Data, Inequality, and Oil

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschaftswissenschaften

durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von

Luis Calderon
aus Miami, FL, USA

2026

Dekan:	Prof. Dr. Martin Böse
Erstreferent:	Prof. Dr. Christian Bayer
Zweitreferent:	Prof. Dr. Benjamin Born
Tag der mündlichen Prüfung:	7. Juli 2026

Acknowledgments

I would like to express my sincere gratitude to my supervisors, Prof. Dr. Christian Bayer, Prof. Dr. Benjamin Born, and Prof. Farzad Saidi.

I'd like to first give a special acknowledgement to Christian. He was there from the beginning, and he believed in me when no one else did, even myself sometimes. His patience is like no other. Whenever people ask me about him, I always say nice things, which backfire once they attend his Macro II course. Jokes aside, he gave me the time to find my path and for that I am eternally grateful.

To Farzad: When I met you, I felt that I had found someone who could relate to my upbringing, and through that I could come to you for anything and express it in the manner that I was taught. You were always direct, which I appreciated. I am a social, emotional person, but can comfortably take critique or whatever because it's my job. It's my livelihood. And it's for the love of science and the recognition that you mean well.

I also thank the faculty, staff, and friends at my institute, the Institute for Macroeconomics and Econometrics. It's been my home since 2019 and hopefully for a long time to come.

A shoutout to my friends, of whom I have too many, for their unwavering support, but a few warrant a special acknowledgement. Thank you, Annie, for being there en las buenas y en las malas—and thank your mom for being there too. You are so dear to me and know I'll always be there for you unconditionally. Thank you, Nigar, for being there as well and helping me disconnect when I needed it. You have a special place in my heart and that will never change. Thank you, City Girl, for just being who you are, always happy, smiling, and positive—an energy that very few have, and one I do not take for granted. Thank you, Crawfish and Tinafish, for also just being there and, at times, providing some much-needed guidance. I love you both. Thank you, butt, for always being there; you were there from the very, very beginning and dealt with me and my nonsense. You still do.

And finally, my dog Bobby, of course, has to be mentioned. Unfortunately, many people here have not had the pleasure of meeting him. He's the best dog I know and no one he's ever met disagrees.

Contents

Introduction

This thesis consists of three essays in distributional macroeconomics. Each essay studies how the joint distribution of household consumption, income, and wealth interacts with macroeconomic dynamics in the U.S. context, combining functional-data methods, Bayesian time-series models, and structural identification.

The **first chapter**, *Data and Inequality* (joint with Christian Bayer and Moritz Kuhn), develops a method for reconstructing high-frequency synthetic distributions of consumption, income, and wealth. We treat the joint distributional data as a time series of functions whose underlying factor structure follows a state-space model estimated with Bayesian techniques. The method can incorporate microdata sources of different frequencies and variable coverage, and we use a wide range of U.S. surveys to document high-quality quarterly distributional data. These data are informative about modern theories of macroeconomic dynamics and provide the empirical backbone for the analyses in the remainder of the thesis.

The **second chapter**, *Oil Supply Shocks and Inequality*, documents how exogenous disruptions to oil supply propagate through the joint distribution of U.S. household income, consumption, and wealth. I compare two structural identifications of the oil supply shock side-by-side: the realized production-shortfall shock of Baumeister and Hamilton (2019) and the oil supply news shock of Känzig (2021). The two shocks have similar aggregate footprints but operate through distinct transmission channels and have opposite distributional consequences: production shortfalls raise inequality, whereas oil supply news mostly reduces it. The framework is a Bayesian VAR that embeds the distributional factors from the first chapter alongside standard oil-market and macroeconomic variables, and the wedge between the two shocks is traced through the causal-mediation decomposition of Dufour and Wang (2024). The substantive conclusion is that the type of oil supply shock matters for who bears the cost, sharing Kilian and Lewis's (2011) view that policy responses should depend on the underlying causes of oil price shocks.

The **third chapter**, *Oil Supply Shocks and Monetary Policy*, asks how the macroeconomic and distributional consequences of an oil supply shock would have changed under alternative monetary policy rules. Building on the sufficient-statistics counterfactual framework of Caravello, McKay, and Wolf (2024), I estimate a Bayesian VAR that jointly identifies a Känzig (2021) oil supply news shock alongside six identified monetary policy innovations. I evaluate three counterfactual rules: a nominal rate peg, strict inflation targeting, and an optimal dual-mandate rule that minimises a quadratic loss in inflation and the unemployment gap. The rules are enforced over a finite constraint window, with the systematic

Fed reaction function continuing to operate thereafter. Counterfactual responses of the distributional factors follow by linear superposition, so the counterfactual income, wealth, and consumption distributions can be reconstructed for any rule that admits a linear constraint on outcomes. The three counterfactual rules deliver macro paths broadly within the credible band of the realised response; on the cross-section, the dual-mandate rule tracks the realised path closely while the rate peg and strict inflation targeting both compress inequality further, with strict-π the larger amplifier on every Gini margin. The mechanism is a redistribution toward the bottom of the distribution rather than top-tail destruction.

Chapter 1

Data & Inequality

1 Introduction

Understanding the dynamics of the joint distribution of consumption, income, and wealth is central to understanding macroeconomic dynamics, the transmission of monetary and fiscal policy, and their cross-sectional effects.[1] The scarcity of high-frequency information on the joint distribution is a significant limitation in this endeavor. We propose a novel and general technique based on functional data analysis and Bayesian time-series methods to obtain high-frequency estimates of this (or other) joint distribution(s). This technique is flexible enough to combine distributional data from different microdata sources with aggregate data, even when these data are of mixed frequency and only one micro-dataset contains all variables of interest and others only a subset.

The challenge is that the joint distributions of consumption, income, and wealth are functional data and hence, in principle, infinite-dimensional objects. Our novel method, however, projects the infinite-dimensional objects onto a *finite* set of basis functions and then exploits standard statistical dimensionality reduction techniques. These techniques reduce the dynamics of the distribution to a factor model that has a state-space representation. This step is motivated by the heterogeneous agent macroeconomics literature which suggests that a small number of factors is sufficient to approximate distributional dynamics. Intuitively, this mirrors the fact that a small set of aggregate prices shapes household decisions and therefore affects the distribution of consumption, income, and wealth (Auclert, Bardóczy, Rognlie, and Straub 2021; Bayer, Born, and Luetticke 2024). In fact, more reduced form inequality research (Di Maggio, Kermani, and

1. See, e.g., Mian, Straub, and Sufi (2020), Holm, Paul, and Tischbirek (2021), Andersen, Johannesen, Jørgensen, and Peydró (2023), Bhandari, Evans, Golosov, and Sargent (2021), and Bayer, Luetticke, Pham-Dao, and Tjaden (2019).

Majlesi 2020; Chodorow-Reich, Nenov, and Simsek 2021; Kuhn, Schularick, and Steins 2020) has so far found much support for the price-distribution nexus, too. These prices are closely tied to the aggregate economy, which itself has a low dimensional factor structure (Stock and Watson 2002).

This set of findings from the macroeconomic literature has two implications for the joint evolution of aggregate and distributional data: First, the dynamics of the distributional data can be represented by a medium-size state-space model. Second, the states of this model are driven by a small set of aggregate factors and unobserved distributional shocks. The combination of these two facts is the key innovation to overcome the challenge of dealing with multidimensional functional data. In practice, we use factor analysis to uncover the lower-dimensional state-space representation of the distributional dynamics and its aggregate drivers. Given these factor structures for the aggregate and distributional data, we estimate the time-series behavior of the functional, i.e., distributional, data using Bayesian techniques and link aggregate and distributional data without imposing a fully-structural macroeconomic model. This makes the method general such that it can be applied to a wide array of distributional dynamics.

The state-space representation lends itself naturally to the use of the Kalman-filter for Bayesian estimation of the state-space model. This has several important advantages. It allows us to use and merge numerous micro-datasets that refer to the same economic object but with different operationalized measures, e.g., differences in the sources of income covered. These different operationalizations, alongside sampling uncertainty, are captured by the measurement error in the observation equation of the model. The observation equation also allows for the combination of micro-datasets with different sampling frequencies and also allows us to exploit information from microdata that contain only a subset of the variables of interest (Schorfheide and Song 2015; Durbin and Koopman 2012).

Finally, we overcome the limited availability of high-frequency distributional information and construct estimates of business cycle fluctuations for joint distributions at any point in time, including periods where microdata on distributions are unavailable. The estimated state-space model allows us to construct synthetic high-frequency distributional data by means of the Kalman smoother. The synthetic data itself then incorporate the information of various micro- and aggregate data sources. Although the synthetic distributions are originally functional, they can be expressed approximately in the form of repeated cross sections of microdata with consumption, income, and wealth of synthetic households.

To demonstrate the power of the novel method, we apply the new estimation technique to a rich set of U.S. household microdata from the *Panel Study of*

Income Dynamics (PSID), the *Survey of Consumer Finances* (SCF), the *Consumption Expenditure Survey* (CEX), the *Survey of Income and Program Participation* (SIPP), and the *Current Population Survey* (CPS). Using the various microdata jointly, our method overcomes three existing challenges. First, only the PSID contains all three variables of interest: consumption, income, and wealth. In the other microdatasets, at least one of the variables is absent. At the same time, *all* of the microdatasets contain information on the joint distribution of consumption, income, and wealth. Second, all of these datasets are available at different frequencies. Third, they differ in sampling approaches and details of measurement concepts. Our method deals with all three challenges. We complement these microdata with a comprehensive set of macroeconomic time series.

From the estimation of the state-space model on these data, we then construct high-frequency synthetic distributional data represented by groups of households. Each group is defined by a particular combination of quantiles of consumption, income, and wealth. Over time, the conditional expectations for each quantile change and so do the consumption, income, and wealth of each group. The population weight reflects how likely it is to observe combinations of quantiles, and therefore also the weight changes over time. Thus, the dynamics of the population weights induce the dynamics of the cross-sectional correlations in the three variables. We construct the detrended business cycle variations of the joint distribution in consumption, income, and wealth from 1962 to 2024.

We carefully validate each step of the estimation procedure. First, we show that the factor representation of the distributional data imposes almost no loss of information compared to the information provided by the microdata when observed. Second, we show that the model closely predicts the distributional data, even when unobserved, through significant comovement with aggregates. Specifically, we show this for the consumption distributions of the CEX and the wealth distributions of the SCF. Finally, we use simulated data from a theoretical HANK model as a data generating process and show that the method can closely track the true high frequency movements of the distribution given the typical sample sizes and frequencies of micro data. Thus, we conclude that the imputed distributional data stemming from our statistical model are reliable.[2]

Finally, we illustrate the usefulness of the method in two applications. First,

2. In addition, we also validate the choice of priors in Bayesian estimation, particularly with respect to measurement error, and show that the state-space model is consistent with the sampling uncertainty of the observed distributional data at the sampling points. We further compare the prediction for the dynamics of income and wealth distribution with that implied by the distributional flow of funds methodology (DFA, see Batty, Bricker, Briggs, Friedman, Nemschoff, Nielsen, Sommer, and Volz 2020) and that estimated by the World Inequality Database (Alvaredo, Atkinson, Chancel, Piketty, Saez, and Zucman 2016; Piketty, Saez, and Zucman 2018). These re-

we use the estimated distributional state-space model to quantify what drives distributional fluctuations. A forecast error variance decomposition shows that aggregate shocks account for the bulk of the variation in the distributional factors (at least 70% for every factor and typically above 85%), and similarly for distributional moments such as conditional means of consumption, income, and wealth. Then, we show consumption dynamics along the joint distribution of income and wealth across the last three U.S. recessions. This application provides a natural stepping stone for future research to better understand particular distributional channels. Furthermore, it provides novel empirical estimates on various groups of households in the joint distribution of consumption, income, and wealth that remain unobserved in existing data—providing new empirical model targets to guide the modeling of business cycles with heterogeneous agents.

The observed dynamics are also economically interesting. We find overwhelming evidence that the impact of macroeconomic shocks on household consumption (relative to the aggregate) is not uniform across the household groups or across recessions. We see this as a complementary work to Bilbiie, Galaasen, Gürkayna, Mæhlum, and Molnar (2025), showing for Norway that automatic stabilizers *on average* render household heterogeneity largely irrelevant for aggregate dynamics. Our analysis for the United States, however, shows a more nuanced, state-dependent role for heterogeneity, where the nature of the prevailing shock in a recession is a critical determinant of how distributional dynamics affect the macroeconomy. As we expand on this later, these findings show that the joint distribution of income and wealth is crucial for understanding consumption dynamics around recessions and policy decisions.

