On the Effects of Monetary Policy Shocks on Income and Consumption Heterogeneity

Minsu Chang*
Seoul National University

Frank Schorfheide
University of Pennsylvania,
CEPR, PIER, NBER

This Version: August 29, 2025

Abstract

We use the functional vector autoregressive framework of Chang, Chen, and Schorfheide (2024) to examine the effects of conventional monetary policy shocks on the cross-sectional distribution of U.S. earnings (Current Population Survey), consumption (Consumer Expenditure Survey), and financial income (Piketty, Saez, and Zucman (2018)'s Distributional National Accounts). We find that an expansionary monetary policy shock reduces earnings inequality, primarily by moving individuals out of unemployment, increases consumption inequality, mainly through spending on durable goods, and raises financial income inequality. The credible intervals for some of these distributional effects, however, are wide. We also demonstrate how an internal instrument can be used to identify the central bank's reaction function and estimate its response to earnings inequality. (JEL C11, C32, C52, E32)

Key words: Consumption Distribution, Earnings Distribution, Financial Income Distribution, Functional Vector Autoregressions, Monetary Policy Rule, Monetary Policy Shocks

* Correspondence: M. Chang: Department of Economics, Seoul National University, Seoul, South Korea. Email: minsuchang@snu.ac.kr (Chang). F. Schorfheide: Department of Economics, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu (Schorfheide). We thank Xiaohong Chen, Stephanie Ettmeier, Eva Janssens, Chi Hyun Kim, and participants at various seminars and conferences for helpful suggestions. This work was supported by the New Faculty Startup Fund from Seoul National University.

1 Introduction

Traditionally, the effects of monetary policy interventions have been studied through the lens of models that abstract from micro-level heterogeneity, such as structural vector autoregressions (VARs) specified in terms of macroeconomic aggregates or as representative agent dynamic stochastic general equilibrium (DSGE) models. However, in view of concerns about rising inequalities in advanced economies in the aftermath of the global financial crisis, there is growing interest in the distributional impacts of conventional and unconventional monetary policies. The contribution of this paper is to extend and apply the functional VAR framework of Chang, Chen, and Schorfheide (2024), henceforth CCS, to study the effects of monetary policy shocks on the cross-sectional distribution of earnings, consumption, and financial income.

The literature has considered two related, but distinct questions. First, what is the effect of a monetary policy shock on the cross-sectional distributions of, say, income and consumption, and inequality measures associated with it? Second, how does income or consumption of particular households or groups of households respond to a monetary policy shock? Answers to the first question provide guidance to central banks that are concerned about distributional effects of their actions. Answers to the second question shed light on the shock propagation mechanism. In principle, household- or agent-level responses could be aggregated to derive the response of the cross-sectional distribution, but this requires very accurate measurement of unit-level responses, which often is obstructed by data limitations and nonlinearities of unit-level time series.[1] Our functional VAR approach is designed to provide a direct answer to the first question. In addition, we also examine whether the central bank systematically responds to fluctuations in the labor earnings distribution.

Starting point for our analysis is the monetary policy shock instruments (high-frequency interest rate and stock price surprises) provided by Jarocinski and Karadi (2020), henceforth JK, which we use as internal instruments in a structural VAR. Our contribution to the literature is to provide empirical evidence on the distributional response of labor earnings, consumption and its components, and financial income. CCS provided a framework for including a cross-sectional distribution in a VAR with aggregate variables. In each time period t the cross-sectional distribution is represented by a log probability density function

[1] Research by Ettmeier, Kim, and Schorfheide (2025) examines to what extent functional and unit-level modeling approaches can deliver the same estimates of distributional responses in the context of German administrative data.

(pdf), which in turn is approximated by a linear sieve. In a nutshell, the sieve coefficients are estimated in each period based on cross-sectional data by maximum likelihood. The coefficient estimates are then added to the vector of macroeconomic variables in a VAR. The resulting time series model can be estimated with standard techniques. Finally, one can compute IRFs for the sieve coefficients, which can be converted into IRFs for the cross-sectional densities and implied summary statistics, such as percentiles or inequality measures. The data requirement of the functional approach used in this paper is relatively weak: repeated cross-sections of micro-level observations suffice.

The empirical analysis in this paper requires us to extend the functional VAR framework in several dimensions. First, we show how one can identify the monetary policy rule in a structural VAR with an internal instrument so that we can assess the systematic reaction of the central bank to fluctuations of cross-sectional distributions. Second, we show how to stack multiple marginal densities in the vector of observables. Third, we introduce observation weights. Finally, we modify the state-space framework of CCS to allow for mixed-frequency observations.

Our empirical analysis generates the following findings. First, representing the labor earnings distribution from the Current Population Survey (CPS) as a mixture of a continuous part, capturing positive earnings of employed individuals, and a point mass at zero that corresponds to the unemployed individuals, we find that an expansionary monetary policy shock reduces earnings inequality mostly because the unemployment rate falls and individuals with no previous earnings receive positive earnings (employment channel). If we focus solely on the continuous part of the earnings distribution, then the effect on inequality is small and short-lived. Thus, the employment channel dominates. The estimated effects are broadly consistent with the heterogeneous agent New Keynesian model (HANK) with indivisible labor studied by Ma (2023) and with findings previously reported in the literature, e.g., Coibion, Gorodnichenko, Kueng, and Silvia (2017), Furceri, Loungani, and Zdzienicka (2018), Lenza and Slacalek (2024), and Mitman, Broer, and Kramer (2022). A subset of these studies also emphasizes that the effect is driven by the extensive margin, i.e., workers transitioning out of unemployment, which is consistent with our findings.

Second, to visualize the central bank's systematic reaction to the shape of the labor earnings distribution, we use the estimated monetary policy rule to compute an alternative interest rate path by holding the earnings distribution fixed at its 2006:M1 level. The counterfactual earnings density has more mass on earnings that are less than the labor share of GDP per capita, compared to the actual earnings densities post 2006:M1. According to

the period-by-period posterior medians, an increase in probability mass to the left of one (shift toward lower earnings) leads the central bank to lower interest rates. However, the credible bands are wide, covering positive and negative interest rate effects.

Third, we generate impulse response functions (IRFs) for the consumption distribution, obtained from the Consumer Expenditure Survey (CEX). These IRFs capture direct effects, i.e., lower interest rates tend to increase household spending, and indirect effects, i.e., increased aggregate demand raises wages and employment, of an expansionary monetary policy shock.[2] We find that an expansionary policy shock slightly increases consumption inequality measures at the posterior median, because 80th and 90th percentiles of the consumption distribution respond slightly stronger than the 10th and 20th percentiles. A decomposition of consumption in durables, nondurables, and services reveals that the distributional response of total consumption closely mirrors the response of the durable consumption distribution.

With respect to consumption inequality, the results in the existing literature have been mixed, which is consistent with our fairly wide posterior credible intervals that span positive and negative responses. Coibion, Gorodnichenko, Kueng, and Silvia (2017) find that consumption inequality decreases in response to an expansionary monetary policy shock. Cloyne, Ferreira, and Surico (2020) document in their pseudo-panel analysis that when interest rates fall, households with a mortgage increase their spending considerably, while outright homeowners without mortgage debt do not change their expenditure at all. Using an indirect approach, McKay and Wolf (2023) find that the overall consumption response is quite even in the cross section, but there is heterogeneity in regard to the channels that lead to the consumption response. Holm, Paul, and Tischbirek (2021) show that in their Norwegian administrative data the consumption response has the same U-shape as disposable income with the biggest gains accruing at the bottom and top end of the liquid asset distribution. Andersen, Johannesen, Jorgensen, and Peydro (2023) use a very limited measure of consumption, namely car purchases. The higher the household income, the larger the effect of an expansionary monetary policy shock on car purchases.

Fourth, we construct density estimates at annual frequency from the Distributional National Accounts microfiles provided by Piketty, Saez, and Zucman (2018), henceforth PSZ data. We find that the density responses imply a fairly persistent increase in financial income inequality. With regard to financial income, the finding in the literature, e.g., Holm, Paul,

[2]Kaplan, Moll, and Violante (2018) emphasize that a decomposition into direct and indirect effects is useful for the analysis of the propagation of monetary policy shocks in HANK models.

and Tischbirek (2021), Amberg, Jansson, Klein, and Rogantini Picco (2022), Andersen, Johannesen, Jorgensen, and Peydro (2023), Lenza and Slacalek (2024), is generally that in absolute terms high-wealth individuals benefit substantially from monetary expansions. A drop in interest rate triggers a rise in dividends, capital gains, and business income. Because these households own a large fraction of the capital stock, their financial income increases in absolute terms much more than the financial income of low-wealth households. In terms of percentage changes of financial income, the effect is less clear. In the left tail of the wealth distribution, households may adjust their portfolio, e.g., pay back credit card debt or start saving, which in percentage terms can generate large changes in financial income. On the other hand, to the extent that capital income tends to respond more strongly than interest rate income, the portfolio return of high-wealth households may benefit from a larger share of firm equity. The aforementioned papers typically focus on the response of groups of individuals, whereas we provide results on the response of the entire cross-sectional distribution.

The aforementioned papers use a variety of methodological approaches that coexist in the empirical literature. (i) Coibion, Gorodnichenko, Kueng, and Silvia (2017) and Furceri, Loungani, and Zdzienicka (2018) directly include inequality statistics in a standard VAR. (ii) Del Canto, Grigsby, Qian, and Walsh (2023), McKay and Wolf (2023), and Lenza and Slacalek (2024) used an indirect approach that extracts information from the micro data about individual-level income and portfolio compositions or consumption shares, and then combines this information with IRFs for aggregate wages, hours worked, interest rates, asset returns, consumption, and so forth, to simulate micro-level outcomes forward. (iii) Cloyne, Ferreira, and Surico (2020) and Mitman, Broer, and Kramer (2022) have constructed pseudo-panels to examine how different groups of individuals respond to shocks. (iv) Finally, some countries make administrative panel data sets available that can be used, for instance, to estimate panel local projections with observed group heterogeneity; e.g., Holm, Paul, and Tischbirek (2021), Amberg, Jansson, Klein, and Rogantini Picco (2022), and Andersen, Johannesen, Jorgensen, and Peydro (2023).

Each approach has advantages and disadvantages, which is why they co-exist in the literature and are, to some extent, complementary. (i) is straightforward to implement, but simulation evidence in CCS shows that it can be inefficient compared to our functional approach (wider credible intervals). (ii) has weak data requirements in that it only requires a single cross-section, but relies on the assumption that changes in portfolio compositions are of second-order importance for distributional dynamics. This is not consistent with our estimated impulse responses of percentiles near zero which are large in percentage terms.

The pseudo-panels in (iii) average over within-group heterogeneity which makes it difficult to convert results into statements about cross-sectional distributions and inequality statistics. While the use of administrative panel data sets in (iv) is desirable, for many countries they are not available. Moreover, modeling dynamics at a unit level, can be challenging. The functional approach directly targets the first question posed at the beginning: how does a cross-sectional distribution respond to a policy shock?

Applications of functional data analysis in macroeconometrics are growing. From a methodological perspective, the main difference across implementations is the use of pre-specified basis functions as in CCS versus empirical basis functions estimated through functional principal component analysis, e.g., Chang, Kim, and Park (2016), Hu and Park (2017), Bjornland, Chang, and Cross (2023). In terms of applications, Ettmeier (2024) studies distributional responses to fiscal policy shocks. Meeks and Monti (2023) study a version of the New Keynesian Phillips curve in which a scalar expected inflation is replaced by a distribution of inflation expectations. Inoue and Rossi (2021) treat the monetary policy shock as a functional object that captures yield curve surprises. Marcellino, Renzetti, and Tornese (2024b) uses functional versions of MIDAS regressions in a mixed-frequency framework to generate nowcasts of the U.S. labor earnings distribution. Finally, Marcellino, Renzetti, and Tornese (2024a) and Bayer, Calderon, and Kuhn (2025) started to develop functional models for multivariate distributions.

The remainder of this paper is organized as follows. Section 2 describes the functional VAR setup, reviewing important material from CCS to make this paper self-contained. Section 3 shows how the monetary policy rule can be recovered from a structural VAR with an internal instrument, provides an overview of the Bayesian estimation, and presents the methodological extensions developed in this paper. An overview of the empirical analysis and a discussion of the IRFs of the aggregate variables to a monetary policy shock are provided in Section 4. Results for the responses of the labor earnings, consumption, and financial income distributions are presented in Sections 5, 6, and 7, respectively. Finally, Section 8 concludes and an Online Appendix contains supplemental information on the econometric methodology, the data sets used, and additional empirical results.

2 A Functional VAR for Cross-Sectional Data

To make this paper self-contained, we provide a summary of the functional VAR framework developed in CCS. The variables in the functional model comprise an $n_y \times 1$ vector of macroeconomic aggregates Y_t and a cross-sectional density $p_t(x)$. Rather than working with the densities directly, we express the model dynamics in terms of (potentially unnormalized) log densities $\ell_t(x)$. The advantage of using log pdfs, instead of level pdfs, cumulative distribution functions (cdfs), or quantiles, is that log densities do not have to satisfy monotonicity or non-negativity restrictions. Thus, they can be easily and coherently propagated using a linear law of motion and then *ex post* normalized to integrate to one in each period.

2.1 The Dynamic Model

We assume that in every period $t = 1, \ldots, T$ an econometrician observes the macroeconomic aggregates Y_t as well as a sample of N_t draws x_{it}, $i = 1, \ldots, N_t$ from the cross-sectional density $p_t(x)$. In practice, N_t is likely to vary from period to period, but for the subsequent exposition it will be convenient, notationally, to assume that $N_t = N$ for all t. We collect the time t cross-sectional observations in the vector $X_t = [x_{1t}, \ldots, x_{Nt}]'$. The likelihood function for the functional model is constructed under the assumption that the draws x_{it} are independently and identically distributed (iid) over the cross-section and independent over time conditional on $p_t(x)$. The measurement equation for the cross-section observations takes the form

$$x_{it} \mid p_t(x) \overset{iid}{\sim} p_t(x) = \frac{\exp\{\ell_t(x)\}}{\int \exp\{\ell_t(\tilde{x})\} d\tilde{x}}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where the denominator on the right-hand side ensures that $p_t(x)$ is properly normalized. The iid assumption is used to construct the (quasi)-likelihood function of $(x_{1t}, \ldots, x_{Nt})$ as a product of the marginal densities $p_t(x_{it})$, $i = 1, \ldots, N$. Spatial correlations of unit-level outcomes across individuals in a given period t can be captured by suitable definitions of x_{it}, e.g., by dividing unit-level labor income by GDP or through shifts of $p_t(\cdot)$ in response to aggregate shocks. In a panel setting, there could be correlations between x_{it} and x_{it+1} through time-invariant or highly-persistent unit-specific effects. However, we are working with repeated cross-sections or frequently rotating panels in which the distortion from temporal correlation of the x_{it}s can be regarded as negligible. Most importantly, the functional modeling approach does not require the econometrician to make assumptions about the evolution of x_{it} at the level of an individual, a household, or a firm.

We assume that Y_t and ℓ_t evolve according to a joint autoregressive law of motion:

$$\begin{aligned} Y_t &= B_{yy}Y_{t-1} + \int B_{yl}(\tilde{x})\ell_{t-1}(\tilde{x})d\tilde{x} + B_{y0} + u_{y,t} \\ \ell_t(x) &= B_{ly}(x)Y_{t-1} + \int B_{ll}(x,\tilde{x})\ell_{t-1}(\tilde{x})d\tilde{x} + B_{l0}(x) + u_{l,t}(x). \end{aligned} \tag{2}$$

Here $u_{y,t}$ is mean-zero random vector with covariance Ω_{yy} and $u_{l,t}(x)$ is a random element in a Hilbert space with covariance function $\Omega_{ll}(x,\tilde{x})$. We denote the covariance function for $u_{y,t}$ and $u_{l,t}(x)$ by $\Omega_{yl}(x)$. For now, (2) should be regarded as a reduced-form fVAR in which $u_{y,t}$ and $u_{l,t}(x)$ are one-step-ahead forecast errors. For the empirical analysis below we add more lags to the system. (1) can be interpreted as a measurement equation that links the unobserved log density $\ell_t(x)$ to the cross-sectional observations X_t.

2.2 Three Simplifications

Unfortunately, the estimation of the model defined by (1) and (2) is impractical. Thus, we will simplify it in three steps. First, we replace the infinite-dimensional objects by finite-dimensional objects. Second, we turn the nonlinear measurement equation (1) into a linear one. Third, we let the implied measurement error variance tend to zero.

A Finite-Dimensional Approximation. We replace $\ell_t(x)$ by a collection of finite-dimensional representations, indexed by the superscript (K). Let

$$\ell_t^{(K)}(x) = \sum_{k=1}^{K} \alpha_{k,t}\zeta_k(x) = \big[\zeta_1(x),\ldots,\zeta_K(x)\big] \cdot \begin{bmatrix} \alpha_{1,t} \\ \vdots \\ \alpha_{K,t} \end{bmatrix} = \zeta'(x)\alpha_t, \tag{3}$$

where $\zeta_1(x),\zeta_2(x),\ldots$ is a sequence of basis functions. We dropped the (K) superscripts from the vectors $\zeta(x)$ and α_t to simplify the notation. We can now write a K'th order representation of the joint density of X_t as follows:

$$\begin{aligned} p^{(K)}\left(X_t|\alpha_t\right) &= \exp\left\{N\mathcal{L}^{(K)}(\alpha_t|X_t)\right\}, \\ \mathcal{L}^{(K)}(\alpha_t|X_t) &= \left(\frac{1}{N}\sum_{i=1}^{N}\zeta(x_{it})'\right)\alpha_t - \ln\int \exp\left\{\zeta'(x)\alpha_t\right\}dx. \end{aligned} \tag{4}$$

We approximate the functions $B_{yl}(\tilde{x})$, $B_{ly}(x)$, and $B_{l0}(x)$, the kernel $B_{ll}(x,\tilde{x})$, and the functional innovation $u_{l,t}(x)$ that appear in (2) using the same vector of basis functions $\zeta(x)$

as in (3):

$$\begin{aligned} B_{yl}^{(K)}(x) &= B_{yl}\xi(\tilde{x}), \quad B_{ly}^{(K)}(x) = \zeta'(x)B_{ly}, \quad B_{l0}^{(K)}(x) = \zeta'(x)B_{l0}, \\ B_{ll}^{(K)}(x,\tilde{x}) &= \zeta'(x)B_{ll}\xi(\tilde{x}), \quad u_{l,t}^{(K)}(x) = \zeta'(x)u_{\alpha,t}, \end{aligned} \tag{5}$$

where $\xi(x)$ is a second $K\times 1$ vector of basis functions and $u_{\alpha,t}$ is a $K\times 1$ vector of innovations. The matrix B_{yl} is of dimension $n_y \times K$, B_{ly} is $K \times n_y$, B_{l0} is $K \times 1$, and B_{ll} is $K \times K$. Because we used the same basis functions for the approximation of $\ell_t(x)$ and the functional objects in the law of motion (2), combining Equations (2), (3), and (5) yields a vector autoregressive system for the macroeconomic aggregates Y_t and the sieve coefficients α_t (omitting K superscripts):

$$\begin{bmatrix} Y_t \\ \alpha_t \end{bmatrix} = \begin{bmatrix} B_{yy} & B_{yl}C_\alpha \\ B_{ly} & B_{ll}C_\alpha \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ \alpha_{t-1} \end{bmatrix} + \begin{bmatrix} B_{y0} \\ B_{l0} \end{bmatrix} + \begin{bmatrix} u_{y,t} \\ u_{\alpha,t} \end{bmatrix}, \tag{6}$$

where $C_\alpha = \int \xi(\tilde{x})\zeta'(\tilde{x})d\tilde{x}$. Let $u_t' = [u_{y,t}', u_{\alpha,t}']$. We subsequently assume that the innovations are Gaussian:

$$u_t \sim \mathcal{N}(0, \Sigma). \tag{7}$$

To obtain a more compact representation, we define the $n_w \times 1$ vector $W_t = [Y_t', \alpha_t']'$ and absorb the matrix C_α into a general regression coefficient matrix Φ_1. Moreover, we add additional lags to (6) to make the model useful for empirical work. This leads to

$$W_t = \sum_{j=1}^{p} \Phi_j W_{t-j} + \Phi_0 + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma). \tag{8}$$

A Finite-Dimensional Linear Approximation. To avoid the use of a nonlinear filter for the likelihood evaluation, one can "linearize" the measurement equation (4) by taking a second-order Taylor approximation of $\ln p^{(K)}(X_t|\alpha_t)$ around the maximum likelihood estimator (MLE) $\hat{\alpha}_t$. This approximation can be written as a linear Gaussian measurement equation:

$$\hat{\alpha}_t = \alpha_t + N^{-1/2}\eta_t, \quad \eta_t \sim \mathcal{N}\big(0, \hat{V}_t^{-1}\big), \tag{9}$$

where $\hat{V}_t$ is the negative inverse Hessian associated with the log-likelihood function evaluated at the MLE. Note that the observations X_t enter the measurement equation indirectly through the MLE $\hat{\alpha}_t$.

A Finite-Dimensional VAR. If N is large relative to K, then the measurement error $N^{-1/2}\eta_t$ is close to zero, and $\alpha_t \approx \hat{\alpha}_t$. Thus, for the empirical analysis we simply replace α_t in (6) by $\hat{\alpha}_t$ and estimate a VAR in the macroeconomic variables and the estimated sieve coefficients. The estimation can be conveniently implemented in two steps:

1. For each period $t = 1, \ldots, T$ estimate the log-spline density model for X_t by maximizing the log-likelihood function in (4). This leads to the sequence $\hat{\alpha}_t$.

2. Estimate a version of the VAR in (8), replacing the "true" sieve coefficients α_t in the definition of W_t with $\hat{\alpha}_t$.

CCS provide rates at which (N, T, K) are allowed to tend to infinity to ensure that the likelihood functions of the three finite-dimensional models are asymptotically equivalent. In this paper, we are considering an application in which the cross-sectional dimension N is large, and we will work with the finite-dimensional VAR approximation.

3 Identification and Estimation

In Section 3.1 we show how one can recover the monetary policy rule in an internal instrument setting. We use this result in the empirical analysis to examine the central bank's reaction to the cross-sectional distribution of labor earnings. To make the paper self-contained, Section 3.2 reviews the Bayesian estimation of the functional model and some key implementation steps previously presented in CCS. Finally, in Section 3.3 we present a number of novel extensions of the CCS framework that are used in the empirical analysis.

3.1 Monetary Policy IRFs and Policy Rule

To identify the effects of a monetary policy shock, we include an instrumental variable (IV) in the definition of Y_t (and hence W_t), assuming that the instrument is ordered first. This is referred to as an *internal* instrument in the SVAR literature. For now, we assume that there is a single policy shock and a single instrument. If the instrumental variable satisfies a version of the relevance and exclusion conditions, then the impact effect of the monetary policy shock is given by the first column of the low-triangular Cholesky factor of the covariance matrix Σ of the reduced-form forecast errors u_t; see, for instance, Ramey (2011), Anderson, Inoue, and Rossi (2016), Jarocinski and Karadi (2020), and Plagborg-Møller and Wolf (2021). The novel part of the subsequent derivations is to show how the monetary policy rule can also be recovered. This allows us to assess the central bank's reaction to cross-sectional distributions.

The starting point of our derivations is (8). We denote the vector of structural innovations by ϵ_t and express the reduced-form innovations u_t as a linear function of ϵ_t:

$$u_t = \Phi_\epsilon \epsilon_t. \tag{10}$$

Now partition

$$W_t' = \begin{bmatrix} W_{0t} & W_{1t} & W_{2t}' \end{bmatrix}, \quad u_t' = \begin{bmatrix} u_{0t} & u_{1t} & u_{2t}' \end{bmatrix}, \quad \epsilon_t' = \begin{bmatrix} \epsilon_{0t} & \epsilon_{1t}' & \epsilon_{2t}' \end{bmatrix},$$

where W_{0t} is the policy shock instrument, W_{1t} is the nominal interest rate, and W_{2t} are the remaining elements of Y_t and α_t. Moreover, ϵ_{0t} is the shock instrument innovation, ϵ_{1t} is a monetary policy shock innovation, and ϵ_{2t} is a vector of private-sector structural shocks. We make the following additional assumptions:

Assumption 1 (Identification)
(i) The structural innovations satisfy: $\mathbb{E}[\epsilon_t \epsilon_t'] = I$.
(ii) The monetary policy shock, denoted by ν_{1t} *can be expressed as a linear function of the instrument innovation and the monetary policy shock innovation:*

$$\nu_{1t} = \gamma_{10}\epsilon_{0t} + \gamma_{11}\epsilon_{1t}, \quad \gamma_{10} > 0. \tag{11}$$

(iii) The economy responds to ϵ_{0t} *and* ϵ_{1t} *only through* ν_{1t}*:*

$$\Phi_\epsilon = \begin{bmatrix} \Phi^\epsilon_{00} & 0 & 0 \\ \Phi^\epsilon_{11}\gamma_{10} & \Phi^\epsilon_{11}\gamma_{11} & \Phi^\epsilon_{12} \\ \Phi^\epsilon_{21}\gamma_{10} & \Phi^\epsilon_{21}\gamma_{11} & \Phi^\epsilon_{22} \end{bmatrix}. \tag{12}$$

Note that (11) resembles the equation $x_i = \gamma z_i + \eta_i$ in a linear instrumental variable model. The instrument innovation ϵ_{0t} creates movements in the monetary policy shock. In addition, the monetary policy shock innovation ϵ_{1t} creates additional fluctuations in unanticipated deviations from the monetary policy rule. The assumption $\gamma_{10} > 0$ ensures instrument relevance, and Assumption 1(i) implies instrument validity. Assumption 1(iii) in combination with (ii) implies that the first column of the Cholesky factor of $\Sigma = \Phi_\epsilon \Phi_\epsilon'$ recovers the impact effect of the monetary policy shock ν_{1t} up to the scale factor γ_{10}, which is the standard implementation in the literature on SVARs with internal instruments.[3]

[3]If Assumption 1(ii) is replaced by the assumption that the instrument provides a noisy measure of the monetary policy shock, then the implied zero restrictions on Φ_ϵ change and the estimation can no longer be implemented with the Cholesky factorization. For Bayesian implementations see Caldara and Herbst (2019), Drautzburg (2020), and Arias, Rubio-Ramirez, and Waggoner (2022).

The novel part of our analysis is the construction of the monetary policy rule. If one plugs (11) into (8) and pre-multiplies by Φ_ϵ^{-1}, then the SVAR can be expressed as a system-of-equations model, in which one of the equations, namely the one associated with the monetary policy shock, is the monetary policy rule:

$$\Phi_\epsilon^{-1} W_t = \Phi_\epsilon^{-1} \left(\sum_{j=1}^{p} \Phi_j W_{t-j} + \Phi_0 \right) + \epsilon_t. \tag{13}$$

We now show how the coefficients of the monetary policy rule equation can be determined.

Step 1. Merge the "1" and "2" partitions of Φ_ϵ in (12) into a "v" partition by defining

$$\Phi_{v0}^\epsilon = \begin{bmatrix} \Phi_{11}^\epsilon \gamma_{10} \\ \Phi_{21}^\epsilon \gamma_{10} \end{bmatrix}, \quad \Phi_{vv}^\epsilon = \begin{bmatrix} \Phi_{11}^\epsilon \gamma_{11} & \Phi_{12}^\epsilon \\ \Phi_{21}^\epsilon \gamma_{11} & \Phi_{22}^\epsilon \end{bmatrix}, \quad W_{vt} = \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}, \quad \epsilon_{vt} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}. \tag{14}$$

Using a QR decomposition of $\Phi_{vv}^{\epsilon'}$ we can write $\Phi_{vv}^\epsilon = \Phi_{vv,tr}^\epsilon \Omega_{vv}$, where $\Phi_{vv,tr}^\epsilon$ is lower triangular and Ω_{vv} is an orthogonal matrix. This leads to

$$\Phi_\epsilon = \begin{bmatrix} \Phi_{00}^\epsilon & 0 \\ \Phi_{v0}^\epsilon & \Phi_{vv}^\epsilon \end{bmatrix} = \underbrace{\begin{bmatrix} \Phi_{00}^\epsilon & 0 \\ \Phi_{v0}^\epsilon & \Phi_{vv,tr}^\epsilon \end{bmatrix}}_{\Phi_{tr}^\epsilon} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & \Omega_{vv} \end{bmatrix}}_{\Omega}. \tag{15}$$

Combining (8), (10), and (15), and premultiplying by $\left(\Phi_{tr}^\epsilon\right)^{-1}$ yields the following system:

$$\left(\Phi_{tr}^\epsilon\right)^{-1} W_t = \left(\Phi_{tr}^\epsilon\right)^{-1} \left(\sum_{j=1}^{p} \Phi_j W_{t-j} + \Phi_0 \right) + \begin{bmatrix} 1 & 0 \\ 0 & \Omega_{vv} \end{bmatrix} \begin{bmatrix} \epsilon_{0t} \\ \epsilon_{vt} \end{bmatrix}. \tag{16}$$

This is the system that we will estimate equation by equation in Section 3.2. We proceed by inverting the partitioned matrix

$$\left(\Phi_{tr}^\epsilon\right)^{-1} = \begin{bmatrix} \left(\Phi_{00}^\epsilon\right)^{-1} & 0 \\ -\left(\Phi_{vv,tr}^\epsilon\right)^{-1} \Phi_{v0}^\epsilon \left(\Phi_{00}^\epsilon\right)^{-1} & \left(\Phi_{vv,tr}^\epsilon\right)^{-1} \end{bmatrix}.$$

Dropping the equation for the shock instrument, the remainder of the system can be written as

$$-\left(\Phi_{vv,tr}^\epsilon\right)^{-1} \Phi_{v0}^\epsilon \left(\Phi_{00}^\epsilon\right)^{-1} W_{0t} + \left(\Phi_{vv,tr}^\epsilon\right)^{-1} W_{vt} = \left(\text{lag terms and intercept}\right) + \Omega_{vv} \epsilon_{vt}. \tag{17}$$

Once we multiply the system by the transpose of the first column of Ω_{vv} we have identified the monetary policy rule because $[\Omega_{vv}]_{\cdot 1}' \Omega_{vv} \epsilon_{vt} = \epsilon_{1t}$.

