better than alternatives, such as dynamic factor models, even when

[1] This model has seen several other applications. For instance, Lenza and Savoia (2024) model the distribution of firms' revenue.

[2] A recent related approach is seen in Ettmeier et al. (2024), who also link micro-level panel data with a VAR. Our approach, as set out below, allows for greater cross-unit heterogeneity at the individual level but at the price of requiring a smaller N. Our approach is also, unlike theirs, specifically designed to work with repeated cross-sections.

the VAR has more than 500 equations. But, anticipating use of a model on the micro earnings data used by Chang et al. (2024), simply including each individual's data as a new equation in the VAR would lead to micro-macro VARs that are much, much larger. This would challenge even state-of-the-art "big data" Bayesian VAR methods.

The second problem is that many relevant micro data sets are often not panels, but rather repeated cross-sections.[3] The individual-level earnings data from the Current Population Survey (CPS) used by Chang et al. (2024), and in much previous research on income heterogeneities in the US (for example, see Katz and Autor (1999)), have this feature. These CPS data are the basic household survey used to estimate official US unemployment and other important labor market statistics. They have the advantage, relative to administrative data, that they are publicly available and, critical for our paper, they provide information on an individual's characteristics, such as their age, gender, and education. The CPS sample size is much larger than alternative surveys, such as the aforementioned PSID. The CPS micro data on labor earnings have also been found to line up well with official aggregate (US) data; see Heathcote et al. (2023).

Following the panel data literature, this paper shows how both problems can be addressed through the construction of a *pseudo* panel; see Deaton (1985). This is a panel made up of pseudo individuals – groups of individuals who share a set of observable characteristics.[4] This strategy of creating pseudo individuals has been used before by Anderson et al. (2016), Parker and Vissing-Jorgensen (2010), and Cloyne et al. (2020), but not in a VAR context allowing for cross-sectional dependencies between the pseudo individuals. It involves grouping the N actual individual observations into N^* pseudo individuals by characteristic. At its simplest, the strategy involves, at each point in time, t, constructing a summary measure across individual survey responses of those within each pseudo individual group (who by construction possess the characteristics that define the group).

Let x^*_{jt} denote the earnings of pseudo individual j at time t and let $\boldsymbol{x}^*_t$ be the vector of earnings across all N^* pseudo individuals. We then work with a large VAR, with $\boldsymbol{x}^*_t$ and $\boldsymbol{y}_t$ as dependent variables. We refer to such a VAR as a "pseudo VAR." Note that the pseudo VAR not only solves the second problem described above, but also the first one if we also choose the pseudo individuals such that $N^* \ll N$. The pseudo VAR can then be used to model the earnings of the pseudo individuals, with defined and known characteristics, that can then

[3]There are notable exceptions, particularly in Scandinavian countries, where individual-level panel data are available via administrative sources; see Holm et al. (2021), Amberg et al. (2022), and Andersen et al. (2023). In contrast, in the US, while the Michigan Panel Study on Income Dynamics (PSID) does provide panel data on earnings back to 1968, its sample size is much smaller.

[4]We draw a distinction between pseudo cohorts in the Deaton (1985) sense and pseudo individuals, where the former are defined with reference to year of birth (among other characteristics), while pseudo individuals replace year of birth with age.

be transformed back into estimates of the entire earnings distribution. The choice of pseudo individuals is crucial, not only to ensure that N^* is of manageable size, but also to ensure that the earnings distribution of the pseudo individuals matches up well with the observed earnings distribution. We develop methods in this paper to achieve this. The pseudo VAR may still be of very high dimension and, accordingly, we use Bayesian methods to ensure parsimony.

The pseudo VAR can be used for impulse response analysis and conditional forecasting. We propose two types of impulse response. The first measures the dynamic impact of a macro shock on the earnings of pseudo individuals. This enables us to address issues relating to the heterogeneity of responses across different types of individuals. For example, we can calculate different impulse responses for individuals (or collections of individuals) with different characteristics. We refer to these as individual (or group-level) impulse response functions (IRFs). The second type of IRF measures the dynamic impact of a macro shock on the earnings distribution itself. We refer to these as "distributional" IRFs. These can be used to address issues relating to heterogeneity across different regions of the earnings distribution.

Our empirical application investigates the heterogeneous impacts on earnings of a business cycle shock. Specifically, to motivate our modeling approach, we revisit the data set and application of Chang et al. (2024). Using the proposed pseudo VAR model, we extend their functional VAR-based analysis to "look within" the micro density. This lets us identify whether the observed distributional effects of the business cycle shock are driven by particular types of micro units. We focus on looking at the effects of a "main business-cycle"– type shock. But we emphasize that the pseudo VAR can be used to study the distributional effects of other types of macro shocks.[5] This focus, as in Chang et al. (2024), on how macro shocks affect distributional outcomes at business cycle frequencies, complements the wider literature on the structural drivers of trend movements in inequality. This includes skilled-biased technological change (e.g., Acemoglu and Autor (2011)), international trade (e.g., Helpman et al. (2017)), institutions including changes in union structure (e.g., Katz and Autor (1999)) and changes in labor supply, including female participation (e.g., Ngai and Petrongolo (2017)), and government policy.

The remainder of the paper is structured as follows. In Section 2, we set out the pseudo VAR and discuss specification choices. Section 3 introduces the micro and macro data used in the application. Section 4 compares the proposed pseudo VAR with the functional VAR of Chang et al. (2024), showing how the pseudo VAR can be used to understand that it is

[5]These macro shocks can be reduced-form or structural. If structural, the researcher can achieve identification by focusing only on the macro block of the pseudo VAR. This means they can, in effect, identify the macro shock in the same way that they would in a macro-only VAR. Inclusion of the micro data in the pseudo VAR does raise the possibility of exploiting their cross-sectional heterogeneity to identify structural macro shocks. This is left for future research.

the earnings of the young and of lower-earning men and those with below college levels of education that benefit most from positive business cycle shocks. It is the increased sensitivity of these types of individuals that explains the finding, common to Chang et al. (2024) and this paper, that positive business cycle shocks narrow the earnings distribution.

2 The Pseudo VAR Model

2.1 Bayesian Inference in the Pseudo VAR

Our pseudo VAR is a VAR and, thus, our econometric methods are standard ones for large Bayesian VARs. The dependent variables in our VAR are both the micro and the macro variables. To be specific, our pseudo VAR is:

$$\mathbf{B}_0\mathbf{z}_t = \mathbf{A}_1\mathbf{z}_{t-1} + \cdots + \mathbf{A}_p\mathbf{z}_{t-p} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), \tag{1}$$

where $\boldsymbol{z}_t = (\boldsymbol{y}_t', \boldsymbol{x}_t^{*'})'$,[6] so that:

$$\begin{bmatrix} \mathbf{B}_{0,11} & \mathbf{B}_{0,12} \\ \mathbf{B}_{0,21} & \mathbf{B}_{0,22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_t \\ \mathbf{x}_t^* \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{1,11} & \mathbf{A}_{1,12} \\ \mathbf{A}_{1,21} & \mathbf{A}_{1,22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{t-1} \\ \mathbf{x}_{t-1}^* \end{bmatrix} + \ldots + \begin{bmatrix} \mathbf{A}_{p,11} & \mathbf{A}_{p,12} \\ \mathbf{A}_{p,21} & \mathbf{A}_{p,22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{t-p} \\ \mathbf{x}_{t-p}^* \end{bmatrix} + \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{xt} \end{bmatrix}.$$

The number of dependent variables in the pseudo VAR is $M + N^*$. Following papers such as Chan (2022), we write the VAR in structural form with $\mathbf{D}$ being a diagonal matrix, thus allowing for equation-by-equation estimation, which vastly reduces the computational burden. Popular VAR priors, such as Chan (2022), assume $\mathbf{B}_0$ to be lower triangular with ones on the diagonal. Our pseudo VAR can be estimated under this lower triangularity assumption, but other choices are possible and, in our empirical work, we make a different choice for $\mathbf{B}_0$ as discussed below. It is worth stressing that our pseudo VAR is simply a large (potentially restricted) VAR, and any of the standard VAR priors and associated posterior simulation algorithms can be used. In our empirical work, we will make particular choices, as discussed below. Furthermore, the researcher is open to identifying structural macro shocks using their preferred identification strategy.

The pseudo VAR model in (1) allows for general patterns of static and dynamic dependence both between and within the macro and micro blocks. In practice, the researcher may be interested in imposing specific economically motivated restrictions on this model to draw out the important dimensions of heterogeneity across individuals and to understand if and how heterogeneity affects the transmission of shocks to the macro economy. They may also need

[6]For estimation, we add an intercept.

to impose restrictions, especially as N^* becomes large, to reduce the number of parameters that have to be estimated. In our empirical section, we will justify and impose a particular set of restrictions, but we note here that a wide range of restrictions could be imposed depending on the empirical application under study. Our pseudo VAR nests, for example, the "one-at-a-time" pseudo VAR model that would involve, repeatedly for each group in turn, estimating the macro VAR along with that single pseudo individual. Such a restricted VAR amounts to the model used by Anderson et al. (2016) to analyze the effects of macro (fiscal) shocks on pseudo individuals. Such a model is more restrictive than the one we work with below, in that it does not allow pseudo individuals to be affected by other pseudo individuals, beyond any common factor effects induced via dependence on the macro variables.

Where we depart from standard Bayesian VAR methods is in our treatment of the micro variables in the pseudo VAR, $\boldsymbol{x}_t^*$. These are the earnings of the pseudo individuals. Two issues arise. The first is that the earnings of any pseudo individual may be made up by the earnings of a heterogeneous group of actual individuals. The second is that the number of individuals making up each group differs. Given a careful choice of pseudo individuals, this heterogeneity is likely to be small. And for some empirical purposes, it may be acceptable to ignore it. For instance, if the researcher is directly interested in IRFs of specific groups of individuals (say, the effect of a shock on the average earnings of young, college-educated women in a certain region), then it is appropriate to replace $\boldsymbol{x}_t^*$ with average earnings in that group. However, in this paper we adopt a different strategy that addresses both of these issues.

When undertaking predictive simulation (either for impulse response analysis, forecasting, or conditional forecasting), we take the number of predictive draws for pseudo individual j in period t to be proportional to the number of actual individuals that make up this group, j.[7] These draws are taken from a parametric approximation to the earnings distribution of pseudo individual i, although a non-parametric approach could be used instead. For simplicity, given that average earnings in a group might be expected to be normally distributed, for each pseudo individual we take draws from a Gaussian distribution. But we need to accommodate the fact that our earnings data are top-coded. We proceed as follows. We assume the distribution is Gaussian, but we only observe a truncated sample from this distribution. Given observed moments (the sample mean and variance based on the truncated observations), we use textbook formulae for the moments of the truncated Gaussian distribution to solve for the moments of the (non-truncated) underlying Gaussian distribution.[8] Within our MCMC

[7] Since this number, N_{jt}, can vary over time, when doing full sample analysis we use N_{jT} as the number of observations in pseudo individual j, even though earlier periods may depart somewhat from this value.

[8] Another strategy would be to make a parametric assumption, say, Pareto, and use this to model the unobserved upper tail of the earnings distribution for pseudo individual, j. Upper tails of earnings distributions can be difficult to estimate; see, for instance, Feenberg and Poterba (2000) and Heathcote et al. (2023)

algorithm, we take earnings draws for each pseudo individual at different points in time via inverse transform sampling from this distribution. This lets us relate the p-th percentile draw of earnings from pseudo individual i at time t to the p-th percentile draw at time $(t-1)$.

Note that this sampling strategy both corrects for the different numbers of individuals making up each pseudo individual (by taking the number of draws proportional to N_{jT}) and acknowledges the within-pseudo-individual heterogeneity in a way that simply using average earnings, say, of each pseudo individual would not.

2.2 Impulse Response Analysis

We distinguish IRFs indicating how macro shocks affect specific pseudo individuals, or groups of them, with distributional IRFs that trace out the dynamic effects of the shock on the earnings distribution itself. Individual IRFs can be calculated in a straightforward fashion using standard structural VAR methods and share similarities with what is done in papers such as Amir-Ahmadi et al. (2022). Suppose, for instance, the researcher is interested in the response of earnings to a business cycle shock. The pseudo VAR will provide the impulse response of every pseudo individual to this shock, meaning there are N^* impulse responses, potentially a large number. Averages across the groupings defined by $\boldsymbol{Z}$ could be taken to reduce this number. For instance, an average of impulse responses across all pseudo individuals in a specific region could be used to produce region-specific impulse responses. Note that our strategy of taking N_{jT} draws for each pseudo individual means that, when we calculate group IRFs that average across all pseudo individuals within a group, we are taking a weighted average. The weights accurately reflect the number of individuals making up each pseudo individual within the wider group.