The remainder of this chapter is organized as follows: Section 2 provides an overview of the relevant literature. Section 3 develops our estimation method and Section 4 evaluates its quality. Section 5 provides an application of the novel method. Finally, Section 6 concludes the chapter. An appendix follows.

2 Literature

Methodological Antecedents. The paper most closely related to ours is Chang, Chen, and Schorfheide (2024), which develops a Bayesian state-space approach to estimate the coevolution of aggregate variables and the marginal distribution of earnings. Our work differs in three key aspects. First, we extend their approach to model the evolution of joint distributions over time, e.g., distributions of consumption, income, and wealth with also a richer set of aggregates. This

sults can be found in Appendix G.

allows us to capture the dependence structure across variables within the distribution, rather than focusing solely on marginal distributions, and to understand the interplay between the distribution and aggregates. Second, instead of directly estimating the distributional dynamics, we follow Kneip and Utikal (2001) and Tsay (2016) by additionally projecting the functional data on an ideal basis through PCA (see also Meeks and Monti 2023, for a macroeconomic application). We then estimate a state-space model in the lower-dimensional factor space—not possible with the functional data alone. Third, our chapter focuses on generating high-frequency synthetic distribution data, addressing challenges related to missing observations and the mixed frequencies of aggregate and microdata.[3]

More broadly, our method builds on the literature that formulates temporal disaggregation in a state-space framework (see e.g., Harvey and Chung 2000) and on functional data methods in economics (see e.g., Kneip and Utikal 2001; Diebold and Li 2006; Chang, Kim, and Park 2016). The key advantage over static methods such as Chow and Lin (1971) and Fernandez (1981) is that the state-space representation is inherently dynamic and lends itself to the Kalman filter for estimation and diagnostics.

Macroeconomic motivation. Our contribution relative to the growing body of work linking macroeconomic aggregates with distributional dynamics (see e.g., Baumeister et al. 2025; Koop et al. 2026; Ettmeier, Hyun Kim, and Schorfheide 2024) is to treat the *joint* distribution—marginals and their dependence structure—as the state variable itself, recovering its dynamics from multiple partially overlapping surveys at mixed frequencies. This provides the still-missing descriptions of the short-run dynamics of the consumption, income, and wealth distribution to the theoretical literature on heterogeneity and aggregate dynamics (Kaplan, Moll, and Violante 2018; Bayer et al. 2019; Bayer, Born, and Luetticke 2024) and extends the rapidly growing empirical work on the effects of policy shocks on marginal distributions.[4]

Through our application, we also speak to a central debate on macroeconomic amplification from household inequality. Auclert (2019) and Bilbiie (2020) formalize the idea that amplification depends on the covariance between the

3. While there is a nascent literature (see, e.g., Ettmeier, Hyun Kim, and Schorfheide 2024) emphasizing the role of unit-level dynamics in properly addressing the mixed-frequency problem, we abstract from this consideration in the present chapter and follow the standard mixed-frequency approach.

4. See, for example, Berger, Bocola, and Dovis (2023), Coibion et al. (2017), Cloyne, Ferreira, and Surico (2020), Holm, Paul, and Tischbirek (2021), Chang and Schorfheide (2024), and Bartscher et al. (2022). McKay and Wolf (2023) surveys this literature.

marginal propensity to consume and the cyclicality of income. Patterson (2023) finds this covariance to be strongly positive for U.S. labor earnings, creating a powerful "matching multiplier"; recent evidence from Bilbiie et al. (2025) for Norway provides a counterpoint, showing that once disposable income is considered, automatic stabilizers render the covariance negligible over *average* business cycles. Our synthetic data reveal that consumption distribution dynamics are potentially *very* business-cycle-specific, suggesting that neither finding generalizes unconditionally.

Distributional measurement. The chapter also contributes to a large empirical literature, beginning with Piketty and Saez (2003), measuring trends and fluctuations in inequality. Most of this work produces high-frequency data for *marginal* distributions—income or wealth separately—but not for their joint distribution. Blanchet, Saez, and Zucman (2022) construct monthly income and wealth distributions for the United States; Smith, Zidar, and Zwick (2023), using administrative data and the capitalization method of Saez and Zucman (2016), provide high-frequency wealth estimates concentrated on the upper tail. The Distributional Financial Accounts (Batty et al. 2020) produces quarterly balance-sheet estimates since 1989 by combining SCF microdata with the Financial Accounts through Chow-Lin/Fernández-type models. Our state-space approach offers three advantages over these methods: it treats the underlying microdata as noisy samples of a time series of distribution functions, explicitly handling sampling uncertainty in a dynamic setting; it can combine multiple microdata sources with different operationalizations of the same economic concepts; and it recovers the *joint* distribution of consumption, income, and wealth—not just individual marginals—going back to the 1960s.

3 Method

This section describes a general method for generating high-frequency estimates of joint distributions of economic variables of interest over a large number of micro-units. The method proceeds in four main steps, see Figure 1. First, starting from microdata, we estimate and represent joint distributions by their quantile functions coupled with a copula, isolating marginal shapes from their dependence structure (Section 3.1), projecting these functions onto Legendre polynomials (Section 3.2.1). Second, we reduce dimensionality using functional principal component analysis (PCA) (Section 3.2.2). Third, we formulate the dynamics of

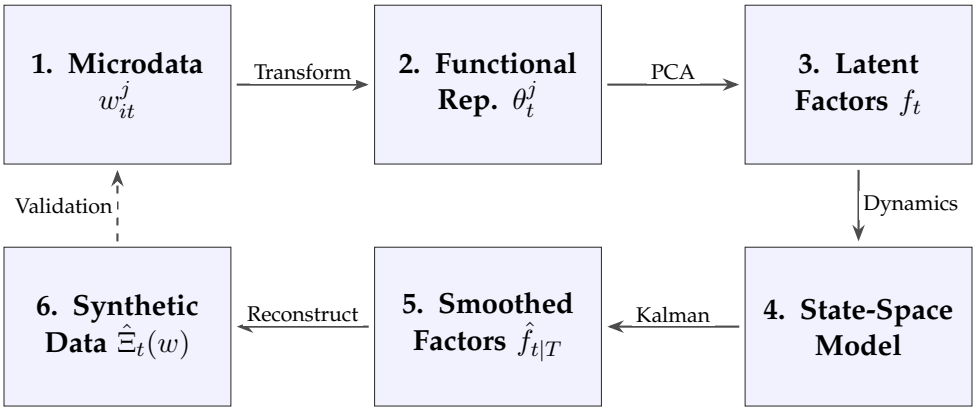


Figure 1: Visual Roadmap of the Methodology

these latent factors as a state-space model that handles mixed frequencies (Section 3.2.3). Fourth, we estimate the system using Bayesian techniques and recover the full distributional time series via the Kalman smoother (Section 3.3). The reverse provides us with the high-frequency synthetic distributional data.

This method uses microeconomic and aggregate data as inputs. It requires only the joint observation of the microeconomic variables in at least one dataset over several, but potentially infrequent, time periods. The developed method treats the distributional data as functional data in a time-series state-space framework with unobserved states. In the following, we describe the method using the example of the joint distribution of consumption, income, and wealth, an important macroeconomic application.

3.1 Distributions as time series of functional data

We consider a sequence $\Xi_t(w)$ of multidimensional distribution functions (cdfs), indexed by $t \in \mathbb{T} := \{1 \dots T\}$, defined over a d-dimensional *vector* $w \in \mathbb{R}^d$. In the case of our application, we have $d = 3$, where w is a vector of consumption, income, and wealth at the household level. In addition to this sequence of distribution *functions*, there is a sequence of real-valued vectors Y_t of stationary aggregate data. In the following exposition, we assume that Y_t is observed at all times $t \in \mathbb{T}$. The extension to missing observations in Y_t is standard.

From the distributions $\Xi_t(w)$, we observe only randomly drawn samples. We allow these samples to come from different sampling procedures or to have different operationalizations of the underlying theoretical variables. For example, the PSID and the SCF use different sampling procedures and slightly different concepts of wealth and income. We index each of the sampling procedures/datasets by $j = 1 \dots J$. All of these different datasets are typically not

observed for all time periods. Instead, dataset j is only observed in a particular subset $\mathcal{T}^j \subset \mathbb{T}$. In addition, not all samples contain all variables of interest, but may contain only a subset $\mathcal{D}^j \subseteq \mathbb{D} := \{1, \dots, d\}$ of variables. For example, the Current Population Survey (CPS) provides only income information but neither wealth nor consumption. However, at least one dataset, j, must contain all variables of interest for $\mathcal{D}^j = \mathbb{D}$. In our application, this dataset is the PSID, which contains information on consumption, income, and wealth (for some years).

Our goal is to obtain estimates of the joint distribution functions, $\hat{\Xi}_t(w)$, $\forall t \in \mathbb{T}$, by efficiently combining information from the various related microdata sources and aggregate information, Y_t. We assume that there is a time-series structure such that the density $d\Xi_t$ evolves according to the functional difference equation

$$d\Xi_{t+1} = G\left(d\Xi_t, Y_{t-\ell:t+1}\right) + \epsilon_{t+1}, \tag{1}$$

where $Y_{t-\ell:t+1}$ are observed contemporaneous and lag aggregate data; G determines the dynamics of the system, and ϵ_t are the corresponding shocks to the functional equation.[5] This time-series structure arises naturally in so-called HANK models (see e.g. Kaplan, Moll, and Violante 2018; Bayer, Born, and Luetticke 2024). It is assumed that Ξ_t is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$ and thus, $d\Xi_t$ exists and is continuous. Moreover, we require that the operator G is Lipschitz-continuous in its first argument, a condition satisfied mechanically under the proposed linear model. We provide details below.

Viewing the J sampling procedures as capturing the same fundamental object Ξ_t but with some measurement error, ν_t, allows us to combine the data in a systematic way. This means that a dataset gives us an estimate

$$d\tilde{\Xi}_t^j = \int_{\mathbb{D}\setminus\mathcal{D}^j} d\Xi_t + \nu_t \qquad \text{for} \quad t \in \mathcal{T}^j, \tag{2}$$

where the integral $\int_{\mathbb{D}\setminus\mathcal{D}^j}$ reflects that those variables unobserved in dataset j have been integrated out.[6] The measurement error then captures sampling uncertainty, differences in sampling procedures, and differences in operationalization of economic concepts.

5. To ensure Ξ_{t+1} is a valid distribution: $\int G(d\Xi, \cdot)(w), dw = 1$, $\int \epsilon_t(w), dw = 0$, and $G(d\Xi, Y_{t-\ell:t+1})(w) \geq -\epsilon_t(w)$ (non-negativity). Non-negativity applies to $d\tilde{\Xi}$ and ν in (2).

6. For clarification, the tilde ($d\tilde{\Xi}$) is to denote that this is an estimate from the data vs. an estimate generated from the model, which will be denoted by a hat ($d\hat{\Xi}$).

3.2 Implementing the Estimation

Estimating (1) directly is not feasible because it is an infinite-dimensional nonlinear functional difference equation and, of course, we only observe samples of the distribution functions, not the functional data itself. Our innovation is to overcome this challenge by making it possible to estimate (1) using traditional Bayesian techniques and a Kalman-filter (Section 3.2.4). This requires transforming (1) into a linearized (infinite-dimensional) state-space model (Section 3.2.3), which is estimable once we render it finite and of reduced dimensionality under an appropriate factor representation (Section 3.2.2). However, first, we need to operationalize the measurement of the distribution functions as they appear in (2) by transforming the microdata samples into estimates of the distribution functions themselves (Section 3.2.1). In doing so, we have to account for changes in the effective support of the distributions and deal with the unobservability of some of the micro-variables in some datasets. In the following, we cover these points, which ultimately define our estimation procedure.

3.2.1 Transforming the microdata

Handling changes in scale One challenge in working with distributional data is that the magnitude of the variables of interest in w, and thus the support of Ξ, changes over time. We deal with this in two ways. First, to deal with level changes, we rescale the vector w observed for individual i in the micro-dataset j at time t, w_{it}^j, by its dataset- and time-specific mean $\bar{w}_t^j$.[7] Second, to deal with changes in the width of the support, we decompose the distributional data into its marginals and a copula. Copulas, by definition, have a constant support (hypercubes of $[0,1]$). Representing the marginals by their quantile functions (i.e., the inverse of the marginal cumulative distribution function) again achieves constant support by construction. This quantile-and-copula representation contains the same information as Ξ, but makes the support of all functions time-invariant.