Step 2. We have to determine $[\Omega_{vv}]_{\cdot 1}$. Using the definitions of Φ^{ϵ}_{v0} and Φ^{ϵ}_{vv} in (14) and the factorization in (15) we obtain the following system of $2(n_w - 1)$ equations

$$\begin{bmatrix}\Phi^{\epsilon}_{11}\\ \Phi^{\epsilon}_{21}\end{bmatrix}\gamma_{10} = \Phi^{\epsilon}_{v0}, \quad \begin{bmatrix}\Phi^{\epsilon}_{11}\\ \Phi^{\epsilon}_{21}\end{bmatrix}\gamma_{11} = \Phi^{\epsilon}_{vv,tr}[\Omega_{vv}]_{\cdot 1}$$

that can be used to determine the $n_w - 2$ free elements of the unit-length vector $\Omega_{vv,\cdot 1}$, the $n_w - 1$ elements of Φ^{ϵ}_{11} and Φ^{ϵ}_{21}, and the two scalars γ_{10} and γ_{11}, conditional on estimates of Φ^{ϵ}_{v0} and $\Phi^{\epsilon}_{vv,tr}$. In turn,

$$[\Omega_{vv}]_{\cdot 1} = \frac{\gamma_{11}}{\gamma_{10}}\big(\Phi^{\epsilon}_{vv,tr}\big)^{-1}\Phi^{\epsilon}_{v0}, \quad \frac{\gamma_{10}}{\gamma_{11}} = \left\|\big(\Phi^{\epsilon}_{vv,tr}\big)^{-1}\Phi^{\epsilon}_{v0}\right\|. \tag{18}$$

Here the ratio γ_{10}/γ_{11} is determined from the unit length restriction for $[\Omega_{vv}]_{\cdot 1}$.

Step 3. Use the first equation in (18) to replace $\big(\Phi^{\epsilon}_{vv,tr}\big)^{-1}\Phi^{\epsilon}_{v0}$ on the left-hand side of (17) and then premultiply the resulting equation by $[\Omega_{vv}]'_{\cdot 1}$. This leads to

$$-\frac{\gamma_{10}}{\gamma_{11}}\big(\Phi^{\epsilon}_{00}\big)^{-1}W_{0t} + [\Omega_{vv}]'_{\cdot 1}\big(\Phi^{\epsilon}_{vv,tr}\big)^{-1}W_{vt} = \big(\text{lag terms and intercept}\big) + \epsilon_{1t}.$$

The first equation in (16) implies that $\big(\Phi^{\epsilon}_{00}\big)^{-1}W_{0t} = (\ldots) + \epsilon_{0t}$. Using the formula for the monetary policy shock ν_{1t} in Assumption 1(ii) we obtain the following representation of the monetary policy rule:

$$\gamma_{11}[\Omega_{vv}]'_{\cdot 1}\big(\Phi^{\epsilon}_{vv,tr}\big)^{-1}W_{vt} = \gamma_{11}\big(\text{lag terms and intercept}\big) + \nu_{1,t}. \tag{19}$$

In the empirical analysis we then choose γ_{11} to normalize the coefficient on the time t interest rate to one.

3.2 Bayesian Estimation and Implementation

Likelihood Function. Let D be a diagonal matrix with elements D_i that are given by the reciprocals of the diagonal elements of $\big(\Phi^{\epsilon}_{tr}\big)^{-1}$. Thus, $A = D\big(\Phi^{\epsilon}_{tr}\big)^{-1}$ is a lower-triangular matrix with ones on the diagonal. Let $B_j = D\big(\Phi^{\epsilon}_{tr}\big)^{-1}\Phi_j$ for $j = 0, 1, \ldots, p$. Then the triangular system in (16) can be rewritten as follows:

$$AW_t = \sum_{j=1}^{p} B_j W_{t-j} + B_0 + \eta_t, \quad \eta_t \sim \mathcal{N}(0, D). \tag{20}$$

Using the lower-triangular structure of the A matrix, define the $(i-1)\times 1$ vectors

$$\mathcal{A}_i = [A_{i,1}, \ldots, A_{i,i-1}], \quad \tilde{W}_{<i,t} = -[W_{1,t}, \ldots, W_{i-1,t}]', \quad i = 2, \ldots, n.$$

Moreover, let $k_i = k + i - 1$ and define the $k_i \times 1$ vectors

$$Z_{it} = \big[\tilde{W}'_{<i,t}, W'_{t-1}, \ldots, W'_{t-p}, 1\big], \quad \beta_i = \big[\mathcal{A}'_i, B'_{i\cdot,1}, \ldots, B'_{i\cdot,p}, B'_{i\cdot,0}\big]',$$

where $B_{i\cdot,j}$ is the ith row of the matrix B_j. Finally, define W_i to be the $T \times 1$ vector with elements W_{it}, Z_i the $T \times k_i$ matrix with rows Z'_{it}, and η_i the $T \times 1$ vector with elements η_{it}. Then we can write the ith equation in matrix form as

$$W_i = Z_i\beta_i + \eta_i. \tag{21}$$

Because A is lower-triangular with ones on the diagonal, the likelihood function for the system is the product of the likelihood functions for each variable i. Let $\beta = (\beta_1, \ldots, \beta_n)$. Then:

$$p(W|\beta, D) \propto \prod_{i=1}^{n} |D_i|^{-1/2} \exp\left\{-\frac{1}{2D_i}(W_i - Z_i\beta_i)'(W_i - Z_i\beta_i)\right\}. \tag{22}$$

Prior Distribution. Chan (2022) proposes a prior distribution that assumes that parameters are independent across equations and allows for some asymmetry across equations. This implies in combination with (22) that the model can be estimated equation-by-equation, speeding up the Bayesian computations in high-dimensional settings considerably. The prior takes the form

$$p(\beta, D|\lambda) = \prod_{i=1}^{n} p(\beta_i|D_i, \lambda)p(D_i|\lambda), \tag{23}$$

where λ is a vector of hyperparameters. For each pair (β_i, D_i) we use a Normal-Inverse Gamma (NIG) distribution of the form

$$\beta_i|(D_i, \lambda) \sim \mathcal{N}\big(\underline{\beta}_i, D_i\underline{V}_i^{\beta}\big), \quad D_i|\lambda \sim IG\big(\underline{\nu}_i, \underline{S}_i\big). \tag{24}$$

The prior distribution loosely follows that of a Minnesota prior. The mean vectors $\underline{\beta}_i$ are chosen such that the implied prior for the reduced form parameters Φ is centered at univariate random walk specifications for those elements of Y_t that correspond to log levels of macroeconomic aggregates. For other elements of Y_t and the elements of the α_t vector the priors are centered at zero. We introduce a hyperparameter λ_1 that controls the overall prior precision $[\underline{V}_i^{\beta}]^{-1}$ and a hyperparameter λ_2 that controls the relative precision of coefficients that capture the effect of lagged α_ts on current Y_ts. As $\lambda_2 \longrightarrow \infty$, the cross-sectional density does not Granger-cause the aggregate variables and the system becomes block-triangular. Further details on the specification of $\underline{\beta}_i$, $\underline{V}_i^{\beta}$, $\underline{\nu}_i$, and $\underline{S}_i$, and the posterior distribution are provided in the Online Appendix.[4]

[4] As in Jarocinski and Karadi (2020), we modify the VAR in (8) to impose that $\Phi_{1\cdot} = 0$.

Posterior Sampling. The conjugate form of the prior implies that the posterior distribution of (β, D) also belongs to the NIG family. Thus, we can generate posterior draws of (β, D) by direct sampling and because both likelihood and prior factorize in terms of (β_i, D_i), $i = 1, \ldots, n$, we can sample the parameters for each equation separately. For each draw (β^s, D^s), $s = 1, \ldots, N_{sim}$ we iterate (20) forward to obtain a draw from the posterior distribution of the $W_t = [Y_t', \alpha_t]'$ IRFs at horizons $h = 0, 1, \ldots$. The α_h^s IRF draws are converted into density IRFs using

$$p^{(K)}(x|\alpha_h^s) = \frac{\exp\{\zeta'(x)\alpha_h^s\}}{\int \exp\{\zeta'(\tilde{x})\alpha_h^s\}\, d\tilde{x}}. \tag{25}$$

Based on the draws from the IRF posterior we compute summary statistics that are then plotted in the figures.

Model Selection. Our model for $(Y_{1:T}, X_{1:T})$ depends on the sieve approximation order K, the number of lags p in the VAR (8), and the hyperparameter vector λ for the prior distribution in (23). We select these model features by maximizing the Bayesian MDD defined as:

$$p^{(K)}(Y_{1:T}, X_{1:T}|\lambda) = \int \left(\prod_{t=1}^{T} p^{(K)}(Y_t, X_t|Y_{1:t-1}, X_{1:t-1}, \theta) \right) p^{(K)}(\theta|\lambda) d\theta, \tag{26}$$

where $\theta = (\beta, D)$ collects the VAR parameters. CCS show that the MDD can be approximated as follows:

$$\begin{aligned} p_*^{(K)}(Y_{1:T}, X_{1:T}|\lambda) \;=\; & \int \left(\prod_{t=1}^{T} p_G^{(K)}(Y_t, \alpha_t = \hat{\alpha}_t|Y_{t-p:t-1}, \alpha_{t-p:t-1} = \hat{\alpha}_{t-p:t-1}, \theta) \right) p^{(K)}(\theta|\lambda) d\theta \\ & \times \left(\prod_{t=1}^{T} \exp\left\{ N\mathcal{L}^{(K)}(\hat{\alpha}_t|X_t) \right\} \left(\frac{2\pi}{N} \right)^{K/2} |\hat{V}_t|^{1/2} \right). \end{aligned} \tag{27}$$

The integral in the first line is the standard MDD associated with a $W_t = [Y_t', \hat{\alpha}_t']'$ VAR and can be evaluated analytically under our prior distribution. The terms in the second line turn the MDD for $W_{1:T} = (Y_{1:T}, \hat{\alpha}_{1:T})$ into a MDD for $(Y_{1:T}, X_{1:T})$. The expression $\exp\left\{ N\mathcal{L}^{(K)}(\hat{\alpha}_t|X_t) \right\}$ is the maximized likelihood function for X_t, ignoring the dynamic features of the model. It is non-decreasing in K. The second part, $(2\pi/N)^{K/2}|\hat{V}_t|^{1/2}$, is a penalty for the dimensionality K, and hence avoids overfitting in the cross-sectional dimension.

Further Implementation Details. As explained in detail in CCS, a few additional steps are required for the implementation. First, we conduct a preliminary data transformation of the form $x = g(z|\vartheta)$, where z are the raw data and x are the transformed data. We explain in each of the empirical sections, which transformation we apply. Second, we need to choose

basis functions $\zeta_1(x), \ldots, \zeta_K(x)$. We use a cubic spline basis and provide further details below. Third, some of the cross-sectional data are top-coded, which means that the likelihood function in (4) needs to be adjusted accordingly. This adjustment is described in detail in the Online Appendix. Fourth, the cross-sectional data are typically not seasonally adjusted, whereas the aggregate data are. If empirically necessary, we project the $\hat{\alpha}_t$ estimates on seasonal dummies.

Fifth, there might be linear dependencies in the estimated $\hat{\alpha}_t$ vectors. We use Principal Components Analysis (PCA) to remove perfect collinearities from $\hat{\alpha}_t$, $t = 1, \ldots, T$ which leads to a vector $\hat{a}_t$ that satisfies $\hat{\alpha}_t' = \hat{\alpha}_*' + \hat{a}_t'\Lambda$. We replace $\hat{\alpha}_t$ by $\hat{a}_t$ in the definition of W_t. We treat Λ, which is obtained as an output of the (PCA), as fixed and use the relationship to adjust the affected model equations. If there are no exact linear dependencies, this step does not lead to a dimensionality reduction but it normalizes the log density coefficients to have unit variance. Sixth, unless otherwise noted, we undo the change of variables $x = g(z)$ and report densities for the original data z and not the transformed data x.

3.3 Extensions

We extend the CCS estimation described in Section 3.2 in several dimensions to handle various complications that arise in the subsequent application.

Stacked (Marginal) Densities. In most cases we combine the aggregate variables with one univariate cross-sectional density. However, in a robustness analysis we stack multiple marginal densities, which extends (2) to

$$\begin{aligned} Y_t &= B_{yy}Y_{t-1} + \sum_{m=1}^{J} \int B_{yl_m}(\tilde{x}^m)\ell_{t-1}^m(\tilde{x}^m)d\tilde{x}^m + B_{y0} + u_{y,t} \\ \ell_t^j(x^j) &= B_{l_jy}(x^j)Y_{t-1} + \sum_{m=1}^{J} \int B_{l_jl_m}(x^j, \tilde{x}^m)\ell_{t-1}^m(\tilde{x}^m)d\tilde{x}^m + B_{l_j0}(x) + u_{l_j,t}(x^j) \end{aligned}$$

for $j = 1, \ldots, J$. Suppose one approximates the log densities by $\ell^j(x^j) = \zeta^{j'}(x^j)\alpha_t^j$. Following the same calculations that lead from (2) to (6) and (8), we deduce multiple marginal densities can be accommodated by redefining $W_t = [Y_t', \alpha_t^{1'}, \ldots, \alpha_t^{J'}]'$. Because our micro-level observations are drawn from different databases, we do not model the joint distribution.[5]

[5]Two recent papers that model dynamics of joint distributions are Marcellino, Renzetti, and Tornese (2024a) and Bayer, Calderon, and Kuhn (2025).

Note that stacking densities increases the dimension of W_t and the VAR parameter space considerably.

Alternative Prior. Under Chan (2022)'s prior the prior covariance matrix for the elements of the B_j matrices in (20) are derived from a prior covariance matrix for the Φ_j elements in the one-step-ahead forecasting representation in (8). This construction makes it difficult to shrink coefficients in the lower-triangular parts of the B_j matrices. As an alternative, we are constructing an alternative prior in which we directly specify the covariance matrix for the B_j elements. This allows us to shrink spillover effects across densities in the stacked version of the model to zero.

Weighted Observations. We use financial income data x_{it} that are associated with weights ω_{it}. To account for the weights we replace $\mathcal{L}^{(K)}(\alpha_t|X_t)$ in (4) by

$$\left(\frac{1}{\sum_{i=1}^N \omega_{it}} \sum_{i=1}^N \omega_{it}\zeta(x_{it})'\right)\alpha_t - \ln \int \exp\left\{\zeta'(x)\alpha_t\right\} dx. \tag{28}$$

Mixed Frequency Analysis. The financial income data are available at annual frequency and we will combine them with quarterly macroeconomic time series. While there is a well-established state-space framework for mixed-frequency (MF) analysis of aggregate time series, e.g., Schorfheide and Song (2015) and references cited therein, low-frequency cross-sectional observations create additional complications. Suppose that the period t is a quarter and the annual cross-sectional observations are available in periods $t \in \mathcal{T}_A = \{4, 8, 12, \ldots\}$. A generative model would deliver joint dynamics for (Y_t, X_t) which can be converted into (Y_t, ℓ_t) dynamics. The standard MF-VAR framework can be easily adapted to settings in which one has either an estimate of ℓ_t based on X_t in periods $t \in \mathcal{T}_A$, say, or an estimate of $\frac{1}{4}\sum_{\tau=0}^3 \ell_{t-\tau}$. However, in practice one often has information on the cross-sectional densities of temporal averages $\frac{1}{4}\sum_{\tau=0}^3 x_{i,t-\tau}$ (or sums), which are different from averages of (log) densities. Carefully handling the time-aggregation problem requires assumptions about or knowledge of the time series law of motion of x_{it} and is beyond the scope of this paper.

Instead, we take the shortcut of postulating a joint vector autoregressive law of motion for Y_t and the density of $\frac{1}{4}\sum_{\tau=0}^3 x_{i,t-\tau}$.[6] We replace (1) by

$$\left(\frac{1}{4}\sum_{\tau=0}^3 x_{i,t-\tau}\right) \;\Big|\; p_t(x) \overset{iid}{\sim} p_t(x) = \frac{\exp\{\ell_t(x)\}}{\int \exp\{\ell_t(\tilde{x})\}d\tilde{x}}, \quad i = 1, \ldots, N, \quad t \in \mathcal{T}_A$$

[6]The MIDAS regressions of Marcellino, Renzetti, and Tornese (2024b) take the same shortcut.

and maintain (2). Recall that after replacing ℓ_t by its finite-dimensional approximation and "linearizing" the measurement equation, we obtain (9). To handle the MF setup we modify the equation by re-scaling the measurement error variance:

$$\hat{\alpha}_t = \alpha_t + N^{-1/2}\eta_t, \quad \eta_t \sim \mathcal{N}(0, \kappa_t \hat{V}_t^{-1}), \quad \kappa_t = \begin{cases} 1 & \text{if } t \in \mathcal{T}_A \\ 1\text{E}4 & \text{otherwise} \end{cases}. \tag{29}$$

Notice that for periods $t \in \mathcal{T}_A$ we can compute $\hat{\alpha}_t$ based on the available X_t observations. To generate pseudo-observations for the periods in which cross-sectional data are not available, we set $\hat{\alpha}_{t-\tau} = \hat{\alpha}_t$, $\tau = 1, 2, 3$, and $t \in \mathcal{T}_A$. Because for periods $t \notin \mathcal{T}_A$ the measurement error is scaled by $\kappa_{t-\tau} = 1\text{E}4$ the padded values $\hat{\alpha}_{t-\tau}$ have essentially no effect on the inference on α_t.[7] The advantage is that we can directly use the state-space routines developed in CCS to implement the estimation. The algorithms are reproduced in the Online Appendix.

4 Overview of Empirical Analysis and Aggregate IRFs

High-frequency Instruments. We use the internal instrument approach described in Section 3.1 to identify monetary policy surprises and the monetary policy rule. The instrumental variables are taken from Jarocinski and Karadi (2020), henceforth JK, who consider two surprise variables that allow them to separate unanticipated changes in monetary policy (monetary policy shocks) from the central bank's revelation of information about the state of the economy that is conveyed through interest rates (information shocks). The variables are surprises in the three-month fed funds futures ($ff4\_hf$) and surprises in the S&P 500 stock market index ($sp500\_hf$). It turned out that in our empirical analysis the use of just the three-month fed funds futures surprises and the combination of fed funds futures and stock market surprises yielded very similar IRFs for "regular" monetary policy shocks. Thus, we will only use the fed funds future surprises when estimating the monetary policy rule in Section 5.2 because we can avoid the challenges related to writing down a monetary policy rule with information shocks.

For the IRFs reported subsequently in Sections 5.3, 5.4, 6, and 7 we follow JK and use sign restrictions to separate the two shocks of interest. It is assumed that a contractionary monetary policy shock generates an interest rate increase and a drop in stock prices, whereas a positive information shock is associated with an increase in both interest rates and stock

[7] A common approach in the mixed-frequency state-space literature is to pad the low-frequency series with zeros and set the measurement error variance to infinity; e.g., Giannone, Reichlin, and Small (2008).

prices.[8] As in JK, to separate monetary policy shocks from information shocks one has to introduce a 2×2 orthogonal matrix Ω that links the orthogonalized instrument innovations to the structural shocks. Our Bayesian analysis requires a prior distribution $p(\Omega|\beta, D)$. Because Ω does not affect the likelihood function, conditional on the reduced-form VAR parameters, its prior does not get updated. We use a prior for Ω that is uniform over the identified set defined by the sign restrictions, conditional on the reduced-form parameters.

Observation Frequency. Ideally, we would like to conduct the empirical analysis at monthly frequency, as in JK, using a single data source for the micro-level observations on labor earnings, consumption, and financial income. Unfortunately, that is not feasible. We only use labor earnings and hours worked data from the CPS at monthly frequency. Consumption data are obtained from the CEX. Although interviews about consumption expenditures are conducted at monthly frequency, we time-aggregate to quarterly frequency to smooth out unit-level expenditures and reduce reporting errors. Our financial income data are at annual frequency and taken from the Distributional National Accounts microfiles provided by Piketty, Saez, and Zucman (2018). More details on these data sets are provided in Sections 5 to 7 below and in the Online Appendix. When working with quarterly and annual distributional data we have two choices: we can either time-aggregate the macro time series or we can employ the mixed-frequency framework described in Section 3.3. We do the former for the analysis of the consumption data after verifying that the time-aggregated quarterly macro variables deliver similar IRFs as the monthly macro time series.[9] For the analysis of the financial data we combine quarterly macro variables with annual micro data in the mixed-frequency framework.

Aggregate Variables. Following JK, all of our empirical models include the following macroeconomic variables, either at monthly or quarterly frequency: the average of the one-year constant-maturity Treasury yield serves as the monetary policy indicator. The advantage of using a one-year rate is that it remains a valid measure of monetary policy stance also when the federal funds rate is constrained by the zero lower bound (ZLB). The average of the S&P 500 stock price index is included log levels. Real GDP and GDP deflator are also included in log levels, and for monthly frequency interpolated based on Stock and Watson (2010). The excess bond premium (EBP) as an indicator of financial conditions. In

[8] As a robustness exercise we also considered the instruments from Nakamura and Steinsson (2018). The empirical results turned out to be very similar to the ones reported in this paper.

[9] Jacobson, Matthes, and Walker (2024) show that temporal aggregation of high-frequency policy shocks and price data can lead to adversely-signed responses of inflation to monetary policy shocks. In our application the inflation response does not change sign when we aggregate monthly data to quarterly frequency.

addition, we include an unemployment rate constructed from micro data. For the functional VAR with hours worked data we also use the fraction of units working 40 hours/week. For the functional VARs with cross-sectional consumption data we include aggregate per capita consumption from the National Income and Product Accounts (NIPA). For the functional VAR with the financial income distribution we add the fraction of units with zero financial income. The vector Y_t in (2) stacks the high-frequency instruments and the macro variables.

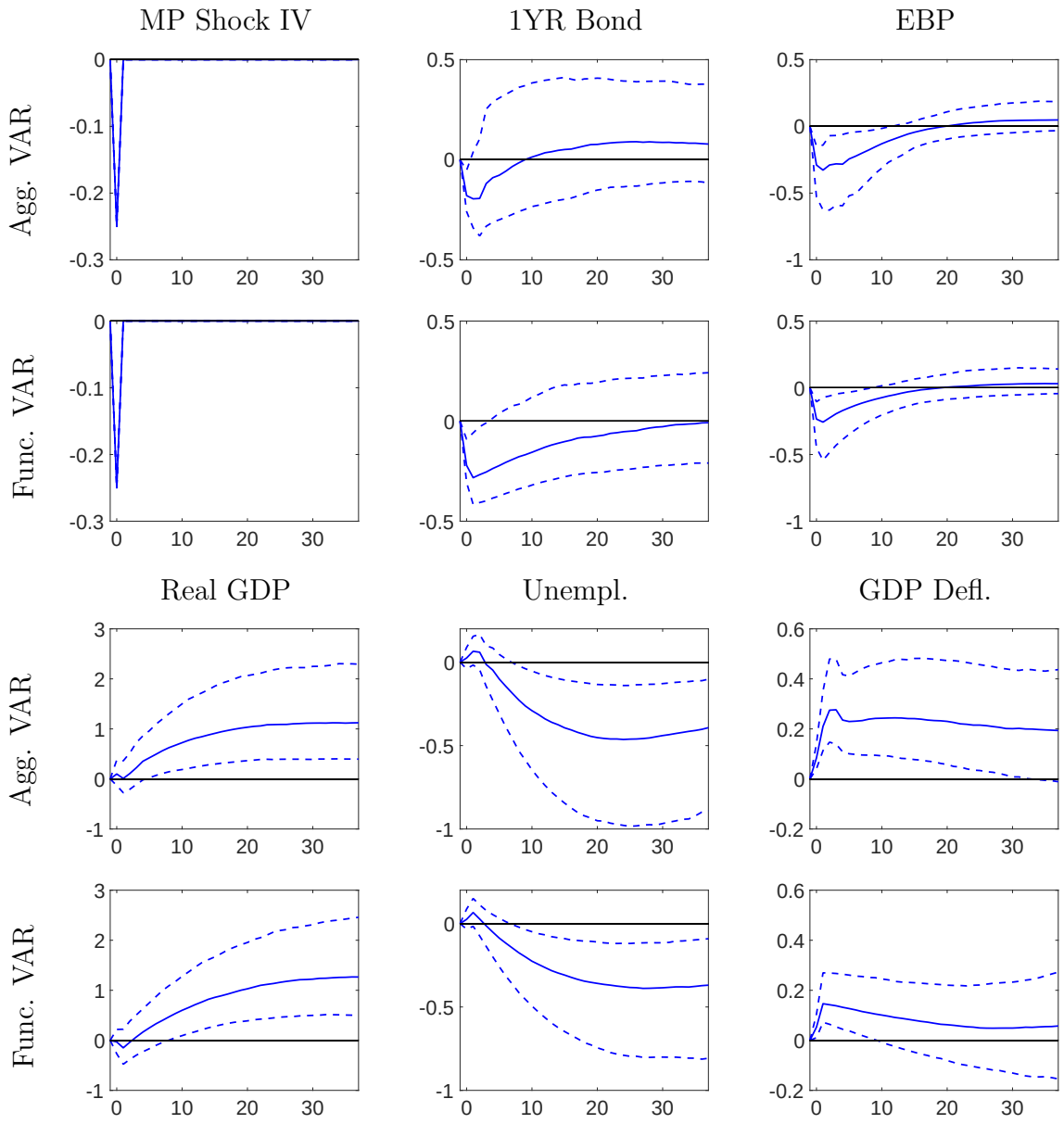
Single versus Multiple Univariate Densities. For the most part, we are combining Y_t with a vector of basis functions for a single cross-sectional density. In Section 6 we do discuss results from stacking the earnings and consumption densities. It turned out, though, that they are very similar to results from the corresponding models that only use one of the two densities. The number of parameters grows linearly with the number of densities included in the model, and we do not have enough time series observations in our sample to estimate the spillover coefficients precisely.

Sample Periods. The functional VARs with earnings or hours worked data in Section 5 are estimated based on monthly data, using the sample period 1990:M2 to 2016:M12. This sample has one missing value (the financial market disruption after the 9/11 terrorist attack in 2001:M9), which is replaced with zero. We use quarterly data from 1990:Q2 to 2016:Q4 for the analysis of the consumption and financial income distribution dynamics. For the earnings and consumption distribution we conduct a robustness analysis using pre-QE samples ranging from 1990:M2 to 2008:M10 (monthly) and 1990:Q2 to 2008:Q3 (quarterly), respectively. By and large, we obtained similar estimates which are discussed in the Online Appendix. While there are some subtle differences in some of the posterior median responses, the main effect of shortening the sample is wider credible intervals.

Splines, Knot Placement, and Model Selection. We fix the shape of the spline segments, and place the knots at percentiles of the cross-sectional observations pooled across time periods. For instance, for $K = 4$ we used the 25th, 50th, and 75th percentile. For $K = 6$ we added the 10th and the 90th percentile, and so forth. The MDD was used to select the dimensionality of the spline approximation K, the number of lags p, and the hyperparameter vector λ. For each of the empirical specifications in Sections 5 to 7 we report the selected model and the grid over which the selection was done in the Online Appendix.