The second type of impulse response is to the earnings distribution itself. Such distributional impulse responses can be rationalized as generalized IRFs; see Koop et al. (1996). To be specific, we take the difference between two conditional forecasts of the pseudo individuals' earnings. One is an h-step-ahead forecast conditional on a business cycle shock; the other is conditioned on the shock being set to zero. We produce these conditional forecasts using the methods of Chan et al. (2025). These are conditional forecasts of earnings for every pseudo individual that can be used to produce estimates of the associated earnings distribution. The difference between the two distributions is the impulse response at horizon h. This type of IRF can be used to address questions such as: "What is the effect of the business cycle shock on the 10th percentile of the earnings distribution h-steps-ahead?"

2.3 Choosing the Pseudo Individuals

Pseudo individuals are defined according to K discrete variables, $\boldsymbol{Z}$, known to the econometrician. These variables capture observed socioeconomic and demographic characteristics elicited by the micro survey. As discussed below, in many applications, including ours, data limitations constrain the possibilities for $\boldsymbol{Z}$. That is, micro surveys typically collect only a subset of characteristics of survey respondents. Even so, many surveys provide information on respondents' gender, age, and education, for example. Given that each of these K characteristics can, in turn, typically be broken down into a range of sub-characteristics, such as age brackets, there end up being many ways in which the researcher might define pseudo individuals. This can lead to a large N^*.

At a high level, there is a trade-off when choosing N^*. When N^* is large, actual individuals better match each pseudo individual. But each pseudo individual is based on fewer observations, increasing estimation error as sample group averages deviate from population group averages.

On the other hand, when N^* is small, each pseudo individual's earnings are better estimated, but the estimation may not match up so well with that for the actual individuals. That is, if the group of individuals making up a given pseudo individual is too large, it may be too heterogeneous to be reasonably represented by a single pseudo individual. These considerations hold in any empirical application involving pseudo panels. With a pseudo VAR we have the additional issue that the choice of N^* determines the VAR dimension and, hence, the computational complexity of estimating the model.

In our empirical work, the limited data restrict us and we simply make standard, economically sensible choices involving age, education, region, and gender. But it is worth noting that one may employ other statistical methods for choosing the pseudo individuals. First, one could adopt the viewpoint that the individuals making up each pseudo individual should be as similar as possible in terms of their earnings (that is, there should be as little heterogeneity within a pseudo individual as possible). There are numerous clustering algorithms that could be used to choose pseudo individuals so as to minimize heterogeneity. An issue with using a statistical metric such as this to define pseudo individuals is that the resulting definitions could be counter-intuitive and lack economic intuition.

A second way to proceed would be to start by estimating a pseudo VAR with as large a value of N^* as feasible and then see if homogeneity restrictions can be imposed. For instance, if the equations for pseudo individuals i and j are found to have the same coefficients (and thus the same IRFs), then pseudo individuals i and j can be merged together into one pseudo individual and the pseudo VAR dimension reduced by one. This general insight could be implemented in several ways, for example, via use of variable selection priors or sequential testing methods. We leave such an implementation to future research.

Empirically, for a given N^*, one can – and we will in our empirical application below – assess the fit of the pseudo VAR model-implied micro density against the observed micro data histogram. Such a comparison can indicate whether N^* is too small. For example, imagine that the pseudo VAR is estimated in just $N^* = 2$ pseudo individuals, say, men and women. If there is considerable within-gender variation, then the micro density constructed from averages for men and women will be unable to generate enough heterogeneity to match the actual micro data histogram. This would serve as evidence that a higher value of $N^* = 2$ should be used.

3 The Micro and Macro Data

To demonstrate the utility of the proposed pseudo VAR model, and facilitate direct comparison with the functional VAR model, we follow Chang et al. (2024) and study the joint dynamics of a set of macro variables and micro data on earnings. To do so we focus on the same sample period, from 1989Q1 through 2017Q3, as Chang et al. (2024).

The micro data on weekly earnings[9] come from the CPS which, like Chang et al. (2024), we access via the National Bureau of Economic Research MORG.[10] The CPS provides a nationally representative sample of the US workforce, surveying around 50–60,000 households per month. The CPS has been subject to increasing non–response over time, particularly since the COVID-19 pandemic, but it remains a robust and broadly representative source for analyzing income dynamics in the US. Its stratified sample design and benchmarking to population controls help anchor estimates against major representativeness concerns. Recent research incorporating administrative data into survey weights shows that the effect of non–response bias on headline measures, such as participation and unemployment rates, is modest and typically within the margin of sampling errors (Eggleston et al., 2024). As discussed above, the CPS earnings data are top-coded. The percentage of top-coded individuals varies over time, but is never more than 5 percent in a given quarter.

We follow Chang et al. (2024) in treating the CPS as a repeated cross-section. This is because the CPS's panel component, namely, the fact that the earnings question is only asked of a given household twice, is too small to exploit in our context. We aggregate the CPS data across the three months in each calendar quarter. Weekly earnings are scaled to annual earnings by multiplying by 52. As in Chang et al. (2024), we construct the macro measure of the unemployment rate directly from the CPS data. This is one minus the

[9]We use the CPS variable on weekly earnings, which reports earnings per week before deductions and includes any overtime, commissions, or tips that are usually received.

[10]https://www.nber.org/research/data/current-population-survey-cps-merged-outgoing-rotation-group-earnings-data.

fraction of individuals with non-zero weekly earnings.[11] To remove the common trend in the cross-sectional earnings data, we also follow the recommendation of Chang et al. (2024), and each quarter we standardize individual-level earnings by (2/3) of nominal per capita GDP. The 2/3 figure is aimed at approximating the labor share of GDP. Finally, we take an inverse hyperbolic sine transformation of the earnings data to mitigate the skewness of the earnings/GDP distribution. The earnings distribution has been shown to be highly non–normal (see Guvenen et al. (2021)). But we reverse this transformation post-estimation of the VAR model before examining the IRFs. Notwithstanding this transformation, the selling point of the functional and pseudo VAR models that we use is that they do not involve having to make a parametric assumption about the overall earnings distribution: The data are left to decide its shape.

A key question is how to define the pseudo individuals. As discussed above, there is a trade–off between wanting individual survey respondents comprising each pseudo individual to be homogeneous (minimizing within-pseudo-individual variation) and wanting our pseudo individuals to be heterogeneous. At the same time, there are computational implications of having more pseudo individuals. We proceed by defining our pseudo individuals across four dimensions that capture commonly understood factors explaining differences in earnings, namely, sex, education, age, and location. These, along with race, represent common characteristics by which earnings in the US are often broken down.[12]

Based on these characteristics, as summarized in Table 1, we define 400 pseudo individuals: 2 sexes $\times$ 2 education levels $\times$ 10 age groups $\times$ 10 regions. Individuals' education levels are classified into two categories: those with a college education or higher and those with a lower level of education.[13] Age is defined with respect to the age of the respondent in the survey, in contrast to pseudo panel approaches where pseudo individuals are defined on the basis of their year of birth. This fixed-age-group type approach has been used elsewhere, notably in studies in population economics where the emphasis is on the effect of different factors on the fertility of specific age groups (see, for example, Chatterjee and Vogl (2018)). Given our focus on assessing the effect of shocks on individuals with specific characteristics, rather than following a group of pseudo individuals through time, we split our sample into pseudo individuals based

[11]The employed include both those individuals "working" and those "with a job not at work." Our focus is modeling the sample with the unemployed, and hence, our earnings data have what Chang et al. (2024) call a point mass at zero.

[12]See, for example, Table A-6 from the US Census Bureau: https://www.census.gov/library/publications/2024/demo/p60-282.html?utm_campaign=20240911acos4&utm_medium=email&utm_source=govdelivery, and breakdowns provided by the Bureau of Labor Statistics: https://www.bls.gov/cps/earnings.htm.

[13]Definitional changes in the CPS, as discussed by Jaeger (2003), affect the consistency of this educational characteristic over time. See the Data Appendix for more details. For robustness, we also experimented with a shorter sample period, ending in 2014, that is less affected by these definitional changes. We find results similar to those reported below.

on their age at the time of each survey. We consider 10 age groupings. Region is defined using the Office of Management and Budget's definition of the 10 "regions" of the US.[14] On average over time and across the 400 pseudo individuals, this results in pseudo individuals comprising 105 actual individuals.

Table 1: The Micro Data

Characteristic	Number	Definition
Region	10	Agency administrative regions (Office location: Boston, New York City, Philadelphia, Atlanta, Chicago, Dallas, Kansas City, Denver, San Francisco, Seattle)
Sex	2	Male, Female
Age	10	≤ 25, $(25,30)$, $(30,35)$, $(35,40)$, $(40,45)$, $(45,50)$, $(50,55)$, $(55,60)$, $(60,65)$, > 65
Education	2	Below college, College and above

Our macro data set uses the same three macroeconomic variables as in Chang et al. (2024) and we use their data transformations too. This means that we model total factor productivity (TFP) growth, GDP growth, and the unemployment rate.[15] The pseudo VAR model then augments these three variables with the 400 pseudo individual observations. This is a large VAR, but manageable using the methods we set out below.

4 Empirical Application Revisiting the Distributional Effects of Business Cycle Shocks

4.1 Implementation Details for the Pseudo VAR

Large VARs can be heavily over–parameterized and require prior shrinkage and/or the imposition of restrictions on the VAR coefficients or the error covariance matrix. Chang et al. (2024) use the asymmetric conjugate prior of Chan (2022) and we adopt a similar approach but adapt the prior to acknowledge that $\mathbf{B}_0$ is not assumed to be lower triangular. We want to impose different restrictions, as explained below. But the theoretical properties derived in Chan (2022), such as order invariance, still hold and the posterior simulation algorithm retains the

[14] https://www.gao.gov/assets/ggd-75-52.pdf. These definitions are provided in Appendix Table A.1, alongside the city in which the head office is based, which is how we name these regions.

[15] Full definitions, along with the data transformation and FRED code used, for these macro variables are listed in Table A.2 in the Data Appendix. We also follow Chang et al. (2024) and when plotting the macro IRFs we do so for the employment, rather than the unemployment, rate.

same form. This prior is a version of the familiar Minnesota prior. An important advantage of this prior is that the conjugacy means that analytical posterior results are available, thus reducing the computational burden. In addition, the marginal likelihood is available in closed form. Chang et al. (2024) use this to estimate the prior hyperparameters in their functional VAR (as do we in our pseudo VAR and the functional VAR we compare it to, thus ensuring that parameters that are common to both models are the same). Prediction is done iteratively using simulation methods. Posterior simulation methods are used to calculate impulse responses.

In experimentation, we have found it practical to estimate this unrestricted VAR even when N^* runs to several hundred. However, estimating VARs of this dimension poses challenges in terms of the computational burden, identifying structural shocks and ensuring precision and stability of empirical results (for example, ensuring that the posterior does not produce any explosive draws can be difficult to impose and computationally burdensome in large VARs). Accordingly, the researcher may wish to impose restrictions. These restrictions could involve either contemporaneous effects (that is, restrictions on the impact matrix $\boldsymbol{B}_0$) or lagged effects (that is, restrictions on $\boldsymbol{A}_j$ for $j = 1 \ldots p$). And they can involve either relationships between the micro and macro variables or relationships between different micro variables.

The choice of restrictions will be application dependent. In our application, to achieve parsimony in an otherwise large VAR, we impose four sets of restrictions. Importantly, these restrictions still allow for rich micro-macro dynamics. We first assume $\mathbf{B}_{0,12} = \mathbf{0}$, so that the macro variables affect the micro variables contemporaneously but not the other way round. Second, we assume that $\mathbf{B}_{0,22}$ is a diagonal matrix, so that the micro units are contemporaneously unrelated. We stress that these restrictions are imposed on the structural VAR, namely, the VAR specified via $\boldsymbol{B}_0$. The corresponding reduced-form VAR error covariance matrix, $\boldsymbol{\Sigma}$, remains full. Intuitively speaking, there will still be non-zero error covariances between different pseudo individuals, but this dependence is driven solely by the (common) macro variables. Third, we assume $\mathbf{A}_{1,12}, ..., \mathbf{A}_{p,12} = \mathbf{0}$, so that lags of the micro variables do not affect the macro variables. Fourth, we assume $\mathbf{A}_{1,22}, ..., \mathbf{A}_{p,22}$ are diagonal matrices, so that the micro units are only affected by lagged values of their own micro unit (as well as lags of the macro variables).