This relies on Sklar's Theorem (Sklar 1973), which states that any multivariate cumulative distribution function $\Xi_t(w)$ can be written in terms of its marginal distributions $\Xi_{mt}(w)$, along the dimension $m \in \mathbb{D}$, and a copula, $C_t : [0,1]^d \longrightarrow [0,1]$ with uniform marginals. The copula captures the dependence structure between the random variables in w and is invariant to monotone transformations in w. Intuitively, the copula isolates the dependence structure (e.g., whether high-

7. Estimating relative to dataset-specific time means also lets us flexibly align the synthetic distributions with per-capita or per-household aggregate targets: we simply rescale using the desired aggregate target rather than the dataset mean. A consensus business-cycle component can be obtained by rescaling with average fixed effects and omitting any trend.

income households also tend to be high-wealth) from the shape of the individual distributions (e.g., how skewed income is). For our application, the copula is

$$\begin{aligned} C_t(u_1, \ldots, u_d) &= P_t(U_1 \leq u_1, \ldots, U_m \leq u_m, \ldots U_d \leq u_d) \\ &= P_t\left(w_1 \leq \Xi_{1t}^{-1}(u_1), \ldots, w_m \leq \Xi_{mt}^{-1}(u_m), \ldots, w_d \leq \Xi_{dt}^{-1}(u_d)\right) \qquad \forall t \in \mathbb{T} \end{aligned} \tag{3}$$

where $U_m \sim U[0,1]$ for $m \in \mathbb{D}$ are the uniform marginals generated by taking the probability integral transform of each component m s.t. $U_m = \Xi_m(w_m) \sim U[0,1]$. The second line highlights the quantile functions or the inverse transform of the univariate CDFs, $\Xi_{mt}^{-1}(w_m)$, where:

$$\Xi_{mt}^{-1}(u_m) = \inf\{w_m \in \mathbb{R} : \Xi_m(w_m) \geq u_m\} \qquad \forall t \in \mathbb{T}, m \in \mathbb{D}. \tag{4}$$

Finally, for C_t to be a copula, the constraint must hold for all $m \in \{1 \ldots d\}$ that when integrating out all but one dimension (the marginal distribution) m, the copula is identical to the value of the marginal distribution:

$$C_t(1, \ldots u_m, \ldots 1) = u_m. \tag{5}$$

Dealing with the infinite-dimensionality of the distribution functions The second challenge in our endeavor is dealing with the infinite-dimensionality of the distribution functions, while maintaining a functional approach. To do so, we will rely on a series estimator for both the copula and quantile function. Series estimators project the functions of interest onto some infinite-dimensional space of polynomials, namely the space of squared-integrable functions $\mathcal{L}^2$. In the following, we will treat the copula densities dC_t and (transformed) quantile functions Ξ_{mt}^{-1} as elements of $\mathcal{L}^2$, and project them onto a space of shifted orthonormal Legendre polynomials $\{Q_o : o \in \mathbb{N}_0\}$ for some order o, satisfying

$$\int_{[0,1]} Q_o(x)\, Q_k(x)\; dx = \delta_{ok}, \tag{6}$$

with $\delta_{ok} = 1$ if $o = k$ (reflecting the normalization condition) and 0 otherwise by orthogonality.[8] Classical Legendre polynomials are defined on $[-1, 1]$, so the

8. One could alternatively project log-densities, as in Chang, Chen, and Schorfheide (2024), which avoids imposing non-negativity and monotonicity constraints in a linear model, but would forfeit the simple series estimators we use. Instead, we verify these constraints ex post. For quantiles, we check monotonicity by evaluating the Legendre approximation on a fine grid (10,000 points) and find no violations in our application. For the copula, the estimator satis-

shift is to the same domain as the copula densities and quantile functions—in the support $[0, 1]$. The orthonormal shifted Legendre polynomials Q_o themselves are defined on $[0, 1]$ by

$$Q_o(x) = \sqrt{2o+1}\, L_o(2x-1), \tag{7}$$

where L_o are the classical Legendre polynomials defined by

$$L_0 = 1, \qquad L_1(x) = x, \qquad (o+1)L_{o+1}(x) = (2o+1)\, x\, L_o(x) - oL_{o-1}(x). \tag{8}$$

One concern is that the approximation method is poorly chosen, as the underlying population functions are themselves not completely square-integrable. For example, empirical evidence for distributions of consumption, income, and wealth points to tail behaviour that violates the $\mathcal{L}^2$ condition (see e.g., Vermeulen 2018).[9] For the copula density, the same issue arises—density at the tails may fail to lie in the fulldomain space $\mathcal{L}^2([0, 1]^d)$ for copulas exhibiting non-zero tail dependence e.g., the Gumbel copula.

To ensure that these underlying functions lie in $\mathcal{L}^2$, rendering the projection apropos, we first treat the data and apply the inverse-hyperbolic-sine transformation on them. This regularizes the upper tail of the distribution. This treatment is similar in nature to working with log densities, as in Chang, Chen, and Schorfheide (2024).[10] We then estimate our quantile functions on these transformed values. Appendix A provides evidence that by using the transformed percentile functions, the projection performs well at capturing variation in the tails. For the copula, we impose a *local* $\mathcal{L}^2$ condition on a trimmed domain $[\varepsilon, 1-\varepsilon]^d$, for some very small $\varepsilon > 0$. Inside that interior cube, $dC_t(\mathbf{u})$ is bounded and hence square-integrable. Appendix A also addresses the treatment of measures in the presence of non-negligible zero-mass.

fies uniform margins by construction (Sections 2–3 Bakam and Pommeret 2023); we additionally check positivity ex post. Any negative values are negligible (never below -10^{-6}), so the estimates satisfy copula-density constraints for practical purposes.

9.
$$\int_a^b \big[Q(p)\big]^2\, dp = \int_a^b \Big[\underbrace{x_m\,(1-p)^{-1/\alpha}}_{\substack{\text{Pareto}\\ \text{quantile}}}\Big]^2\, dp = x_m^2 \int_a^b (1-p)^{-2/\alpha}\, dp. \tag{9}$$

In particular, if the uppertail of the income distribution on $p \in [a, b]$ follows a Pareto law with shape α, then its quantile function is $Q(p) = x_m(1-p)^{-1/\alpha}$. This lies in $\mathcal{L}^2[a, b]$ (i.e. $\int_a^b Q(p)^2\, dp < \infty$) if and only if $\alpha > 2$. When $\alpha \le 2$, the integral diverges and the second moment does not exist.

10. For notional convenience, we use $\Xi_{mt}^{-1}(u_m)$ as the treated quantile function.

Altogether, this means the following representation:

$$\Xi_{mt}^{-1}(u_m) = \sum_{o \in \mathbb{N}_0} \xi_{mot} Q_o(u) \tag{10}$$

$$dC_t(u_1, u_2, ...u_d) = \sum_{(o_1,...,o_d) \in \mathbb{N}_0^d} \kappa_{(o_1,...,o_d),t} \prod_{m=1}^{d} Q_{o_m}(u_m), \quad u_m \in [\varepsilon, 1-\varepsilon] \tag{11}$$

where both functions are represented as infinite sums of polynomials over the order of the polynomials $o \in \mathbb{N}_0$. Because of orthonormality (6) and using (10), for some fixed o, we obtain:

$$\int_0^1 \Xi_{mt}^{-1} Q_o(u) du = \int_0^1 \sum_{k \in \mathbb{N}_0} \xi_{mkt} Q_k(u) \, Q_o(u) du = \sum_{k \in \mathbb{N}_0} \delta_{ko} \xi_{mkt} = \xi_{mot}. \tag{12}$$

This means that the coefficients can be obtained as the inner products of the functions and the Legendre polynomials of some order o. The same holds true for the copula density:

$$\int_{[0,1]^d} \prod_{m=1}^{d} Q_{o_m}(u_m) \, dC_t \; du_1, \ldots, du_d = \kappa_{(o_1,...,o_d),t}. \tag{13}$$

For the estimation of these coefficients in practice, we rely on the uniformity of ranks and that ranks are within $[0, 1]$ and replace the inner product by sample averages:

$$\hat{\xi}_{mot} := N_j^{-1} \sum_i w_{mit} Q_o(u_{mit}) \tag{14}$$

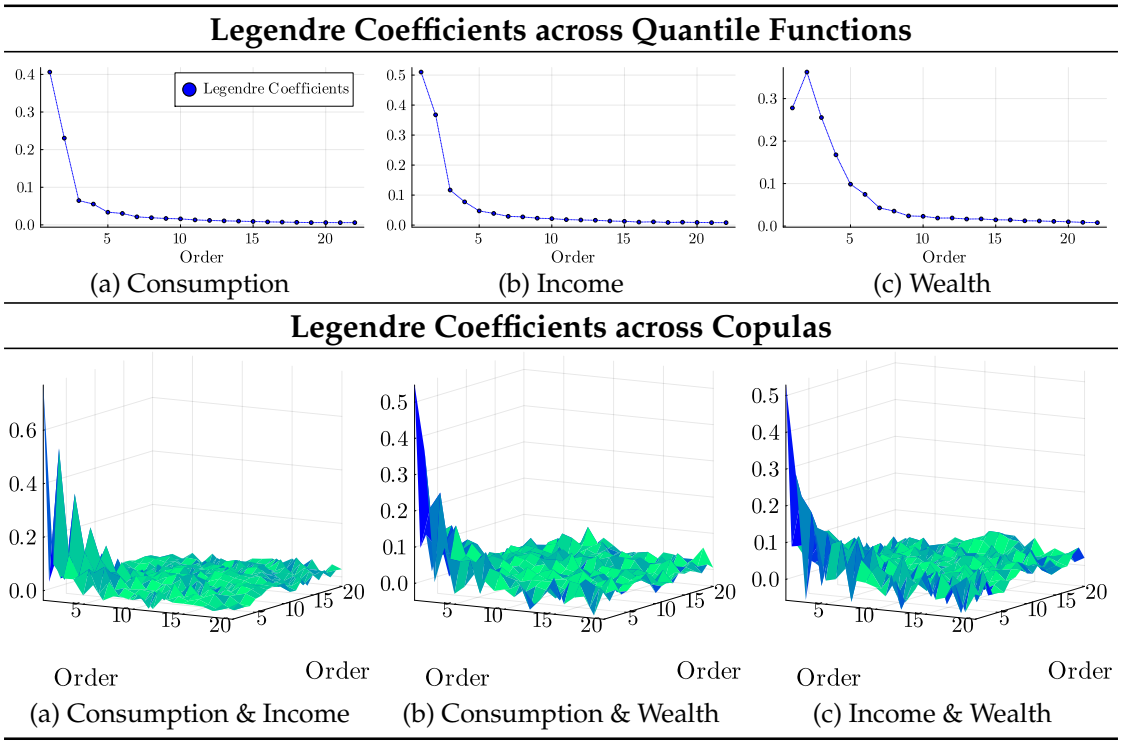
$$\hat{\kappa}_{(o_1,...,o_d),t} := N_j^{-1} \sum_i \left(\prod_{m=1}^{d} Q_{o_m}(u_{mit}) \right) \tag{15}$$

where u_{mit} is the data rank of w_{mit}, the sample analogue of Ξ_{mt}^{-1} for observation i.

In the estimation, we truncate the sums in (10) and (11) at a given maximal order O and by orthonormality of the polynomials, the kept coefficients are not affected by the truncation. The coefficients in our case indeed decrease rapidly with each polynomial as we show in Figure 2. The figure plots the absolute coefficient value as a function of the order of the polynomial term's order. At around order 10, the coefficients all become very small.[11] Evidence from Appendix A supports this, showing a truncation to $O = 11$ captures the variation observed in the data quite well.

11. A small coefficient in absolute value does imply the contribution of the corresponding polynomial to be small in an R^2 sense.

Figure 2: Legendre Coefficients Across the Distributional Data



Notes: Figure presents two rows of Panels. The first row presents the coefficients (in dots) on the Legendre polynomials from estimating the quantile function, in increasing order. The second row presents the coefficients (as a surface) on the Legendre polynomials from estimating the copula density in lexicographic order. Data are from the 2019 PSID.

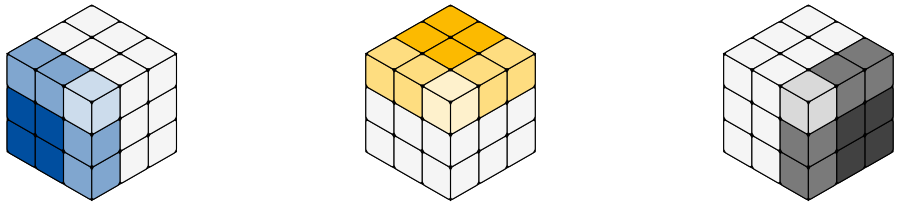
Dealing with partial unobservability of the microdata Another difficulty is that the microdata may not always contain the entire vector w, but only a subset. When w is incompletely observed, we cannot estimate the full d-dimensional copula directly. However, we can still estimate copulas with the unobserved dimensions integrated out—that is, lower-dimensional copulas. We show below that these lower-dimensional copulas are *slices* of the higher-dimensional copula of interest, and thus remain informative about the full d-dimensional object.