Response of Aggregate Variables to a Monetary Policy Shock. Figure 1 depicts impulse responses of the aggregate variables to an unanticipated monetary policy shock for a monthly VAR estimated on Y_t only (Aggregate VAR) and for the monthly VAR that

Figure 1: Responses of Aggregate Variables to Monetary Policy Shock



Notes: Responses to a 25 basis point monetary policy shock. The Aggregate VAR sets $W_t = Y_t$, whereas the functional VAR uses $W_t = [Y_t', \hat{\alpha}_t']'$. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The plots depict 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior. GDP defl. and real GDP responses are percentage deviations from the steady state, whereas the other responses are absolute percentages. x-axis horizon is months.

includes the labor earnings density (see Section 5). The responses are normalized such that the surprise reduction in the three-month federal funds rate is 25 basis points (bp). The horizons on the x-axes of the plots correspond to months.

The responses obtained from the aggregate VAR and the functional VAR are very similar to each other and also quantitatively similar to the responses reported in Figure 2 of Jarocinski and Karadi (2020) after the latter are scaled by minus five. The presence of the cross-sectional variables has no substantial effect on the impulse response inference. The main difference is that the GDP deflator response in the functional VAR is slightly weaker. The width of the bands depends on the number of lags and the hyperparameters selected by the MDD. The results for the aggregate VAR are based on $p = 4$, whereas the functional VAR (earnings) uses $p = 1$. At the posterior median, the one-year bond rate moves approximately one-for-one with the federal funds rate surprise and then reverts back to steady state. The expansionary monetary policy leads to an increase of real GDP by slightly more than 1% and a reduction of the unemployment rate by 0.3 percentage points after three years. Prices increase by about 0.25% upon impact in the aggregate VAR and 0.15% in the functional VAR and slightly fall subsequently.

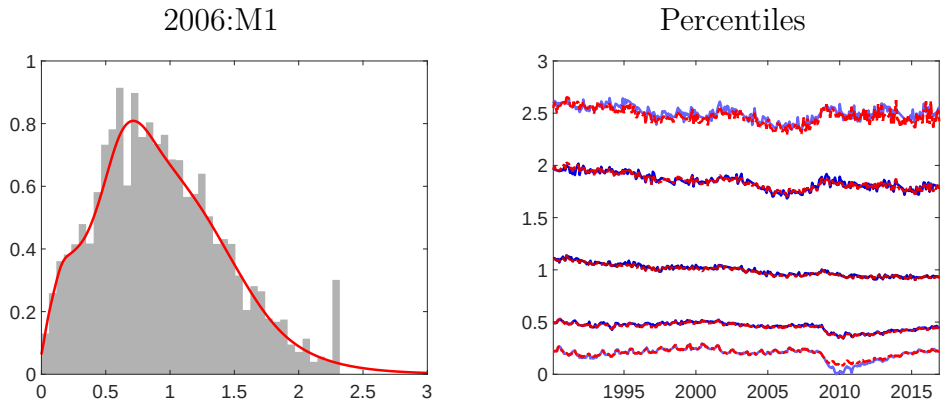
5 Monetary Policy and the Earnings Distribution

In the first part of the empirical analysis we examine the interaction of monetary policy and the distribution of labor earnings. The micro data are discussed in Section 5.1. In Section 5.2, we study the central bank's reaction to fluctuations in the labor earnings distribution. Monetary policy shock IRFs for the earnings distribution and statistics derived from it are presented in Section 5.3. Finally, Section 5.4 provides a comparison to estimates previously reported in the literature and presents IRFs for the hours worked.

5.1 Micro Data and Density Representation

The micro-level earnings data are constructed in the same way as in the application in CCS: weekly earnings (PRERNWA) are obtained from the monthly CPS through the website of the National Bureau of Economic Research (NBER) and scaled to annual earnings by multiplying with 52. Based on the CPS variable PREXPLF "Experienced Labor Force Employment" we generate an employment indicator which is one if the individual is employed and zero otherwise. This indicator is used to compute an aggregate unemployment rate which we include in the Y_t vector; see Section 4.

Figure 2: Fit of Estimated Labor Earnings Densities



Notes: Left panel: histogram (gray bars) and estimated density (red line, $K = 10$) for z. The continuous part of the density integrates to one minus the unemployment rate. Right panel: percentiles (10, 20, 50, 80, 90) computed from the estimated densities (red) and directly from the cross-sectional observations z (blue).

We standardize individual-level earnings by (2/3) of nominal per-capita GDP and apply the inverse hyperbolic sine transformation

$$x = g(z|\theta) = \frac{\ln(\theta z + (\theta^2 z^2 + 1)^{1/2})}{\theta} = \frac{\sinh^{-1}(\theta z)}{\theta}, \quad z = \frac{\text{Earnings}}{(2/3) \cdot \text{per-capita GDP}} \tag{30}$$

with $\theta = 1$. For small values of z the function is approximately equal to z and for large values of z it is equal to $\ln(z) + \ln(2)$. Below we will refer to x as transformed data and to z as original data. We assume that the transformed earnings are located on the interval $[0, \bar{x}]$ and use cubic splines as basis functions. We construct the spline from $x = \bar{x}$ to $x = 0$, using a linear element for the right tail:

$$\zeta_K(x) = \max\{\bar{x} - x, 0\}, \quad \zeta_k(x) = \left[\max\{x_{k-1} - x, 0\}\right]^3 \text{ for } k = K-1, \ldots, 1. \tag{31}$$

In the left panel of Figure 2 we plot the estimated density for the non-transformed earnings data z_{it}, see (30), for the period 2006:M1. We overlay a histogram constructed from the raw data. We regard the earnings distribution as a mixture of a point mass at zero and a continuous part. The point mass equals the unemployment rate and the continuous part integrates to one minus the unemployment rate. The density estimate smoothes out the histogram and distributes the point mass at the top-coded value into the right tail of the distribution. In the right panel we overlay empirical percentiles with percentiles computed from the estimated densities to show that the two measures are very similar.

5.2 Monetary Policy Rule and Earnings Distribution

We begin by examining the systematic reaction of the central bank to the cross-sectional distribution of earnings, which enters the monetary policy rule through the K-dimensional vector $\hat{\alpha}_t$. The reaction function coefficients that interact with $\hat{\alpha}_t$ are difficult to interpret for two reasons. First, unlike in the case of an inflation or output gap, we have K coefficients instead of a single coefficient. Second, fluctuations in $\hat{\alpha}_t$ have no direct economic meaning because the log-density is determined by the product of $\hat{\alpha}_t$ and the basis functions stacked in the vector $\zeta(x)$. Moreover, several transformations have been applied to the original $\hat{\alpha}_t$ vector; see Section 3.2.

To communicate the effect of the cross-sectional density on the policy rate R_t, we evaluate the policy rule under a time-invariant counterfactual $\hat{\alpha}^c$ and the observed $\hat{\alpha}_t$. The difference between the counterfactual interest rate and the actual interest rate measures the reaction of the central bank to the cross-sectional distribution. To illustrate the analysis consider the following stylized example. Suppose that the central bank sets the interest rate (we are omitting constants) according to the rule

$$R_t = \psi_0 + \psi_p(\pi_t + \ln P_{t-1}) + \rho_p \ln P_{t-1} + \rho_r R_{t-1} + \sigma_r \epsilon_{r,t}, \tag{32}$$

which represents a simplified version of (19). Fix the parameters at the posterior mean estimates $(\bar{\psi}_p, \bar{\rho}_p, \bar{\rho}_r, \bar{\sigma}_r)$, and denote the residuals that reproduce R_t by $\bar{\epsilon}_{r,t}$, $t = 1, \ldots, T$.[10] A counterfactual interest rate can be generated by replacing the observed inflation rate π_t with a counterfactual inflation rate π^c, and computing

$$R_t^c = \bar{\psi}_0 + \bar{\psi}_p(\pi^c + \ln P_{t-1}) + \bar{\rho}_p \ln P_{t-1} + \bar{\rho}_r R_{t-1} + \bar{\sigma}_r \bar{\epsilon}_{r,t}. \tag{33}$$

We can then compute the difference between the actual and the counterfactual interest rate. It is given by

$$R_t^c - R_t = \bar{\psi}_p(\pi^c - \pi_t). \tag{34}$$

The subsequent empirical results are generated by replacing the posterior mean estimates with draws from the posterior distribution.

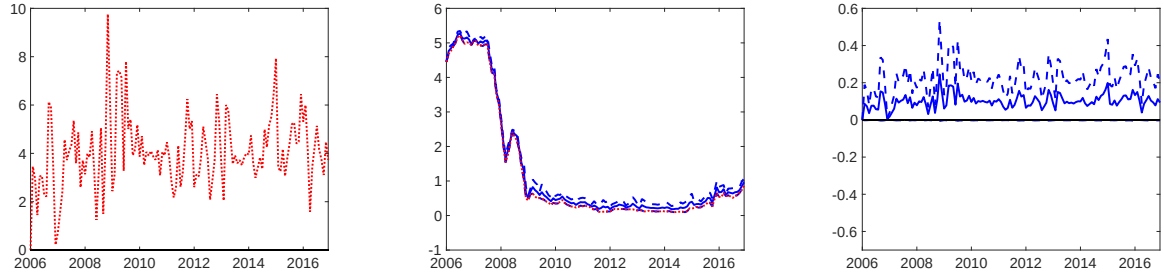
Results are presented in Figure 3. The counterfactuals are based on the state of the economy in 2006:M1. To fix ideas, we compute a counterfactual interest policy rate in Panel (i), assuming that the inflation rate had remained at its 2006:M1 level. The annualized

[10] In practice, the posterior mean estimates can be replaced by draws from the posterior distribution.

Figure 3: Counterfactual Interest Rates

Panel (i): Inflation Rate Fixed at 2006:M1

$\pi^c - \pi_t$ R_t^c and R_t $R_t^c - R_t$



Panel (ii): Earnings Distribution Fixed at 2006:M1

$p^c(x) - p_t(x)$ R_t^c and R_t $R_t^c - R_t$

2010:M1, 2014:M1 (dashed)

Notes: The base period for the counterfactual analysis is 2006:M1. Bottom left: we show difference between the cross-sectional earnings density in periods 2010:M1, 2014:M1 (actuals) and 2006:M1 (counterfactual). Other panels: actual (red), and 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution.

M-o-M inflation rate in 2006:M1 was 5.7%. Setting the counterfactual inflation rate to the 2006:M1 level implies a roughly 4% increase relative to the historical average between 2006:M2 and 2016:M12; see top left panel. The center plot of Panel (i) shows the posterior median and 80% bands for the counterfactual interest rate and overlays the actual interest rate. Uncertainty about the counterfactual interest rate is due to posterior uncertainty about the VAR parameters. Compared to real activity, inflation did not change much between 2006 and 2016, which is why the difference between R_t^c and R_t induced by the inflation counterfactual is small compared to the overall drop from 5% to almost 0% over this period. The difference $R_t^c - R_t$ is shown in the top left plot of Figure 3. According to a Taylor rule with a coefficient of $(1 - \rho_r)1.5$ with an interest rate smoothing parameter of $\rho_r = 0.95$, a

400 basis point increase in inflation, ceteris paribus, should raise interest rates by 30 basis points. This is quantitatively consistent with the 80th percentile of the posterior distribution of $R_t^c - R_t$.

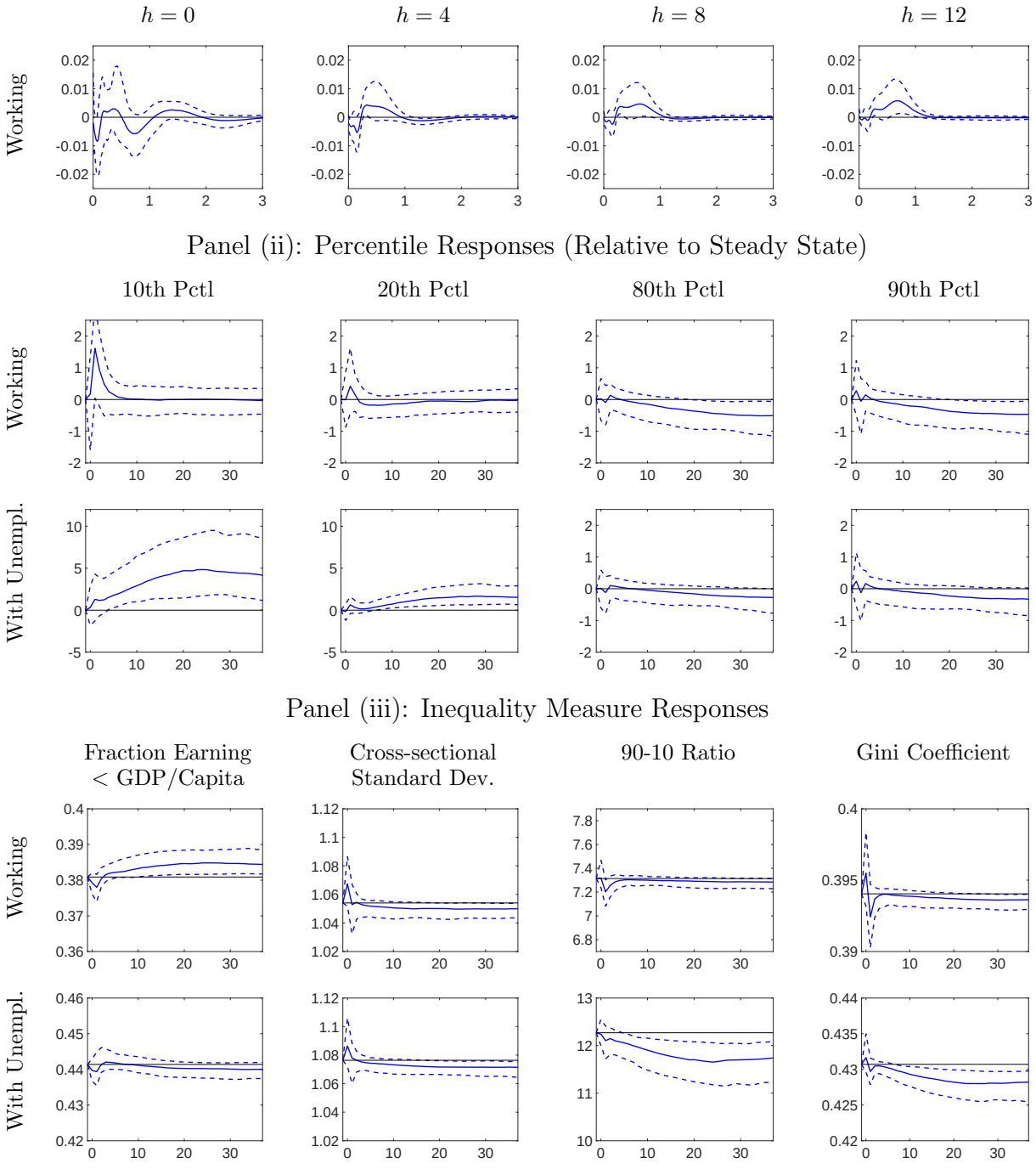
Results for the earnings distribution counterfactual are presented in Panel (ii) of Figure 3. The counterfactual focuses on the continuous part of the density represented by the $\hat{\alpha}_t$ coefficients. In the left plot we show density differentials $p_c(x) - p_t(x)$ for periods 2010:M1 and 2014:M1. The counterfactual density (2006:M1) has more mass on earnings that are less than the labor share of GDP per capita. The density differential is associated with a slightly lower interest rate. In the center plot the interest rate differential is visually swamped by the level change of the interest rate between 2006 and 2016. In the right panel we show $R_t^c - R_t$. The posterior median of the differential between 2008 and 2015 is negative, indicating that the increase in probability mass to the left of one, leads the central bank to lower interest rate. The magnitude of the posterior median interest rate change is similar to the inflation counterfactual. However, measurement is imprecise: the 10th percentile of the posterior distribution is consistently below zero, whereas the 90th percentile is above zero.

5.3 Earnings Distribution Response

Panel (i) of Figure 4 depicts the response of the continuous part of the earnings distribution to a monetary policy shock. The plots show the difference between the steady state earnings density and the shocked density for $h = 0$ (impact of the shock), $h = 4$, $h = 8$, and $h = 12$. The x-axes in these plots correspond to the level of earnings z. Recall that a value of one means that the earnings of the individual are equal to 2/3 (approximately the labor share) of GDP per capita. As mentioned previously, the earnings densities are normalized to integrate to one minus the unemployment rate. Because the unemployment rate drops in response to an expansionary monetary policy shock, the probability mass increases along the density response, relative to the steady state density.

To understand density differential IRFs, consider the following two examples. First, suppose that an aggregate shock shifts the cross-sectional distribution to the right, e.g., the steady state distribution is centered at $\bar{x} = 1$ and the shocked distribution is centered at $\tilde{x} = 2$. If the distributions are unimodal, then there is a value x_* between $\bar{x}$ and $\tilde{x}$ such that the differential is negative for $x < x_*$ and positive for $x > x_*$. Second, suppose the shock increases the dispersion of the cross-sectional distribution, but does not affect the location $\bar{x}$. Then the density differential will be negative near $\bar{x}$ and positive further away from $\bar{x}$.

Figure 4: Response of Earnings to Monetary Policy Shock



Notes: The system is in steady state at $h = -1$ and a 25bp monetary shock occurs at $h = 0$. 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution. Panel (i): Differences between the shocked and the steady state cross-sectional density (continuous part, normalized to $1 - UR_t$) of z. Panels (ii) and (iii): horizon h (x axes) is in months..

The density differential patterns in Panel (i) are more complicated than in the two examples. The 10th, 20th, 50th, 80th, and 90th percentiles of the baseline labor earnings distribution (without point mass at zero) are 0.34, 0.55, 1.05, 1.90, 2.53, respectively. Upon impact of the shock ($h = 0$) at the posterior median there is some loss in probability mass for income ratios near zero and between 0.50 and 1.08. This mass is added to income levels between 0.15 and 0.50, and between 1.08 and 1.95, respectively. However, the 80% bands are wide and the sign of the responses is mostly ambiguous. Beyond $x = 2$ the probability mass stays essentially the same. The pattern changes as the monetary policy shock propagates through the economy. For $h = 8$ and $h = 12$ the dominant effect is an increase in probability mass between $z = 0.21$ and $z = 1.11$, generated by the increase in employment.

The density response can be converted into IRFs for statistics derived from the earnings distribution. Panel (ii) of Figure 4 shows the responses of the percentiles of the earnings distribution as a function of the horizon h. Results reported in the row labeled (conditional on) "Working" are based on percentiles of the continuous part of the earnings distribution, i.e., earnings conditional on working. The percentile plots in the row "With Unempl." account for a point mass of zero labor earnings, capturing the individuals that are unemployed. The responses are reported as percentage changes relative to the respective baseline levels. According to the top left response in Panel (ii), there is a 1.6% increase of the 10th percentile upon impact, which raises labor earnings from 0.346 to 0.352 times the (approximate) average aggregate labor share. This amount may appear small, but it is of the same order of magnitude as the increase of aggregate GDP after about 20 months; see Figure 1. However, the effect dies out quickly. Moreover, in the long-run the 80th and 90th percentile distributions drop by 0.51%.[11]

The bottom row of Panel (ii) contains percentile responses for the case in which the point mass at zero is included. We start from a baseline unemployment rate of 6%, which is the average in our sample. The inclusion of the point mass lowers the baseline percentiles to 0.20, 0.45, 0.99, 1.84, and 2.47, respectively. The magnitude and patterns of the responses of the 10th and 20th percentiles are markedly different now. Notice that the plots for these two IRFs are scaled differently in Panel (ii). Including the point mass, the 10th percentile roughly corresponds to the 4th percentile of the earnings distribution conditional on working, which is about 0.20 times 2/3 of GDP. The monetary policy shock now has two effects. First,

[11] The pointwise posterior medians of the density differentials do not generate the posterior medians of the percentile responses. In the Online Appendix we overlay the responses generated from the posterior mean parameter estimates, which look very similar to the posterior medians plotted in Figure 4.

it moves the 4th percentile conditional on working. Second, because the unemployment rate falls by about 0.5 percentage points, the 10th percentile now corresponds (approximately) to the 4.5th percentile of the shocked earnings distribution conditional on working. In combination, the 10th percentile in the bottom left plot of Panel (ii) increases by 4.7% at the posterior median. For the 20th percentile the response is a 1.5% increase after 20 months. In comparison, the responses of the 80th and 90th percentiles are flat.

We proceed by computing four measures of earnings inequality from the cross-sectional densities without ("Working") and with ("With Unempl.") the point mass at zero: the fraction of individuals earning less than the labor share of GDP, the Gini coefficient, the ratio of the 90th and the 10th percentile of the income distribution (90-10 ratio), and the cross-sectional standard deviation. Impulse responses for the inequality measures are depicted in Panel (iii) of Figure 4. Instead of centering the IRFs at zero, we center them at their steady state levels. Note that under "Working" the steady state numbers are smaller than under "With Unempl." because conditioning on being employed eliminates zero earnings and reduces inequality.

Under "Working" the monetary policy shock increases the fraction of units earning less than GDP per capita over the medium run, which is consistent with a model of sorting: it is mostly low productivity individuals who transition from unemployment into employment in response to the expansion. Except for some variability upon impact, conditional on working, the other earnings inequality measures stay essentially unchanged in the medium run. The pattern changes if the computation of inequality measures includes the point mass at zero earnings for the unemployed. Now the response of the fraction of earnings below the labor share of GDP per capita stays essentially flat, which is consistent with the newly employed earning comparatively low wages. Because some unemployed individuals transition from zero earnings to positive earnings, the 90-10 ratio and the Gini coefficient fall over time. For instance, at the posterior median, the 90-10 ratio drops from 12.3 to 11.7 after 36 months. The path of the Gini coefficient resembles the path of the 90-10 ratio, falling from 0.431 to 0.428.

An expansionary monetary policy lowers the real interest rate temporarily, which creates a disincentive to save, and stimulates economic activity in the current period. Labor demand rises and earnings increase. This increase is most pronounced at the 10th percentile and operates along the extensive margin: low productivity workers transition from unemployment into employment. This monetary policy shock responses are consistent with the results in Figure 11 of CCS, where the strong positive response of 10th percentile of the labor earnings

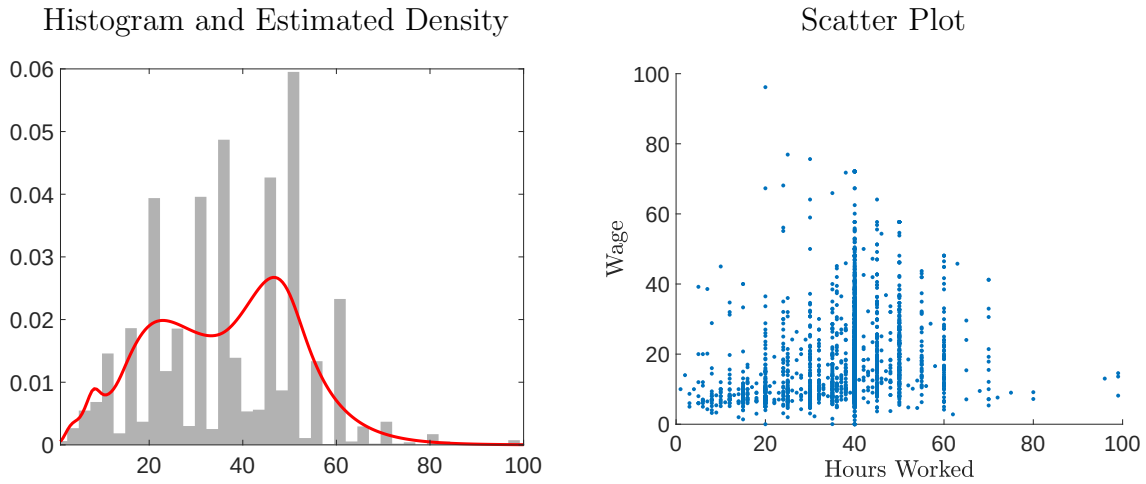
distribution to a technology shock vanished after the point mass at zero was removed from the percentile calculation.

From a policy making perspective, our results suggest tracking the effect on employment is of first-order importance for earnings inequality, whereas a more granular inspection of shape changes of the earnings distribution is of secondary importance. To support this claim we recomputed the IRFs of the percentiles and inequality measures reported in the "With Unempl." rows of Panels (ii) and (iii) of Figure 4, holding the shape of the continuous part of the earnings distribution (determined by that $\hat{\alpha}$ responses) fixed. Thereby we isolate the effect of the point mass as zero, which corresponds to the unemployment rate, on the percentiles and inequality measures. The resulting IRFs are reported in the Online Appendix. The percentile responses are visually indistinguishable from the ones reported in Figure 4 and so is the IRF of the 90-10 ratio. The only aspect that is not fully captured by the simplified analysis is a slight reduction in the cross-sectional standard deviation. Thus, according to our estimates, a policy maker can do the following back-of-the-envelope calculation: start from a baseline cross-sectional distribution. Then compute the response of statistics of interest by using an unemployment response to shift the point mass at zero earnings and update the normalization of the continuous part of the distribution.

5.4 Further Discussion

Empirical Findings in Other Studies. There are a number of earlier studies that examined the effect of monetary policy shocks on earnings inequality. Some studies included the inequality measures directly into a VAR or linear projections. For instance, Coibion, Gorodnichenko, Kueng, and Silvia (2017) report IRFs to a 100 bp increase in the monetary policy rate, estimated on U.S. data. They find that in the medium run the Gini coefficient on earnings rises by about 0.0025. Adjusting for the different shock size, their estimate is slightly smaller than ours. Furceri, Loungani, and Zdzienicka (2018) consider a panel of 32 advanced and emerging market economies. They report an estimate (converted into our scale) of 0.005, which is larger, but in the same order of magnitude as our estimates. Other studies did not aggregate their results into the response of inequality statistics, but they also find that monetary policy shocks predominately affect the left tail of the earnings distribution, and that the effects are driven to a large extent by transitions into and out of employment, e.g., Amberg, Jansson, Klein, and Rogantini Picco (2022), Mitman, Broer, and Kramer (2022), and Lenza and Slacalek (2024).

Figure 5: Hours Worked and Wages in 2006:M1



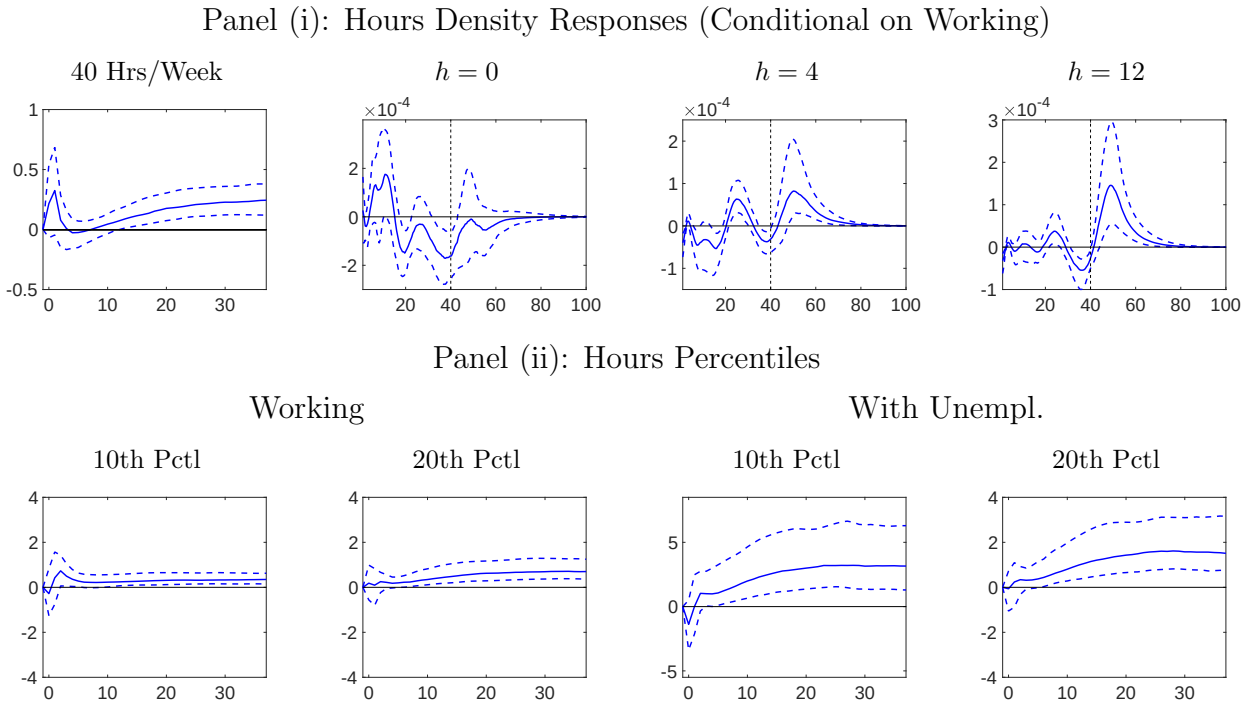
Notes: Left panel: histogram (gray bars) of hours worked and estimated density (red line, K=10). Right panel: wages (dollars/hour) and hours worked for a randomly selected set of 3,000 units.