In a standard macro VAR, the identified structural shock will only reflect the macro variables. In the unrestricted version of our pseudo VAR, the micro variables can effect the macro variables and, thus, the identified shock. This means that the structural shock will not be the same as the one produced by a macro VAR. If the researcher finds this undesirable, then additional restrictions can be imposed on our pseudo VAR to identify a structural shock that is the same as that in the macro VAR. However, one might argue that in most applications it is desirable to allow for the micro variables to have an impact on the macro shock; see, for example, Lenza and Savoia (2024). With our particular choice of restrictions, we allow for micro-macro linkages contemporaneously (that is, via non-zero elements in $\mathbf{B}_{0,21}$) and this will have a dynamic effect (that is, $\mathbf{B}_{0,21}$ will appear in the calculation of IRFs at all horizons). In many applications, this degree of flexibility might suffice, particularly since structural shocks are often identified through their impact effects. But if this is too restrictive, micro-macro linkages could be made more flexible dynamically by relaxing the assumption that $\mathbf{A}_{1,12}, ..., \mathbf{A}_{p,12} = \mathbf{0}$. But, given the high dimensionality of these matrices, this would lead to a substantial loss in parsimony. A more parsimonious restriction that does allow more flexibility in terms of lagged micro-macro interactions might involve, for instance, restrictions that imply that lags of average earnings (over all pseudo individuals) appear on the right-hand side of the macro equations. Alternatively, Baumeister et al. (2025) enter lagged quantiles of the micro distribution on the right-hand side of the macro equations as a way of allowing distributional information about the micro data to impact the macro shocks.

Our prior is similar to the asymmetric conjugate prior of Chan (2022), but minor adaptations are required due to the restricted nature of our VAR. The prior is inverse Gamma for the variances of each of the VAR errors. The asymmetric conjugate prior chooses the prior mean of the error variance in equation i to be s_i^2, which is the sample variance of the residuals of an AR(4) model for variable i. We also make this choice for the prior mean. The asymmetric conjugate prior chooses a prior degrees of freedom starting from an inverse-Wishart prior for the reduced-form error covariance matrix. This is chosen to be $N+2$, which is necessary in the unrestricted VAR to ensure that the prior mean exists. This implies very strong prior shrinkage when N is large. With the restrictions imposed on our VAR, it is not possible to use an inverse-Wishart prior for the reduced-form error covariance matrix. In addition, the number of right-hand-side variables in our restricted VAR is always fairly small and does not increase with N. Accordingly, we use a much less informative prior for our error variances and set the prior degrees of freedom to k_i, which is the number of right-hand-side variables in equation i. In our model, k_i will be small and will take on only two values: one for the macro block and one for the micro block.

For the coefficients, since we estimate the VAR in structural form, there are two sets of parameters: the impact coefficients and the VAR coefficients. The priors are conjugate and constructed following Chan (2022). Specifically, the priors for the impact coefficients are normal, with hyperparameters set equal to those in Chan (2022). We set the VAR lag length to $p = 2$, and impose a Minnesota prior on the VAR coefficients. The hyperparameters are chosen to ensure consistency across all model specifications. Further details are provided in Appendix C. When estimating the functional VAR model of Chang et al. (2024), we adopt the same priors described in Appendix C, focus on their specification with 10 knots for the spline approximation, and follow Chang et al. (2024) by censoring the likelihood function of the functional VAR to acknowledge that the micro earnings data are top-coded.

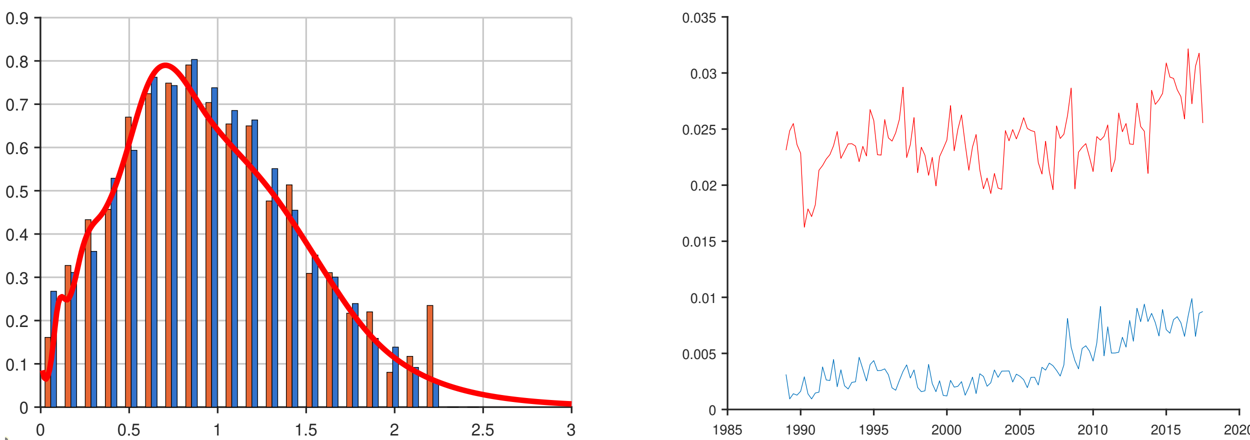
4.1.1 Micro Density Estimation

The functional and pseudo VAR models offer different ways of fitting the micro/cross-sectional density. We can assess their ability to reproduce the time path of the observed micro earnings/GDP data by comparing their cross-sectional fit, over time, against actual earnings. Panel (a) of Figure 1 provides a snapshot for 2009Q1, indicating that both the functional and the pseudo VAR models capture well the micro data. In panel (b), we plot the Cramér–von Mises distance between the actual earnings data and the two model-based fitted densities over time. We again see that both models provide a good fit, with low Cramér–von Mises values. But the pseudo VAR model consistently offers better cross-sectional fit of the micro data than the functional VAR.

Figure 1: Cross-sectional density fit of the functional and pseudo VAR models

(a) Orange: Actual micro data (2009Q1). Blue: pseudo VAR. Red: functional VAR.

(b) Cramér–von Mises distance relative to actual micro data. Blue: pseudo VAR. Red: functional VAR.



Notes: Actual earnings/GDP cross-sectional distribution, alongside fitted densities from the functional VAR (in red) and pseudo VAR (in blue) models.

4.2 Overview of Empirical Results

To illustrate the insights that the pseudo VAR can provide, we structure the next three sections around different sets of results. The first compares the IRFs we obtain for the macro variables in our model to those of Chang et al. (2024). The second presents distributional IRFs showing how the business cycle shock affects the earnings distribution, again comparing the pseudo VAR against the functional VAR of Chang et al. (2024). The final set of results demonstrate the additionality that our model provides in allowing us to directly “look within” the earnings density to ascertain who is most affected by macro shocks. This is harder to do with a functional VAR. The best a functional VAR can do is produce distributional IRFs for a specific pseudo individual (or group of pseudo individuals) by estimating the functional VAR on a subset of groups. This risks losing interdependencies between the pseudo individuals.

4.3 IRFs for the Macro Variables

We begin by presenting the IRFs for the macro variables. As well as letting us assess whether movements in the earnings distribution affect how macro shocks impact aggregate variables, this exercise lets us compare how our pseudo VAR fits the macro data relative to the functional VAR of Chang et al. (2024). The IRFs for the macro variables are produced in response to a macro shock to TFP. Following Chang et al. (2024), this is identified via Cholesky orthogonalization, with TFP growth ordered first, GDP growth second, and the unemployment rate third, and we order the macro ahead of the micro variables. In online Appendix B, we show that we obtain results similar to those presented in the three sections below when, rather than via orthogonalization, we identify the macro shock using the maximum-share-of-variance approach. Echoing Angeletos et al. (2020), this is defined to be the shock that maximizes the variability of the unemployment rate over business cycle frequencies (6–32 quarters) and can be interpreted as a “main business cycle” shock.[16] Following Chang et al. (2024), all of our IRFs measure the effect of a three-standard-deviation shock, signed so that positive values of the macro shock are associated with expansionary phases of the business cycle.

We compare the IRFs produced by our pseudo VAR with two benchmark models: first, the functional VAR of Chang et al. (2024), using the same choice of knots for the spline approximation, and, second, the aggregate VAR model estimated using only the three macro variables. The functional VAR model is estimated using the Bayesian methods described in Chang et al. (2024). We use the same data set, data transformations, and prior (for all common parameters) in these models. All three VARs have a lag length of 2, a choice also made in Chang et al. (2024).

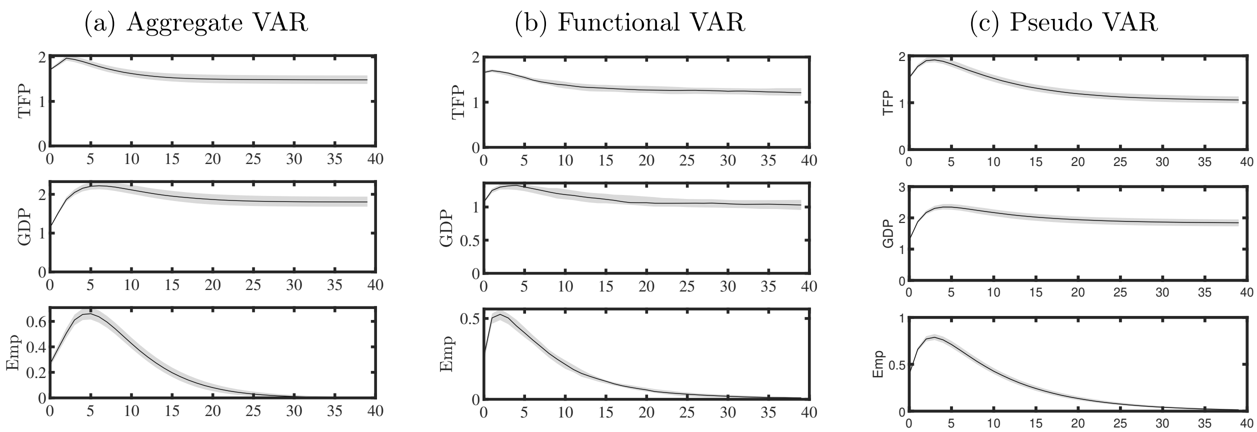
Figure 2 compares the IRFs from these different models. At a high level, qualitatively, all three VAR models deliver similar IRFs. TFP, employment, and output all move together over business cycle frequencies, with the effects of the macro shock on GDP and employment peaking around a year after impact. The TFP shock raises the level of TFP and GDP

[16] We emphasize that, consistent with the notion of a main business cycle shock, we obtain similar IRFs when, rather than maximizing the variability of the unemployment rate, we maximize the variability of GDP growth instead.

permanently across the three models, and creates a temporary employment boom.

In short, both the functional and pseudo VAR models support the view that the rich are, in effect, scaled versions of the poor. If one posits that, although the rich have more wealth, their consumption and saving behavior is similar to that of the poor, one should not expect to see the dynamics of the macro aggregates affected by the micro data, as we find is the case empirically. Our results support the view that the representative agent assumption holds, as seen in the "perfect aggregation" environment of the Krusell and Smith (1998) heterogeneous-agent model. Like the functional VAR of Chang et al. (2024), the pseudo VAR suggests that micro heterogeneity does not matter for (macro) business cycle fluctuations.

Figure 2: IRFs of the macro variables to the aggregate shock



Notes: Responses of macro variables at 1 through 40 quarters to a 3-SD TFP shock at time 0, orthogonalized via Cholesky factorization. Panels depict responses of the log level of TFP and GDP, scaled by 100, and responses of the employment rate (Emp) in percentages. Shaded areas represent 68 percent credible intervals.

4.4 Distributional IRFs

The second set of results is the distributional IRFs showing how the business cycle shock affects the entire earnings distribution. Again, we compare the pseudo VAR against the functional VAR of Chang et al. (2024). We use heat maps to illustrate distributional IRFs, showing how an aggregate shock affects different percentiles of the earnings-to-GDP distribution.

The distributional impulse responses can be interpreted as generalized IRFs; see Koop et al. (1996). Specifically, for each draw from the fitted normal distribution, we obtain a set of pseudo individuals, denoted by $\mathbf{x}_t^*$, and construct the VAR model. For each MCMC draw from the VAR, we compute two conditional forecasts of their future earnings: one conditional on the realization of the shock of interest, and one conditional on the same shock being set to zero.

Under each conditioning scenario, we directly compute the forecasted earnings of the pseudo individuals, who are weighted proportionally to the number of actual individuals that they represent.[17] We then pool the forecasted earnings of all pseudo individuals to construct the implied cross-sectional distribution at each horizon h. The impulse response at horizon h is defined as the difference between the corresponding percentile of these two conditional distributions. For example, the response of the 10^{th} percentile is computed as the difference between the 10^{th} percentiles of the two forecasted earnings distributions.