The representation in the form of (Legendre) polynomials is very useful in this respect. First, we need to show that the density of the higher-dimensional copula must be equal to the lower-dimensional one when we integrate out the "missing" dimension d:

$$\int_0^1 dC(u_1, \ldots, u_d)\, du_d \overset{!}{=} dC(u_1, \ldots, u_{d-1}) \tag{16}$$

In this effort, we write out the integrand using (11) and make use that the first

Figure 3: Geometric Representation of Partially Observed Copula Coefficients



(a) Only Variables 1& 3 (b) Only Variables 2& 3 (c) Only Variables 1& 2

Notes: Figure shows three cubes. A cube can be interpreted as an array of copula coefficients $\kappa^j_{(o_1,\ldots,o_d),t}$ for some dataset j at time t, for $d = 3$. Each cube corresponds to a scenario where one variable is missing in the estimation of the copula density. The light edge denotes the (1,1,1) coordinate. For each scenario, the white boxes are coefficients we cannot estimate. The slightly colored boxes correspond to the immutable coefficients, which have fixed values independent of data. The darker colored boxes are scenario specific and correspond to (time-varying) coefficients that need to be estimated.

(shifted) Legendre polynomial integrates to one and all others to zero to obtain:

$$\begin{aligned}\int_0^1 dC(u_1,\ldots,u_d)\,du_d &= \int_0^1 \sum_{o_1}\cdots\sum_{o_d}\kappa_{(o_1,\ldots,o_d)}\prod_{m=1}^{d}Q_{o_m}(u_m)\,du_d \\ &= \sum_{o_1}\cdots\sum_{o_{d-1}}\sum_{\mathbf{o_d}}\kappa_{(o_1,\ldots,o_{d-1},\mathbf{o_d})}\left(\prod_{m=1}^{d-1}Q_{o_m}(u_m)\right)\int_0^1 Q_{\mathbf{o_d}}(u_d)\;du_d \\ &= \sum_{o_1}\cdots\sum_{o_{d-1}}\kappa_{(o_1,\ldots,o_{d-1},1)}\left(\prod_{m=1}^{d-1}Q_{o_m}(u_m)\right). \qquad (17)\end{aligned}$$

In words, the polynomial coefficients of the lower-dimensional copula density are identical to the leading "slice" of the higher-dimensional copula. Therefore, a dataset that contains only two out of the three variables of interest still provides a measurement of a subset of the coefficients; see Figure 3.[12]

3.2.2 Dealing with the Curse of Dimensionality

Vectorizing the coefficients for each time period, t, leaves us with sequences of coefficients, $\theta^j_t = (\xi^j_{mot},\ldots,\kappa^j_{(o_1,\ldots,o_d),t})$, for each cross-sectional dataset, j. For example, with $d = 3$ dimensions in our application—consumption, income, and wealth and using polynomials up to order eleven ($O = 11$) to organize the data—the copula density would be represented by a vector with $(O+1)^d-(d\times O+1) = 1694$ variable entries for each time point and $d \times O + 1 = 34$ invariable. These invariable entries are not included in the estimation. In addition, there would be $d\times(O+1) = 36$ coefficients of the polynomials (including a constant) representing

12. By the same line of argument, a copula requires $\kappa_{(1,\ldots,1)} = 1$ and $\kappa_{(1,\ldots,j,\ldots,1)} = 0$.

the quantile functions, which we collect in θ_t^j as well.

This example makes apparent that even for a modest polynomial order, the dimensionality of $\theta_t^j \in \mathbb{R}^N, N = (O+1)^d + (d-1)$ for a dataset with d variables is large. Equally important is the frequency of missing observations across the N entries; in our application, the data structure is irregular and includes many gaps. Taken together, this makes it impractical to specify and estimate a time-series model directly on the raw θ_t^j coefficients. For this purpose, we postulate (and then estimate) a dynamic factor model for θ. This is where another advantage of the *orthonormal* polynomial representation is advantageous: The variance (over time) of a coefficient is proportional to its contribution to fluctuations of the function (in the $\mathcal{L}^2$ sense). Put simply: The polynomial coefficient provides a useful form of standardization that has a natural metric and allows us to uncover the factor structure behind the time-series changes in the distributions. This factor structure finally allows us to overcome the curse of dimensionality in the distributional data.

For this purpose, all (variable) coefficients of the polynomial representation of dC_t^j (and separately of the quantiles) are horizontally concatenated:

$$\boldsymbol{\theta}^j = \begin{bmatrix} \theta_{1,1}^j & & \theta_{1,T}^j \\ & \ddots & \\ \theta_{N,1}^j & & \theta_{N,T}^j \end{bmatrix}.$$

Define $\boldsymbol{\theta}$ as the matrix obtained by horizontally concatenating the individual blocks $\boldsymbol{\theta}^j$, retaining only those time periods for which every coefficient is available.[13] A PCA (a singular value decomposition) is then performed, which nonparametrically reduces the dimensionality of the data (Breitung and Eickmeier 2006; Chen, Er, and Wu 2005). For the PCA, we define $\tilde{\boldsymbol{\theta}}$, which is then standardized representation of $\boldsymbol{\theta}$. The standardization is done by distribution object $\zeta \in \{1, \ldots, d+1\}$ (d quantile functions and a copula).[14] We standardize by object (and not by coefficient) because the polynomial coefficients represent

13. Before the concatenation, each coefficient is detrended using an HP-filter. This takes care of dataset-specific effects. We store the information needed to transform $\tilde{\boldsymbol{\theta}}^j$ back to the originally observed objects to obtain source-specific predictions. For example, the income quantiles (that would be one object ζ) in the SCF and the PSID (two sources j) may be permanently different due to differences in sample design and operationalization.

14. The normalization of coefficient n in object ζ in dataset j is given by $\tilde{\theta}_{nt}^j = \left(\frac{\theta_{nt}^j - \mu_n^j}{\sigma_{\zeta(n)}^j}\right)$ where μ_n^j is the coefficient-specific mean and $\sigma_{\zeta(n)}^j$ is the standard deviation of *all* coefficients (rows) pertaining to object ζ. This removes dataset- and coefficient-specific fixed effects, μ_n^j. This can capture, e.g., permanent differences in sampling procedure and operationalization.

standardized contributions to the object's structure (e.g., the quantile function or copula). As a result, additional standardization within the same object is unnecessary, since their relative magnitudes are naturally balanced by the properties of the polynomial basis. This leaves us with the standardized observation $\tilde{\theta}^j_{nt}$ of coefficient n in data source j at time t.

The PCA of $\tilde{\boldsymbol{\theta}}$ provides us with a preliminary projection matrix $\tilde{\Gamma} \in \mathbb{R}^{N\times r}$, with full column rank r, that projects $r << N$ factors into the N dimensional distributional data. We then normalize $\tilde{\Gamma}$ by setting $\hat{\Gamma} := \tilde{\Gamma} D^{-1/2}$, where D is the diagonal matrix of eigenvalues of $\tilde{\Gamma}'\tilde{\Gamma}$ in decreasing order. Throughout we work with the normalized matrix $\hat{\Gamma}$.

More specifically, we decompose $\tilde{\boldsymbol{\theta}}$ into latent orthogonal factors $\begin{bmatrix} \hat{f}'_\Gamma & \hat{f}'_\gamma \end{bmatrix}'$ divided into "important" and "unimportant" factors according to their contribution to the total variance (measured by their singular value). Their respective time-invariant loadings are $\begin{bmatrix} \hat{\Gamma} & \hat{\gamma} \end{bmatrix}$, for $\hat{f}_\Gamma \in \mathbb{R}^{r\times T}$ and $\hat{f}_\gamma \in \mathbb{R}^{(N-r)\times T}$. This decomposition is unique up to the scale of each factor, which allows us to normalize the loadings so that all factors have unit variance. Altogether, we have:

$$\tilde{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\Gamma} & \hat{\gamma} \end{bmatrix} \begin{bmatrix} \hat{f}_\Gamma \\ \hat{f}_\gamma \end{bmatrix} \tag{18}$$

where $\hat{f}_\Gamma$ represents the r important factors, which capture almost all of the variation in the data, and $\hat{f}_\gamma$ the $N-r$ less important factors, which can be interpreted as some measurement noise. From this step, we identify an ideal functional basis f_Γ (see Kneip and Utikal 2001; Tsay 2016), which determines the *size* of the model. We then use $\hat{\Gamma}$ in the model as the linear mapping between the high-dimensional $\tilde{\theta}$ and basis f_Γ. The use of $\hat{\Gamma}$ will be a necessary condition for the proposed procedure of estimating our distributional factors. This will be elaborated on in Section 3.2.3.

The estimation of $\hat{\Gamma}$ requires time observations where every coefficient is observed, yet some data sources will not satisfy this criteria *at all*—for example, the SCF does not measure consumption ever. Observations from such incomplete sources are therefore excluded from the PCA that yields $\hat{\Gamma}$. In Appendix B, we consider alternative estimators that accommodate partial unobservability in $\tilde{\boldsymbol{\theta}}^j$, which we detail in Appendix B. Following standard practice, we select among these alternatives the estimator with the highest marginal data density.

3.2.3 Factor State Space Model and Measurement

With this preprocessing of the data, we can turn to specifying the state-space model that captures the evolution of the r latent distributional factors. Due to the possible mixed-frequency nature of the data, we postulate a mixed-frequency state space model as in Schorfheide and Song (2015), with the following augmented state vector

$$F_t = \left[f_t' \; f_{t-1}' \; \dots \; f_{t-p+1}' \right]' \in \mathbb{R}^{rp \times 1}, \tag{19}$$

where F_t collects the r latent distributional factors f_t, in addition to its $p-1$ lags. For the application of our methodology in Section 4, we set $p = 4$, with the high-frequency data of interest being quarterly. To facilitate the reading across sections, we provide the example of this quarterly case herein, though the methodology easily generalizes to other frequencies.

State Equation Stacking the lags directly with the q aggregate factors Y_t[15] in the state gives the following representation

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + \epsilon_t = \begin{bmatrix} \Phi_{FF} & \Phi_{FY} \\ \Phi_{YF} & \Phi_{YY} \end{bmatrix} \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \, \Omega) \tag{20}$$

and for the case of quarterly data

$$F_{t-1} = \begin{bmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \end{bmatrix}, \Phi_{FF} = \begin{bmatrix} A & 0 & 0 & 0 \\ I_r & 0 & 0 & 0 \\ 0 & I_r & 0 & 0 \\ 0 & 0 & I_r & 0 \end{bmatrix}, \Phi_{FY} = \begin{bmatrix} B \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Phi_{YF} = \begin{bmatrix} C \\ 0 \\ 0 \\ 0 \end{bmatrix}', \Phi_{YY} = D,$$

where distributional factors f_t are assumed to follow an AR(1) process, summarized by the law of motion matrix $A \in \mathbb{R}^{r \times r}$. In addition, they are assumed to co-move with aggregate factors Y_t, whose relationship is summarized by loading matrix $B \in \mathbb{R}^{r \times q}$ and $C \in \mathbb{R}^{q \times r}$. To represent the correlation structure between the aggregate factors, we posit $D \in \mathbb{R}^{q \times q}$ and assume it is diagonal, i.e., no cross-correlation between aggregate factors. All other matrices A, B and C are left unrestricted. As in Schorfheide and Song (2015), the remaining rows of Φ are defined to deliver the identities $f_{t-l} = f_{t-l}$ for $l = 1, ..., p-1$—hence the I_r in

15. Recall that the set of aggregate factors at time t depends on an information set that captures both contemporaneous and past aggregate variation.

the example of a quarterly estimation. While it is customary to restrict the innovations, ϵ_t, to be uncorrelated, both serially and between factors, so that Ω is a diagonal matrix, we do estimate both the variances and covariances of Ω. This allows for a shock structure where innovations to the distributional factors can be *contemporaneously* correlated with innovations to the aggregate factors.