What Does a Calibrated Model with Indivisible Labor Supply Imply? Our findings are broadly consistent with a heterogeneous agent model with indivisible labor supply as in Chang and Kim (2006). This class of models generates a negative correlation between idiosyncratic productivity and reservation wage. In turn, low-skill workers enter the labor market during booms, when the demand for labor is sufficiently high such that the wage per efficiency unit exceeds their reservation wage. At this point the labor earnings switch from zero to a positive value, which reduces labor earnings inequality. Ma (2023) incorporates this mechanism into a HANK model and shows that in his model an expansionary monetary policy shock raises wages and more low productivity individuals are starting to work, which raises earnings in the left tail of the distribution. Under his calibration the Gini coefficient for labor earnings (on a scale from 0 to 1) drops by approximately 0.001 upon impact.[12] In our estimated VAR the drop is between 0.001 and 0.003, which is very similar. As in Ma (2023)'s HANK model, the reduction in the earnings inequality is mainly driven by the fall in unemployment.

Hours Worked. The micro-level labor earnings were constructed as the product of "hours per week at one's main job" and "hourly earnings." We have previously seen that much of the effect of monetary policy shocks operates through unemployed individuals entering the

[12]See Figure 3 in Ma (2023). He considers a 100 bp monetary policy shock and measures the Gini coefficient on a scale from 0 to 100. Thus, $-0.4/(4 \cdot 100) = -0.001$.

Figure 6: Responses of Hours Worked to MP Shock



Notes: The system is in steady state at $h = -1$ and a 25bp monetary shock occurs at $h = 0$. 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution. Panel (i) Hours worked density is normalized so that it integrates to one minus the unemployment rate and the fraction of units working 40 hours/week. 40 hours/week is marked by the vertical lines. Panel (ii): horizon h (x axes) is in months.

employment state. We now examine to what extent the monetary policy shock affects the distribution of hours worked. The left panel of Figure 5 shows a histogram of hours worked for 2006:M1 after we took out a 56% point mass at 40 hours/week. We overlay the log spline density estimates. They smooth out the bars of the histogram which mostly reflect rounding of respondents. Thus, similar to earnings, we treat hours as a mixture of a continuous and discrete distribution. The discrete part covers the point mass at 40 hours/week, which we add to the vector Y_t, and, if we include the unemployed units in the analysis, a second point mass at zero hours. The right panel of Figure 5 shows a scatter plot of wages (labor earnings divided by hours, measured in dollars/hours) and hours worked for 3,000 units randomly selected from the 2006:M1 sample. Among units working less than 40 hours/week there is a positive correlation between wages and hours worked. This correlation turns slightly negative for units working 40 to 60 hours.

Impulse responses of the hours worked distribution are plotted in Figure 6. In Panel (i)

we show the IRF of the fraction of units working 40 hours/week and density differentials for the continuous part of the hours worked distribution, normalized so that it scales to one minus the unemployment rate. The fraction of units working 40 hours/week increases upon impact, drops back to zero, and then increases again from $h = 5$ onwards. The density response is easiest to interpret for $h = 12$. Here probability mass shifts from the left of 40 hours/week to the right, with units slightly increasing their hours worked. At $h = 0$ the density differential is positive below 20 hours/week and then turns negative. Above 40 hours/week the median differential is close to zero. The positive differential is consistent with part-time workers entering the employment state. The subsequent negative differential, in particular for hours worked between 30 and 40 in combination with a rise in the point mass at 40 hours, is consistent with individuals shifting toward a 40 hour work week. Units working 55 hours or more do not adjust their work week.

The plots in Panel (ii) depict the responses of the 10th and 20th percentiles of the hours worked distribution. In the two columns labeled (conditional on) "Working" the percentiles are computed from the continuous part of the distribution and the point mass at 40 hours per week. The baseline values of the two percentiles are 23.6 and 36.0. We observe a drop on impact and then a subsequent rise to about 1% for the 10th percentile and 0.5% for the 20th percentile. While the units make adjustments along the intensive margin, the hours worked responses cannot fully explain the movements in the labor earnings distribution. Conditioning on working, upon impact of the shock, $h = 0$, at the 10th percentile hours worked fall by about 1.5%, while labor income rises by approximately the same amount relative to GDP, but GDP does not yet respond. After two periods hours worked are about 1% above their initial level, whereas the percentiles of relative labor earnings are close to zero again. At the same time, real GDP is starting to rise. In the long-run GDP is higher, and there are more units working 40 hours/week, but the 10th and 20th percentiles (conditional on working) are back to their initial values.

The third and fourth columns of Panel (ii) of Figure 6 titled "With Unempl." include the point mass of zero hours, given by the fraction of unemployed. The baseline values of the 10th and 20th percentiles are by construction lower: 16.8 and 31.1, respectively. If the point mass at zero is included in the percentile calculation, then the long-run responses of the 10th (plotted with a different y-axis range) and 20th percentiles are driven by the increase in unemployment. We did not show results for the 80th and 90th percentiles because they do not respond to a monetary policy shock.

6 Consumption Distribution Response

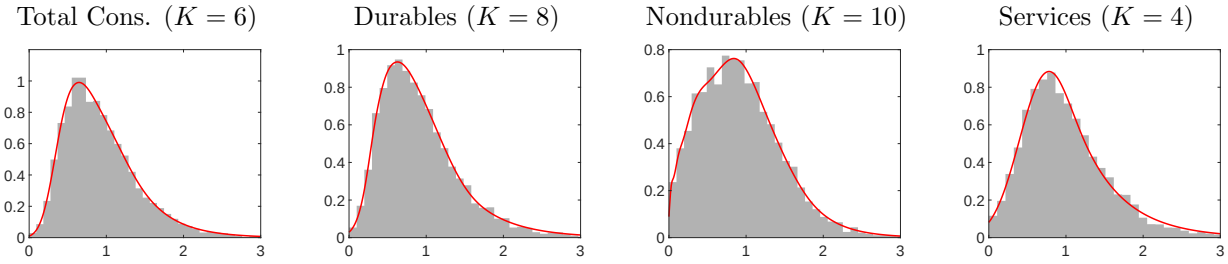
In the second part of the empirical analysis we estimate the response of the cross-sectional consumption distribution to a monetary policy shock. We first discuss the micro data and then present the baseline distributional responses. We proceed by disaggregating consumption into components, and finally stack earnings and consumption distributional data in a single model. As mentioned in Section 4 the vector of aggregate variables Y_t now also includes an aggregate personal consumption expenditures per capita series.

Micro Data. The consumption data are obtained from the CEX, conducted by the Bureau of Labor Statistics. Although interviews about consumption expenditures are conducted at monthly frequency, we aggregate to quarterly frequency to smooth out unit-level expenditures and reduce reporting errors. Each consumption unit (CU) provides expenditure data for three consecutive months. If a CU was surveyed in 1990Q1, consumption responses could refer to one of three possible combination of months: (i) Oct. 1989, Nov. 1989, Dec. 1989; (ii) Nov. 1989, Dec. 1989, Jan. 1990; or (iii) Dec. 1989, Jan. 1990, Feb. 1990. To convert this information into quarterly expenditure data, we add the three monthly values and assign the sum to the quarter that covers at least two of the months for which responses were obtained. Lastly, we use the count of males and females over 16 years old within a CU to compute per capita expenditures. Following this approach, we compile cross-sectional observations for the periods 1990:Q1 to 2016:Q4.

We construct four measures of consumption: (i) consumption of nondurable goods includes food and beverages, clothing and footwear, gasoline and other fuel, personal care, reading materials, and tobacco. (ii) Consumption of services encompasses child-care spending, hospital and nursery services, household utilities and energy, recreation services, financial services, accommodations, telecommunication, transportation. (iii) Total consumption encompasses consumption of nondurables and services and also covers spending on items such as recreational goods, furniture and furnishing, jewelry and watches, housing and rent (or imputed rental value), health care and insurance, education, cash support for college students, vehicle purchases. (iv) Finally, we define consumption of durables as total consumption minus consumption of nondurables and services.[13]

[13]Starting point of the definition of consumption and its components is Coibion, Gorodnichenko, Kueng, and Silvia (2017), but then we made some adjustments to add items that were excluded from their analysis. Detailed information on the data construction can be obtained from our replication files.

Figure 7: Estimated Consumption Densities for 2006:Q1



Notes: Histograms (gray bars) and estimated densities (red lines) for z.

The micro-level consumption data are scaled by aggregate nominal consumption from NIPA, divided by the population over 16 years old. It is well known that CEX consumption per capita does not aggregate to NIPA consumption per capita. Thus, we assume that only a fraction $\tilde{c}$ of total individual consumption c_{it}^* is reported to the CEX: $c_{it} = \tilde{c}c_{it}^*$. We compute separate constants for each consumption category: total (T), non-durables (N), services (S), and durables (D):

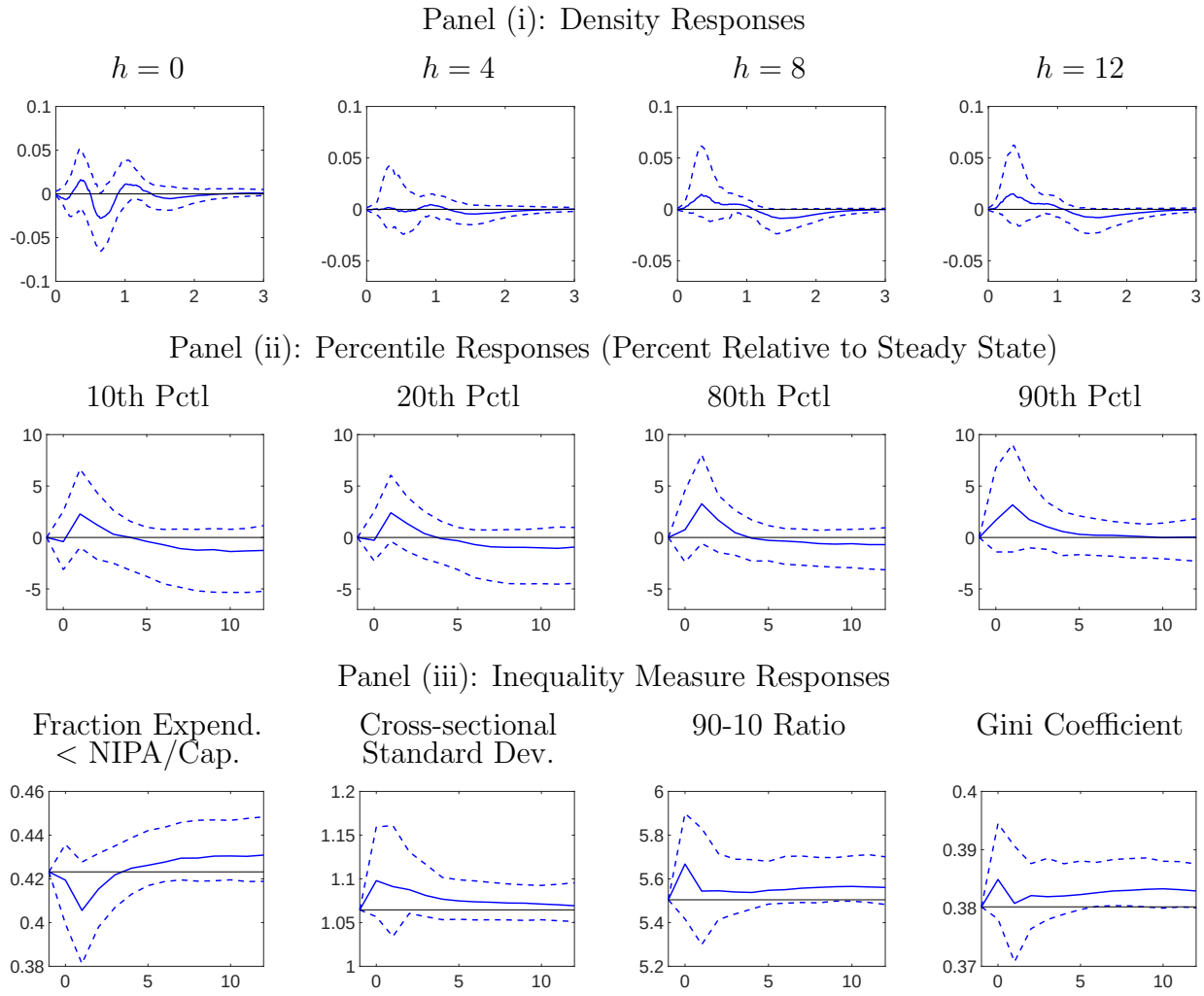
$$\tilde{c}^s = \frac{1}{T}\sum_{t=1}^{T} \text{median}(c_{1t}^s, \ldots, c_{Nt}^s)/C_t, \quad s \in \{T, N, S, D\}. \tag{35}$$

The scaling constants are $\tilde{c}^T = 0.46$, $\tilde{c}^D = 0.02$, $\tilde{c}^S = 0.07$, $\tilde{c}^N = 0.05$ and capture both the underreporting and the fraction of the components N, S, D as part of total consumption. We then define $z_{it}^s = c_{it}^s/(\tilde{c}^s \cdot C_t)$ such that a value of one means that the unit approximately consumes at the aggregate per capita level.[14] Histograms and the density estimate for total consumption and its three components for 2006:Q1 are plotted in Figure 7. Finally, as for the earnings data, we apply the inverse hyperbolic sine transformation to obtain x_{it} before fitting the log spline densities.

Baseline Consumption Distribution Responses. Figure 8 summarizes the response of the cross-sectional consumption distribution to a monetary policy shock. Panel (i) shows density differentials, Panel (ii) depicts the responses of percentiles of the consumption distribution, and Panel (iii) documents the response of inequality measures. The differentials

[14]Suppose there is no underreporting and the median unit i consumes at the aggregate per capita levels C_t^s, $s \in \{T, N, S, D\}$. Then $\tilde{c}^s = T^{-1}\sum_{t=1}^{T} C_t^s/C_t$, which is the average share of consumption of category s among total consumption. In turn, our definition of z_{it}^s implies that a value of 1 for category s means that the unit is consuming the aggregate share of category s in total consumption.

Figure 8: Response of Consumption to Monetary Policy Shock



Notes: Responses to a 25 basis point monetary policy shock. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The plots depict 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution. Panel (i): we depict differences between the shocked and the steady state cross-sectional density of z at various horizons. Panels (ii, iii): horizon h (x axes) is in quarters.

in Panel (i) are with respect to a steady state distribution with percentiles 0.44 (10th), 0.58 (20th), 1 (median), 1.73 (80th), and 2.4 (90th). According to the pointwise posterior median estimates of the density differentials there is a decrease in the fraction of units consuming less than per capita consumption at horizon $h = 0$ and an increase at horizons $h = 8$ and $h = 12$; see also left-most plot in Panel (iii).

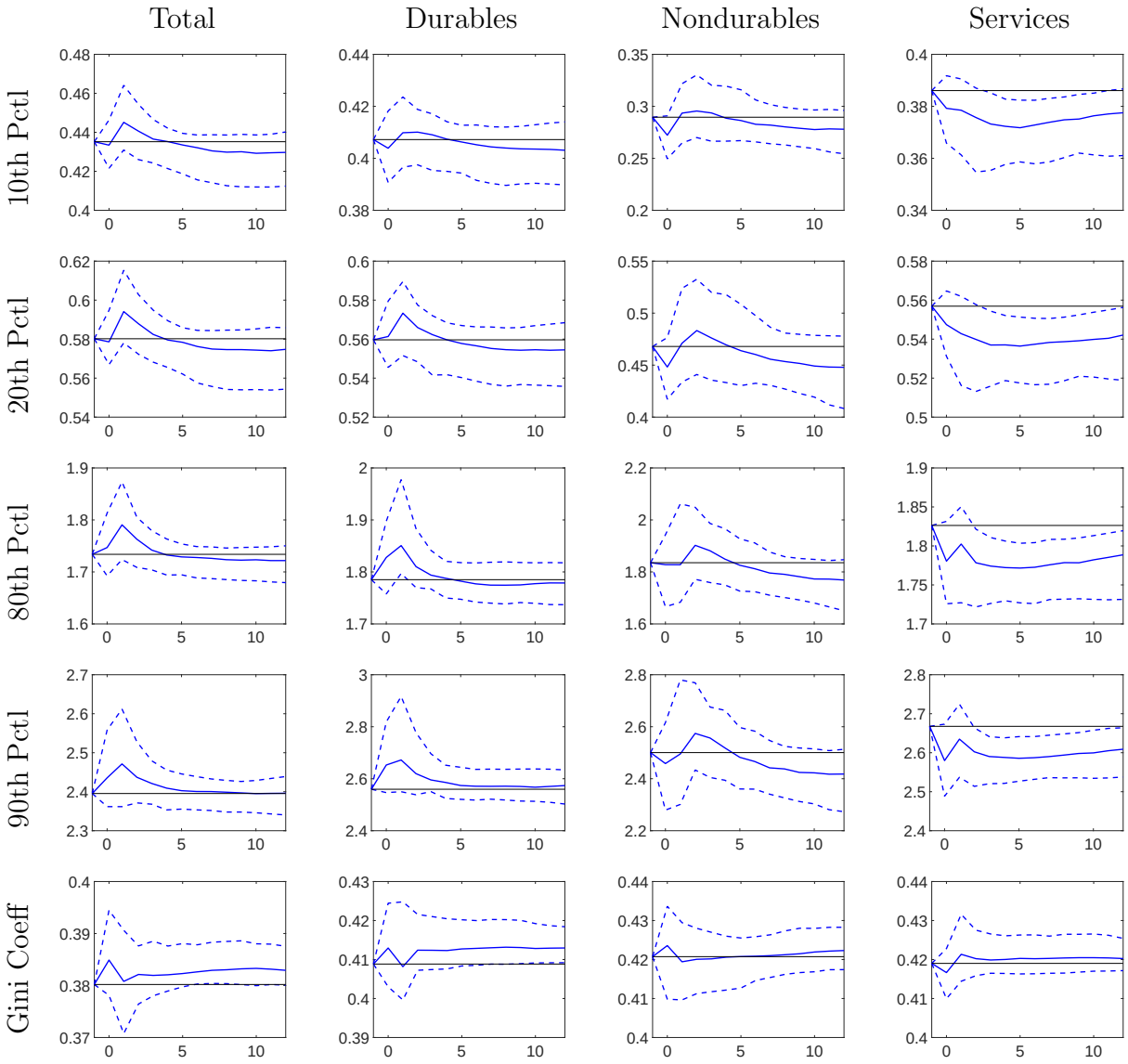
The percentile responses reported in Panel (ii) are in percentage deviations from per-

centiles of the steady state distribution. For instance, a 5% increase at the median would imply that consumption rises from 1 to 1.05. Based on the posterior median the 10th and 20th percentiles rise above their steady state values at $h = 1$, before they revert back to zero four quarters after impact. This is consistent with the $h = 4$ posterior median of the density differential which is essentially zero. Over the subsequent quarters the responses remain 0.94% (20th percentile) to 1.26% (10th percentile) below their baseline values. The responses of the 80th and 90th percentiles are similar in shape but quantitatively slightly larger, which is most apparent from the credible bands. Because our unit-level consumption is expressed as a ratio relative to aggregate per capita consumption, see the definition of z_{it}^s following (35), the percentage change of the percentile of the non-normalized consumption distribution is the sum of the response in Figure 8 and the response of aggregate consumption (in the Online Appendix). Consider the response of the 10th and 90th percentiles after 10 quarters. Combining the posterior median estimates from the two figures, we deduce that in absolute terms consumption at the 10th percentile rises by $0.65\% - 0.14\% = 0.51\%$, whereas it rises by $0.65\% + 0.01\% = 0.66\%$ at the 90th percentile.

Panel (iii) of Figure 8 depicts the implied responses of our four inequality measures. As previously discussed, the posterior median response of the fraction of households consuming less than the aggregate per capita amount falls initially, but stays above its steady state value from horizon $h = 5$ onwards. The other three inequality measures – the cross-sectional standard deviation, the 90-10 ratio, and the Gini coefficient peak upon impact ($h = 0$) and stay above their steady state levels subsequently. Thus, according to the posterior median estimates, consumption inequality generally increases in response to an interest rate cut. However, there is a considerable amount of uncertainty as the credible bands cover both positive and negative values.

Consumption Components. In Figure 9 we compare the responses of the distribution of consumption components: durables, nondurables, and services. The first column of the figure reproduces the percentile and Gini coefficient responses for total consumption from Figure 8, except that we now report the percentile responses in levels rather than in percentage changes relative to the steady state distribution. Because of the normalization constants defined in (35), the steady state levels for the percentiles of the consumption components are roughly the same. The responses of the inequality statistics associated with the consumption of durables look very similar to the responses of total consumption. The responses of the distribution of nondurables and services, on the other hand, are quite different. Even at the posterior median, the Gini coefficient responses are essentially zero. The posterior medians

Figure 9: Responses of Consumption Distribution By Components



Notes: Responses to a 25 basis point monetary policy shock. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The plots depict 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution. Horizon h (x axes) is in quarters. IRFs in the first column are identical to those in Figure 8.

of the percentile responses of nondurable consumption oscillate at short horizons and then fall below the steady state level. The service percentiles drop upon impact below their steady state values and remain low for the subsequent ten quarters, indicating a shift from services to durables. Recall, however, from Figure 1 that total per capita consumption,

used to normalize unit-level consumption, also rises, which mitigates the drop in service consumption.

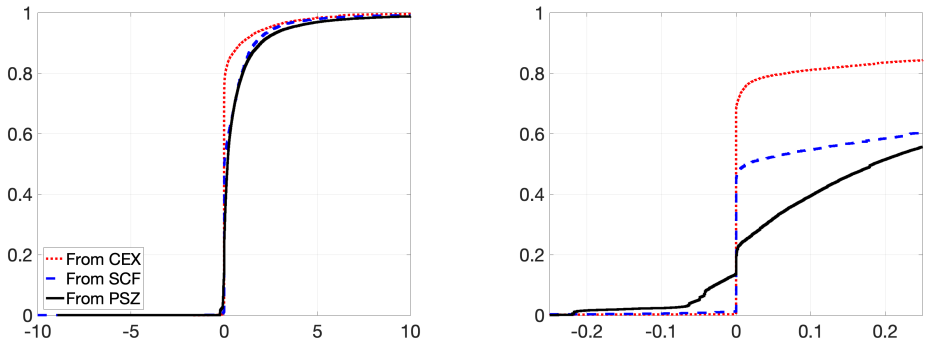
Relative Prices. A caveat to our analysis is that we are using value measures of consumption (prices times quantities) to aggregate across different product groups. If units are consuming products in different proportions, then a shock to relative prices can change the distribution of the value of consumption without changing the distribution of consumption quantities.

Stacking Earnings and Consumption Data. So far, we used two separate VARs for the analysis of the labor earnings and consumption densities. As discussed in Section 3.3 it is straightforward to combine the aggregate variables Y_t with two marginal densities, by stacking the associated $\hat{\alpha}_t$ coefficients. The advantage is that one can capture dynamic spillovers between the two densities. The disadvantage is that the model complexity increases substantially. If each density is approximated by K basis functions, then the number of regressors increases from $1 + p(n_y + K)$ to $1 + p(n_y + 2K)$. To implement the stacked estimation we time-aggregate the earnings data to quarterly frequency and we re-do the MDD-based model determination. Impulse responses from this analysis are reported in the Online Appendix. It turned out that stacking the densities leads to essentially the same IRFs as the estimation of separate functional models for each of the densities. We also considered an alternative prior that allows us to shrink blocks of the B coefficient matrix directly to zero. We report results from this alternative prior in the Online Appendix. It turned out that they are essentially identical to the baseline prior.

7 Financial Income Distribution Response

The slight increase in consumption inequality measured in the previous section may appear puzzling in view of the decrease in earnings inequality documented in Section 5. The consumption response is determined by the income response in combination with the marginal propensity to consume (MPC) and low-income households, i.e., those that experience a strong rise in labor earnings, tend to be those with high MPCs. However, in Section 5 we only considered labor income. Wealthy households typically have a significant amount of financial income, which may be directly affected by the monetary policy intervention, e.g., rising stock and bond prices in response to an interest rate cut. For instance, applying panel

Figure 10: Financial Income Distribution in 2012: PSZ vs. CEX vs. SCF

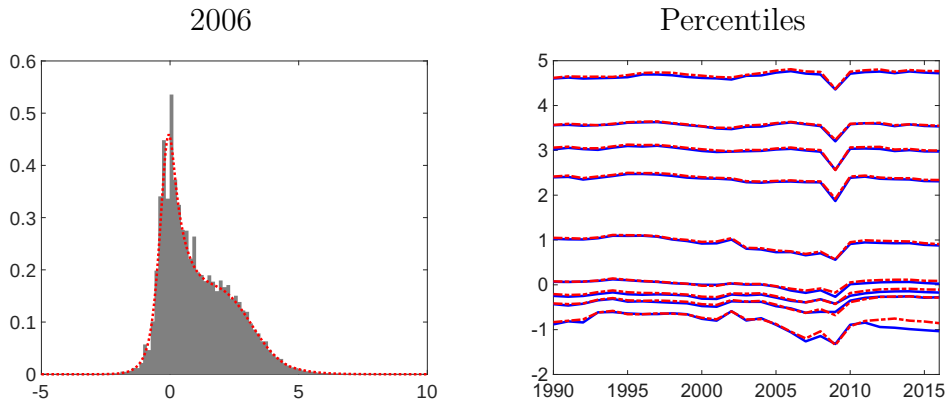


Notes: The figure depicts empirical cdfs for z, including negative values for financial income. Financial income from the three sources is standardized by 1/3 of GDP per capita to obtain z.

local projections using the Swedish administrative data Amberg, Jansson, Klein, and Rogantini Picco (2022) find that high-income individuals experience a substantial rise of capital income, more so than other individuals. This channel is also present in a calibrated DSGE model by Lee (2021), who finds that consumption inequality rises in his model in response to an expansionary monetary policy shock because it boosts profits and equity prices, which benefits wealthy households with high level of consumption. In view of these considerations, we now examine the financial income distribution response to a monetary policy shock.

Micro Data. While there are several potential micro-level data sources for the U.S., many of them either have limited cross-sectional coverage or they are only available at a low frequency that is not suitable for business cycle analysis. For instance, the Survey of Consumer Finances (SCF) is conducted every three years and the Panel Study of Income Dynamics (PSID) collects data bi-annually. In the working paper version Chang and Schorfheide (2024) of this article we analyzed household level financial income data from the public-use income data of the CEX family interview files. However, relative to the SCF, the CEX misses high-financial-income households who may benefit from the interest cut through stock market gains or dividend payouts. In view of these concerns we decided to use the annual Distributional National Accounts microfiles provided by Piketty, Saez, and Zucman (2018). These authors distribute NIPA financial income across synthetic individuals based on income tax records. We use a broad measure of personal factor capital income that includes housing asset income, equity asset income, interest income, business asset income, pension and insurance asset income, interest payments, sales and excise taxes allocated to capital and subsidies allocated to capital.

Figure 11: Fit of Estimated Labor Earnings Densities



Notes: Left panel: histogram (gray bars) and estimated density (red line, $K = 11$) of transformed data x (excluding zero incomes). Right panel: percentiles (1, 5, 10, 20, 50, 80, 90, 95, 99) for x computed from the estimated densities (red) and directly from the cross-sectional observations (blue).

We plot the empirical cdf of financial income for the year 2012 in Figure 10. We standardize the financial income by 1/3 (approximate capital share) of GDP per capita to construct z. The graph overlays cdfs for the PSZ, SCF, and CEX data. The left panel shows the cdfs for the range -10 to 10, whereas the right panel zooms into the range -0.2 to 0.2. All three data sets exhibit a large fraction of units with (near) zero financial income. For the PSZ data set this fraction is about 4.5%. The cdfs illustrate that PSZ covers more households with high financial income compared to the CEX. For instance, in the CEX sample less than 20% of the households have a financial income that exceeds 0.2 times the capital share of per capita GDP. The PSZ data set includes more than 40% of such households.