This conditional forecast procedure allows us to characterize the dynamic effects of shocks on the entire earnings distribution. In our implementation, we perform 2,000 *simulations* of the conditional forecast procedure.[18] For each simulation, impulse responses are computed from 10,000 MCMC draws. We then pool the 10,000 responses across the 2,000 simulations, and report the median response as well as the 16^{th} and 84^{th} percentiles to form credible intervals. If the 16^{th}–84^{th} percentile interval contains zero, the median response is set to zero. To highlight statistically meaningful effects and improve readability, statistically insignificant values are colored white, while significant positive and negative responses are colored green and red, respectively.

Figure 3 shows the distributional IRFs for both the functional and pseudo VARs.[19] The two VARs can again be seen to offer broadly similar distributional IRFs, in the sense of having positive values (green) at lower percentiles of the earnings distribution and negative values (red) at higher percentiles. That is, both models imply that in response to a positive macro shock, the poor get richer and the rich get poorer, relative to GDP. This empirical fact is consistent with the wider literature that finds that the distribution of wages tightens as the economy strengthens (see Okun (1973), Katz and Krueger (1999), and Aaronson et al. (2019)).[20] We emphasize that these effects on earnings are relative to GDP. GDP is itself growing after the macro shock, as already seen in Figure 2.

The distributional effects of the macro shock are strongest in the immediate aftermath of the macro shock and slowly die out thereafter, with effects becoming statistical insignificant after about 8 or 9 years. We observe in both models somewhat stronger effects above median

[17] Since the number is different over time, we use the end of sample, 2017Q3, value.

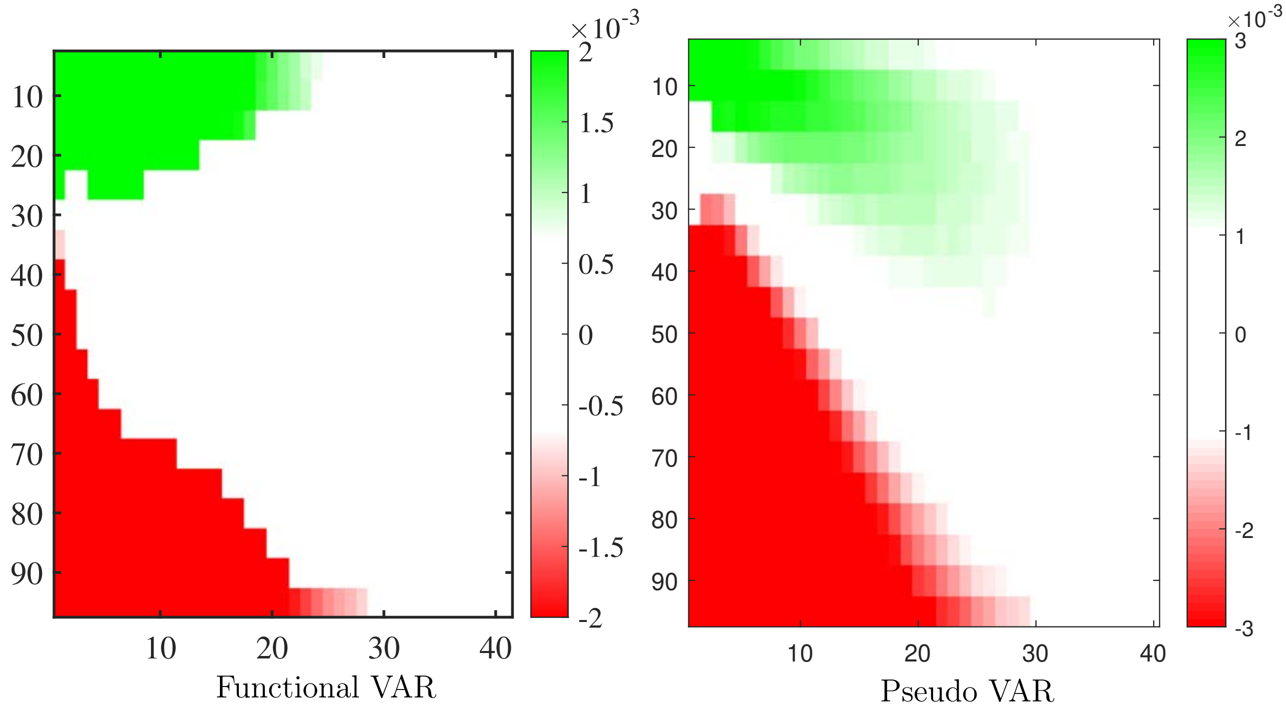
[18] Each *simulation* consists of randomly drawing from the set of pseudo densities fitted to $\mathbf{x}_t^*$, estimating the VAR model, and computing the conditional forecasts using 10,000 MCMC draws of the VAR parameters.

[19] In their paper, Chang et al. (2024) did not flesh out the impact of the business cycle shock across all percentiles of the earnings distribution. They focus on two specific percentiles, one in the upper tail and one in the lower tail.

[20] Declining inequality after a positive aggregate shock is also consistent with studies that have looked at the distributional effects of monetary policy shocks. Amberg et al. (2022), for example, find that the labor-income effects of expansionary monetary policy shocks are strongest for lower-income individuals.

earnings. But, supporting the findings in Heathcote et al. (2023), we also observe strong cyclical dynamics of income below the median too.

Figure 3: Distributional IRFs to the aggregate shock



Notes: Micro distributional responses at 1 through 40 quarters, across percentiles of the earnings/GDP distribution, to a 3-SD TFP shock, orthogonalized via Cholesky factorization, at time 0. Green areas denote statistically significant positive percentiles, red areas denote statistically negative percentiles, and white areas denote statistically insignificant percentiles as judged by the 16th - 84th credible interval of the percentile containing zero.

4.5 IRFs for Pseudo Individuals: Group and Distributional

Our final set of results demonstrates the additionality that our model provides in enabling us to "look within" the earnings density to ascertain who is most affected by macro shocks. The pseudo VAR can thereby provide stylized facts about how certain types of individuals react to macro shocks. Such an analysis is not possible with the functional VAR. We contrast, for each of the four demographic factors, the IRFs obtained by grouping together, by characteristic, the IRFs across the 400 pseudo individuals with those obtained when we construct distributional IRFs for the same group.

To illustrate the additional insights that the pseudo VAR enables, relative to functional VAR models, we contrast the group-level and distributional IRFs for each demographic group in turn.

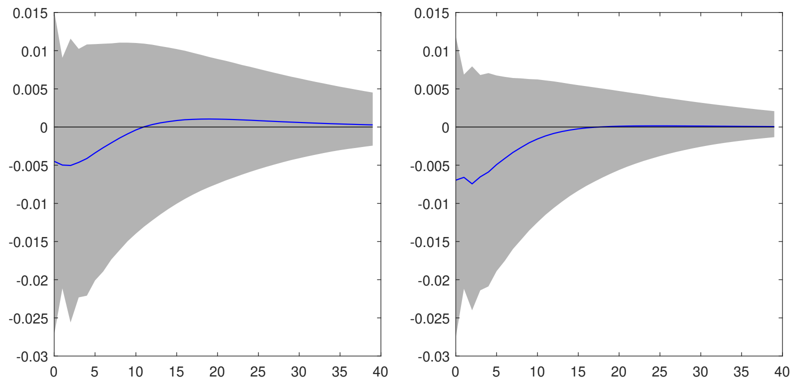
4.5.1 Men versus Women

We start by looking at Figure 4, which shows the effects of the macro shock on the earnings of men and women. For both men and women we see that, relative to GDP, earnings fall on average for about three years after the macro shock. This is consistent with the labor share being countercyclical. Figure 4 evidences that the effects on the "average" man are slightly more positive than the effects on the "average" woman. This is, in turn, consistent with the stylized fact that the labor earnings of men are more exposed to macro fluctuations than those of women (Guvenen et al. (2017)). But it is only when we turn from Figure 4 to Figure 5 to "look within" the earnings distribution that we gain a fuller picture. From Figure 5 we learn that the effects on men at the lower end of the earnings distribution are stronger than the effects on women. We also see that, over time, the positive effects of the macro shock seen at lower percentiles percolate up the earnings/GDP distribution. After three or four years, we see that all men with earnings in the bottom third of the earnings distribution gain. By contrast, the positive effects on women are both much more contained, over time and across the earnings distribution, and weaker – as evidenced by the middle panel in Figure 5.

Our findings in Figure 5 therefore support the literature documenting that (pre-COVID) recessions are "mancessions," in which men's earnings, certainly those who earn less than median earnings, are more cyclically sensitive than women's. This is because men are typically more exposed to sectors, such as construction and manufacturing, that employ more men; see Albanesi and Şahin (2018) and Alon et al. (2022).

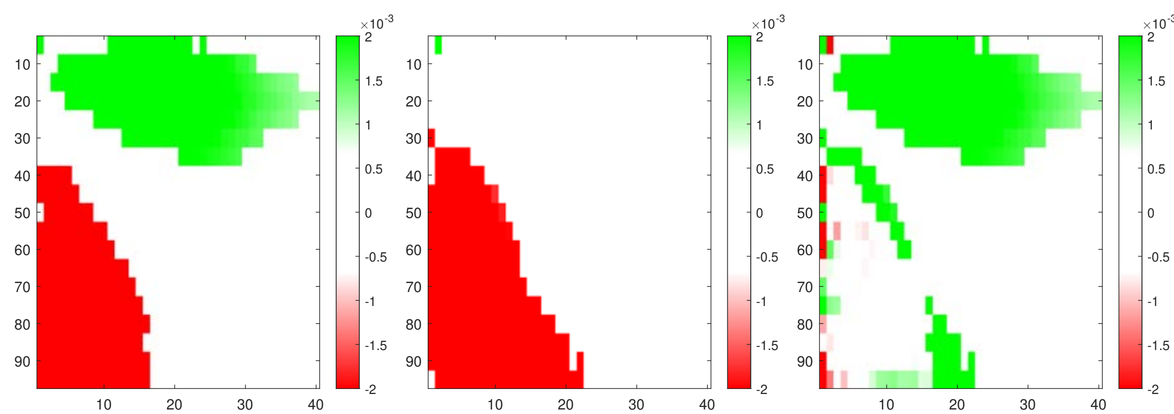
Figure 4: IRFs for men and women

(a) Men (b) Women



Notes: IRFs for men and women due to a 3-SD TFP shock, orthogonalized via Cholesky factorization. We perform 2,000 simulations of the conditional forecast procedure. For each pseudo individual draw, impulse responses are computed for men and women across all MCMC draws taking a weighted average of the pseudo individual IRFs, where the weights are proportional to the number of actual individuals represented (based on 2017Q3). We then pool, by gender, these weighted responses across the 10,000 MCMC draws and the 2,000 simulations and report the median and the 16^{th} and 84^{th} percentiles to form credible intervals.

Figure 5: Distributional IRFs for men and women



Men Women Men minus women

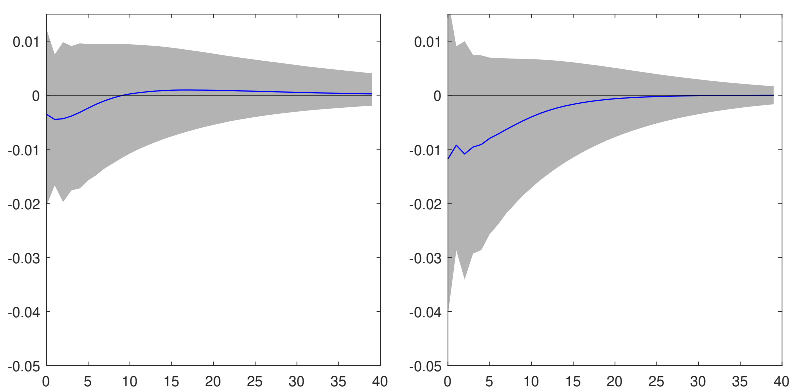
Notes: Micro distributional responses on men and women, across percentiles of the earnings/GDP distribution, to a 3-SD TFP shock, orthogonalized via Cholesky factorization. Green areas denote statistically significant positive percentiles, red areas denote statistically negative percentiles, and white areas denote statistically insignificant percentiles as judged by the 16th - 84th credible interval of the percentile containing zero. Right-hand-side panel tests whether the observed differences between men and women are statistically significant.