Measurement Equation Since the factors are not directly observable, we complement the factor model with an observation equation for each dataset j:

$$\tilde{\theta}_t^j = H_t^j \left(\alpha^j \hat{\Gamma}^{MF} F_t + \nu_{F,t}^j \right), \quad \nu_{F,t}^j \sim \mathcal{N}\left(\mathbf{0}, \hat{\Sigma}_j^{1/2} S^j \hat{\Sigma}_j^{1/2\prime} \right) \tag{21}$$

This observation equation maps factors F_t, through the loading matrix $\alpha^j \hat{\Gamma}^{MF}$, to the standardized coefficients $\tilde{\theta}_t^j$, allowing for a Gaussian measurement error $\nu_{F,t}^j$. The selector matrix $H_t^j \in \{0,1\}^{N \times N}$ indicates whether (parts of) the distribution are observed in data source j at time t, following Durbin and Koopman (2012). The Gaussian error term $\nu_{F,t}^j$ has a variance-covariance matrix decomposed into S^j, a diagonal scaling matrix with diagonal entries s_ζ^j varying by object ζ and $\hat{\Sigma}_j$, a positive semi-definite (covariance) matrix.

The diagonal matrix $\alpha^j \in \mathbb{R}^{N \times N}$ and the matrix $\hat{\Gamma}^{MF}$ are constructed to take time aggregation into account. Specifically, these matrices map the high-frequency factors to the lower-frequency observations. For example, when the frequency of interest is quarterly, as it is in the application, $\alpha^j_{\zeta(n),ii}$ takes one of two values:

$$\alpha^j_{\zeta(n),ii} = \begin{cases} \frac{1}{4} \\ 1 \end{cases} \quad \text{and} \quad \hat{\Gamma}^{MF}_{\zeta(n)} = \begin{cases} \left[\hat{\Gamma}_{\zeta(n)}, \hat{\Gamma}_{\zeta(n)}, \hat{\Gamma}_{\zeta(n)}, \hat{\Gamma}_{\zeta(n)}\right] & \text{annual flow,} \\ \left[\hat{\Gamma}_{\zeta(n)}, \mathbf{0}, \mathbf{0}, \mathbf{0}\right] & \text{quarterly flow or stocks.} \end{cases}$$

First, it can take a value of $\frac{1}{4}$, which means the distributional object corresponds to an *annual flow* (e.g. the PSID income / consumption quantile function). The $\frac{1}{4}$ can be interpreted as averaging over the four quarterly factors $f_t, \ldots, f_{t-3}$ and thus loads on all four quarters i.e., $\hat{\Gamma}^{MF}_{\zeta(n)} = \left[\hat{\Gamma}_{\zeta(n)}, \hat{\Gamma}_{\zeta(n)}, \hat{\Gamma}_{\zeta(n)}, \hat{\Gamma}_{\zeta(n)}\right]$. For *stock variables* (e.g. any wealth object for which the date is known), $\alpha^j_{\zeta(n),ii} = 1$ and loads exclusively on the contemporaneous factor block and the lagged blocks are set to zero i.e., $\hat{\Gamma}^{MF}_{\zeta(n)} = \left[\hat{\Gamma}_{\zeta(n)}, \mathbf{0}, \mathbf{0}, \mathbf{0}\right]$, for $\mathbf{0}$ the same size as $\hat{\Gamma}_{\zeta(n)}$. The same holds for the third case, *quarterly flows* (e.g., SIPP income, CEX), which again loads on the contemporaneous block only. In terms of how α^j varies by object, it is straightforward for the quantile functions, since each quantile function corresponds to one variable at some point in time t; however, for the copula, since multiple variables are necessary for its construction, the different measurements of the

variables could be mixed *within* e.g., annual income flows and stock wealth. In these cases, for some point in time t, the lowest frequency is chosen among the set of measurements.

Measurement Error The measurement error for the distributional data, $\nu^j_{F,t}$, is composed of sampling uncertainty and other errors that reflect the fact that a given dataset has its specific operationalizations of the common economic variables being measured. Differences in operationalizations can not only shift the level of a particular measurement (which we capture through fixed effects, see Footnote 14), but can also become differentially important over time.[16] In principle, to fully capture these time-varying differences, we would need to estimate a measurement error variance for each coefficient and dataset. However, this would be too large a number of parameters to estimate within the time-series model; a common problem when estimating approximate dynamic factor models with maximum likelihood (see e.g., Bai and Li 2016). The standard approach in the literature uses quasi-maximum-likelihood with EM algorithms (Doz, Giannone, and Reichlin 2012; Babura and Modugno 2014; Barigozzi and Luciani 2019), but convergence to the global optimum is not guaranteed in high-dimensional settings (Wu 1983; Balakrishnan, Wainwright, and Yu 2017).

For this reason, we do two things, the second of which will be discussed in Section 3.2.4, when we cover our sampling procedure. First, we assume that the correlation structure of all measurement errors is proportional to the correlation structure for sampling uncertainty. Under this assumption, the matrix Σ_j can be estimated outside the time-series model using bootstraps or the supplied replication weights to estimate the covariance from sampling uncertainty by data source j.[17] This achieves differently the diagonality assumed in the aforementioned literature. This leaves N diagonal elements of S^j to be estimated within the time-series model. To reduce the parameter count further, we constrain these elements to vary only by dataset j and variable ζ, yielding just $d+1=4$ parame-

16. For example, the PSID and SCF differ in the way they ask respondents about their business wealth (Pfeffer, Schoeni, Kennickell, and Andreski 2016). PSID and CPS differ in the sampling unit, making household/family income sensitive to labor supply patterns (Gouskova, Andreski, and Schoeni 2010). Similarly, differences in the propensity to sample business owners between the different datasets make income sensitive to relative changes in business and labor income (Kim and Stafford 2000). Finally, the CEX and the PSID differ in the consumption categories covered in the survey, with the PSID being much coarser (Insolera, Simmert, and Johnson 2021).

17. We draw bootstrap samples / use the supplied replication weights for each dataset j, $\{\tilde{\boldsymbol{\theta}}^j_{t,b}\}^B_{b=1}$, for each period t. Then, we demean the bootstrap samples b for each j and t and compute the average within-time variance-covariance matrix $\hat{\Sigma}_j$ pooling the demeaned bootstrap samples of the dataset j. If an object ζ is unobserved in dataset j, we set the covariance terms to zero and the diagonal elements to one to still be able to compute $\Sigma_j^{-1/2}$.

ters per dataset. Consequently, each diagonal element s_ζ^j indicates by what factor the sampling-based standard error for ζ must be inflated (or deflated) in dataset j. Therefore, the overall measurement error variance in (21) is $\hat{\Sigma}_j^{1/2} S^j (\hat{\Sigma}_j^{1/2})'$.

Thus, similarly to feasible generalized least squares, we transform observations $\tilde{\theta}_t^j$ by pre-multiplying the coefficients with the external estimate $\hat{\Sigma}_j^{-1/2}$

$$\begin{aligned} \hat{\Sigma}_j^{-1/2}\tilde{\theta}_t^j &= H_t^j(\alpha^j \hat{\Sigma}_j^{-1/2}\,\hat{\Gamma}^{MF} F_t + \tilde{\nu}_{F,t}^j) \qquad \tilde{\nu}_{F,t}^j \sim \mathcal{N}(0, S^j) \\ \tilde{\tilde{\theta}}_t^j &= H_t^j(\alpha^j \tilde{\hat{\Gamma}}_j^{MF} F_t + \tilde{\nu}_{F,t}^j), \qquad \tilde{\hat{\Gamma}}_j^{MF} := \hat{\Sigma}_j^{-1/2}\hat{\Gamma}^{MF}. \end{aligned} \tag{22}$$

With the measurement error treatment in place, denoted by the additional tilde, we can stack the J measurement equations, and with the aggregates, arrive at the final measurement equation

$$\begin{bmatrix} \tilde{\tilde{\theta}}_t \\ \hat{Y}_t \end{bmatrix} = H_t \left(\begin{bmatrix} \alpha \tilde{\hat{\Gamma}}^{MF} & \mathbf{0} \\ \mathbf{0} & I_Y \end{bmatrix} \begin{bmatrix} F_t \\ Y_t \end{bmatrix} + \begin{bmatrix} \tilde{\nu}_{F,t} \\ \nu_{Y,t} \end{bmatrix} \right) \qquad \nu_{Y,t} \sim \mathcal{N}(\mathbf{0}, \Sigma_Y). \tag{23}$$

Note that the terms are free of the j index to indicate they have been vertically stacked and note the additional appendage for the aggregates Y_t. We treat the aggregate factors to be observed with some noise Σ_Y, but still precise, since factors are constructed from $N \approx 2000$ series (details in Section 3.2.4).

In sum, the state-space model (20) and (23) forms a linear system of equations that can be estimated using standard Bayesian techniques and a Kalman-filter.

3.2.4 Bayesian Estimation

We need to estimate seven objects. From the state equation, we need to estimate the factor-autoregression matrix A, the loading matrix B from aggregate controls to distributional factors, the loading matrix C from distributional factors to aggregate factors, the vector autocorrelation of the aggregate factors themselves, the diagonal of D, the variance-covariance matrix of the shocks to the factors Ω, the variance-covariance matrices $\Sigma_j^{1/2} S^j \Sigma_j^{1/2'}$ and Σ_Y of the measurement errors. In a first step, the covariance structure $\Sigma^{1/2}$ is estimated outside the time-series model, as noted above, leaving only the scaling matrices S^j to be estimated within the model. Given the size of the A, B, C, D, Ω, S^j, and Σ_Y matrices, we use a Bayesian approach to estimate the system. This imposes regularization via priors on the parameters, as well as hyperpriors on the hyperparameters, to avoid overfitting the data.

The estimation is then one of a fully hierarchical Bayesian state-space model with mixed-frequency data. We collect all parameters and hyperparameters into

a single parameter vector $\psi = (\psi_{\text{par}}, \psi_{\text{hyper}})$, estimated jointly. The remainder of this section details the specific prior distributions on ψ, first for the state equation parameters and subsequently for those of the measurement equation.

State Equation For the state equation, we define the following prior:

$$\begin{pmatrix} vec(A) \\ vec(B) \\ vec(C) \\ diag(D) \end{pmatrix} \sim \mathcal{MN}(\mu_{\text{Minn}}, V_{\text{Minn}}) \qquad \begin{aligned} L_{chol} \quad &\sim LKJChol(r+q, \eta) \\ T^{-1}(\Omega_{F,ii}) \quad &\sim \mathcal{N}(\mu_{\Omega_F}, \sigma_F^2) \\ T^{-1}(\Omega_{Y,ii}) \quad &\sim \mathcal{N}(\mu_{\Omega_Y}, \sigma_Y^2) \end{aligned} \tag{24}$$

First, we impose a Minnesota prior on A, B, C, and $diag(D)$. This means we set the vector of expected values μ_{Minn} so that all *but* the autocorrelation terms in A and D have an expected value of zero. This implies that B and C are shrunk to zero and so are the off-diagonals of A if not needed to describe the data. The remaining values, the expected values for the autocorrelations (main diagonal of A and D), are governed by hyperparameters described later in the section. For the prior on Ω, we impose a *separation strategy* (see e.g., Barnard, McCulloch, and Meng 2000) and decompose Ω into a diagonal of standard deviations Ω_σ and a correlation matrix Ω_{corr} i.e., $\Omega = \Omega_\sigma \times \Omega_{corr} \times \Omega_\sigma{}'$. This avoids any apriori correlation between variances and covariances, normally imposed in Inverse-Wishart setups (Barnard, McCulloch, and Meng 2000). For the shocks, Ω_σ is further decomposed into Ω_F and Ω_Y, imposing different priors on the standard deviations of aggregate shocks and those of distributional shocks. It will be on the diagonal elements $\Omega_{\sigma,ii}$ for which we specify the priors.

For the parameters $\Omega_{\sigma,ii}$, we first draw from a set of Normal distributions and map these parameters to $\mathbb{R}^+$ using the soft-plus transformation, denoted as $T(\cdot)$. This provides flexible exploration for the optimizer by avoiding exponential jumps from small perturbations in these parameters and avoids acceptance rates being driven by out of bound draws (Chapter 55, Stan User Guide). For $\Omega_{F,ii}$, the mean μ_{Ω_F} is defined such that the apriori long-run variance of each distributional factor is $1.0 = \frac{\mu_{\Omega_F}}{1-\kappa_3^2}$, consistent with our prior for autocorrelation (in the matrix A) and factor normalization to unit variance. The same is done for the standard deviations of the shocks to the aggregate factors, $\Omega_{Y,ii}$, only using $1.0 = \frac{\mu_{\Omega_Y}}{1-\kappa_4^2}$. Hyperparameters κ_3 and κ_4 are discussed in more detail later.