We include the fraction of zero financial income in the vector Y_t and proceed by modeling the continuous part of the financial income distribution. Because the financial income can be negative and its distribution is fat tailed, we extend the inverse hyperbolic sign transformation to negative z values as follows:

$$x = \begin{cases} \frac{1}{\theta}\ln(\theta\kappa z + (\theta^2(\kappa z)^2 + 1)^{1/2}) & \text{if } z \geq 0 \\ -\frac{1}{\theta}\ln(\theta|\kappa z| + (\theta^2(\kappa z)^2 + 1)^{1/2}) & \text{if } z < 0 \end{cases}. \tag{36}$$

The parameter κ can be used to scale the financial income near zero in the linear part of the transformation. This dampens the mode of the x density. We use $\kappa = 5$ and $\theta = 1$.

In the left panel of Figure 11 we plot a histogram and the estimated density of the transformed (and detrended) financial income x for the year 2006, excluding the point mass

at zero. The density plot shows the same features as the cdfs in Figure 10. A fraction of units have negative financial income, a large number of units have financial income close to zero, and the distribution has fat tails, which are mitigated through the approximately logarithmic transformation of large (in absolute terms) z values. The right panel of Figure 11 shows empirical and estimated percentiles, which are generally very close to each other. At the bottom is the 1st percentile and at the top the 99th percentile (excluding the point mass at zero) of the x distribution, which are -1.05 and 4.77 in 2006, respectively. Reversing the transformation for the two percentiles leads to -0.25 and 11.7 (times 1/3 GDP per capita), respectively. From a time series perspective, the most striking feature is the drop in financial income during the Great Recession. After 2011 the 1st percentile for the estimated model is slightly larger than the empirical percentile.[15]

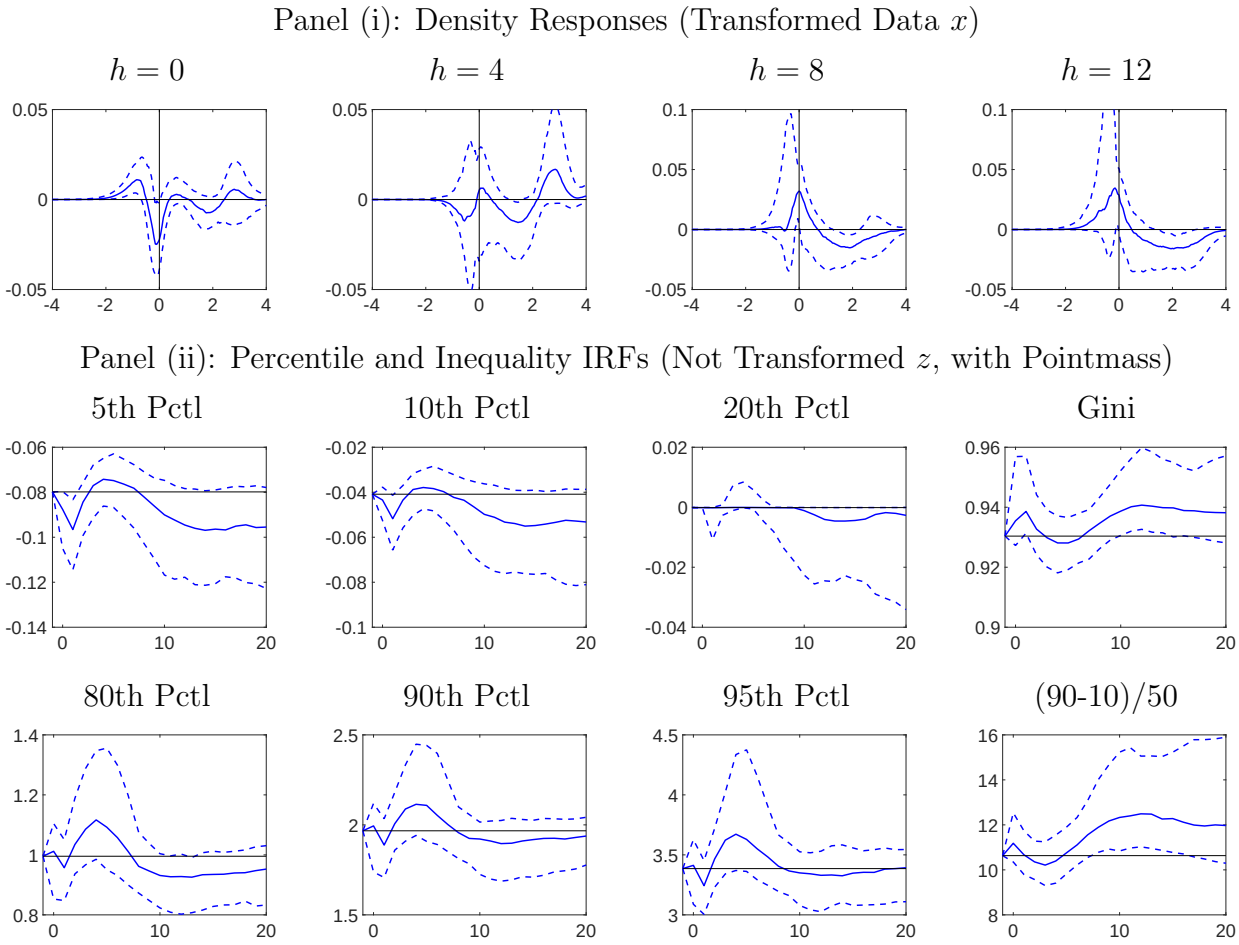
Financial Income Distribution Response. We use two of the extensions to the basic econometric framework described in Section 3.3. First, because the PSZ financial income data are weighted we modify the likelihood function for the cross-sectional data accordingly. Second, we combine the annual financial income distributions with quarterly macroeconomic observations Y_t employing the mixed-frequency state-space framework.

The results are plotted in Figure 12. Panel (i) shows density differentials for the continuous part of the distribution of transformed financial income x. Recall that $x = 0$ if and only if $z = 0$. Upon impact, $h = 0$, there is some increase in probability mass between -1.9 and -0.5 and a drop between -0.5 and zero. At the same time the probability mass at zero rises (not shown in the figure). For positive incomes, the density increases between 0.4 and 1.1 and beyond 2.5. At $h = 4$ the density increase for incomes of $x > 2$ is even more pronounced. At longer horizons the pattern changes a bit: density differentials are positive for negative incomes between -1 and zero and mostly negative for positive differentials.

Panel (ii) shows responses of selected percentiles and two inequality measures: the Gini coefficient and the symmetric 90-10 ratio defined as $(P90 - P10)/P50$. We use the symmetric percentile ratio because the 10th percentile of the financial income distribution is negative. Moreover, whereas we plotted density differentials for the transformed data x in Panel (i), the percentiles and inequality measures are computed for the not-transformed z data and account for the response of the fraction of zero financial incomes. Focusing on the posterior median of the responses, after an initial drop the response of the percentiles at horizon $h = 4$ is positive. For the 5th and 10th percentiles the initial drop is more pronounced than the

[15]This discrepancy vanishes if an additional knot is added in the left tail of the distribution, but the penalized fit of the functional model deteriorates.

Figure 12: Responses of Financial Income to MP Shock



Notes: Responses to a 25 basis point monetary policy shock. The system is in steady state at $h = -1$ and the shock occurs at $h = 0$. The plots depict 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution. Panel (i) densities for transformed data (x) are normalized so that they integrate to one. Panel (ii) The x-axes is IRF horizon in quarters. Percentile and inequality responses are computed from the distribution of not-transformed financial income z and the point mass at zero.

subsequent rise, whereas for the 80th, 90th, and 95th percentiles the initial drop is very small and the subsequent increase above the initial level is comparatively large. After 5 quarters, the percentile responses start to fall again. The medium-run responses (10 to 20 quarters) of individual financial income to 1/3 GDP per capita are negative for the 5th, 10th, and 20th percentiles and close to zero for the 80th, 90th, and 95th percentiles. Overall, this leads to an increase in the two inequality measures: the Gini coefficient and the symmetric 90-10 ratio.

The existing literature that uses panel local projections on administrative data, or in-

direct calculations based on fixed portfolio shares and return responses for different asset classes emphasizes that wealthy individuals benefit more strongly from expansionary monetary policy because they hold most of the assets. Suppose a low-wealth household receives 10 dollars in dividends per year, and a high-wealth household receives 10,000 dollars. If the monetary policy shock raises dividends by 1% the financial income of the low wealth household increases by 10 cents and that of the high-wealth household by 10 dollars. Our percentile IRFs are consistent with this "absolute magnitude" effect, because the absolute magnitudes of the responses in the right tail of the distribution are much larger than in the left tail.[16] But, our results also imply that the movements of the 5th and 10th percentiles are substantially larger in terms of percentages than the responses of the 90th and 95th percentile. In fact, the largest percentage changes occur at the 20th percentile which is close to zero. This means that the portfolio shares react to the monetary policy shock as well, in particular for some of the low wealth individuals, e.g., those who start to pay back credit card debt or begin to save and earn interest of savings after an increase in labor income.

8 Conclusion

We estimated a functional VAR that stacks macroeconomic aggregates with the cross-sectional distributions of earnings, consumption, and financial income to provide semi-structural evidence on the distributional effects of monetary policy shocks and on the central bank's reaction to fluctuations in earnings inequality. While much of the existing literature examines how specific groups of households or individuals are affected by shocks, our analysis implicitly aggregates unit-level responses and provides quantitative evidence on changes in cross-sectional distributions that can inform monetary policy, deepen our understanding of business cycle fluctuations, and help evaluate heterogeneous-agent DSGE models.

References

Amberg, N., T. Jansson, M. Klein, and A. Rogantini Picco (2022): "Five Facts about the Distributional Income Effects of Monetary Policy Shocks," *AER: Insights*, 4(3), 289–304.

[16]The range of the y-axis in the 95th percentile plot is about 20 times larger than the range for the 5th percentile IRF.

Andersen, A. L., N. Johannesen, M. Jorgensen, and J.-L. Peydro (2023): "Monetary Policy and Inequality," *Journal of Finance*, 78(5), 2945–2989.

Anderson, E., A. Inoue, and B. Rossi (2016): "Heterogeneous Consumers and Fiscal Policy Shocks," *Journal of Money, Credit, and Banking*, 48(8), 1877–1888.

Arias, J. E., J. Rubio-Ramirez, and D. Waggoner (2022): "Inference in Bayesian Proxy-SVARs," *Journal of Econometrics*, 225(1), 88–106.

Bayer, C., L. Calderon, and M. Kuhn (2025): "Distributional Dynamics," *Manuscript, University of Bonn*.

Bjornland, H. C., Y. Chang, and J. Cross (2023): "Oil and the Stock Market Revisited: A Mixed Functional VAR Approach," *Manuscript, BI Norwegian Business School and Indiana University*.

Caldara, D., and E. Herbst (2019): "Monetary Policy, Real Activity, and Credt Spreads: Evidence from Bayesian Proxy SVARs," *American Economic Journal: Macroeconomics*, 11(1), 157–192.

Chan, J. C. C. (2022): "Asymmetric Conjugate Priors for Large Bayesian VARs," *Quantitative Economics*, 13(3), 1145–1169.

Chang, M., X. Chen, and F. Schorfheide (2024): "Heterogeneity and Aggregate Fluctuations," *Journal of Political Economy*, 132(12), 4021–4067.

Chang, M., and F. Schorfheide (2024): "On the Effects of Monetary Policy Shocks on Income and Consumption Heterogeneity," *NBER Working Paper*, 32166.

Chang, Y., C. S. Kim, and J. Park (2016): "Nonstationarity in Time Series of State Densities," *Journal of Econometrics*, 192, 152–167.

Chang, Y., and S.-B. Kim (2006): "From Individual to Aggregate Labor Supply: A Quantitative Analysis Based on a Heterogeneous Agent Macroeconomy," *International Economic Review*, 47, 1–27.

Cloyne, J., C. Ferreira, and P. Surico (2020): "Monetary Policy When Households Have Debt: New Evidence on the Transmission Mechanism," *Review of Economic Studies*, 87, 102–129.

Coibion, O., Y. Gorodnichenko, L. Kueng, and J. Silvia (2017): "Innocent Bystanders? Monetary Policy and Inequality," *Journal of Monetary Economics*, 88, 70–89.

Del Canto, F., J. Grigsby, E. Qian, and C. Walsh (2023): "Are Inflationary Shocks Regressive? A Feasible Set Approach," *Manuscript, Columbia University*.

Drautzburg, T. (2020): "A Narrative Approach to a Fiscal DSGE Model," *Quantitative Economics*, 11, 801–837.

Ettmeier, S. (2024): "No Taxation without Reallocation: The Distributional Effects of Tax Changes," *Manuscript, University of Bonn*.

Ettmeier, S., C. H. Kim, and F. Schorfheide (2025): "Measuring the Effects of Aggregate Shocks on Cross-sectional Distributions: Functional vs. Panel Approach," *Manuscript, University of Pennylvania*.

Furceri, D., P. Loungani, and A. Zdzienicka (2018): "The Effects of Monetary Policy Shocks on Inequality," *Journal of International Money and Finance*, 85, 168–186.

Giannone, D., L. Reichlin, and D. Small (2008): "Nowcasting: The Real-Time Informational Content of Macroeconomic Data," *Journal of Monetary Economics*, 55, 665–676.

Holm, M. B., P. Paul, and A. Tischbirek (2021): "The Transmission of Monetary Policy under the Microscope," *Journal of Political Economy*, 129(10), 2861–2904.

Hu, B., and J. Y. Park (2017): "Econometric Analysis of Functional Dynamics in the Presence of Persistence," *Manuscript, Department of Economics, Indiana University*.

Inoue, A., and B. Rossi (2021): "The Effects of Conventional and Unconventional Monetary Policy: A New Approach," *Quantitative Economics*, 12(4), 1085–1138.

Jacobson, M., C. Matthes, and T. Walker (2024): "Temporal Aggregation Bias and Monetary Policy Transmission," *Working Paper*.

Jarocinski, M., and P. Karadi (2020): "Deconstructing Monetary Policy Surprises - The Role of Information Shocks," *Amercian Economic Journal: Macroeconomics*, 12, 1–43.

Kaplan, G., B. Moll, and G. L. Violante (2018): "Monetary Policy According to HANK," *American Economic Review*, 108(3), 697–743.

Lee, D. (2021): "The Effects of Monetary Policy on Consumption and Inequality," *Manuscript, Federal Reserve Bank of New York*.

Lenza, M., and J. Slacalek (2024): "How Does Monetary Policy Affect Income and Wealth Inequality? Evidence from Quantitative Easing in the Euro Area," *Journal of Applied Econometrics*, 39(5), 746–765.

Ma, E. (2023): "Monetary Policy and Inequality: How Does One Affect the Other," *International Economic Review*, 64(2), 691–725.

Marcellino, M., A. Renzetti, and T. Tornese (2024a): "Firm Heterogeneity and Macroeconomic Fluctuations: a Functional VAR Model," *Manuscript, Bocconi University*.

——— (2024b): "Nowcasting Distributions: A Functional MIDAS Model," *arXiv Working Paper 2411.05629*.

McKay, A., and C. K. Wolf (2023): "Monetary Policy and Inequality," *Journal of Economic Perspectives*, 37(1), 121–144.

Meeks, R., and F. Monti (2023): "Heterogeneous Beliefs and the Phillips Curve," *Journal of Monetary Economics*, 139, 41–54.

Mitman, K., T. Broer, and J. Kramer (2022): "The Curious Incidence of Monetary Policy Across the Income Distribution," *CEPR Discussion Paper*, 17589.

Nakamura, E., and J. Steinsson (2018): "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics*, 133, 1283–1330.

Piketty, T., E. Saez, and G. Zucman (2018): "Distributional National Accounts: Methods and Estimates for the United States," *Quarterly Journal of Economics*, 133(2), 553–609.

Plagborg-Møller, M., and C. K. Wolf (2021): "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, 89(2), 995–980.

Ramey, V. (2011): "Identifying Government Spending Shocks: It's All in the Timing," *Quarterly Journal of Economics*, 126(1), 1–50.

Schorfheide, F., and D. Song (2015): "Real-Time Forecasting with a Mixed-Frequency VAR," *Journal of Business and Economic Statistics*, 33(3), 366–380.

Stock, J. H., and M. W. Watson (2010): "Monthly GDP and GNI – Research Memorandum," *Manuscript, Princeton University*.

Online Appendix: On the Effects of Monetary Policy Shocks on Earnings and Consumption Heterogeneity

Minsu Chang, and Frank Schorfheide

This Appendix consists of the following sections:

A Density Estimation

A.1 Top Coding

Likelihood Function with Censoring. We define the censoring point c_t as

$$c_t = \max_{i=1,\ldots,N} x_{it}$$

Moreover, we let

$$N_{t,max} = \sum_{i=1}^{N} \mathbb{I}\{x_{it} = c_t\}.$$

If $N_{t,max} = 1$, we assume that the observed sample is not constrained by the top-coding and use the standard likelihood function described in the main text. If $N_{t,max} > 1$ we use a likelihood function that assumes that any earnings value exceeding c_t is coded as c_t.

Recall that in the main text we ignored the dependence of the cross-sectional sample size N on t in the notation and defined $p^{(K)}(X_t|\alpha_t) = \exp\left\{N\mathcal{L}^{(K)}(\alpha_t|X_t)\right\}$, where

$$\mathcal{L}^{(K)}(\alpha_t|X_t) = \bar{\zeta}'(X_t)\alpha_t - \ln\int_0^\infty \exp\left\{\zeta'(x)\alpha_t\right\}dx, \quad \bar{\zeta}(X_t) = \frac{1}{N}\sum_{i=1}^{N_t}\zeta(x_{it}).$$

We introduce the unknown parameter $\pi_t = \mathbb{P}\{x_{it} \geq c_t\}$. We drop the top-coded observations from the definition of $\bar{\zeta}(X_t)$ und make the time dependence explicit in the notation. Let

$$\bar{\zeta}_t(X_t) = \frac{1}{N_t}\sum_{i=1}^{N_t}\zeta(x_{it})\mathbb{I}\{x_{it} < c_t\}. \tag{A.1}$$

The log-likelihood function is obtained as follows: the sample contains $N_{t,max}$ top-coded observations where the probability of sampling a top-coded observation is π_t. The probability of samping an observation that is not top-coded is $(1-\pi_t)$. Conditional on not being top-coded, the observation $x_{it} < c_t$ is sampled from a continuous density with a domain that is truncated at c_t. Thus, dividing the log-likelihood by the sample size N_t, we obtain

$$\begin{aligned}\mathcal{L}^{(K)}(\alpha_t, \pi_t|X_t) &= \frac{N_{t,max}}{N_t}\ln\pi + \frac{N_t - N_{t,max}}{N_t}\ln(1-\pi_t) \\ &\quad + \bar{\zeta}_t'(X_t)\alpha_t - \frac{N_t - N_{t,max}}{N_t}\ln\int_0^{c_t}\exp\left\{\zeta'(x)\alpha_t\right\}dx.\end{aligned} \tag{A.2}$$

Notice that regardless of the value of α_t, the MLE of π_t is

$$\hat{\pi}_t = \text{argmax}_{\pi\in[0,1]}\, \mathcal{L}^{(K)}(\alpha_t, \pi_t|X_t) = N_{t,max}/N_t. \tag{A.3}$$

Moreover, regardless of the value of π_t, the MLE of α_t is given by

$$\begin{aligned}\hat{\alpha}_t &= \text{argmax}_{\alpha_t}\, \mathcal{L}^{(K)}(\alpha_t, \pi_t|X_t) \\ &= \text{argmax}_{\alpha_t}\, \bar{\zeta}_t'(X_t)\alpha_t - \frac{N_t - N_{t,max}}{N_t} \ln \int_0^{c_t} \exp\left\{\zeta'(x)\alpha_t\right\} dx.\end{aligned} \tag{A.4}$$

The objective function for α_t is almost identical to what we had without top coding, except for a definition of $\bar{\zeta}_t(X_t)$ that drops the top-coded observations in the summation and the factor of $(N_t - N_{t,max})/N_t$ in front of the normalization constant of the density.

Recovering the Density for Uncensored Observations. To reconstruct the full density we can use

$$p(x|\alpha_t) = \frac{\exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\}}{\int_0^\infty \exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\} dx}. \tag{A.5}$$

Note that here we drop the censoring indicator function and the integration is now from 0 to ∞. Once the α_t's have been estimated based on the censored observations, we work with the full density in the functional state-space model and its K-dimensional approximation.

Modification of Hessian Matrix. We now re-compute the score and the Hessian. Dropping the (K) superscript we obtain the following first derivatives with respect to α_k for $k = 1, \ldots, K$:

$$\mathcal{L}_k^{(1)}(\alpha_t|\pi_t, X_t) = \bar{\zeta}_{t,k}(X_t) - \left(\frac{N_t - N_{t,max}}{N_t}\right) \int_0^{c_t} \zeta_k(x)\bar{p}(x|\alpha_t)dx,$$

where

$$\bar{p}(x|\alpha_t) = \frac{\exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\}}{\int_0^{c_t} \exp\left\{\sum_{k=1}^K \alpha_{k,t}\zeta_k(x)\right\} dx} \mathbb{I}\{x < c_t\}.$$

We can now deduce from our previous calculations that

$$\begin{aligned}&\mathcal{L}_{kl}^{(2)}(\alpha_t|\pi_t, X_t) \\ &= -\left(\frac{N_t - N_{t,max}}{N_t}\right) \int_0^{c_t} \left(\zeta_k(x) - \int_0^{c_t} \zeta_k(x)\bar{p}(x|\alpha_t)dx\right) \\ &\qquad \times \left(\zeta_l(x) - \int_0^{c_t} \zeta_l(x)\bar{p}(x|\alpha_t)dx\right) \bar{p}(x|\alpha_t)dx.\end{aligned} \tag{A.6}$$

Thus, compared to the standard case, the limits of integration change and there is an additional factor $(N_t - N_{t,max})/N_t$.

A.2 Transformations of the $\hat{\alpha}_t$s

Compression/Standardization. The vector $\hat{\tilde{\alpha}}_t = \hat{\alpha}_t - \alpha_*$ may exhibit collinearity. Even though K basis functions may be necessary to approximate the cross-sectional densities, the time variation might be concentrated in a lower-dimensional space, because, for instance, only the means of the cross-sectional distributions are varying over time. This feature can be captured by assuming that the time-variation is captured by a $\tilde{K} < K$ dimensional factor a_t:

$$(\alpha_t - \alpha_*)' = a_t'\Lambda, \tag{A.7}$$

where Λ is a $\tilde{K} \times K$ matrix. As is well known from the factor model literature, Λ and a_t are only identified up to a $\tilde{K} \times \tilde{K}$ dimensional invertible matrix. In principle, the matrix Λ and the sequence of vectors a_t, $t = 1, \ldots, T$ have to be estimated simultaneously under this factor structure,

To avoid the simultaneous estimation of the cross-sectional densities, we take the following short cut. First, we compute the $\hat{\alpha}_t$s period-by-period without imposing any restrictions. Second, conditional on α_* we compute the demeaned (and potentially seasonally adjusted) MLEs $\hat{\tilde{\alpha}}_t = \hat{\alpha}_t - \alpha_*$ and arrange them in a $T \times K$ matrix $\hat{\tilde{\alpha}}$ with rows $\hat{\tilde{\alpha}}_t'$. Third, we conduct a principal components analysis which is based on the eigenvalue decomposition of the sample covariance matrix $\hat{\tilde{\alpha}}'\hat{\tilde{\alpha}}/T$. Let $\hat{M}$ be $K \times \tilde{K}$ matrix of eigenvectors associated with the $\tilde{K}$ non-zero eigenvalues (in practice greater than 10^{-10}).[17] Then, let

$$\hat{a} = \hat{\tilde{\alpha}}\hat{M}, \quad \hat{\Lambda} = (\hat{a}'\hat{a})^{-1}\hat{a}'\hat{\tilde{\alpha}}, \tag{A.8}$$

where $\hat{a}$ is the $T \times \tilde{K}$ matrix with rows $\hat{a}_t'$. Even if $\tilde{K} = K$ this operation standardizes the basis function coefficients α_t. To evaluate the MDD formula (27), we replace K by $\tilde{K}$, $\mathcal{L}^{(K)}(\hat{\alpha}_t|X_t)$ by $\mathcal{L}^{(\tilde{K})}(\alpha_* + \hat{\Lambda}'\hat{a}_t|X_t)$, and we change the term $\sum_{t=1}^T \ln|\hat{V}_t|^{1/2}$ to $\sum_{t=1}^T \ln|(\hat{\Lambda}\hat{V}_t^{-1}\hat{\Lambda}')^{-1}|^{1/2}$.

[17] Because our goal is to eliminate perfect collinearities, we choose an eigenvalue cut-off that yields $\alpha_* + \hat{\Lambda}'\hat{a}_t = \hat{\alpha}_t$.

Seasonal Adjustments. Deterministic seasonal adjustments of the cross-sectional densities can be incorporated in the model, as needed, by replacing the vector of constants $\alpha_* = \alpha_t - \tilde{\alpha}_t$ by a time-varying process. In our application the time period t is either a quarter or a month. For quarterly data we let $\alpha_{*,t} = \sum_{q=1}^{4} \alpha_{q,t} s_q(t)$, where $s_q(t) = 1$ if period t is associated with quarter q and $s_q(t) = 0$ otherwise. For monthly data we we use $\alpha_{*,t} = \sum_{m=1}^{12} \alpha_{m,t} s_m(t)$, where $s_m(t) = 1$ if period t is associated with month m and $s_m(t) = 0$ otherwise.

A.3 Recovering Cross-Sectional Densities

Based on the estimated state-transition equation we can generate forecasts and impulse response functions for the compressed coefficients a_t. However, the dynamics of these coefficients in itself are not particularly interesting. Thus, we have to convert them back into densities using the following steps (which can be executed for each prior/posterior draw of a_t from the relevant posterior distribution). First, use (A.7) with $\Lambda = \hat{\Lambda}$ to transform a_t into α_t. If the estimation is based on a seasonal adjustment, α_* can be replaced by $\alpha_{*,t}$, or, if the goal is to compute impulse responses, one could use the average of the seasonal dummies as intercept. Second, compute

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

B Prior and Posterior Computations

B.1 More Details on the Prior

Recall that the ith equation of the W_t VAR is given by (21). We assume that the parameters (β_i, D_i) are *a priori* independent across equations, i.e.,

$$p(\beta, D) = \prod_{i=1}^{n} p(\beta_i|D_i)p(D_i). \tag{A.9}$$

For each pair (β_i, D_i) we use a Normal-Inverse Gamma (NIG) distribution of the form

$$\beta_i|D_i \sim \mathcal{N}\big(\underline{\beta}_i, D_i\underline{V}_i^{\beta}\big), \quad D_i \sim IG\big(\underline{\nu}_i, \underline{S}_i\big). \tag{A.10}$$

The prior density takes the form

$$\begin{aligned} p(\beta_i, D_i) &= (2\pi)^{-k_i/2}|\underline{V}_i^{\beta}|^{-1/2}\frac{\underline{S}_i^{\underline{\nu}_i}}{\Gamma(\underline{\nu}_i)}D_i^{-(\underline{\nu}_i+1+k_i/2)} \\ &\quad \times \exp\left\{-\frac{1}{D_i}\left[\underline{S}_i + \frac{1}{2}(\beta_i - \underline{\beta}_{\cdot i})'(\underline{V}_i^{\beta})^{-1}(\beta_i - \underline{\beta}_{\cdot i})\right]\right\}. \end{aligned}$$

In the remainder of this subsection we discuss the construction of $\underline{\beta}_{\cdot i}$, $\underline{V}_i^{\beta}$, $\underline{\nu}_i$, and $\underline{S}_i$. The prior is obtained by transforming a prior for the reduced-form parameters (Φ, Σ) into a prior for the quasi-structural parameters $(\beta_1, \ldots, \beta_{n_w}, D)$.

Prior for D_i and the α_i component of β_i. We start from a prior for $\Sigma = A^{-1'}DA^{-1}$:

$$\Sigma \sim IW\big(\underline{\nu}, \underline{S}\big), \quad \underline{S} = \text{diag}(\underline{s}_1^2, \ldots, \underline{s}_n^2). \tag{A.11}$$

Chan (2021) shows that this prior implies

$$D_i \sim IG\left(\frac{\underline{\nu} + i - n}{2}, \frac{\underline{s}_i^2}{2}\right), \quad i = 1, \ldots, n. \tag{A.12}$$

Thus, a comparison with (A.10) indicates that we are setting

$$\underline{\nu}_i = \frac{\underline{\nu} + i - n}{2}, \quad \underline{S}_i = \frac{\underline{s}_i^2}{2}. \tag{A.13}$$

Moreover, (A.11) implies that

$$A_{ij}|D_i \sim \mathcal{N}\left(0, \frac{D_i}{\underline{s}_j^2}\right), \quad 1 \leq j < i, \quad i = 2, \ldots, n, \tag{A.14}$$

which determines the prior for the α_i component of β_i.