4.5.2 Education

We next turn to examining if and how the effects of business cycle shocks on earnings vary by education. In Figure 6, we see that after falling in relative terms for about two years, the macro shock raises earnings for those without a college education. However, the effects are more negative for those with a college education. But, again, looking at these average effects is masking strong differentials across the earnings distribution. Figure 7 reveals that the macro shock has strong positive effects on lower-earning individuals without a college education. The bottom quintile of earnings for those without a college education, in particular, is boosted in the aftermath of the business cycle shock. In contrast, only those college-educated individuals in the bottom decile of the earnings distribution are not negatively (in relative terms) impacted by the business cycle shock.

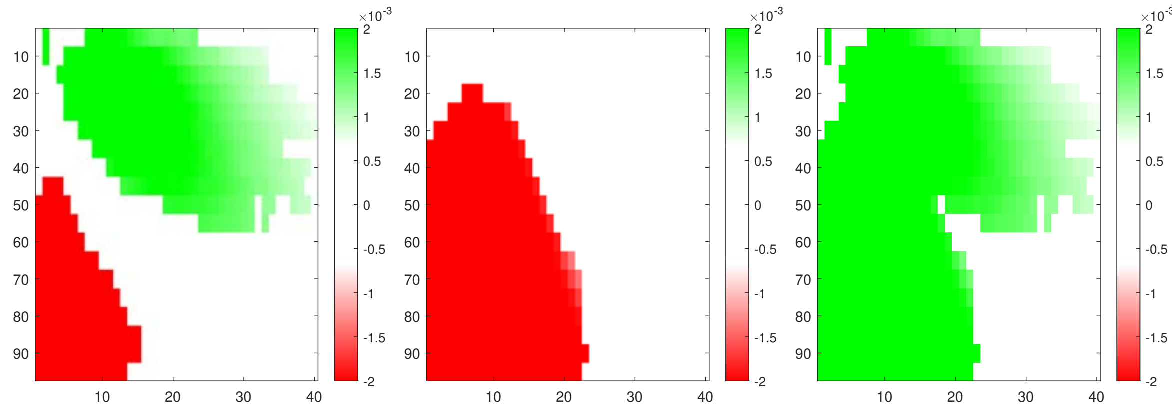
Figure 6: IRFs by education

(a) Below college (b) College and above



Notes: IRFs due to a 3-SD TFP shock, orthogonalized via Cholesky factorization. Shaded areas represent 68 percent credible intervals. IRFs involve taking the weighted average of the IRFs for those pseudo individuals who have and do not have a college education.

Figure 7: Distributional IRFs by education



Notes: Micro distributional responses by education, across percentiles of the earnings/GDP distribution, to a 3-SD TFP shock, orthogonalized via Cholesky factorization. Green areas denote statistically significant positive percentiles, red areas denote statistically negative percentiles, and white areas denote statistically insignificant percentiles as judged by the 16th - 84th credible interval of the percentile containing zero. Right-hand-side panel tests whether the observed differences between the college-educated and the non-college-educated are statistically significant.

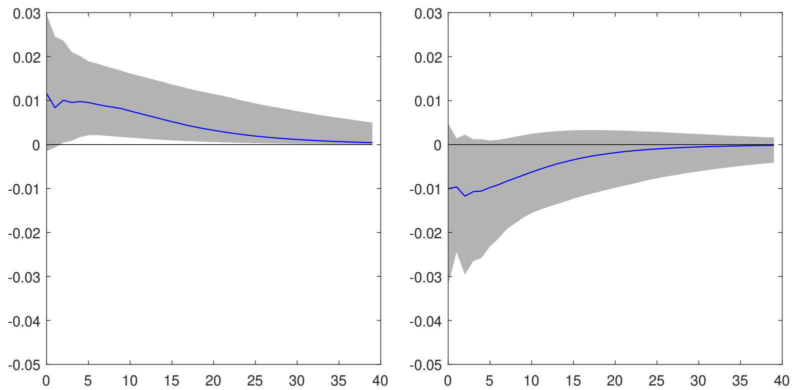
4.5.3 Age

We now turn to examining the effects on earnings of the business cycle shock by age. Figure 8 focuses on those aged 25 years or younger and those aged 50 through 55. Even looking at the age-specific IRFs, we see clear differences. The young are positively and persistently affected, whereas the effects on the older group are negative for four to five years after the macro shock. HANK models, such as those developed by Bardoczy and Velasquez-Giraldo (2024), also emphasize the importance of modeling heterogeneities across age in understanding the dynamic effects of macro shocks. This builds on the significant literature showing how the labor market outcomes of the young, more than other demographic groups, are particularly exposed to cyclical fluctuations (for example, see Clark and Summers (1981) and Guvenen et al. (2017)).

When, in Figure 9, we "look within" the distribution, we see that these averaged age-specific effects hold reasonably well across the earnings distribution. Both low- and high-earning young individuals benefit from the business cycle shocks. This is again consistent with the narrative about higher elasticities of labor supply for younger workers.

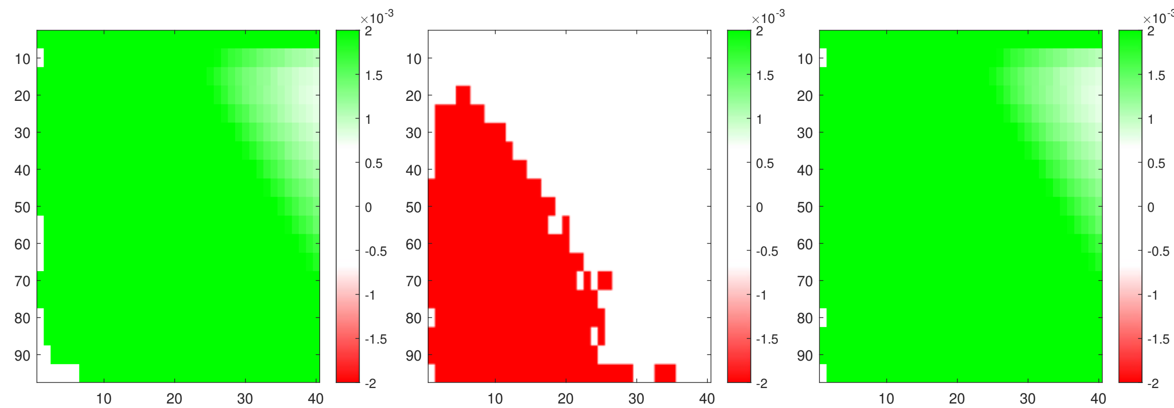
Figure 8: IRFs by age

(a) Age ≤ 25 (b) Age 50–55



Notes: IRFs due to a 3-SD TFP shock, orthogonalized via Cholesky factorization. Shaded areas represent 68 percent credible intervals. IRFs involve taking the weighted average of the IRFs for the younger and older pseudo individuals.

Figure 9: Distributional IRFs by age



Ages ≤ 25 Ages $(50, 55)$ Young minus old

Notes: Micro distributional responses by age, across percentiles of the earnings/GDP distribution, to a 3-SD TFP shock, orthogonalized via Cholesky factorization. Green areas denote statistically significant positive percentiles, red areas denote statistically negative percentiles, and white areas denote statistically insignificant percentiles as judged by the 16th - 84th credible interval of the percentile containing zero. Right-hand-side panel tests whether the observed differences between younger and older individuals are statistically significant.

4.5.4 Region

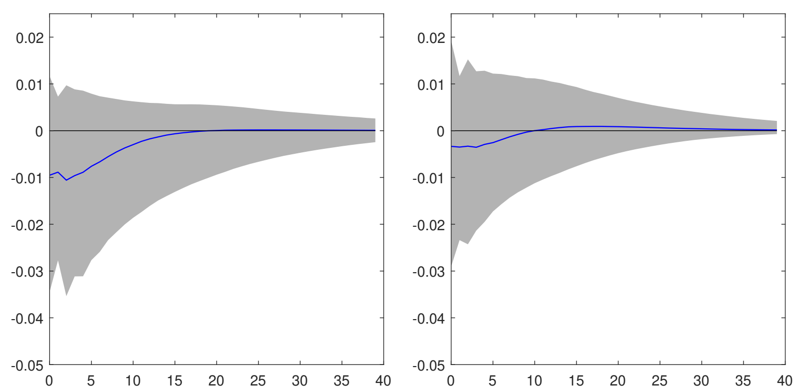
Finally, we illustrate how macro shocks have differential regional effects. We do so by focusing on two regions: Philadelphia and Seattle. We find that these two regions experience the most distinct responses, relative to the US as a whole, after the macro shock. It is important to reiterate that these regions are named after the city in which the office is located, and comprise a much larger area than these cities alone (the Philadelphia region comprises Delaware, Maryland, Pennsylvania, Virginia, Washington, D.C., and West Virginia, while the Seattle region comprises Alaska, Idaho, Oregon, and Washington). See Table A.1 for more detail.

Figure 10 starts by showing the IRFs for these two regions. It shows that the effects are more positive on Seattle than on Philadelphia. Turning to look within the earnings/GDP distribution, we see in Figure 11 that the positive effects on the Seattle region, relative to the US, of the macro shock hold across much of the earnings distribution with the exception of the lower tail. In contrast, in the Philadelphia region they are consistently more negative. This fits with the wider finding that not all business cycles are alike: There is regional heterogeneity, as noted, for example, by Owyang et al. (2005).

Figure 10: IRFs by region

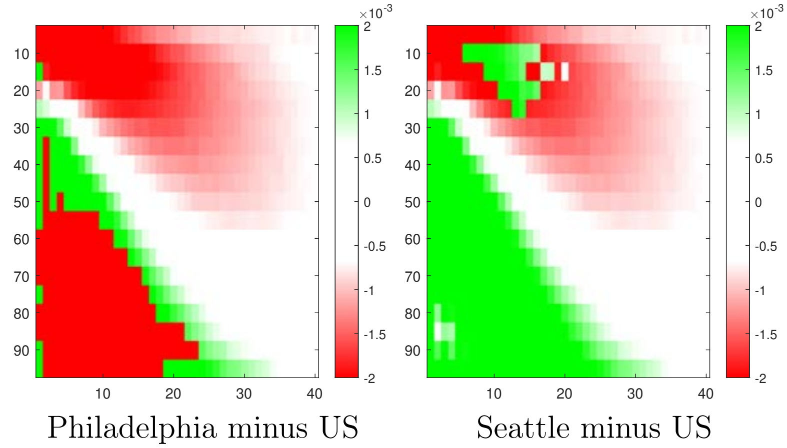
(a) Agency admin region: office location Philadelphia

(b) Agency admin: office location Seattle



Notes: IRFs due to a 3-SD TFP shock, orthogonalized via Cholesky factorization. Shaded areas represent 68 percent credible intervals. IRFs involve taking the weighted average of the IRFs for the different regions.

Figure 11: Distributional IRFs by region



Notes: Micro distributional responses by selected region, across percentiles of the earnings/GDP distribution, to a 3-SD TFP shock, orthogonalized via Cholesky factorization. Green areas denote statistically significant positive percentiles, red areas denote statistically negative percentiles, and white areas denote statistically insignificant percentiles as judged by the 16th - 84th credible interval of the percentile containing zero. Each panel tests whether the observed differences between the stated region and the US as a whole are statistically significant.

5 Conclusions

There is growing interest in the distributional effects of macroeconomic shocks. Much of the existing evidence is derived from structural macroeconomic models with heterogeneous agents; see, for example, Krusell and Smith (1998), Ahn et al. (2018), Kaplan and Violante (2018), Kaplan et al. (2018), Liu and Plagborg-Møller (2023), and Bayer et al. (2024). In contrast, the recent trend of developing "micro-macro" VAR models has provided a framework for comparing the predictions of these structural heterogeneous-agent models with observed data and providing data-based evidence on how much microeconomic heterogeneity matters for the macro business cycle and vice versa.

Contributing to this literature, this paper has developed a new econometric approach to combine macro and micro data in a VAR model, which enables new insights to be derived on the impact that macro shocks have on individuals. By jointly modeling the individual-level data alongside a selection of key macroeconomic variables in a large VAR, we gain great flexibility in modeling the static and dynamic inter-relationships between both micro and macro variables. We have developed a feasible way of implementing this approach. We turn the repeated cross-section of our micro data into a pseudo panel and include the earnings of the pseudo individuals as variables in our VAR. This turns an impractically large VAR into a still large, but feasible, one.

This approach has a number of advantages over directly modeling aggregate moments of both the micro (cross-sectional) distribution and the functional VAR models. In addition to gaining insight into the dynamic impact of the shock on different parts of the income distribution, our approach lets us "look within" the distribution to identify which types of micro units are most or least affected. We illustrate these capabilities in an empirical exercise revisiting a standard set of US macro variables and a micro data set of survey data on individual earnings.