For the estimation of the correlation matrix, Ω_{corr}, the matrix must satisfy several constraints simultaneously which may be difficult to enforce in any optimization: it must be symmetric, have a unit diagonal, and be positive semidef-

inite. To circumvent this, we employ a reparameterization strategy. The key insight is to work with the Cholesky factor, L_{chol} where $\Omega_{corr} = L_{chol}\, L'_{chol}$. Following the methodology of Lewandowski, Kurowicka, and Joe (2009, hereafter: LKJ), we construct the Cholesky factor, L_{chol}, from an unconstrained vector of real numbers, $v \in \mathbb{R}^{(r+q)(r+q-1)/2}$ and it is precisely on this Cholesky factor that we evaluate our prior. The LKJ-Choleksy prior will be governed by a single shape parameter, $\eta \geq 1$, which will control the expected correlation strength between aggregate and distributional shocks.

For the state equation, there are $|\boldsymbol{\psi}_{hyper}| = 6$ hyperparameters that are jointly estimated with $\boldsymbol{\psi}_{par}$. One hyperparameter shapes V_{Minn}; two hyperparameters govern the autocorrelation in A and D; and the remaining three hyperparameters govern Ω. See Appendix C for further information on these hyperparameters.

Measurement Equation First, for the distributional measurements, we specify independent priors for the measurement error variances for each dataset and object, s^j_ζ. Recall that estimating the diagonal of S^j would imply one variance parameter per coefficient n. As such, we impose the restrictions explained in the previous section, assigning one variance parameter s^j_ζ per object ζ and dataset j. This means a maximum (since some measures are unobserved) of $J \times (d+1)$ variance parameters. To ensure strict positivity of s^j_ζ while maintaining efficient sampling, we again parameterize each variance using a softplus transformation and place Normal priors on the unconstrained parameters $T^{-1}(s_\zeta)$. For the distributional data, we assign a weakly informative prior centered such that the expected variance is 1.0, with a standard deviation of 2.0. This says, first, on average, there is only sampling uncertainty and no additional measurement error reflecting conceptual differences, but that, second, $\hat{\Sigma}_j$ is still itself an estimate, additionally perturbing each s^j_ζ away from 1.[18]

$$T^{-1}(s^j_\zeta) \sim \mathcal{N}(0.54, 2), \qquad T^{-1}(\Sigma_{Y,ii}) \sim \mathcal{N}(-7.6, 0.02). \tag{25}$$

For the aggregate factors, we impose highly informative priors on variances $\Sigma_{Y,ii}$, reflecting the fact that these factors have been estimated with high precision. These priors are centered at a near-zero variance with a very narrow standard deviation of 0.02, effectively constraining the model to treat these aggregates as nearly perfectly measured.

18. Identical priors across datasets mean that we do not a priori prioritize a conceptual measure for object ζ in one dataset over another. Setting different priors is generally possible if a particular measurement concept should be prioritized on theoretical grounds.

Likelihood and sampling With this prior on ψ, we obtain the model-likelihood $p(\tilde{\tilde{\boldsymbol{\theta}}}|\psi)$ using a Kalman-filter. The posterior log-likelihood is then calculated as the sum of the hyperprior log-probability, the prior log-probability and the data log-likelihood. To sample from the potentially complex, multi-modal, high-dimensional posterior distribution, we employ the Differential-Independence Mixture Ensemble (DIME) sampler from Boehl (2024). Details and convergence results are in Appendix D.

3.3 Reconstructing and Using the Distributional Data

After finding the posterior mode of our model parameters, we apply the Kalman smoother to recover the full path of latent factors $\{\hat{F}_t\}_{t=1}^T$. Given these smoothed factors, we then reconstruct the sequence of joint distributions, $\hat{\Xi}_t^j$, from the implied standardized series of polynomial coefficients. Appendix E outlines the procedure.

This estimation approach allows practitioners to use the above data in three principal ways: First, in its factor representation, $\hat{F}_t$, second, in its functional representation, $\hat{\theta}_t^j$, or, third, in its observational representation, e.g., quantile functions and copula—$\left(\{\hat{\Xi}_{jmt}^{-1}\}_{m=1}^d,\, d\hat{C}_t^j\right)$.

For example, if one is interested in replicating the exercise of Chang and Schorfheide (2024), one can use the second representation and use the coefficients related to the marginal distributions of interest in their *distribution*-augmented VAR. If one is only interested in augmenting a VAR with a low-dimensional summary containing virtually all variation in the joint distribution of consumption, income, and wealth—in the style of these factor-augmented VARs—one can use the first representation.

Lastly, one can use the coefficients to generate some data of economic interest and work with this economically interpretable/meaningful data directly. One can generate the quantile functions and copula densities, but also other common statistics of interest for the marginal distributions, such as, gini coefficients of consumption, income, or wealth; levels, quantiles, shares, etc. and objects that arise from the correlational structure (the copula density) for example, correlation measures, conditional probabilities (e.g., population shares), kendall's tau, and tail-dependence.

For the application in Section 5, we adopt (3) and use the quantile functions and copula densities to generate a synthetic micro-dataset X_{it}, which will ultimately be a pseudo-panel. From the quantile functions, we obtain average realizations of all variables m for different (representative) households i, each

defined by a range of ranks $u \in U_i^m$ (e.g., deciles). We do this by integrating the quantile functions over U_i^m, i.e., forming conditional expectations. By having also estimated the copula density, we can combine these averages for household i with the household's probability weight by integrating the copula densities over some hyper-rectangular region, defined by ranks $\prod_{m=1}^{d} U_i^m$. Below is precisely this combination, for the example of the distribution of consumption, income, and wealth, which together forms a single row of the pseudo-panel:

$$X_{it}^j = \begin{pmatrix} c_{it}^j \\ y_{it}^j \\ w_{it}^j \\ \omega_{it}^j \end{pmatrix} = \begin{pmatrix} \frac{1}{|U_i^c|} \int_{u \in U_i^c} \hat{\Xi}_{jct}^{-1}(u)\ du \\ \frac{1}{|U_i^y|} \int_{u \in U_i^y} \hat{\Xi}_{jyt}^{-1}(u)\ du \\ \frac{1}{|U_i^w|} \int_{u \in U_i^w} \hat{\Xi}_{jwt}^{-1}(u)\ du \\ \iiint_{(u^c,u^y,u^w) \in U_i^c \times U_i^y \times U_i^w} d\hat{C}_t^j(u^c, u^y, u^w) \end{pmatrix}. \tag{26}$$

In this example, we have a synthetic (representative) household i, where $(U_i^c \times U_i^y \times U_i^w)$ is the quantile combination that defines household i, e.g., the first decile in consumption c, the third decile in income y, and the seventh decile in wealth w. The mass, ω_i, of the households in that cell defines a weight for that synthetic household.[19] The $|U_i^m|$ refers to the Lebesgue measure over the U_i^m to generate the representative (average) household in this interval i.e., $|U_i^m| = b_i^m - a_i^m$ for $a,\ b$ end points of the closed interval U_i^m. This implementation implies that, without the need for sampling, we obtain an output that can be immediately interpreted as synthetic microdata. As illustrated above, X_{it}^j is dataset specific, but one can obtain a consensus estimate across datasets as a simple average of the $d\hat{C}_t^j$ and $\hat{\Xi}_{jmt}^{-1}$ over datasets j or, alternatively, as a weighted average by using the inverse measurement error standard deviations (by object and dataset) as weights.

4 Credibility Checks

Before turning to the application, we evaluate the reliability of our procedure step-by-step using actual (Section 4.1) and simulated data. This addresses two potential concerns: (i) Does the factor representation lose information? (Section 4.2) and (ii) Does the state-space model predict accurately when data are missing? (Section 4.3). We first show that the factor representation preserves the information needed for state-space estimation, validating the approach described in Sections 3.2.1 and 3.2.2. Without this reduction step, we would need to track

19. The integrals can be calculated very efficiently as time varying linear combinations of the time invariant integrals of the basis functions.

and propagate a high-dimensional set of distributional coefficients directly in the state-space model (as in Chang, Chen, and Schorfheide 2024), which is computationally infeasible in our setting. The mapping from these noisy coefficients and latent distributional states then defines our measurement equation, whose priors and hyperparameter choices are validated in Appendix F (validating in part Section 3.2.3 and Section 3.2.4), showing that the estimated path of the functional data falls reasonably well into the bounds given by sampling uncertainty.

With the functional state-space model in place, we perform three further experiments to address the second concern. The first two experiments show that the model predicts well omitted survey waves using actual data. The third experiment evaluates the procedure in a controlled environment with a known data-generating process, using the heterogeneous-agent model of Bayer, Born, and Luetticke (2024). Results from these experiments show the resulting synthetic microdata closely match the omitted observations, indicating that the estimated state-space model successfully captures time-series fluctuations in the distributional data (thereby validating the procedure in Sections 3.2.3 and 3.2.4) and in particular, when the data-generating process is known. In Appendix G, we further show that our reconstructed distributional data agree well with the cyclical fluctuations (for the coverage of measures) found in the World Inequality Database (WID) and the Distributional Financial Accounts (DFAs).[20]

4.1 Data

To test and apply our method, we estimate the joint distribution of consumption, income, and wealth at the household level for the United States from 1962 to 2024. We rely on commonly used microdata for the U.S.: the *Consumer Expenditure Survey* (CEX), *Current Population Survey* (CPS, Flood et al. (2023)), *Panel Study of Income Dynamics* (PSID), *Survey of Consumer Finances* (SCF); including the historical backfiles (SCF+), and the *Survey of Income and Program Participation* (SIPP).[21]

We abstain from any sample selection in all of these datasets. For the CEX and SIPP income, which are at quarterly frequency, we remove seasonality using X-13 ARIMA-SEATS. We date CPS to quarter 4; the PSID is assumed to reflect quarter 2; and the SCF is dated to quarter 3. SIPP income data are aggregated

20. Note that all trends will be fitted by construction, see Section 3.2.1.

21. The methodology accommodates permanent and time-varying differences in measurement, but major survey redesigns require intervention. For the CPS (1992) and SIPP (1996, 2013), we treat pre- and post-break data as separate surveysi.e., distinct measurements of the distribution to avoid spurious dynamics from seam bias.

Table 1: Micro Data Sources and their Sample Periods

Object	CEX	CPS	SCF	PSID	SIPP
Consumption quantiles	1984Q4 - 2021Q4	-	-	1999Q2-2021Q2	-
Income quantiles	1984Q4 - 2021Q4	1967Q4-2022Q4	1962Q3-2022Q3	1968Q2-2021Q2	1983Q3-2022Q4
Wealth quantiles	-	-	1962Q3-2022Q3	1983Q2-2021Q2	1983Q3-2022Q4
Copula densities	1984Q4 - 2021Q4	-	1962Q3-2022Q3	1983Q2-2021Q2	1983Q3-2022Q4

Notes: The table reports sample periods for different micro datasets across the different objects.

to quarterly level and then naturally assigned to the respective quarter; wealth assignment depended on the survey vintage.[22] Further details on the microdata and in particular how the respective measures are defined can be found in Appendix H. Table 1 lists the distributional objects from each dataset, together with the sampling period. Note again that we require at least one dataset j that includes all objects, which in our case is the PSID between 1999 to 2021.

In terms of aggregate data, we use a wide range of standard business cycle data (GDP, consumption, employment, etc.) as well as data on household balance sheets, expectations, asset prices, and interest rates from McCracken and Ng (2021). We include data from 1962Q3 to 2024Q1. The starting point of the aggregate data determines the earliest date for the sample periods of the microdata used. From these aggregate time series, we extract the 11 most important factors. Details on the macro-data can also be found in Appendix H. Further details on factor selection and estimated parameters are available upon request.

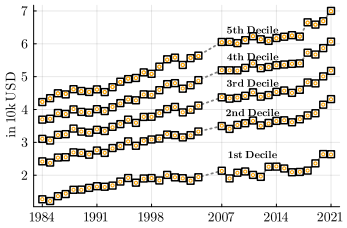
4.2 Precision of the Factor Model

The first step in our procedure is to estimate Legendre polynomial coefficients for the functional representation of the distribution. Then, we estimate the factor structure in these coefficient data. Since we only retain "important" factors, we potentially introduce an approximation error (relative to the data) resulting from forcing "unimportant" factors f_t to take time-averaged values. The size of the approximation error can be controlled by choosing how many factors to keep. We choose to retain the eight most important factors, which explain 99%

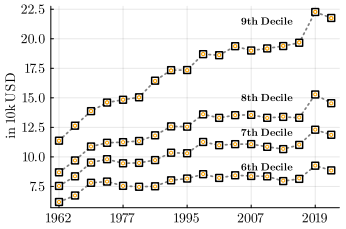
22. Documentation on the timing of the CPS can be found here, pg. 6. For the timing of the SCF, see here, pg. 33. For the PSID, interview dates are provided as variables, with the vast majority of interviews falling in quarter 2. Given the PSID reports annual income and the model is estimated with mixed-frequency, this quarter 2 assignment only affects the timing of wealth, which is always dated to the interview date. For data on consumption and income, which is for the year prior, it needed to be scaled by the growth rate of their respective aggregate to align with the timing of wealth, which facilitates the interpretation of the copula.