Prior for the $B_{\cdot i}$ component of β_i. Recall that $B_{\cdot i}$ consists of np coefficients on lagged elements of y_t and an intercept. The overall dimension of the vector is $k \times 1$. The prior will take the form

$$B_{\cdot i} \sim \mathcal{N}\big(\underline{B}_{\cdot i}, \underline{V}_i^B\big), \tag{A.15}$$

where the $k \times k$ matrix $\underline{V}_i^B$ is assumed to be diagonal.

To specify a prior for $B_{\cdot i}$, we loosely map *a priori* beliefs about $(\alpha_i, \Phi_{\cdot i})$ into beliefs about $B_{\cdot i}$. To simplify the notation a bit, let $\phi_i = \Phi_{\cdot i}$ and suppose that the researcher starts with the belief that

$$\phi_i \sim \mathcal{N}\big(\underline{\phi}_{\cdot i}, D_i\underline{V}_i^{\phi}\big) \tag{A.16}$$

with $\underline{\phi}_i = 0$. Because the macroeconomic variables are in log-level we let for $i = 1, \ldots, n_y$:

$$[\underline{\phi}_i]_j = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases} \qquad j = 1, \ldots, k. \tag{A.17}$$

The $k \times k$ prior covariance matrix $\underline{V}_i^{\phi}$ is assumed to be diagonal with elements $l = 1, \ldots, k$:

$$[\underline{V}_i^{\phi}]_{ll} = \begin{cases} \frac{1}{\lambda_1}\frac{1}{\underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if vars } (i,j) \text{ belong to same block} \\ \frac{1}{\lambda_1}\frac{1}{\lambda_2 \underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if var } i \text{ belongs to } Y \text{ and } j \text{ belongs to } a \\ \frac{1}{\lambda_1}\frac{1}{\lambda_3 \underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if var } i \text{ belongs to } a \text{ and } j \text{ belongs to } Y \\ \frac{1}{\lambda_5} & \text{for the intercept} \end{cases}$$

We now turn the prior for ϕ_i into a prior for $B_{\cdot i}$, utilizing the relationship between the quasi-structural-form coefficients, B, and the reduced form coefficients, Φ. For the coefficients on lagged elements of y_t we obtain:

$$[B_h]_{ij} = [\Phi_h]_{ij} + \sum_{l=1}^{i-1} A_{il}[\Phi_h]_{lj}. \tag{A.18}$$

Likewise, the $n \times 1$ vector of intercepts, B_0, is related to the reduced form intercept via

$$[B_0]_i = [\Phi_0]_i + \sum_{l=1}^{i-1} A_{il}[\Phi_0]_l. \tag{A.19}$$

Taking expectations of (A.18) and (A.19), and using $\mathbb{E}[A_{ij}] = 0$, we deduce that

$$\mathbb{E}\big[[B_h]_{ij}\big] = \mathbb{E}\big[[\Phi_h]_{ij}\big], \quad \mathbb{E}\big[[B_0]_i\big] = \mathbb{E}\big[[\Phi_0]_i\big] \tag{A.20}$$

We use a prior covariance matrix $\underline{V}_i^{B}$ that is diagonal. The entries on the diagonal are specified as follows: we first express the variance of a generic element $[B_h]_{ij}$ in terms of the variances of A_{ij} and $[\Phi_h]_{ij}$:

$$\begin{aligned} \mathbb{V}\big[[B_h]_{ij}\big] &= \mathbb{E}\big[\mathbb{V}\big([B_h]_{ij}\big)|A\big] + \mathbb{V}\big[\mathbb{E}\big([B_h]_{ij}\big)|A\big] \\ &= \mathbb{E}\left[\mathbb{V}\big([\Phi_h]_{ij}\big) + \sum_{l=1}^{i-1} A_{il}^2 \mathbb{V}\big([\Phi_h]_{lj}\big)\right] + \mathbb{V}\left[\mathbb{E}([\Phi_h]_{ij}) + \sum_{l=1}^{i-1} A_{il}\mathbb{E}([\Phi_h]_{lj})\right] \\ &= \mathbb{V}\big([\Phi_h]_{ij}\big) + \sum_{l=1}^{i-1} \mathbb{V}(A_{il})\mathbb{V}\big([\Phi_h]_{lj}\big) + \sum_{l=1}^{i-1} \mathbb{V}(A_{il})\big(\mathbb{E}([\Phi_h]_{lj})\big)^2. \end{aligned}$$

The same calculation for the variance of the intercept leads to:

$$\begin{aligned}
\mathbb{V}\big[[B_0]_i\big] &= \mathbb{E}\big[\mathbb{V}\big([B_0]_i\big)|A\big] + \mathbb{V}\big[\mathbb{E}\big([B_0]_i\big)|A\big] \\
&= \mathbb{E}\left[\mathbb{V}\big([\Phi_0]_i\big) + \sum_{l=1}^{i-1} A_{il}^2 \mathbb{V}\big([\Phi_0]_l\big)\right] + \mathbb{V}\left[\mathbb{E}([\Phi_0]_i) + \sum_{l=1}^{i-1} A_{il}\mathbb{E}([\Phi_0]_l)\right] \\
&= \mathbb{V}\big([\Phi_0]_i\big) + \sum_{l=1}^{i-1} \mathbb{V}(A_{il})\mathbb{V}\big([\Phi_0]_l\big) + \sum_{l=1}^{i-1} \mathbb{V}(A_{il})\big(\mathbb{E}([\Phi_0]_l)\big)^2.
\end{aligned}$$

To arrange the first np $\mathbb{V}\big[[B_h]_{ij}\big]$ terms on the diagonal of the $k \times k$ matrix $\underline{V}_i^B$ we use the index function

$$f(j,h) = (h-1)n + j. \tag{A.21}$$

Here h corresponds to the lag and j is the index for the element of the y_{t-h} vector. Using the definition of the index function, the expressions for $\mathbb{V}[A_{il}]$ and $\mathbb{E}([\Phi_h]_{lj})$ from the Normal distribution in (A.14), and the expression for $\mathbb{V}\big([\Phi_h]_{lj}\big)$ from the Normal distribution in (A.16), we can write

$$\mathbb{V}\big[[B_h]_{ij}\big] = D_i[\underline{V}_i^{\phi}]_{f(j,h)} + \sum_{l=1}^{i-1} \frac{D_i}{\underline{s}_l^2}\left[D_l[\underline{V}_l^{\phi}]_{f(j,h)} + [\underline{\phi}_l]_{f(j,h)}^2\right].$$

For the intercept in equation i we obtain

$$\mathbb{V}\big[[B_0]_i\big] = D_i\frac{D_i}{\lambda_5} + \sum_{l=1}^{i-1} \frac{D_i D_l}{\underline{s}_l^2\lambda_5}.$$

We now replace the variance parameter D_l by the hyperparameter $\underline{s}_l^2$. This ensures that $\underline{V}_i^B$ is not a function of the (unknown) variance parameter D_i and simplifies posterior calculations. Using

$$\mathbb{V}\big[[B_h]_{ij}\big] = D_i[\underline{V}_i^B]_{f(j,h)} \quad \text{and} \quad \mathbb{V}\big[[B_0]_i\big] = D_i[\underline{V}_i^B]_{np+1}$$

we obtain

$$[\underline{V}_i^B]_{f(j,h)} = [\underline{V}_i^{\phi}]_{f(j,h)} + \sum_{l=1}^{i-1}\left([\underline{V}_l^{\phi}]_{f(j,h)} + \frac{1}{\underline{s}_l^2}[\underline{\phi}_l]_{f(j,h)}^2\right). \tag{A.22}$$

$$[\underline{V}_i^B]_{np+1} = \frac{1}{\lambda_5} + \sum_{l=1}^{i-1}\frac{1}{\lambda_5} = \frac{i}{\lambda_5}. \tag{A.23}$$

Table A-1: Hyperparameters for VAR Prior

Parameter	Description
$\underline{\nu} = 2n$	Degrees of freedom for IG distribution
$\underline{s}_i = \text{StDev}(W_i)$	Shape para for IG; use sample standard dev.
λ_1	Overall precision of prior
λ_2	Relative precision for a to Y transmission
$\lambda_3 = 1$	Relative precision for Y to a transmission
$\lambda_4 = 2$	Decay rate for prior variance on lags
$\lambda_5 = 0.001$	Relative precision for intercept

Summary. The overall prior takes the form (A.10). The prior for D_i is given by (A.12). The prior for β_i is obtained by combining (A.14) with (A.15), where mean and variance are given in (A.20) and (A.22), respectively. The hyperparameters for the prior are summarized in Table A-1. We set $\underline{s}_i$ equal to the sample standard deviation of W_i.

B.2 Posterior Sampling and MDD

Model and prior are set up so that the coefficients can be estimated equation by equation:

$$p(W, \beta, D) = \prod_{i=1}^{N} \left((2\pi D_i)^{-1/2} \exp\left\{ -\frac{1}{2D_i} (W_i - Z_i\beta_i)'(W_i - Z_i\beta_i) \right\} p(\beta_i | D_i) p(D_i). \right) \tag{A.24}$$

Because the prior is conjugate, the posterior stays in the NIG family. It takes the form

$$\beta_i | (D_i, W_i) \sim \mathcal{N}\big(\bar{\beta}_i, D_i \bar{V}_i^{\beta}\big), \quad D_i \sim IG\big(\bar{\nu}_i, \bar{S}_i\big), \tag{A.25}$$

Instead of working with covariance matrices, it is more efficient to work with precision matrices. Define:

$$\underline{P}_i^{\beta} = \big(\underline{V}_i^{\beta}\big)^{-1}, \quad \bar{P}_i^{\beta} = \big(\bar{V}_i^{\beta}\big)^{-1}.$$

The updating equations for the posterior take the form

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

When using the JK instruments, we set the coefficients on the lags and the intercept equal to zero: $B_i = 0$ for $i = 1, 2$. Thus, $\beta_1 = 0_{(np+1)\times 1}$ and $\beta_2 = [A_{2,1}, 0_{1\times(np+1)}]'$. The updating equations for the posterior change as follows. For $i = 1$:

$$\begin{aligned}
\bar{P}_1^{\beta} &= N/A \\
\bar{\beta}_1 &= 0_{(np+1)\times 1} \\
\bar{\nu}_1 &= \underline{\nu}_i + T/2 \\
\bar{S}_1 &= \underline{S}_1 + \frac{1}{2} W_1' W_1.
\end{aligned}$$

For $i = 2$ we are regressing W_{2t} on the single regressor W_{1t}. Partition $\beta_2 = [\beta_{1,2}, 0_{1\times(np+1)}]'$ and denote the precision associated with the first element of the β_2 vector by $[P_1^{\beta}]_{11}$. Then, we can write the updating equations as

$$\begin{aligned}
[\bar{P}_2^{\beta}]_{11} &= [\underline{P}_2^{\beta}]_{11} + W_1' W_1 \\
\bar{\beta}_{2,1} &= \big([\bar{P}_2^{\beta}]_{11}\big)^{-1}\big([\underline{P}_2^{\beta}]_{11}\underline{\beta}_{2,1} + W_1' W_2\big) \\
\bar{\nu}_2 &= \underline{\nu}_2 + T/2 \\
\bar{S}_2 &= \underline{S}_2 + \frac{1}{2}\big(W_2' W_2 + [\underline{P}_2^{\beta}]_{11}\underline{\beta}_{2,1}^2 - [\bar{P}_2^{\beta}]_{11}\bar{\beta}_{2,1}^2\big).
\end{aligned}$$

The MDD can be computed analytically as follows:

$$\begin{aligned}
\ln p(W) = &-\frac{Tn}{2}\ln(2\pi) + \sum_{i=1}^{n}\Bigg[\frac{1}{2}\big(\ln|\underline{P}_i^{\beta}| - \ln|\bar{P}_i^{\beta}|\big) \\
&+\underline{\nu}_i \ln|\underline{S}_i| - \bar{\nu}_i \ln|\bar{S}_i| - \ln\Gamma(\underline{\nu}_i) + \ln\Gamma(\bar{\nu}_i)\Bigg],
\end{aligned} \tag{A.26}$$

with the understanding that for $i = 1$ the expression $\ln|\underline{P}_i^{\beta}| - \ln|\bar{P}_i^{\beta}| = 0$ and for $i = 2$ it gets replaced by $\ln[\underline{P}_i^{\beta}]_{11} - \ln[\bar{P}_i^{\beta}|]_{11}$.

C State-Space Representation for Mixed-Frequency Analysis

C.1 Companion Form

To derive the updating equations for the filter and simulation smoother we express the state-transition equation in companion form. We illustrate the companion form notation for $p = 2$. The generalization is straightforward. We define

$$W_t' = [Y_t', Y_{t-1}', \alpha_t', \alpha_{t-1}'] \tag{A.27}$$

and partition W_t' into

$$W_t' = [Z_t', s_t'], \quad Z_t' = [Y_t', Y_{t-1}'], \quad s_t = [\alpha_t, \alpha_{t-1}]'.$$

The companion-form law of motion for w_t can be written as

$$\begin{bmatrix} Y_t \\ Y_{t-1} \\ \alpha_t \\ \alpha_{t-1} \end{bmatrix} = \begin{bmatrix} \Phi_{1,yy} & \Phi_{2,yy} & \Phi_{1,y\alpha} & \Phi_{2,y\alpha} \\ I_y & 0 & 0 & 0 \\ \Phi_{1,\alpha y} & \Phi_{2,\alpha y} & \Phi_{1,\alpha\alpha} & \Phi_{2,\alpha\alpha} \\ 0 & 0 & I_\alpha & 0 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ \alpha_{t-1} \\ \alpha_{t-2} \end{bmatrix} + \begin{bmatrix} I_y & 0 \\ 0 & 0 \\ 0 & I_\alpha \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{y,t} \\ u_{\alpha,t} \end{bmatrix}, \tag{A.28}$$

or, more compactly, as

$$W_t = \Psi W_{t-1} + M u_t. \tag{A.29}$$

We further partition the selection matrix M into

$$M = \begin{bmatrix} M_y & M_\alpha \end{bmatrix}$$

such that

$$M_y' W_t = Y_t, \quad M_\alpha' W_t = \alpha_t.$$

In some instances, we need to extract the subvectors Z_t and s_t from W_t, which is done by the selection matrices

$$\Xi_z' = \begin{bmatrix} I_z & 0 \end{bmatrix}, \quad Z_t = \Xi_z' W_t, \quad \Xi_s' = \begin{bmatrix} 0 & I_s \end{bmatrix}, \quad s_t = \Xi_s' W_t.$$

For the forward iterations of the Kalman filter, it is useful to separate s_t and Z_t. To that end, define

$$\Psi_{ss} = \Xi_s' \Psi \Xi_s = \begin{bmatrix} \Phi_{1,\alpha\alpha} & \Phi_{2,\alpha\alpha} \\ I_\alpha & 0 \end{bmatrix}, \quad \Psi_{sz} = \Xi_s' \Psi \Xi_z = \begin{bmatrix} \Phi_{1,\alpha y} & \Phi_{2,\alpha y} \\ 0 & 0 \end{bmatrix}.$$

Moreover, we define

$$M_{s\alpha} = \Xi_s' M_\alpha = \begin{bmatrix} I_\alpha \\ 0 \end{bmatrix}, \quad \Phi_{yz} = \begin{bmatrix} \Phi_{1,yy} & \Phi_{2,yy} \end{bmatrix}, \quad \Phi_{ys} = \begin{bmatrix} \Phi_{1,y\alpha} & \Phi_{2,y\alpha} \end{bmatrix}.$$

Using this notation, the measurement equation can be written as

$$\hat{\alpha}_t = M_{s\alpha}' s_t + \eta_t, \quad \eta_t \sim N(0, V_t). \tag{A.30}$$

C.2 Forward Filtering

The forward filtering iterations are obtained from a modified Kalman filter that recognizes that y_t is directly observable. Thus, only s_t is a latent state variable.

Recursive Assumption. We start from the assumption that

$$s_{t-1} | (Y_{1:t-1}, \hat{\alpha}_{1:t-1}) \sim \mathcal{N}\big(s_{t-1|t-1}, P_{t-1|t-1}(ss)\big). \tag{A.31}$$

We use the notation $P_{t-1|t-1}(ss)$ to indicate that one can define a larger matrix, conforming with w_t, which is partitioned as follows:

$$P_{t-1|t-1} = \begin{bmatrix} P_{t-1|t-1}(zz) & P_{t-1|t-1}(zs) \\ P_{t-1|t-1}(zs) & P_{t-1|t-1}(ss) \end{bmatrix},$$

with the understanding that

$$P_{t-1|t-1}(zz) = 0, \quad P_{t-1|t-1}(zs) = 0, \quad P_{t-1|t-1}(sz) = 0.$$

Initialization. Just in a regular VAR analysis, we will condition the likelihood function on observations to initialize the lags associated with the determination of (Z_1, s_1). Note that

the initial lags of Y_t are directly observed under suitable definition of the sample period and hence we know Z_0. To initialize the latent state we set

$$s_{0|0} = \begin{bmatrix} \hat{\alpha}_0 \\ \vdots \\ \hat{\alpha}_{-p+1} \end{bmatrix}, \quad \mathrm{P}_{0|0}(ss) = \mathrm{diag}\big[V_0, \cdots, V_{-p+1}\big].$$

Thus, we assume that

$$s_0|(Y_{-p+1:0}, \hat{\alpha}_{-p+1:0}) \sim \mathcal{N}\big(s_{0|0}, P_{0|0}\big).$$

For the smoother below, it will become important to properly account for the initialization because we will also need draws of $s_0 = \alpha_{-p+1:0}$ to set up the Gibbs sampler.

Forecasting. We begin by forecasting (Y_t, s_t) jointly, using (A.28). Let

$$\begin{bmatrix} Y_t \\ s_t \end{bmatrix} \Big| (Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) \sim \mathcal{N}\left(\begin{bmatrix} y_{t|t-1} \\ s_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1}(yy) & P_{t|t-1}(ys) \\ P_{t|t-1}(sy) & P_{t|t-1}(ss) \end{bmatrix} \right), \tag{A.32}$$

where

$$\begin{aligned} Y_{t|t-1} &= \Phi_{yz} Z_{t-1} + \Phi_{ys} s_{t-1|t-1} \\ s_{t|t-1} &= \Psi_{sz} Z_{t-1} + \Psi_{ss} s_{t-1|t-1} \\ P_{t|t-1}(yy) &= \Phi_{ys} P_{t-1|t-1}(ss) \Phi_{ys}' + \Sigma_{yy} \\ P_{t|t-1}(sy) &= \Psi_{ss} P_{t-1|t-1}(ss) \Phi_{ys}' + M_{s\alpha} \Sigma_{\alpha y} \\ P_{t|t-1}(ss) &= \Psi_{ss} P_{t-1|t-1}(ss) \Psi_{ss}' + M_{s\alpha} \Sigma_{\alpha\alpha} M_{s\alpha}'. \end{aligned}$$

We can now factorize the joint distribution of (Y_t, s_t) into a marginal distribution of Y_t and a conditional distribution of $s_t|Y_t$:

$$\begin{aligned} Y_t|(Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) &\sim \mathcal{N}\big(Y_{t|t-1}, P_{t|t-1}(yy)\big) \\ s_t|(Y_t, Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) &\sim \mathcal{N}\big(s_{t|y,t-1}, P_{t|y,t-1}(ss)\big), \end{aligned} \tag{A.33}$$

where

$$\begin{aligned} s_{t|y,t-1} &= s_{t|t-1} + P_{t|t-1}(sy)[P_{t|t-1}(yy)]^{-1}(Y_t - Y_{t|t-1}) \\ P_{t|y,t-1}(ss) &= P_{t|t-1}(ss) - P_{t|t-1}(sy)[P_{t|t-1}(yy)]^{-1} P_{t|t-1}(ys). \end{aligned}$$

Finally, the distribution of $\hat{\alpha}_t$ conditional on y_t and $t-1$ information is

$$\hat{\alpha}_t|(Y_t, Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) \sim N(\hat{\alpha}_{t|y,t-1}, F_{t|y,t-1}), \tag{A.34}$$

where

$$\begin{aligned} \hat{\alpha}_{t|y,t-1} &= M'_{s\alpha} s_{t|y,t-1} \\ F_{t|y,t-1} &= M'_{s\alpha} P_{t|y,t-1}(ss) M_{s\alpha} + V_t. \end{aligned}$$

Updating. The updating step in the Kalman filter is done conditional on Y_t. Generically,

$$\begin{aligned} & p(s_t|\hat{\alpha}_t, Y_t, Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) \\ \propto\ & p(\hat{\alpha}_t|s_t, Y_t, Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}) p(s_t|y_t, Y_{-p+1:t-1}, \hat{\alpha}_{-p+1:t-1}). \end{aligned} \tag{A.35}$$

We can use the standard updating formulas, substituting in the means and variances that are conditional on Y_t:

$$\begin{aligned} s_{t|t} &= s_{t|y,t-1} + P_{t|y,t-1}(ss) M_{s\alpha} F^{-1}_{t|y,t-1} (\hat{\alpha}_t - \hat{\alpha}_{t|y,t-1}) \\ P_{t|t}(ss) &= P_{t|y,t-1}(ss) - P_{t|y,t-1}(ss) M_{s\alpha} F^{-1}_{t|y,t-1} M'_{s\alpha} P_{t|y,t-1}(ss). \end{aligned} \tag{A.36}$$

C.3 Simulation Smoothing

The goal is to generate draws from the distribution of $\alpha_{-p+1:T}$ given $(Y_{-p+1:T}, \hat{\alpha}_{-p+1:T})$. Note that the filter in its final step generates the distribution $s_T|(Y_{-p+1:T}, \hat{\alpha}_{-p+1:T})$. Because of the companion form, a draw s^i_T determines $\alpha^i_T, \ldots, \alpha^i_{T-p+1}$. Thus, the distribution $s_{T-1}|(s_T, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T})$ is degenerate. In the implementation of the smoother, we will draw blocks as follows:

$$s_T, \quad s_{T-p}|s_T, \quad s_{T-2p}|(s_{T-p}, s_T), \quad s_{T-3p}|(s_{T-2p}, s_{T-p}, s_T), \quad \ldots.$$

Because $s_{T-jp} = [\alpha'_{T-jp}, \ldots, \alpha'_{T-(j+1)p+1}]'$, this approach generates the entire sequence $\alpha_{-p+1:T}$. However, special care needs to be given to $\alpha_{-p+1:0}$ and the fact that T may not be divisible by p.

Preliminaries. Assuming for now that T is a multiple of p and that $T = bp$, where b is the number of blocks, the simulation smoother relies on the factorization

$$p(Y_{-p+1:T}, \alpha_{-p+1:T}|Y_{-p+1:0}, \alpha_{-p+1:0}) = \prod_{j=1}^{b} p(Y_{(j-1)p+1:jp}, \alpha_{(j-1)p+1:jp}|Y_{-p+1:(j-1)p}, \alpha_{-p+1:(j-1)p}),$$

where, because of the VAR(p) structure of the state-transition equation,

$$\begin{aligned} & p(Y_{(j-1)p+1:jp}, \alpha_{(j-1)p+1:jp}|Y_{-p+1:(j-1)p}, \alpha_{-p+1:(j-1)p}) \\ = \;& p(Y_{(j-1)p+1:jp}, \alpha_{(j-1)p+1:jp}|Y_{(j-2)p+1:(j-1)p}, \alpha_{(j-2)p+1:(j-1)p}). \end{aligned}$$

The conditional distribution on the right-hand side, can be obtained by iterating the companion form (A.29) p periods forward:

$$W_{t+p} = \Psi^p W_t + M \sum_{h=0}^{p-1} \Psi^h u_{t+p-h}.$$

Thus,

$$W_{t+p}|W_t \sim \mathcal{N}\left(\Psi^p W_t, \sum_{h=0}^{p-1} \Psi^h M \Sigma M' \Psi^{h'}\right). \tag{A.37}$$

Now, set $t = (j-1)p$ and note that

$$W_{(j-1)p} = [Y_{(j-1)p}, \alpha_{(j-1)p}, \ldots, Y_{(j-2)p+1}, \alpha_{(j-2)p+1}],$$

as required.

Generic Smoother. We first modify the generic derivation of the smoother to account for the presence of both $Y_{-p+1:T}$ and $\hat{\alpha}_{-p+1:T}$ in the conditioning set and the block sampling. As before, it is convenient for now to assume that $T = bp$ and the time index t shifts in steps

of p periods, that is, $t = (j-1)p$, where $j = 0, \ldots, b$. Consider the following factorization

$$\begin{aligned}
&p(\alpha_{t-p+1:t}, \alpha_{t+1:T}, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) \\
&\quad = \int p(\alpha_{-p+1:T}, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) d\alpha_{-p+1:t-p} \\
&\quad = \int p(\alpha_{-p+1:t}, Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) \\
&\qquad \times \left(\prod_{\tau=t+1}^{T} p(Y_\tau | Y_{\tau-p:\tau-1}, \alpha_{\tau-p:\tau-1}) p(\alpha_\tau | Y_\tau, Y_{\tau-p:\tau-1}, \alpha_{\tau-p:\tau-1}) p(\hat{\alpha}_\tau | \alpha_\tau) \right) d\alpha_{-p+1:t-p} \\
&\quad = p(\alpha_{t-p+1:t}, Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) \\
&\qquad \times \prod_{\tau=t+1}^{t+p} p(Y_\tau | Y_{\tau-p:\tau-1}, \alpha_{\tau-p:\tau-1}) p(\alpha_\tau | Y_\tau, Y_{\tau-p:\tau-1}, \alpha_{\tau-p:\tau-1}) p(\hat{\alpha}_\tau | \alpha_\tau) \\
&\qquad \times \text{terms without } \alpha_{t-p+1:t}.
\end{aligned}$$

Maintaining the assumption that $T = bp$ and $t = (j-1)p$, we can deduce

$$\begin{aligned}
&p(s_t | s_{t+p}, s_{t+2p}, \ldots, s_T, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) \qquad (A.38) \\
&\quad = p(\alpha_{t-p+1:t} | \alpha_{t+1:T}, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) \\
&\quad \propto p(\alpha_{t-p+1:t}, \alpha_{t+1:T}, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) \\
&\quad \propto p(\alpha_{t-p+1:t}, Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) p(Y_{t+1:t+p}, \alpha_{t+1:t+p} | Y_{t-p+1:t}, \alpha_{t-p+1:t}) \\
&\quad = p(s_t, Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) p(Z_{t+p}, s_{t+p} | Z_t, s_t).
\end{aligned}$$

The first proportionality follows from Bayes Theorem. The second proportionality follows from dropping the factor $p(Y_{t+p+1:T}, \alpha_{t+p+1:T} | Y_{t+1:t+p}, \alpha_{t+1:t+p})$ because it does not depend on $\alpha_{t-p+1:t}$. The last equality uses $s_t = [\alpha'_{t-p+1}, \ldots, \alpha'_t]'$ and $Z_t = [Y'_{t-p+1}, \ldots, Y'_t]'$. Because

$$p(s_t, Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) = p(s_t | Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) p(Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}),$$

it follows that

$$p(s_t | s_{t+p}, s_{t+2p}, \ldots, s_T, Y_{-p+1:T}, \hat{\alpha}_{-p+1:T}) \propto p(s_t | Y_{-p+1:t}, \hat{\alpha}_{-p+1:t}) p(Z_{t+p}, s_{t+p} | Z_t, s_t). \quad (A.39)$$

Smoothing Formulas for the Linear Gaussian Model. We previously established that

$$s_t | (Y_{1:t}, \hat{\alpha}_{1:t}) \sim \mathcal{N}\big(s_{t|t}, P_{t|t}(ss)\big).$$

As previously shown in (A.37), iterating the companion form forward for p periods yields

$$(Z_{t+p}, s_{t+p})|(Z_t, s_t) \sim \mathcal{N}\left(\Psi^p W_t, \sum_{h=0}^{p-1} \Psi^h M \Sigma M' \Psi^{h'}\right). \tag{A.40}$$

Let

$$W_{t|t} = [Z_t', s_{t|t}']'$$

and define

$$P_{t+p|t}(ww) = \Psi^p \Xi_s P_{t|t}(ss) \Xi_s' \Psi^p + \sum_{h=0}^{p-1} \Psi^h M \Sigma M' \Psi^{h'}. \tag{A.41}$$

The joint distribution of (s_t, W_{t+p}) is given by

$$\left[\begin{array}{c} s_t \\ W_{t+p} \end{array}\right] \Bigg| (\cdot) \sim \mathcal{N}\left(\left[\begin{array}{c} s_{t|t} \\ \Psi^p W_{t|t} \end{array}\right], \left[\begin{array}{cc} P_{t|t}(ss) & P_{t|t}(ss)\Xi_s'\Psi^{p'} \\ \Psi^p \Xi_s P_{t|t}(ss) & P_{t+p|t}(ww) \end{array}\right]\right). \tag{A.42}$$

Then, we sample s_t from

$$s_t \sim \mathcal{N}\big(s^i_{t|t+p}, P_{t|t+p}(ss)\big), \tag{A.43}$$

where

$$s_{t|t+p} = s_{t|t} + P_{t|t}(ss)\Xi_s'\Psi^{p\prime} P^{-1}_{t+p|t}(ww)\left(\left[\begin{array}{c} Z_{t+p} \\ s_{t+p} \end{array}\right] - \Psi^p W_{t|t}\right) \tag{A.44}$$

$$P_{t|t+p}(ss) = P_{t|t}(ss) - P_{t|t}(ss)\Xi_s'\Psi^{p\prime} P^{-1}_{t+p|t}(ww)\Psi^p \Xi_s P_{t|t}(ss). \tag{A.45}$$

T is not a multiple of p. Consider the following example. Suppose that $T = 10$ and $p = 3$. In this case we can use the formulas to generate

$$p(s_{10}|Y_{-2:10}, \hat{\alpha}_{-2:10}), \quad p(s_7|s_{10}, Y_{-2:10}, \hat{\alpha}_{-2:10}),$$
$$p(s_4|s_7, s_{10}, Y_{-2:10}, \hat{\alpha}_{-2:10}), \quad p(s_1|s_4, s_7, s_{10}, Y_{-2:10}, \hat{\alpha}_{-2:10}).$$

The last step gives us

$$p(\alpha_{-1:1}|\alpha_{2:10}, Y_{-2:10}, \hat{\alpha}_{-2:10}),$$

but we still lack

$$p(\alpha_{-2}|\alpha_{-1:10}, Y_{-2:10}, \hat{\alpha}_{-2:10}).$$

In order to generate the draws needed to initialize the vector autoregressive law of motion of the state transition equation, we take the following short-cut: (i) discard $\alpha_{-1:0}$, (ii) redraw s_0 from (A.43) by evaluating the formulas $s_{t|t+p}$ and $P_{t|t+p}(ss)$ for $t = 0$.