We show that macro shocks do have business cycle effects on the micro earnings distribution. Expansionary business-cycle-type shocks raise earnings, relative to GDP, at the bottom of the earnings distribution and decrease them at the top. What our pseudo VAR approach is able to tell us, unlike a functional VAR model, is who these people are. We find the most cyclically sensitive individuals to be men in the bottom deciles of the earnings distribution, those individuals with less than a college education in the lower percentiles of the earnings distribution, and the young.

References

Aaronson, Stephanie R., Mary C. Daly, William L. Wascher, and David W. Wilcox (2019). "Okun revisited: Who benefits most from a strong economy?" *Brookings Papers on Economic Activity*, 50(1 (Spring), pp. 333–404. URL https://ideas.repec.org/a/bin/bpeajo/v50y2019i2019-01p333-404.html.

Acemoglu, Daron and David Autor (2011). "Chapter 12 - Skills, tasks and technologies: Implications for employment and earnings." In David Card and Orley Ashenfelter, editors, *Handbook of Labor Economics: Volume 4*, pp. 1043–1171. Elsevier. doi:10.1016/S0169-7218(11)02410-5.

Acemoglu, Daron and Pascual Restrepo (2022). "Tasks, automation, and the rise in U.S. wage inequality." *Econometrica*, 90(5), pp. 1973–2016. doi:10.3982/ECTA19815.

Ahn, SeHyoun, Greg Kaplan, Benjamin Moll, Thomas Winberry, and Christian Wolf (2018). "When inequality matters for macro and macro matters for inequality." *NBER Macroeconomics Annual*, 32, pp. 1–75. doi:10.1086/696046.

Albanesi, Stefania and Ayşegül Şahin (2018). "The gender unemployment gap." *Review of Economic Dynamics*, 30, pp. 47–67. doi:10.1016/j.red.2017.12.005.

Alon, Titan, Sena Coskun, Matthias Doepke, David Koll, and Michèle Tertilt (2022). "From mancession to shecession: Women's employment in regular and pandemic recessions." *NBER Macroeconomics Annual*, 36, pp. 83–151. doi:10.1086/718660.

Amberg, Niklas, Thomas Jansson, Mathias Klein, and Anna Rogantini Picco (2022). "Five facts about the distributional income effects of monetary policy shocks." *American Economic Review: Insights*, 4(3), p. 289â304. doi:10.1257/aeri.20210262.

Amir-Ahmadi, Pooyan, Christian Matthes, and Mu-chun Wang (2022). "What does monetary policy do to different people?" *manuscript*. URL https://cm1518.github.io/files/inequality_monetary_policy_web.pdf.

Andersen, Asger Lau, Niels Johannesen, Mia Jørgensen, and José-Luis Peydró (2023). "Monetary policy and inequality." *The Journal of Finance*, 78(5), pp. 2945–2989. doi:10.1111/jofi.13262.

Anderson, Emily, Atsushi Inoue, and Barbara Rossi (2016). "Heterogeneous consumers and fiscal policy shocks." *Journal of Money, Credit and Banking*, 48(8), pp. 1877–1888. doi:10.1111/jmcb.12366.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas (2020). "Business-cycle anatomy." *American Economic Review*, 110(10), pp. 3030–70. doi:10.1257/aer.20181174.

Banbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010). "Large Bayesian vector autoregressions." *Journal of Applied Econometrics*, 25(1), pp. 71–92. doi:10.1002/jae.1137.

Bardoczy, Bence and Mateo Velasquez-Giraldo (2024). "HANK comes of age." Technical report, Finance and Economics Discussion Series 2024-052. Washington: Board of Governors of the Federal Reserve System. doi:10.17016/FEDS.2024.052.

Baumeister, Christiane, Florian Huber, Pascal Frank, and Gary Koop (2025). "Oil, inflation expectations, and household characteristics: A nonlinear heterogeneous agent VAR approach." *Presentation slides*. URL https://www.ecb.europa.eu/press/conferences/shared/pdf/20250929_inflation_conference/04_Keynote_Baumeister.pdf.

Bayer, Christian, Benjamin Born, and Ralph Luetticke (2024). "Shocks, frictions, and inequality in US business cycles." *American Economic Review*, 114(5), pp. 1211–47. doi:10.1257/aer.20201875.

Bjornland, Hilde C., Yoosoon Chang, and Jamie L. Cross (2023). "Oil and the stock market revisited: A mixed functional VAR approach." CAEPR Working Papers 2023-005, Indiana University Bloomington. URL https://ideas.repec.org/p/inu/caeprp/2023005.html.

Chan, Joshua C.C. (2022). "Asymmetric conjugate priors for large Bayesian VARs." *Quantitative Economics*, 13(3), pp. 1145–1169. doi:10.3982/QE1381.

Chan, Joshua C.C., Davide Pettenuzzo, Aubrey Poon, and Dan Zhu (2025). "Conditional forecasts in large Bayesian VARs with multiple equality and inequality constraints." *Journal of Economic Dynamics and Control*, 173, p. 105061. doi:10.1016/j.jedc.2025.105061.

Chang, Minsu, Xiaohong Chen, and Frank Schorfheide (2024). "Heterogeneity and aggregate fluctuations." *Journal of Political Economy*, 132(12), pp. 4021–4067. doi:10.1086/731411.

Chari, V.V., Patrick J. Kehoe, and Ellen R. McGrattan (2008). "Are structural VARs with long-run restrictions useful in developing business cycle theory?" *Journal of Monetary Economics*, 55(8), pp. 1337–1352. doi:10.1016/j.jmoneco.2008.09.010.

Chatterjee, Shoumitro and Tom Vogl (2018). "Escaping Malthus: Economic growth and fertility change in the developing world." *American Economic Review*, 108(6), pp. 1440–1467. doi:10.1257/aer.20170748.

Clark, Kim B. and Lawrence H. Summers (1981). "Demographic differences in cyclical employment variation." *The Journal of Human Resources*, 16(1), pp. 61–79. doi:10.2307/145219.

Cloyne, James, Clodomiro Ferreira, and Paolo Surico (2020). "Monetary policy when households have debt: New evidence on the transmission mechanism." *The Review of Economic Studies*, 87(1), pp. 102–129. doi:10.1093/restud/rdy074.

Cogley, Timothy and James M. Nason (1995). "Output dynamics in real-business-cycle models." *American Economic Review*, 85(3), pp. 492–511. URL http://www.jstor.org/stable/2118184.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia (2017). "Innocent bystanders? Monetary policy and inequality." *Journal of Monetary Economics*, 88, pp. 70–89. doi:10.1016/j.jmoneco.2017.05.005.

Deaton, Angus (1985). "Panel data from time series of cross-sections." *Journal of Econometrics*, 30(1-2), pp. 109–126. doi:10.1016/0304-4076(85)90134-4.

Eggleston, Jonathan, Yarissa Gonzalez, Carl Lieberman, Tim Trudell, and John Voorheis (2024). "Incorporating administrative data in survey weights for the basic monthly Current Population Survey." Technical report, US Census Bureau Working Paper CES-24-02. URL https://www2.census.gov/library/working-papers/2024/adrm/ces/CES-WP-24-02.pdf.

Ettmeier, Stephanie, Chi Hyun Kim, and Frank Schorfheide (2024). "Measuring the effects of aggregate shocks on unit-level outcomes and their distribution." *Presentation slides*. URL https://bpb-us-w2.wpmucdn.com/web.sas.upenn.edu/dist/e/242/files/2024/10/slides_fvar_csuvar_v4.pdf.

Feenberg, Daniel R. and James M. Poterba (2000). "The income and tax share of very high-income households, 1960-1995." *American Economic Review*, 90(2), p. 264â270. doi:10.1257/aer.90.2.264.

Feldkircher, Martin, Florian Huber, Gary Koop, and Michael Pfarrhofer (2022). "Approximate Bayesian inference and forecasting in huge-dimensional multicountry VARs." *International Economic Review*, 63(4), pp. 1625–1658. doi:10.1111/iere.12577.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song (2021). "What do data on millions of U.S. workers reveal about lifecycle earnings dynamics?" *Econometrica*, 89(5), pp. 2303–2339. doi:10.3982/ECTA14603.

Guvenen, Fatih, Serdar Ozkan, and Jae Song (2014). "The nature of countercyclical income risk." *Journal of Political Economy*, 122(3), pp. 621–660. doi:10.1086/675535.

Guvenen, Fatih, Sam Schulhofer-Wohl, Jae Song, and Motohiro Yogo (2017). "Worker betas: Five facts about systematic earnings risk." *American Economic Review*, 107(5), p. 398â403. doi:10.1257/aer.p20171094.

Heathcote, Jonathan, Fabrizio Perri, Giovanni Violante, and Lichen Zhang (2023). "More unequal we stand? Inequality dynamics in the United States, 1967-2021." *Review of Economic Dynamics*, 50, pp. 235–266. doi:10.1016/j.red.2023.07.014.

Helpman, Elhanan, Oleg Itskhoki, Marc-Andreas Muendler, and Stephen J. Redding (2017). "Trade and inequality: From theory to estimation." *The Review of Economic Studies*, 84(1), pp. 357–405. doi:10.1093/restud/rdw025.

Hersch, Joni, Fernando Mendoza Lopez, and Jennifer Bennett Shinall (2020). "Estimating years of education using the Current Population Survey after 2014." *Economics Letters*, 189, p. 109,058. doi:10.1016/j.econlet.2020.109058.

Holm, Martin Blomhoff, Pascal Paul, and Andreas Tischbirek (2021). "The transmission of monetary policy under the microscope." *Journal of Political Economy*, 129(10), pp. 2861–2904. doi:10.1086/715416.

Huber, Florian, Massimiliano Marcellino, and Tommaso Tornese (2024). "The distributional effects of economic uncertainty." *arXiv*. doi:10.48550/arXiv.2411.12655.

Jaeger, David A. (2003). "Estimating the returns to education using the newest Current Population Survey education questions." *Economics Letters*, 78(3), pp. 385–394. doi:10.1016/S0165-1765(02)00259-8.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante (2018). "Monetary policy according to HANK." *American Economic Review*, 108(3), pp. 697–743. doi:10.1257/aer.20160042.

Kaplan, Greg and Giovanni L. Violante (2018). "Microeconomic heterogeneity and macroeconomic shocks." *Journal of Economic Perspectives*, 32(3), pp. 167–94. doi:10.1257/jep.32.3.167.

Katz, Lawrence F. and David H. Autor (1999). "Chapter 26 - Changes in the wage structure and earnings inequality." In Orley C. Ashenfelter and David Card, editors, *Handbook of Labor Economics: Volume 3*, pp. 1463–1555. Elsevier. doi:10.1016/S1573-4463(99)03007-2.

Katz, Lawrence F. and Alan B. Krueger (1999). "The high-pressure U.S. labor market of the 1990s." *Brookings Papers on Economic Activity*, 30(1), pp. 1–88. URL https://www.brookings.edu/wp-content/uploads/1999/01/1999a_bpea_katz.pdf.

Koop, Gary, M. Hashem Pesaran, and Simon M. Potter (1996). "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics*, 74(1), pp. 119–147. doi:10.1016/0304-4076(95)01753-4.

Krusell, Per and Anthony A. Smith, Jr. (1998). "Income and wealth heterogeneity in the macroeconomy." *Journal of Political Economy*, 106(5), pp. 867–896. doi:10.1086/250034.

Lenza, Michele and Ettore Savoia (2024). "Do we need firm data to understand macroeconomic dynamics?" Working Paper Series 438, Sveriges Riksbank (Central Bank of Sweden). URL https://www.riksbank.se/globalassets/media/rapporter/working-papers/2024/no.-438-do-we-need-firm-data-to-understand.pdf.

Liu, Laura and Mikkel Plagborg-Møller (2023). "Full-information estimation of heterogeneous agent models using macro and micro data." *Quantitative Economics*, 14(1), pp. 1–35. doi:10.3982/QE1810.

Meeks, Roland and Francesca Monti (2023). "Heterogeneous beliefs and the Phillips curve." *Journal of Monetary Economics*, 139, pp. 41–54. doi:10.1016/j.jmoneco.2023.06.003.

Mumtaz, Haroon and Angeliki Theophilopoulou (2017). "The impact of monetary policy on inequality in the UK. An empirical analysis." *European Economic Review*, 98, pp. 410–423. doi:10.1016/j.euroecorev.2017.07.008.

Ngai, L Rachel and Barbara Petrongolo (2017). "Gender gaps and the rise of the service economy." *American Economic Journal: Macroeconomics*, 9(4), pp. 1–44. doi:10.1257/mac.20150253.