Figure 4:
Quantile Functions: Raw vs. Approximation (Important Factors)

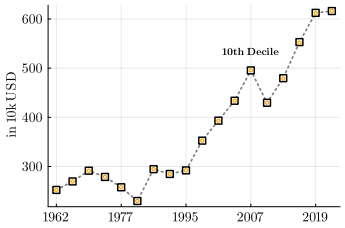
Consumption	*Income*	*Wealth*



(a) 1st to 5th Decile



(b) 6th to 9th Decile

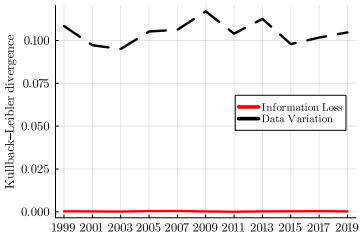


(c) Top Decile

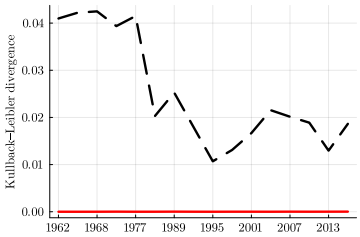
○ Approximation □ Data

Notes: Figure shows the quantile functions (mean within decile) for consumption, income, and wealth deciles from the survey data (squares) and an approximation (circles) using only the fluctuations in the most important factors in (18). Consumption panel shows quantile functions for CEX consumption. Income panel shows quantile functions for CEX income. Wealth panel shows quantile functions for SCF wealth. Dotted lines show linear interpolation between survey waves.

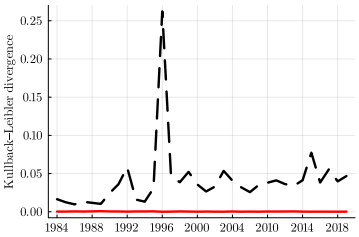
Figure 5: Copulas: Raw vs. Approximation (Important Factors)



(a) C, Y, and W



(b) Y and W



(c) C and Y

Notes: Figure shows the Kullback-Leibler divergence between the raw-data copula and two reference copulas, by survey year. The black dashed line represents the divergence between the time-averaged copula and the raw data copula for each survey year. The red solid line represents the divergence between the copula obtained by allowing only the most important factors in (18) to fluctuate and the raw data copula for each survey year. Panel (a) is from the PSID (Consumption, Income, and Wealth), Panel (b) is from the SCF (Income and Wealth), and Panel (c) is from the CEX (Consumption and Income).

of the (business cycle frequency) variation of the distribution (i.e., of $\tilde{\theta}$ to be precise). The different panels in Figure 4 visualize the approximation error in our application by showing some of the implied deciles. The figure compares the observed conditional decile means for consumption (bottom five), income (next four), and wealth (top) (squares) with their approximated counterparts (circles). The comparison for all deciles and variables looks analogously and is available upon request.

We find that the factor model with its eight main factors is very close to the distributional dynamics over time. The circles are typically entered around the midpoint of the squares. Figure 5 compares the copula over time between the approximation and the raw data. We do this in terms of the Kullback-Leibler divergence. The dashed black line shows how distant the actual distribution is from its long-term average (how much variation is there to capture), and the solid line shows the difference between the actual distribution and the approximation based on the important factors only (how much the factors do not capture). The Kullback-Leibler divergence of the actual copula from its long-run average is between 0.075 and 0.12 (between 1999 and 2019), while the divergence between the approximation and the actual distribution is almost two orders of magnitude smaller. To put this simple: There are significant fluctuations in the copulas over time, but the factors are able to capture them well.

4.3 Predictability of Distributional Data

To validate how well the method can predict distributional dynamics, we perform three experiments. The first experiment fixes the model parameters to that of the baseline and removes some waves of microdata as inputs when running the Kalman smoother. This experiment answers the question how informative are the model structure, aggregate data, and other available microdata for pinning down the distributional series when some survey waves are missing, holding parameters fixed. The second experiment in essence performs the same thing, but *re-estimates* the model. This experiment takes into account that changes in the data also imply changes in the model parameter estimates. In the final experiment, we evaluate our model's ability to recover distributional dynamics generated from a heterogeneous agent model. We show the results of these experiments in Figures 6, 7, and 8.

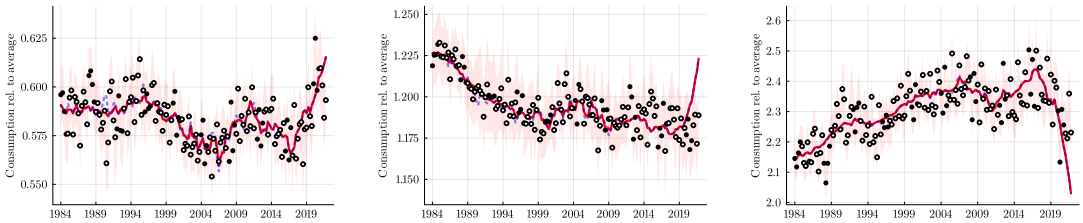
4.3.1 Predicting Omitted Survey Waves

For the first experiment, we first include only every fourth CEX survey year in the estimation, mimicking the fact that countries in Europe only survey consumption every four years and reducing the number of CEX survey years included in the Kalman smoother from 38 to 10. The resulting estimates from this estimation with less data are shown in Figure 6. We show the average of the top 10%, next 40% and bottom 50% for the measures in the CEX: consumption (Panels (a) to (c)) and income (Panels (d) - (f)). Note the large sampling uncertainty

Figure 6: Predictability of Distributional Data (given the Baseline Parameters)

Consumption by Consumption Percentile

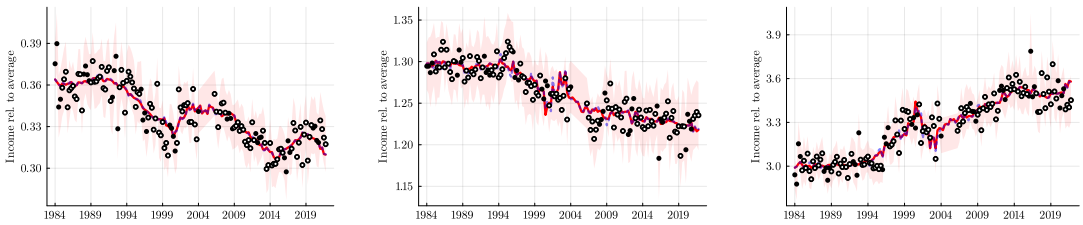
CEX data vs. estimates from using CEX every 4 years



(a) Bottom 50 Percent (b) 50-90 Percent (c) Top 10 Percent

Income by Income Percentile

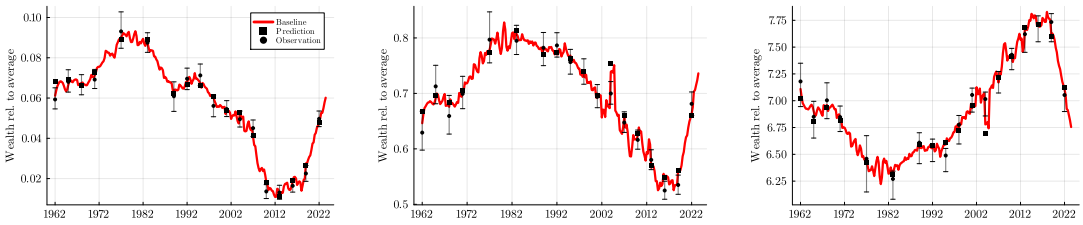
CEX data vs. estimates from using CEX every 4 years



(d) Bottom 50 Percent (e) 50-90 Percent (f) Top 10 Percent

Wealth by Wealth Percentile

Removing one SCF wave at a time



(g) Bottom 50 Percent (h) 50-90 Percent (i) Top 10 Percent

Baseline Less Data

Notes: The figure shows baseline model estimates for consumption (Panels (a) to (c)), income (Panels (d) to (f)), and wealth (Panels (g) to (i)) for different samples. Baseline estimates using all data are always shown as a solid red line. Panel (a) to (f): *less data* (dashed blue line) shows smoothed estimates when CEX microdata enters the smoother only every fourth year (black **solid dots**). **Empty dots** denote the observations removed from the Kalman smoother. Panel (g) to (i): shows smoothed estimates when only a single SCF wave has been dropped in the Kalman smoother. The black squares show the prediction of the dropped data at the survey wave and the dots show the empirical estimate from the survey data in that wave. Error bars/bands in all figures indicate 95% confidence bounds for each individual survey sample.

around the CEX data, whose 95% confidence intervals are displayed as a red error band. For this reason, even in the data-rich specification (with all CEX data), the smoothed estimate regularly deviates from the raw distributional data, with correlations of the two around 95%. The correlation of the smoothed data using only every fourth survey year with the data using every survey year is very high, meaning that the model can predict the consumption distribution well.

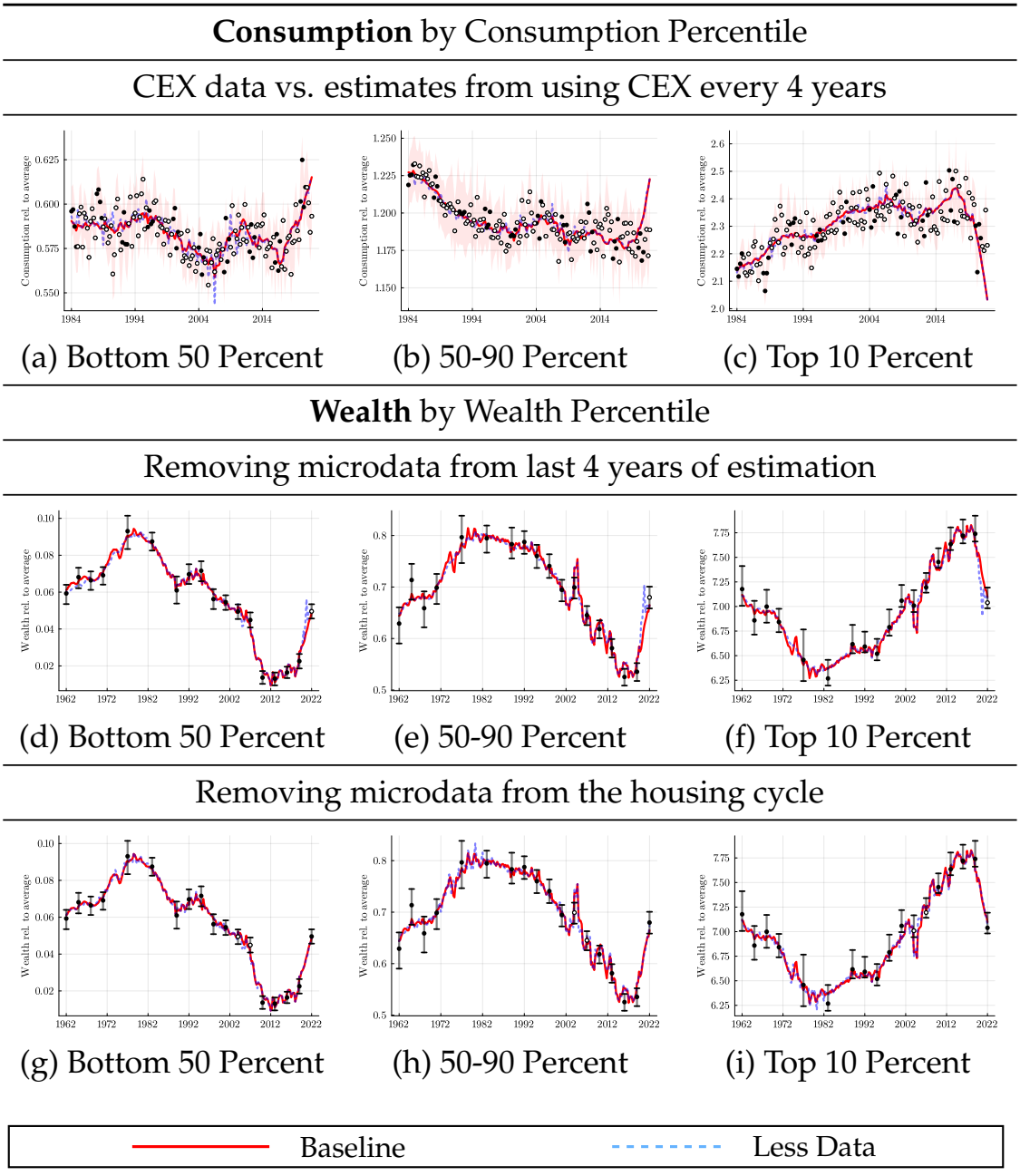
The bottom row of Figure 6, Panels (g) to (i), presents a different exercise. We run 17 mini experiments, each time dropping one SCF wave and generating new smoother estimates. This gives us 17 smoothed distributional data series alongside the one using all SCF waves. The solid lines are estimates from using all the SCF data. The squares represent the smoothed predictions corresponding to the timing of each omitted survey wave. The circle shows the direct empirical estimate of the corresponding survey (with its confidence limits). For example, for Panel (g), the square for 1992 is the prediction for the wealth distribution where the 1992 SCF survey *did not* enter the smoother. The fact that the squares are virtually on top of the solid line implies that, conditional on the model, a single observation of the distributional data has little effect on the smoothed series. In other words, the model structure, aggregate factors, and other surrounding micro-data collaborate well in imputing the missing wave. An exception to highlight happens in 2004 during the housing boom, where the model (squares) predicts greater cyclicality than shown in the baseline (red line — with all data).