D Data Used in the Empirical Analysis

Aggregate Data. Following Jarocinski and Karadi (2020) we use six monthly macroeconomic variables in the empirical model: (i) the monthly average of the one-year constant-maturity Treasury yield serves as the monetary policy indicator. (ii) The monthly average of the S&P 500 stock price index in log levels. (iii,iv) Real GDP and GDP deflator in log levels interpolated to monthly frequency based on Stock and Watson (2010). (v) The excess bond premium (EBP) as indicator of financial conditions. (vi) An aggregate employment rate constructed from the micro data (see below). The functional VAR with micro-level consumption data includes an additional aggregate variable: real personal consumption expenditures per capita in log levels. We use Personal Consumption Expenditures (*PCE* from FRED) and divide it by population level (*CNP16OV* from FRED) to get per capita values. Then we use GDP deflator to get the real values.

Micro-Level Earnings Data. The CPS raw data are downloaded from
http://www.nber.org/data/cps_basic.html.
The raw data files are converted into STATA using the do-files available at:
http://www.nber.org/data/cps_basic_progs.html.
We use the series PREXPLF ("Experienced Labor Force Employment"), which is the same as in the raw data, and the series PRERNWA ("Weekly Earnings"), which is constructed as PEHRUSL1 ("Hours Per Week at One's Main Job") times PRHERNAL ("Hourly Earnings") for hourly workers, and given by PRWERNAL for weekly workers. STATA dictionary files are available at: http://www.nber.org/data/progs/cps-basic/

We pre-process the cross-sectional data as follows. We drop individuals if (i) the employment indicator is not available; and (ii) if they are coded as "employed" but the weekly earnings are missing. In addition, we re-code individuals with non-zero earnings as employed and set earnings to zero for individuals that are coded as not employed. A CPS-based unemployment rate is computed as the fraction of individuals that are coded as not employed. By construction this is one minus the fraction of individuals with non-zero weekly earnings, which is used to normalize the cross-sectional density of earnings. It turns out that the CPS-based unemployment rate tracks the aggregate unemployment rate (*UNRATE* from FRED)

very closely. The levels of the two series are very similar, but the CPS unemployment rate exhibits additional high-frequency fluctuations, possibly due to seasonals that have been removed from the aggregate unemployment rate.

Micro-Level Hours Worked Data. We use the CPS series PEHRUSL1, which was previously used to construct the micro-level earnings. No further transformations are applied.

Micro-Level Consumption Data. We use public use microdata (PUMD) from the CEX conducted by the Bureau of Labor Statistics (BLS). Quarterly expediture (by respondent ID and expenditure category) are computed as the sum over three months in a quarter. We divide the expenditures by the number of individuals with age 16 and over belonging to the consumption unit (i.e. household or family) to obtain per capita expenditures. Because CEX per capita expenditures capture less than 50% of NIPA per capita expenditures we rescale them as follows. Let C_t be aggregate per capita consumption from NIPA. We calculate $\frac{1}{T}\sum_{t=1}^{T} \text{median}(c_{1t}, \ldots, c_{Nt})/C_t \approx 0.46$. We then define $z_{it} = c_{it}/(0.46 \cdot C_t)$. Thus, if $z_{it} = 1$ the individual approximately consumes at the level of aggregate consumption per capita. Finally, as for the earnings data, we apply the inverse hyperbolic sine transformation to obtain x_{it}.

Micro-Level Financial Data. We are using the Distributional National Accounts microfiles provided by Piketty, Saez, and Zucman (2018). We use a broad measure of financial income defined as *fkinc*: personal factor capital income = fkhou (housing asset income) + fkequ (equity asset income , S corporations) + fkfix (interest income) + fkbus (business asset income) + fkpen (pension and insurance asset income) + fkdeb (interest payments) + fkprk (sales and excise taxes allocated to capital) + fksubk (subsidies allocated to capital). We transform the financial income data by first dividing by (1/3 of GDP) and then applying the inverse hyperbolic sign transform.

Inverse Hyperbolic Sign Transformation. We transform the micro data (labor or financial income to GDP ratios, consumption relative to aggregate per capita consumption), denoted by z below, using the inverse hyperbolic sine transformation, which is given by

$$x = g(z|\theta) = \frac{\ln(\theta z + (\theta^2 z^2 + 1)^{1/2})}{\theta} = \frac{\text{sinh}^{-1}(\theta z)}{\theta} \tag{A.46}$$

wth $\theta = 1$. Note that $g(0|\theta) = 0$ and $g^{(1)}(0|\theta) = 1$, that is, for small values of z the transformation is approximately linear. For large values of z the transformation is logarithmic:

$$g(z|\theta) \approx \frac{1}{\theta}\ln(2\theta z) = \frac{1}{\theta}\ln(2\theta) + \frac{1}{\theta}\ln(z).$$

The inverse of the transformation takes the form

$$z = g^{-1}(x|\theta) = \frac{1}{\theta}\sinh(\theta x) = \frac{1}{2\theta}(e^{\theta x} - e^{-\theta x}).$$

Most of the calculations in the paper are based on $p_x(x)$. But in some instances, it is desirable to report for $p_z(z)$. From a change of variables (omitting the θ), we get

$$p_z(z) = p_x(g(z))|g'(z)|,$$

where

$$g'(z) = \frac{1 + \frac{\theta z}{(\theta^2 z^2+1)^{1/2}}}{\theta z + (\theta^2 z^2 + 1)^{1/2}} = \frac{1}{(\theta^2 z^2 + 1)^{1/2}}.$$

Whenever we do convert the estimated densities back from z to x, we recycle the density evaluations at x_j. Thus, we evaluate $p_z(z)$ for grid points $z_j = g^{-1}(x_j)$, which leads to

$$p_z(z_j) = p_x(x_j)\big|g'(g^{-1}(x_j))\big|,$$

where

$$\big|g'(g^{-1}(x_j))\big| = \frac{1}{\left(\frac{1}{4}(e^{\theta x_j} - e^{-\theta x_j})^2 + 1\right)^{1/2}} = \frac{2}{(e^{2\theta x_j} + e^{-2\theta x_j} + 2e^{2\theta x_j}e^{-2\theta x_j})^{1/2}} = \frac{2}{e^{\theta x_j} + e^{-\theta x_j}}.$$

E Knot Placement, Lag length, and Hyperparameters

- Figure 1: the results for the aggregate VAR are generated from the following specification: $\hat{p} = 4$, $\hat{\lambda}_1 = 148.4$, and $\hat{\lambda}_2 = 1$.
- Figure 4 (functional VAR with monthly earnings) is generated with the following settings:
  - Knot placement (percentiles of distribution of pooled across time periods micro-level data):

$$\begin{aligned} K = 4 &: 0.25, 0.50, 0.75 \\ K = 6 &: 0.10, 0.25, 0.50, 0.75, 0.90 \\ K = 8 &: 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95 \\ K = 10 &: 0.01, 0.025, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95 \end{aligned}$$

  - Splines: $x \in [0, 3]$, cubic-cubic-linear constructed from the right.
  - Model specifications for MDD selection: We consider four values of K: 4, 6, 8, and 10. For λ_1 and λ_2, we use 31 equally spaced values of $\ln \lambda_j$ between -10 and 20, while λ_3 is fixed at 1. The lag length p takes four values: 1, 2, 3, and 4.
  - Model selection: $K = 10$, $p = 1$, $\lambda_1 = 54.6$, $\lambda_2 = 54.6$, $\lambda_3 = 1$.
- Figure 6 (functional VAR with monthly hours worked) is generated with the following settings:
  - Knot placement (percentiles of distribution of pooled across time periods micro-level data):

$$\begin{aligned} K = 4 &: 0.25, 0.50, 0.75 \\ K = 6 &: 0.10, 0.25, 0.50, 0.75, 0.90 \\ K = 8 &: 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95 \\ K = 10 &: 0.01, 0.025, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95 \end{aligned}$$

  - Splines: $x \in [1.05, 155.8]$, cubic-cubic-linear constructed from the right.
  - Model selection: $K = 10$, $p = 1$, $\lambda_1 = 54.6$, $\lambda_2 = 20.1$, $\lambda_3 = 1$.

- Figure 8 (functional VAR with quarterly total consumption) is generated with the following settings:

  - Knot placement (percentiles of distribution of pooled across time periods micro-level data):

$$\begin{aligned} K = 4 &: 0.25, 0.50, 0.75 \\ K = 6 &: 0.10, 0.25, 0.50, 0.75, 0.90 \\ K = 8 &: 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95 \\ K = 10 &: 0.01, 0.025, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95 \end{aligned}$$

  - Splines: $x \in [0, 3]$, cubic-cubic-linear constructed from the right.
  - Model selection: $K = 6$, $p = 1$, $\lambda_1 = 403.4$, $\lambda_2 = 0.37$, $\lambda_3 = 1$.

- Figure 9 (functional VARs with quarterly consumption components) is generated with the following settings:

  - Model selection (Total): $K = 6$, $p = 1$, $\lambda_1 = 403.4$, $\lambda_2 = 0.37$, $\lambda_3 = 1$.
  - Model selection (Durables): $K = 8$, $p = 1$, $\lambda_1 = 54.6$, $\lambda_2 = 20.1$, $\lambda_3 = 1$.
  - Model selection (Nondurables): $K = 10$, $p = 1$, $\lambda_1 = 54.6$, $\lambda_2 = 7.4$, $\lambda_3 = 1$.
  - Model selection (Services): $K = 4$, $p = 1$, $\lambda_1 = 403.4$, $\lambda_2 = 0.37$, $\lambda_3 = 1$.

- The functional VAR with stacked earnings and consumption densities is done with the following hyperparameters (based on MDD): $\lambda_1 = 54.6, \lambda_2 = 54.6, K_E = 8, K_C = 6$.

- Figure 12 (mixed-frequency functional VAR with quarterly aggregate variables and annual financial income density) is generated with the following settings:

  - Knot placement (percentiles of distribution of pooled across time periods micro-level data):

$$K = 11 \quad : \quad 0.005, 0.01, 0.05, 0.15, 0.2, 0.3, 0.75, 0.9, 0.95, 0.99.$$

- Splines: $x \in [-5, 10]$, linear-cubic-linear constructed from the left.
- Model choice: $K = 11$, $p = 1$, $\lambda_1 = 148.4$, $\lambda_2 = 2.71$, $\lambda_3 = 1$.

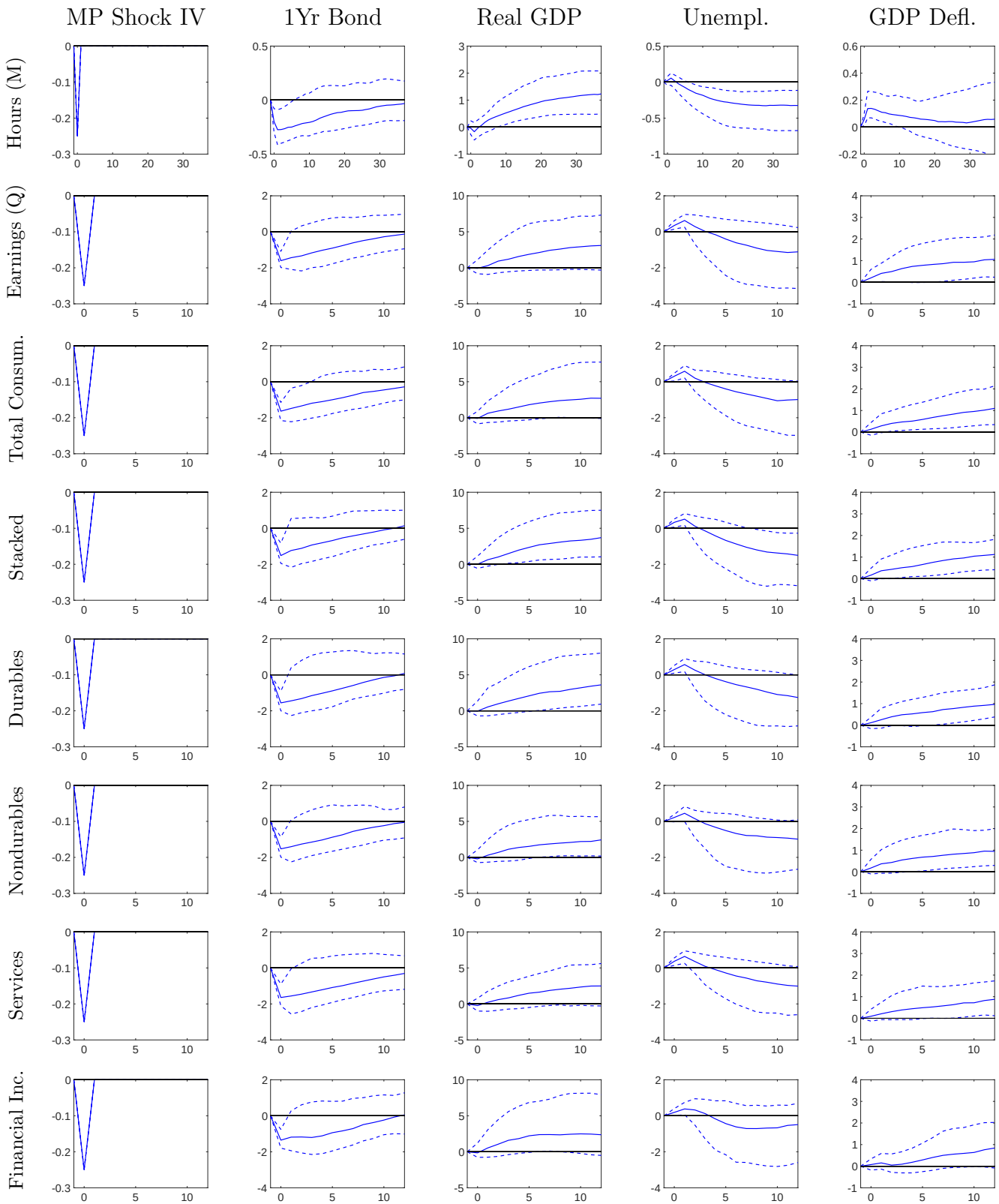
F Additional Empirical Results

F.1 IRFs of Aggregate Variables

In Figure A-1 we are plotting the response of aggregate variables to a monetary policy shock for other model specifications discussed in the main paper. Note that the functional VAR for hours worked (plots in the first row of the figure) is estimated based on monthly data, which is why the horizon ranges from zero to 36 months (3 years). The remaining functional VARs are estimated based on quarterly data and the IRF horizons correspond to quarters, ranging from zero to 12 (3 years).

Qualitatively, the aggregate responses from the quarterly models are similar to the responses from the monthly models. Quantitatively, there is a difference in magnitude. The reason is that a 25 basis point surprise at monthly frequency is only roughly a third of a 25 basis point surprise over an entire quarter. In calculations not reported in the paper we time aggregated the responses from the monthly earnings VAR to quarterly frequency and re-scaled the IRFs such that the instrument moves by 25 basis points over the quarter. After this re-scaling, the magnitude of the aggregate responses from the VARs estimated with monthly and quarterly data, respectively, were indeed very similar.

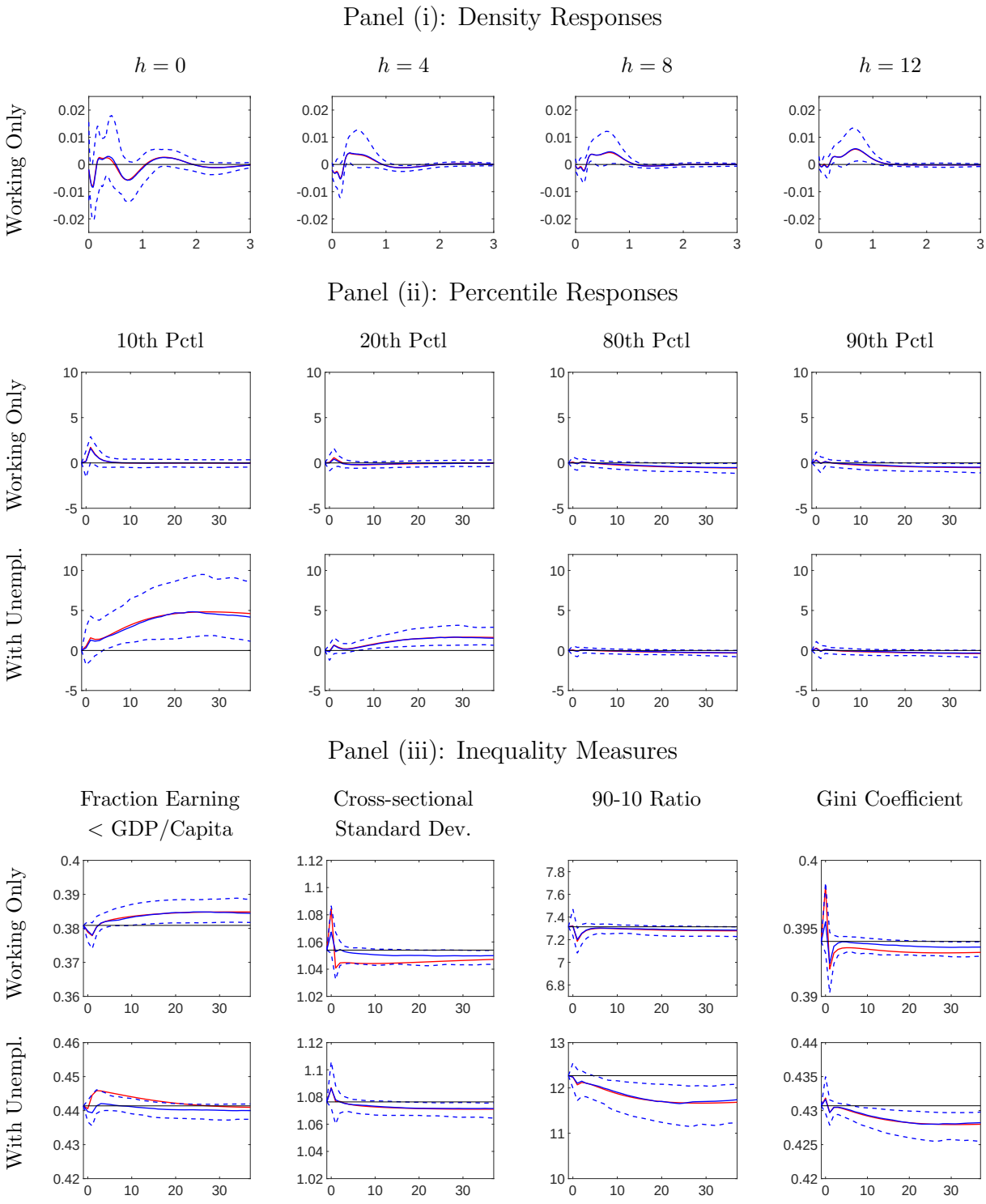
Figure A-1: Responses of Aggregate Variables to MP Shock



F.2 Responses at the Posterior Mean Parameters

In Footnote 11 we mentioned that the pointwise posterior medians of the density differentials do not generate the posterior medians of the percentile responses. In Figure A-2 we overlay the responses generated from the posterior mean parameter estimates, which look very similar to the posterior medians reported in Figure 4 of the paper.

Figure A-2: Response of Earnings to Monetary Policy Shock

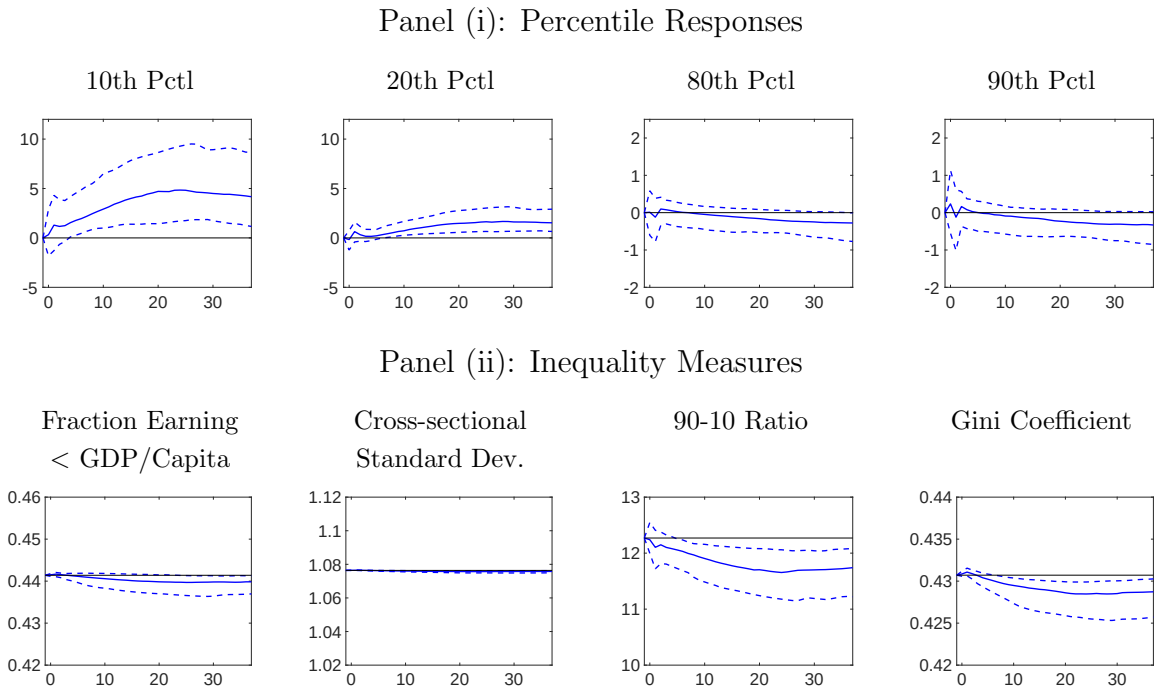


Notes: 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution. The red line corresponds to IRFs that are generated by fixing the parameters at their posterior mean estimates.

F.3 Effect of Unemployment on Earnings Percentile and Inequality IRFs

To justify the claim that the earnings percentile and inequality IRFs are dominated by changes in unemployment we conduct the following exercise. When computing the response of percentiles and inequality statistics we keep the shape of the continuous part of the earnings distribution fixed at its initial level. We only change its normalization to $1 - UR_h$, where UR_h is the unemployment rate response at horizon h. We then compute the percentiles and inequality statistics from the re-normalized continuous part plus the UR_h point mass at zero. The results are plotted in Figure A-3. The percentile responses are visually identical to the "With Unempl." plots in Figure 4 of the main text, which also account for movements in the shape of the continuous part of the earnings distribution. Accordingly the response of the 90-10 ratio is identical as well. The main feature that the simplified calculation does not capture is the slight reduction in the cross-sectional standard deviation.

Figure A-3: Earnings Percentile and Inequality IRFs, Holding Continuous Part Fixed



Notes: The system is in steady state at $h = -1$ and a 25bp monetary shock occurs at $h = 0$. 10th (dashed), 50th (solid), and 90th (dashed) percentiles of the posterior distribution. Horizon h is on the x-axis.

F.4 IRFs from a Model that Stacks Earnings and Consumption Densities

F.4.1 Baseline Analysis

In this section we are comparing the density responses for earnings and consumption from a functional VAR that only includes a single cross-sectional density to a functional VAR that includes two cross-sectional densities simultaneously. The responses are quantitatively very similar. They are plotted in Figures A-4 and A-5.