Okun, Arthur M. (1973). "Upward mobility in a high-pressure economy." *Brookings Papers on Economic Activity*, 4(1), pp. 207–262. URL https://www.brookings.edu/wp-content/uploads/1973/01/1973a_bpea_okun_fellner_greenspan.pdf.

Owyang, Michael T., Jeremy Piger, and Howard J. Wall (2005). "Business cycle phases in U.S. states." *Review of Economics and Statistics*, 87(4), pp. 604–616. doi:10.1162/003465305775098198.

Parker, Jonathan A. and Annette Vissing-Jorgensen (2010). "The increase in income cyclicality of high-income households and its relation to the rise in top income shares." *Brookings*

Papers on Economic Activity, 41(2 (Fall)), pp. 1–70. URL https://www.brookings.edu/wp-content/uploads/2010/09/2010b_bpea_parker.pdf.

Sims, Christopher A. (1989). "Models and their uses." *American Journal of Agricultural Economics*, 71(2), pp. 489–494. doi:10.2307/1241619.

Online Appendix for "Incorporating Micro Data into Macro Models Using Pseudo VARs"
by Koop, McIntyre, Mitchell, and Wu

A Data Appendix

Here we provide additional information on the micro and macro data used. We start with the micro data.

Defining the educational levels in the CPS is complicated by changes in definitions. As explained in Jaeger (2003) and Hersch et al. (2020), changes introduced in 1992 make comparisons of educational attainment before and after this date problematic. Specifically, the CPS went from asking respondents to "report the highest grade that they had completed" to asking them "to report the highest degree that they had received." "The shift from a 'time spent in school' measure to focus on degrees received, which can take varying amounts of time, caused a break in the long time series on educational attainment available in the CPS." (Jaeger, 2003, 385). The NBER now provides as part of its CPS extract a new variable, constructed following Jaeger (2003), which it describes as allowing "researchers to come closer to the 'highest grade completed' measure available before 1992."[21] We use this measure to break individuals into those with below college level education and those with above.[22] As noted earlier, this approach no longer works after 2014 given changes in the underlying CPS questionnaire that removed three key questions on education necessary to implement this approach.

[21] https://cps.ipums.org/cps/resources/earner/cpsxNBER.pdf

[22] Specifically, we calculate educational levels as follows: i) 1979 – 1991 we use "gradeat" with values of 16 – 18 coded = 1; ii) 1992 – 1997 we have "grade92" >= 43 coded = 1; iii) 1998 - 2014 we use "ihigrdc" using a value of 16 – 18 = 1.

Table A.1: Region Definitions

Region	Office Location and States
Region I	Office location: Boston Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
Region II	Office location: New York City New York, New Jersey, Puerto Rico, U.S. Virgin Islands
Region III	Office location: Philadelphia Delaware, Maryland, Pennsylvania, Virginia, Washington, D.C., West Virginia
Region IV	Office location: Atlanta Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee
Region V	Office location: Chicago Illinois, Indiana, Minnesota, Michigan, Ohio, Wisconsin
Region VI	Office location: Dallas Arkansas, Louisiana, New Mexico, Oklahoma, Texas
Region VII	Office location: Kansas City Iowa, Kansas, Missouri, Nebraska
Region VIII	Office location: Denver Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming
Region IX	Office location: San Francisco Arizona, California, Hawaii, Nevada, Guam, Northern Mariana Islands, American Samoa
Region X	Office location: Seattle Alaska, Idaho, Oregon, Washington

Table A.2: The Macro Data

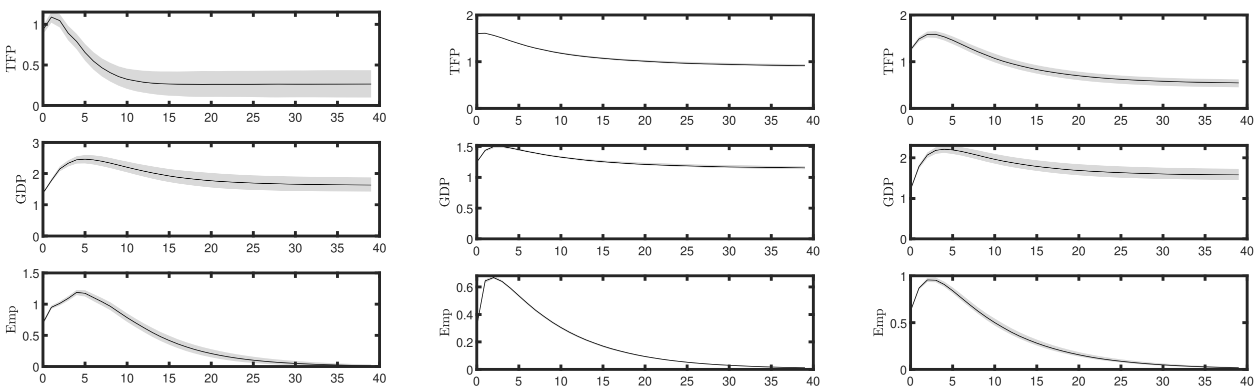
Series	Transformation/Calculation	Source
Unemployment rate	Level	CPS
Real GDP per capita	100 × log(A939RX0Q048SBEA)	https://fred.stlouisfed.org/
TFP	Cumulative sum (DTFPu/4)	https://www.johnfernald.net/TFP

B Additional Empirical Results

Here we provide additional results that identify the aggregate shock using the max-share approach. Specifically, we identify the macro shock using the maximum-share-of-variance approach. As in Angeletos et al. (2020), this is the shock that maximizes the variability of the unemployment rate over business cycle frequencies (6-32 quarters) and can be interpreted as a "main business cycle" (MBC) shock.

Figure A.1: IRFs of the macro variables to the MBC shock

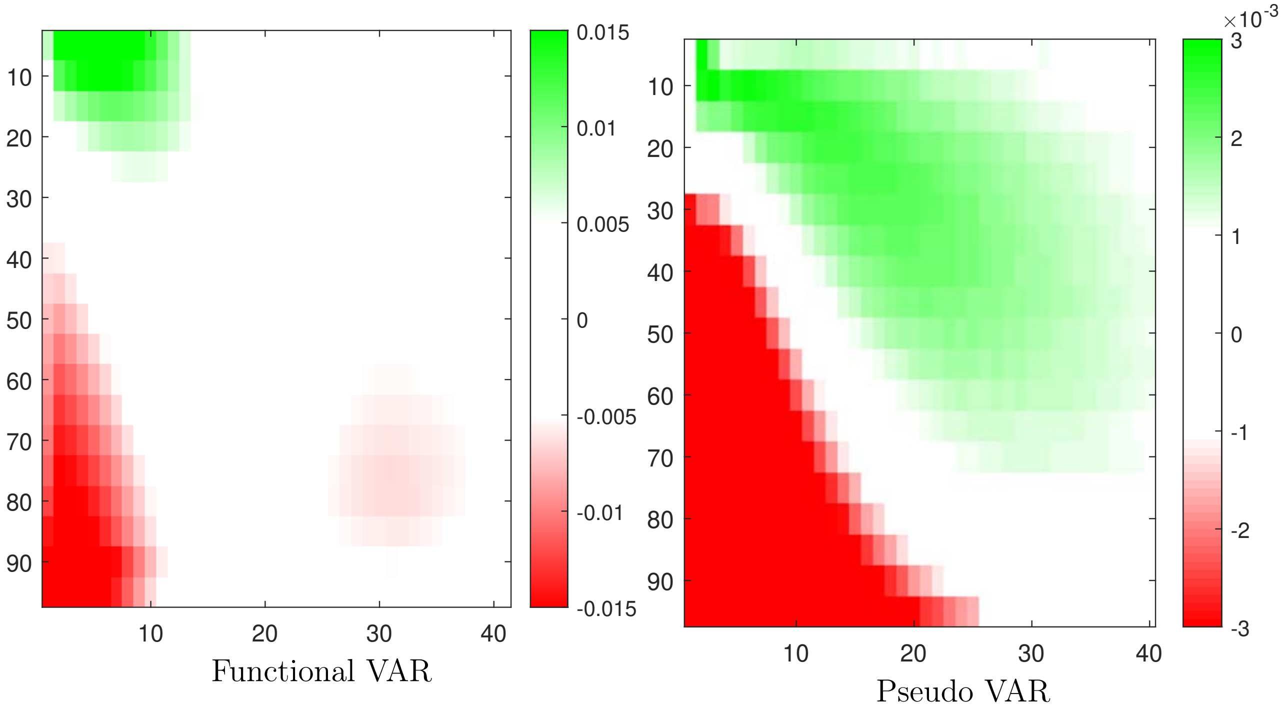
(a) Aggregate VAR (b) Functional VAR (c) Pseudo VAR



Notes: Shaded areas represent 68 percent credible intervals.

B.1 Distributional IRFs to the MBC Shock

Figure A.2: Distributional IRFs to the MBC Shock



Notes: Green areas denote statistically significant positive percentiles, red areas denote statistically negative percentiles, and white areas denote statistically insignificant percentiles as judged by the 16th - 84th credible interval of the percentile containing zero.

B.2 IRFs to the MBC Shock for Pseudo Individuals: Group and Distributional

B.2.1 Men versus women

Figure A.3: IRFs for men and women

(a) Men (b) Women

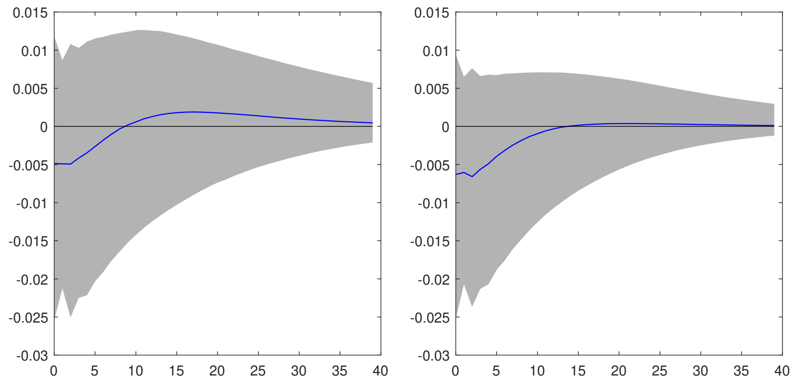
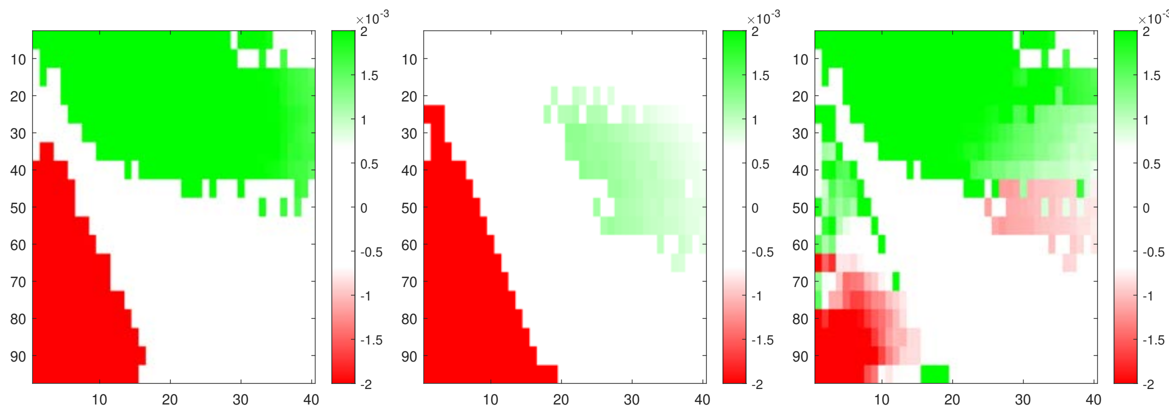