Figure A-4: Responses of Earnings to MP Shock: Stacked vs. Single Density

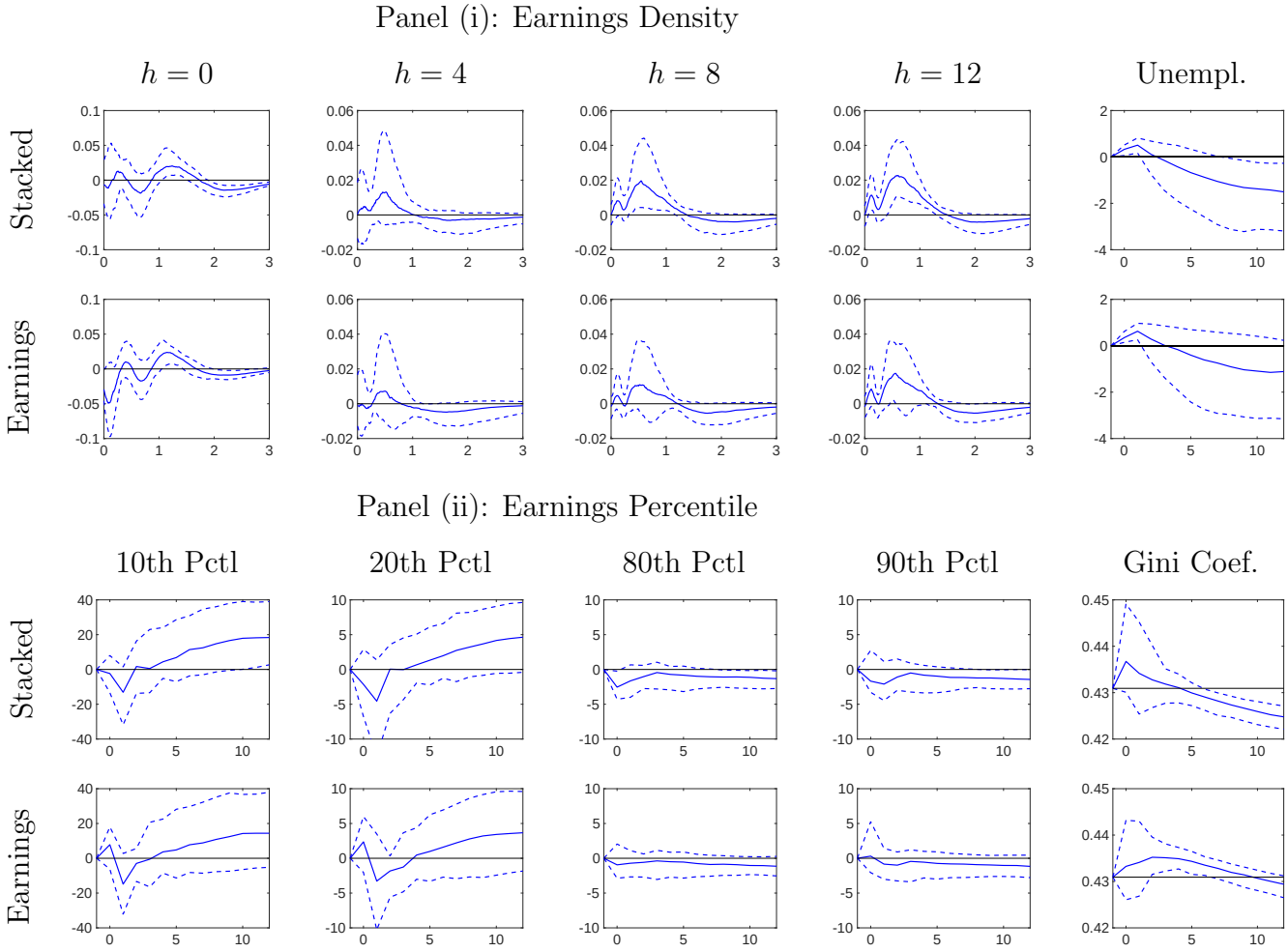
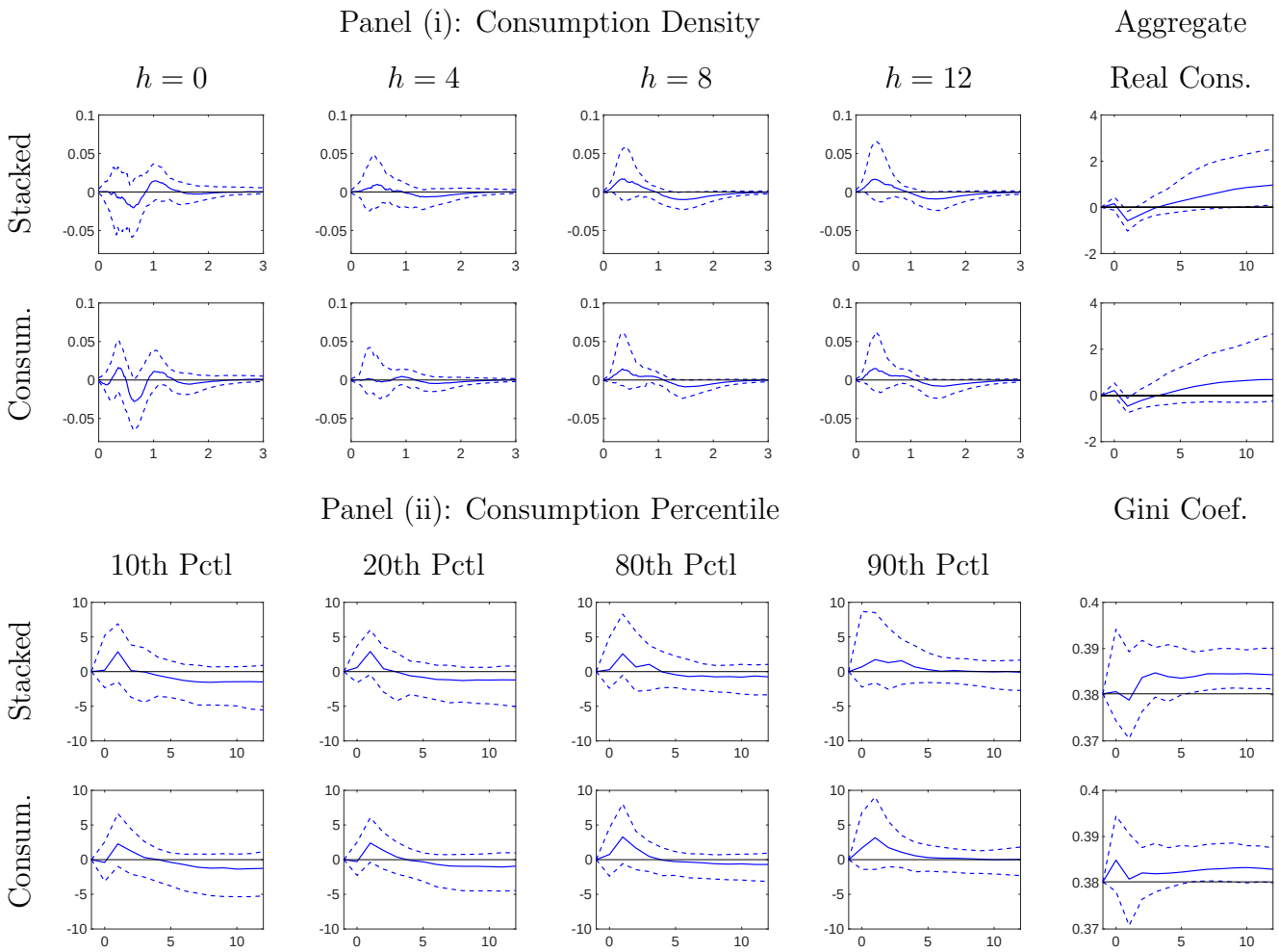


Figure A-5: Responses of Consumption to MP Shock: Stacked vs. Single Density



F.4.2 Estimation under an Alternative Prior

The prior distribution is largely identical to the baseline prior used previously, except for the construction of the matrix $\underline{V}_i^B$. Rather than deriving this matrix from a prior for the reduced-form coefficients, we specify it directly. In particular, the $k \times k$ prior covariance matrix $\underline{V}_i^B$ is assumed to be diagonal with elements $l = 1, \ldots, k$:

$$[\underline{V}_i^B]_{ll} = \begin{cases} \frac{1}{\lambda_1}\frac{1}{\underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if vars } (i,j) \text{ belong to same block} \\ \frac{1}{\lambda_1}\frac{1}{\lambda_{2.y1}\underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if var } i \text{ belongs to } Y \text{ and } j \text{ belongs to } a_1 \\ \frac{1}{\lambda_1}\frac{1}{\lambda_{2.y2}\underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if var } i \text{ belongs to } Y \text{ and } j \text{ belongs to } a_2 \\ \frac{1}{\lambda_1}\frac{1}{\lambda_{2.1y}\underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if var } i \text{ belongs to } a_1 \text{ and } j \text{ belongs to } Y \\ \frac{1}{\lambda_1}\frac{1}{\lambda_{2.2y}\underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if var } i \text{ belongs to } a_2 \text{ and } j \text{ belongs to } Y \\ \frac{1}{\lambda_1}\frac{1}{\lambda_{2.12}\underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if var } i \text{ belongs to } a_1 \text{ and } j \text{ belongs to } a_2 \\ \frac{1}{\lambda_1}\frac{1}{\lambda_{2.21}\underline{s}_i^2 h^{\lambda_4}} & \text{for coeff. on the } h\text{-th lag if var } i \text{ belongs to } a_2 \text{ and } j \text{ belongs to } a_1 \\ \frac{1}{\lambda_5} & \text{for the intercept} \end{cases}$$

MDD Selection and Comparison: The chosen lag length is $p = 1$ and the selected hyperparameters are $\lambda_1 = 3.49$, $\lambda_{2,y1} = 6310.69$, $\lambda_{2,y2} = 485165195.41$, $\lambda_{2,1y} = 148.41$, $\lambda_{2,2y} = 148.41$, $\lambda_{2,12} = 148.41$, $\lambda_{2,21} = 148.41$. The spline dimensions are $K_E = 8$ (labor earnings), and $K_C = 6$ (consumption) based on MDD maximization. The alternative prior leads to an MDD of -1,436,178.44. The baseline prior yields an MDD of -1,436,457.66.

IRFs. The IRFs under the baseline and alternative priors are plotted in Figures A-6 and A-7. The IRFs under the two priors are essentially identical. Despite the slightly higher MDD of under the alternative prior (which has a bit more flexibility in terms of shrinking different blocks of coefficients to zero) we are reporting results for the baseline prior in the main text.

Figure A-6: Responses of Earnings to MP Shock

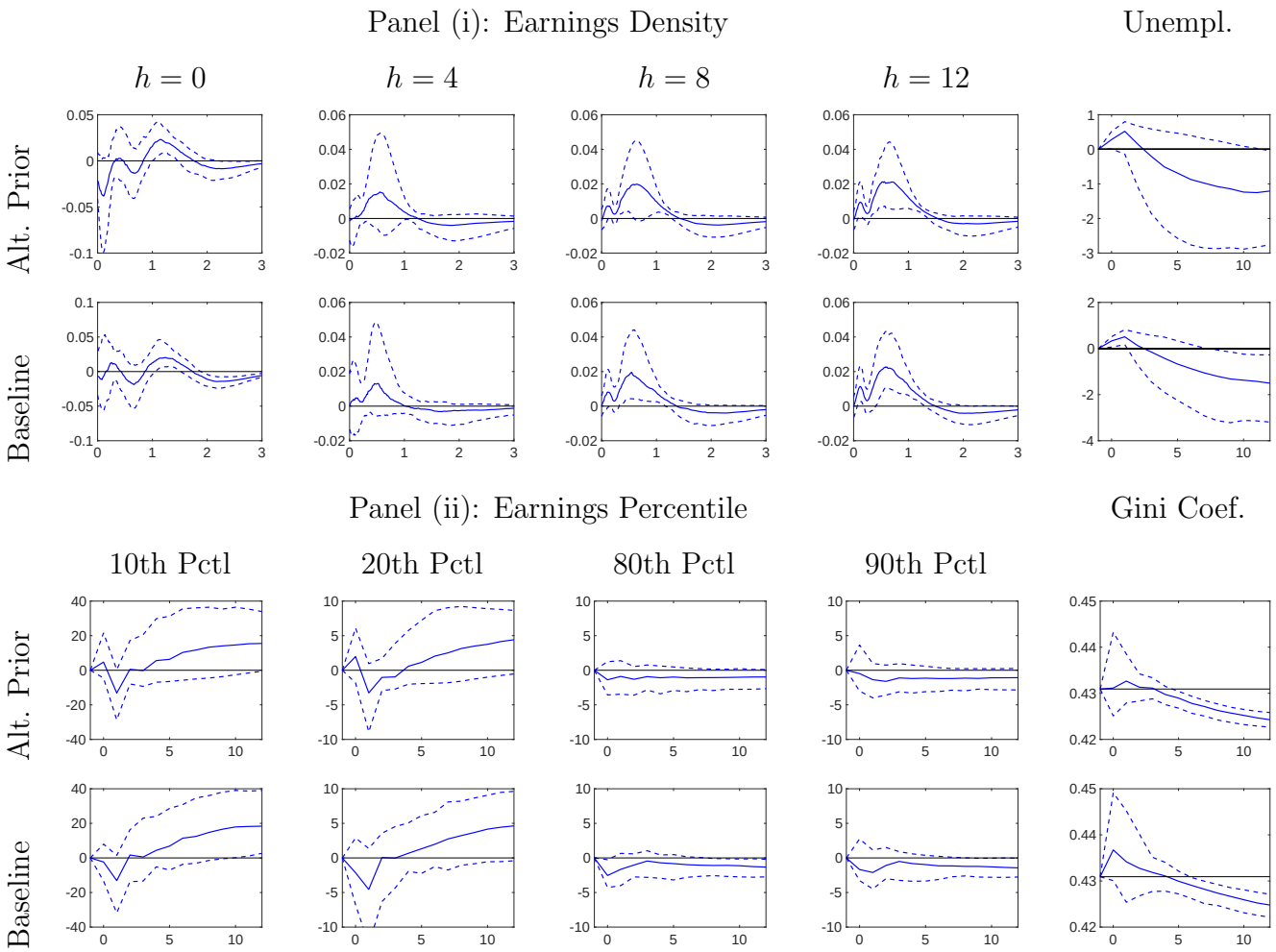
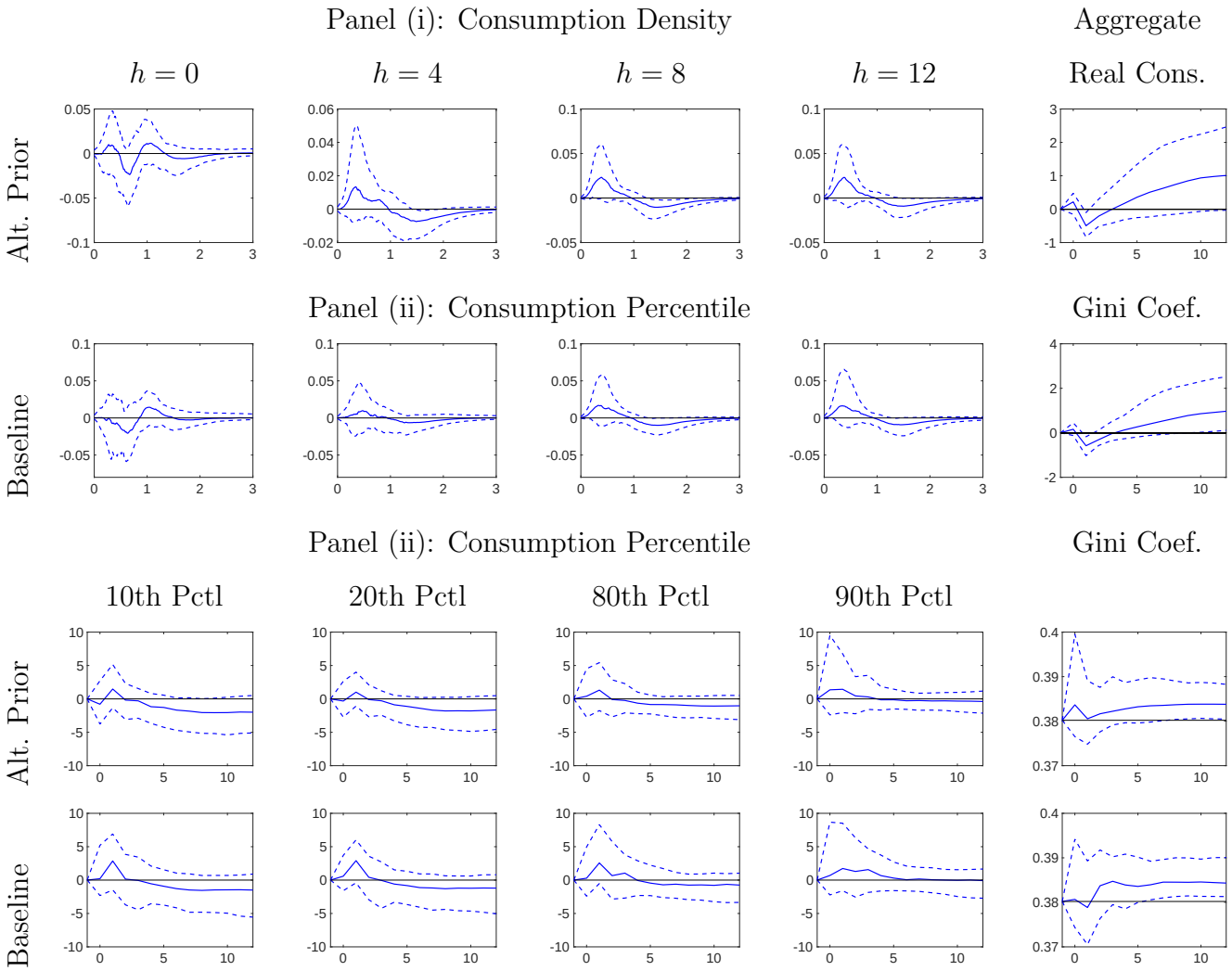


Figure A-7: Responses of Consumption to MP Shock



F.5 Estimation Results Based on a Pre-QE Sample

As a robustness check we re-estimated the functional models for earnings and consumption based on samples that end in 2008:M10 and 2008:Q3, respectively. A comparison between pre-QE and full sample results is provided in Figures A-8 to A-11.

Earnings Distribution. Figures A-8 compares the responses of the aggregate variables and the responses of the density differentials for the full sample and the pre-QE sample. The IRFs of the density differentials are quantitatively very similar for the two sample periods. The responses of the aggregate variables, however, is somewhat different. Foremost, because the pre-QE sample is shorter, the bands are much wider. But also, the unemployment response is more sluggish. Figures A-9 provides a comparison of the responses of the percentiles and the inequality measures. Here the response of the 10th percentile is not quite as strong in the pre-QE sample as in the full sample. This is in part due to differences in the unemployment response, and in part due to subtle difference in the density response. Qualitatively, it remains true that at the posterior median there is a decrease in the inequality measures, but it is not as strong. However, the credible bands do have a lot of overlap. Thus, factoring in the overall level of uncertainty, shortening the sample does not overturn the results from the full sample analysis.

Consumption Distribution. Results are plotted in Figures A-10 and A-11.The responses of the aggregate variables look very similar. In regard to the density responses, the biggest difference is that at horizon $h = 8$ and $h = 12$ the posterior median responses are slightly stronger for the pre-QE sample and the bands are wider. In regard to the percentile responses the main difference is that for $h = 10$ quarters the pre-QE percentile response are negative, whereas they revert back to zero in the full sample. In regard to the inequality measures under the full sample the estimated posterior median increases in inequality (standard deviation, 90-10 ratio, and Gini coefficient) are slightly larger for the pre-QE than the full sample.

Figure A-8: Response of Earnings: Full Sample vs. pre-QE, Part 1

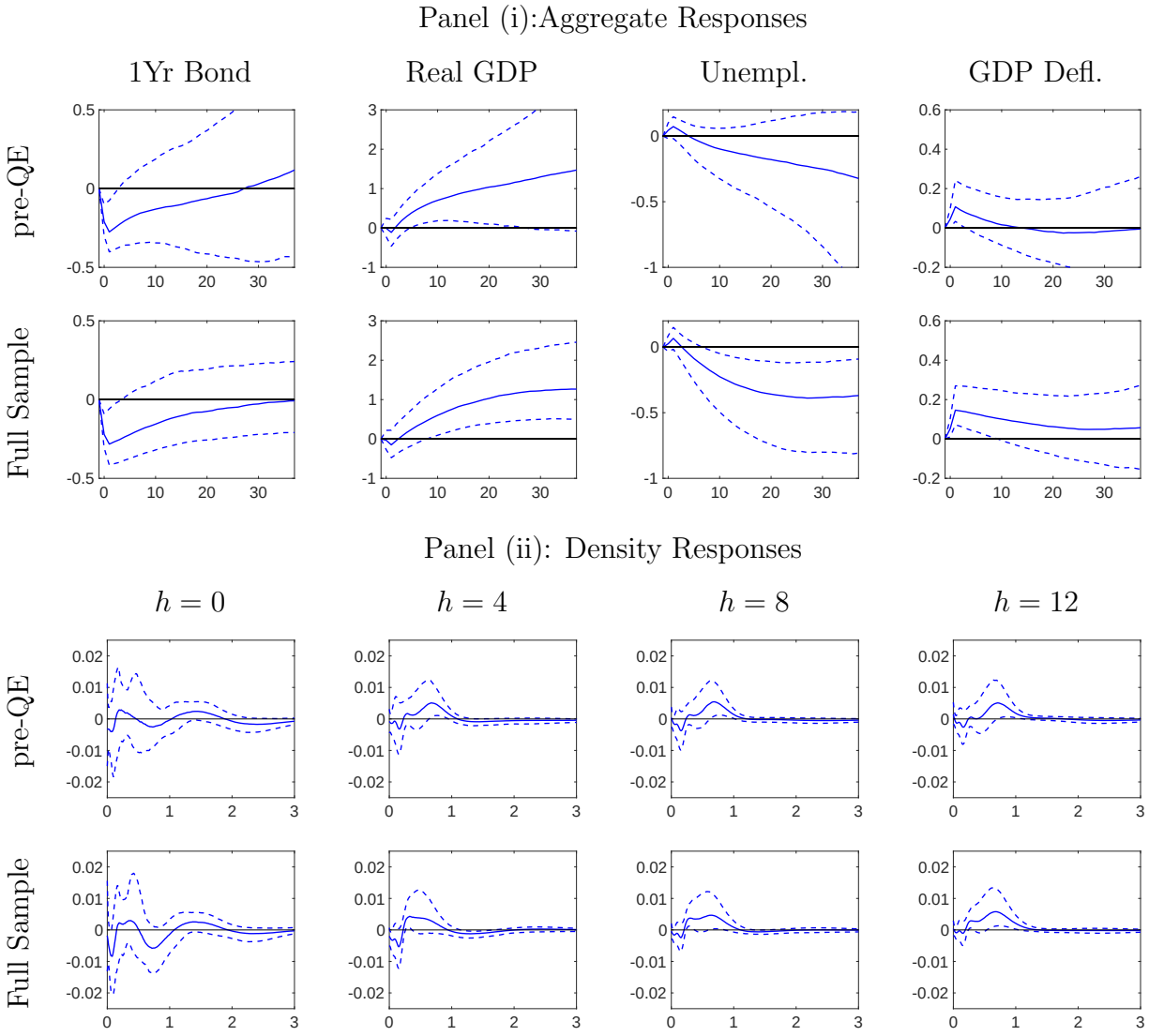


Figure A-9: Response of Earnings: Full Sample vs. pre-QE, Part 2

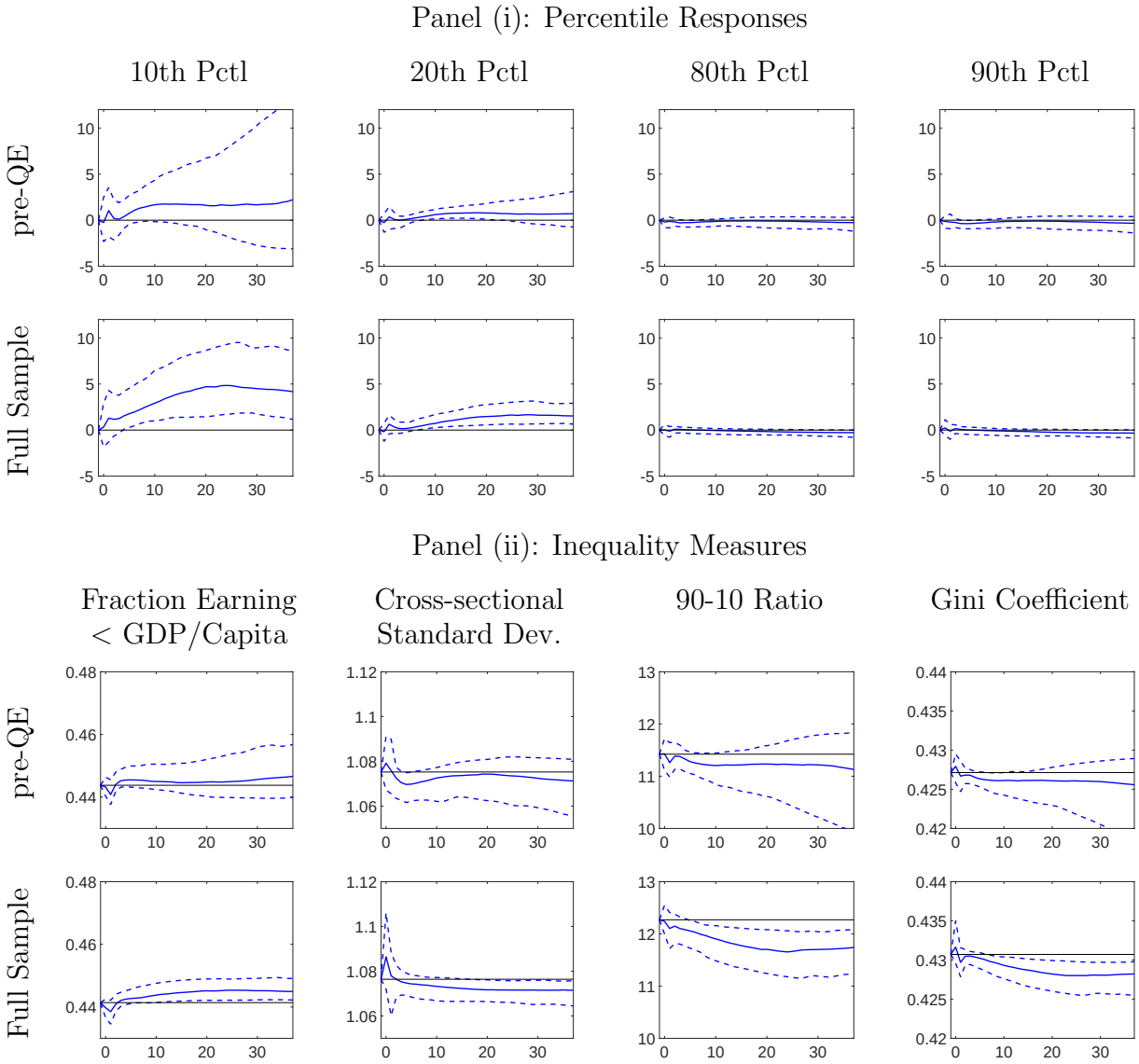


Figure A-10: Response of Consumption: Full Sample vs. pre-QE, Part 1

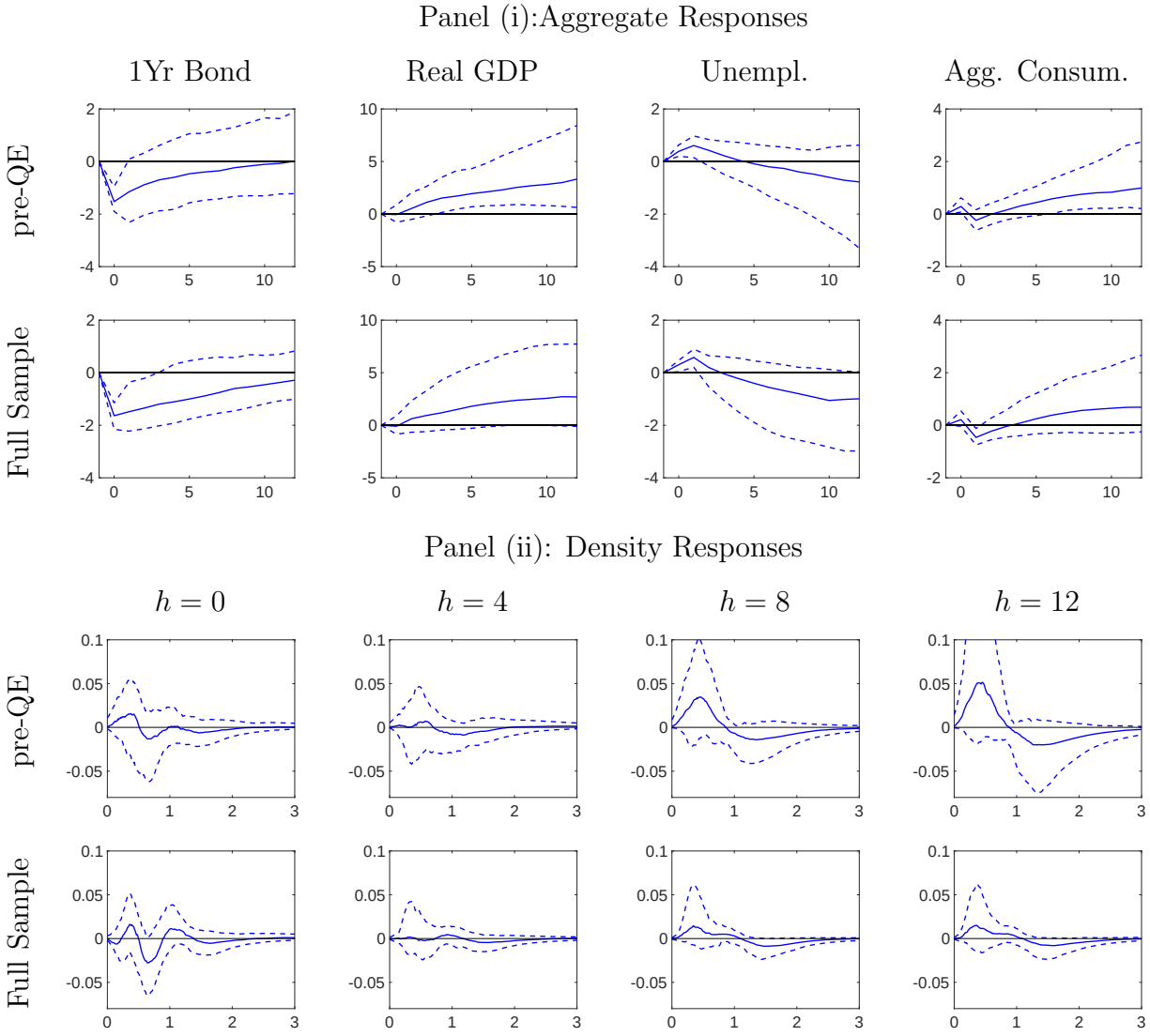


Figure A-11: Response of Consumption: Full Sample vs. pre-QE, Part 2