Figure A.4: Distributional IRFs for men and women



Men Women Men minus women

B.2.2 Education

Figure A.5: IRFs by education

(a) Below college (b) College and above

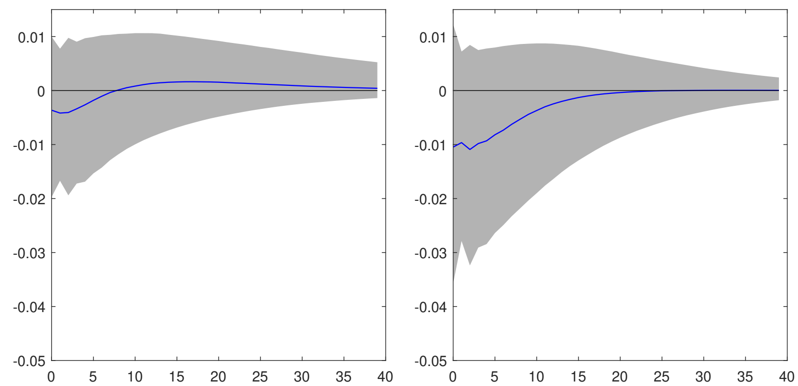
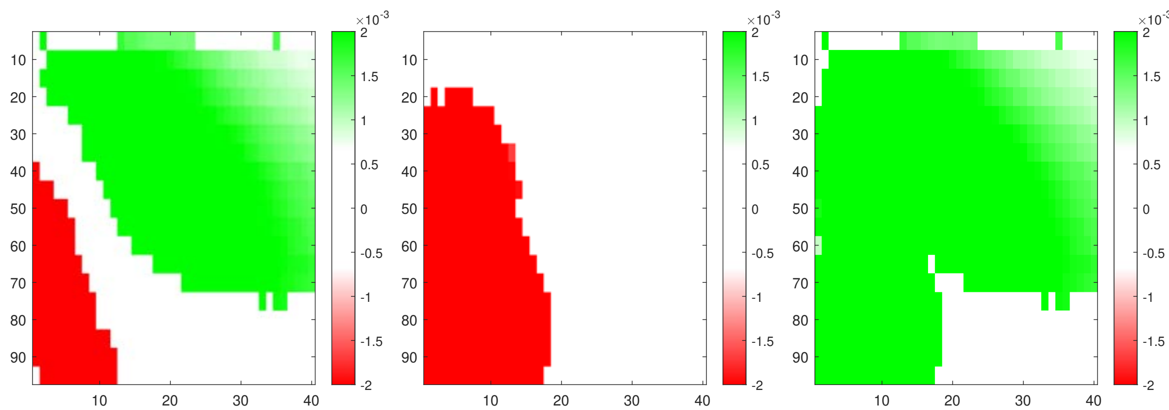


Figure A.6: Distributional IRFs by education



Below college College and above Below minus above

B.2.3 Age

Figure A.7: IRFs by age

(a) Age $\leq$ 25 (b) Age 50–55

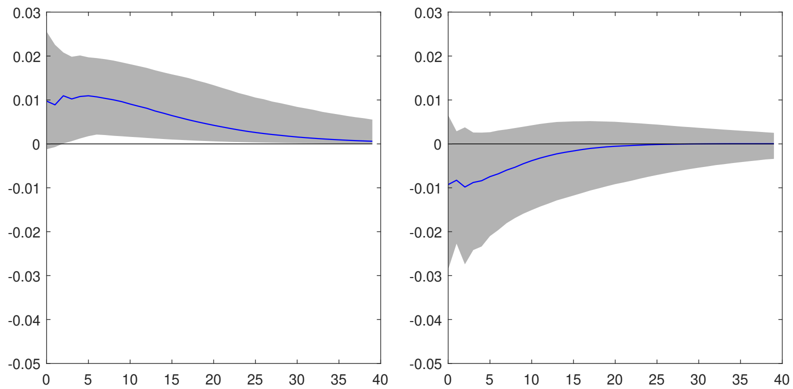
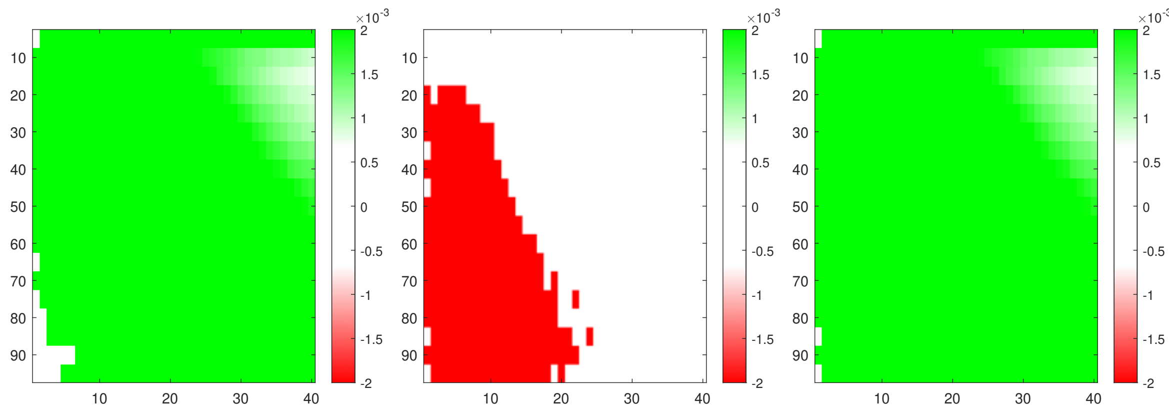


Figure A.8: Distributional IRFs by age



Ages ≤ 25 Ages $(50, 55)$ Young minus old

B.2.4 Region

Figure A.9: IRFs by region

(a) Agency administrative region: office location Philadelphia

(b) Agency administrative region: office location Seattle

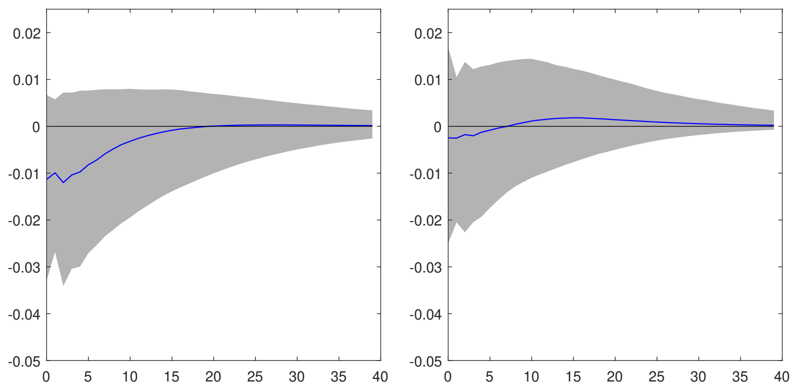
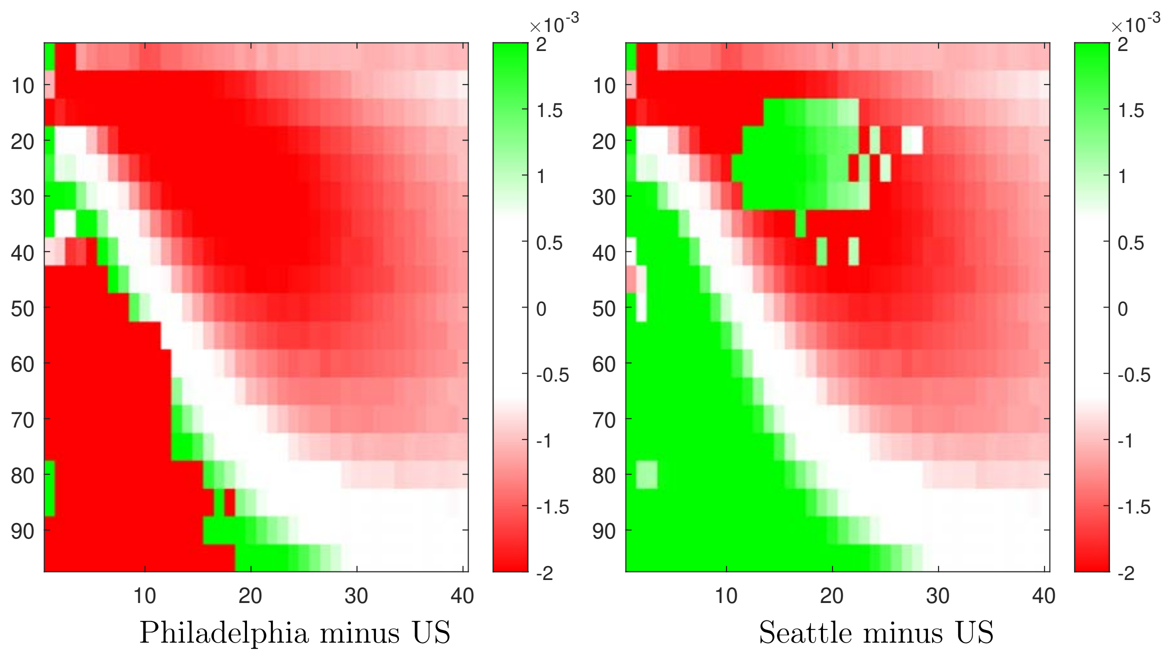


Figure A.10: Distributional IRFs by region



C Construction of Prior Distribution

Here we provide details of our variant of the prior proposed in Chan (2022). Equation (1), repeated below, is our VAR(p) in "structural" form:

$$\mathbf{B}_0\mathbf{z}_t = \mathbf{A}_1\mathbf{z}_{t-1} + \cdots + \mathbf{A}_p\mathbf{z}_{t-p} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), \tag{A.1}$$

where $\mathbf{D} = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{dim}^2)$. This allows us to estimate the VAR equation by equation. As described in Section 2.1, we partition $\mathbf{B}_0$ as $\mathbf{B}_0 = \begin{bmatrix} \mathbf{B}_{0,11} & \mathbf{B}_{0,12} \\ \mathbf{B}_{0,21} & \mathbf{B}_{0,22} \end{bmatrix}$ and impose the restrictions summarized in Section 4.1: $\mathbf{B}_{0,11}$ is full with ones on the diagonal, $\mathbf{B}_{0,12} = \mathbf{0}$, $\mathbf{B}_{0,21}$ is full, and $\mathbf{B}_{0,22}$ is an identity matrix. Under these restrictions, the i-th equation can be written as:

$$z_{it} = \mathbf{w}_{it}'\boldsymbol{\alpha}_i + \mathbf{x}_{it}'\boldsymbol{\theta}_i + \varepsilon_{it}, \quad \epsilon_t \sim N\left(0, \sigma_i^2\right), \tag{A.2}$$

where:

$$\mathbf{w}_{it} = \begin{cases} (-z_{1t}, \ldots, -z_{i-1t}, -z_{i+1t}, \ldots, -z_{Mt})' & i \leq M \text{ (for the macro equation)}, \\ (-z_{1t}, -z_{2t}, \ldots, -z_{Mt})' & i > M \text{ (for the micro equation)}, \end{cases}$$

and $\mathbf{x}_{it} = (1, \mathbf{z}_{t-1}', \ldots, \mathbf{z}_{t-p}')'$.

We use the normal-inverse-gamma prior:

$$\boldsymbol{\alpha_i} \mid \sigma_i^2 \sim \mathcal{N}(0, \sigma_i^2 \mathrm{V}_i^{\alpha}), \tag{A.3}$$

$$\boldsymbol{\theta}_i \mid \sigma_i^2 \sim \mathcal{N}(0, \sigma_i^2 V_i^{\theta}),$$

$$\sigma_i^2 \sim \mathcal{IG}(\frac{v_i}{2}, \frac{s_i^2}{2}),$$

where s_i^2 denotes the sample variance of the residuals from an AR(p) model estimated on variable i.

Following Chan (2022), we set:

$$\mathrm{V}_i^{\alpha} = \begin{cases} \text{diag}(1/s_1^2, \ldots, 1/s_{i-1}^2, 1/s_{i+1}^2, \ldots, 1/s_M^2) & i \leq M \text{ (for the macro equation)}, \\ \text{diag}(1/s_1^2, \ldots, 1/s_M^2) & i > M \text{ (for the micro equation)}. \end{cases}$$

For the VAR coefficients θ_i, the prior covariance matrix depends on V_i^{θ}. It is a Minnesota-type prior and contains three hyperparameters, namely, $\kappa_1, \kappa_2, \kappa_3$, that control the degree of shrinkage for different types of coefficients:

$$\mathrm{V}_i^{\theta} = \begin{cases} \frac{\kappa_1}{l^2 s_i^2} & \text{for the coefficient on the } l\text{th lag of variable } i \text{ (own lags),} \\ \frac{\kappa_2}{l^2 s_j^2} & \text{for the coefficient on the } l\text{th lag of variable } j \text{ (lags of others),} \\ \kappa_3 & \text{for the intercept.} \end{cases}$$

We set $\kappa_1 = 10$, $\kappa_3 = 1$, and:

$$\kappa_2 = \begin{cases} 0.01 & i \leq M \text{ (for the macro equation),} \\ 0.001 & i > M \text{ (for the micro equation).} \end{cases}$$

Finally, the degrees of freedom of the inverse-gamma prior are:

$$v_i = \begin{cases} 20 & i \leq M \text{ (for the macro equation),} \\ 40 & i > M \text{ (for the micro equation).} \end{cases}$$