The first row of Figure 7 repeats the CEX exercise of Figure 6 for the CEX but now re-estimates the entire model. The first row of Figure 7 shows the resulting reconstruction. The difference to Figure 6 is small, reconfirming our claim that the estimation procedure can impute well distributional data. The second and third rows show this in a different vein for wealth, which is observed least frequently. In both rows, we remove a contiguous block of wealth microdata and ask whether the model can still recover distributional dynamics from aggregate data and the remaining (surrounding) microdata. In the second row, we omit the last four years of the sample (2020Q1-2024Q1) to evaluate nowcast performance for the wealth distribution. In the third row, we omit all wealth microdata during the housing cycle (2004Q1-2009Q4), a period with large swings in wealth inequality (Kuhn, Schularick, and Steins 2020).[23]

Across all exercises, the reconstructed series obtained with withheld microdata closely track the baseline estimates, indicating that the model captures wealth

23. We also consider an alternative housing-cycle window (2007Q4-2011Q4) to assess predictability during the recovery phase. Appendix I, Table 9 reports correlations between the baseline estimates and all missing-data estimates, including this alternative window.

Figure 7: Predictability of Distributional Data (Re-estimating the Model)



Notes: The figure shows baseline model estimates for consumption (Panels (a) to (c)) and wealth (Panels (d) to (i)) for different samples. Baseline estimates using all data are always shown as a solid red line. Panel (a) to (c): *less data* (dashed blue line) shows the smoothed estimate that results from a re-estimation of the model (and Kalman smoother) when CEX microdata enters only every fourth year (black **solid dots**). Panel (d) to (f): *less data* (dashed blue line) shows smoothed estimates when the last four years of all microdata have been dropped in model estimation (2020Q1 to 2024Q1) and the Kalman smoother. **Empty dots** denote the observations removed from the re-estimation and Kalman smoother. Panel (g) to (i): same exercise as (d) to (f) but dropping the observations over the house price cycle (2004Q1 to 2009Q4). Error bars in all figures indicate 95% confidence bounds for each individual survey sample.

distribution dynamics even during periods without microdata-consistent with evidence that aggregate factors account for an important share of distributional fluctuations (Kuhn, Schularick, and Steins 2020; Bayer, Born, and Luetticke 2024), and by information from surrounding microdata, since smoother estimates incorporate the full set of observations.

4.3.2 Predicting Dynamics from a simulated HANK model

Thus far, we have examined the model's ability to predict intentionally omitted data. While informative, these tests do not fully resolve a central concern: in empirical settings, the underlying distributional dynamics are only imperfectly observed and agreement with these sparse microdata (or external benchmarks, see Appendix G) is not definitive proof that the model recovers the true latent process. To assess this, we study the model in a controlled environment with a known data-generating process. For this purpose, we simulate a heterogeneous agent business cycle (HANK) model (concretely, from Bayer, Born, and Luetticke 2024) as a data-generating process. This model is well suited to our setting, as it generates business-cycle fluctuations in general equilibrium in response to a set of aggregate shocks and their interaction with the joint distribution of households, also characterized by a copula and a set of marginals. The simulation produces functional distributional data, which we can use to draw micro-data comparable to ones actually available.

We simulate one economy for $T = 2200$ periods (quarters). For each period t, we buffet the HANK model economy with a series of aggregate shocks, whose effects propagate through linearized policy functions—only first-order responses are retained. From the resulting joint distribution, we generate **four** samples: *A*, *B*, *C*, and *D*. Each sample is drawn from the copula histogram, ensuring that the samples preserve the correlational structure between economic measures. Sampling is stratified along each dimension to ensure coverage of the marginal distributions. Each simulated household defines a unit of observation and is characterized by its consumption, income, and wealth and a sampling weight. We abstract from unit-level measurement error.

Samples *A*, *B*, *C*, and *D* are designed to mimic one dataset from the U.S., each representative, but with differences in their degree of missingness and precision. Dataset *A* can be considered a modern PSID-like dataset: 9,000 household observations per survey, containing consumption, income, and wealth, and released every two years. Dataset *B* can be considered CPS-like: 60,000 observations per survey (very precise), with only income, released every year. Dataset *C* can be

considered a CEX-like dataset: 3,000 observations per survey, with income and consumption, released every quarter. Dataset D can be considered an SCF-like dataset: 5,000 observations per survey, with only wealth and income observed, released every three years. We abstract from differences in operationalization across the different sample measurements.

The aggregate shocks driving the economy are also recorded at each t, whose factor representation will be used in our state-space system.[24] Before proceeding, the initial 1200 periods are discarded as burn-in, ensuring that the simulated economy has converged to its ergodic distribution independently of initial conditions. The remaining 1000 periods are kept for the exercise and split into 10 chunks, obtaining 10 economies observed for 100 quarters each. Accordingly, we can run 10 separate estimation exercises that are comparable to the actual data also in terms of the time that they span (25 years, similar to the available PSID waves containing consumption). In the end, the procedure suggests five distributional factors and eight aggregate factors.

We use the four survey-style samples (*A–D*) as microdata inputs to our dynamic factor method alongside the factor representation of the aggregate shocks from the HANK model. The resulting estimates are labeled as *Distributional Dynamics Factor Model (DDFM)* and compared to two benchmarks. The first benchmark are the values we obtain directly from the model simulation, the *Truth*. The second benchmark is a linearly-interpolated series of (conditional) cross-sectional sample moments for some variable and dataset X. The interpolation points would be the survey release dates. This *naive* Sample X benchmark ignores that the samples (*A–D*) are, in fact, just samples of an underlying distribution that has some dynamic structure.

Table 2 compares two benchmarks to the HANK simulated truth. First, column *DDFM* reports correlations between the HANK simulated truth and our DDFM estimates generated from using the incomplete samples. Column *Sample X* reports correlations between the HANK simulated truth and the linearly-interpolated sample $X \in \{C, A, D\}$. These are time correlations, averaged over the ten simulated economies. We find that our DDFM estimates are an improvement over the correlations of the linear-interpolation with the true distributions. In most cases, the correlation between the truth and DDFM implied distributional summary statistics are well above 90%. This suggests that the model ap-

24. The HANK economy is subject to shocks in total factor productivity, investment-specific technology, consumption wedges, labor wedges, technology, monetary policy, government spending, fiscal policy, and household preferences. For a full description of the model, see Bayer, Born, and Luetticke (2024) and the associated `BASEtoolbox.jl` code repository.

Table 2: Time-Series Correlations: Model Estimates vs. HANK Simulated Truth

	Consumption		*Income*		*Wealth*	
Percentile Group	DDFM	Sample C	DDFM	Sample A	DDFM	Sample D
			By Consumption Groups			
Bottom < 50	96	84	98	76	–	–
Middle 50-90	97	84	96	55	–	–
Top > 90	93	42	92	55	–	–
			By Income Groups			
Bottom < 50	98	88	98	72	72	46
Middle 50-90	90	78	96	34	69	33
Top > 90	85	42	98	72	81	61
			By Wealth Groups			
Bottom < 50	–	–	95	50	96	96
Middle 50-90	–	–	94	49	81	79
Top > 90	–	–	81	45	87	31

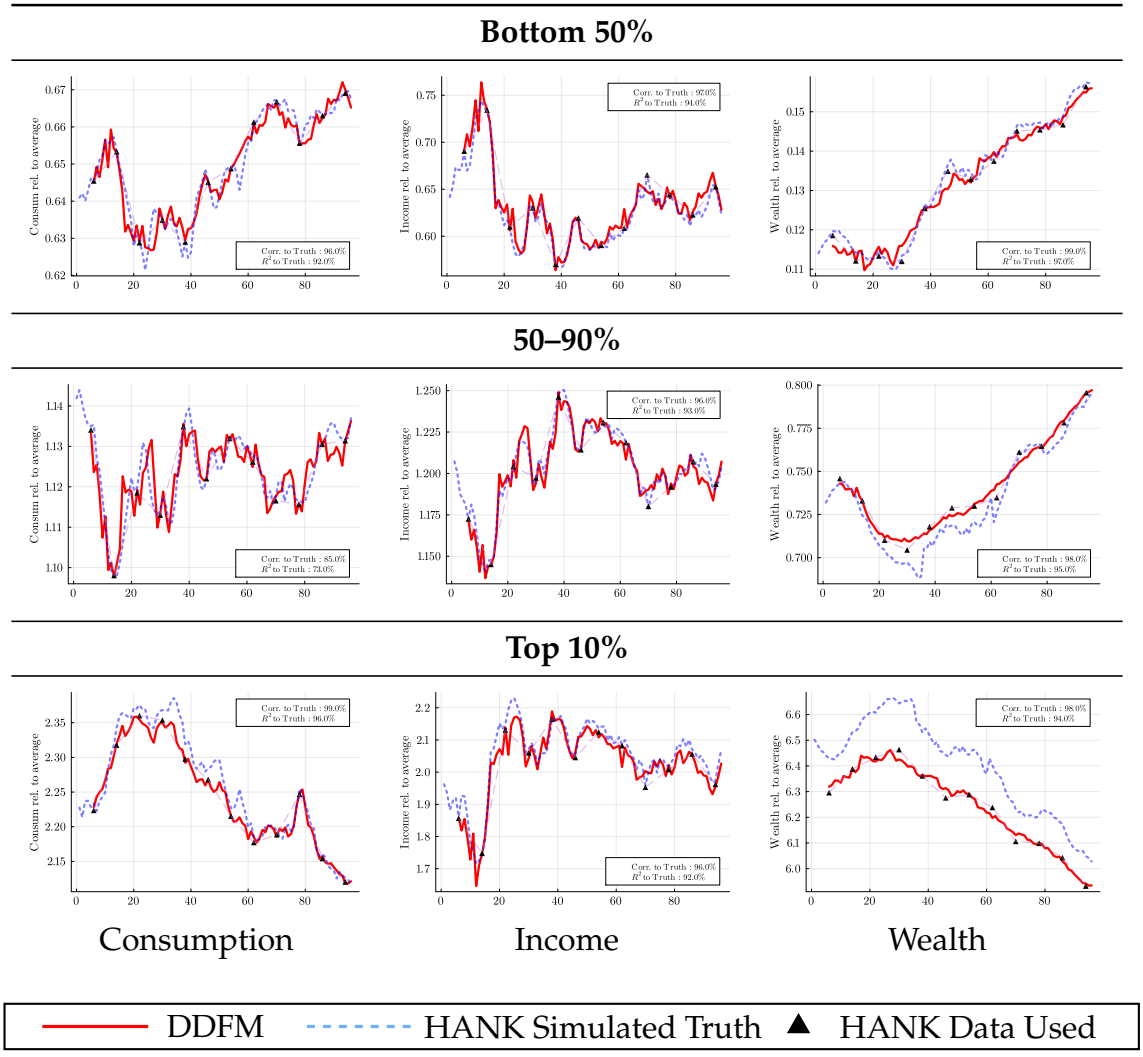
Notes: The table shows correlations between two benchmarks and the HANK simulated truth. Column *DDFM* shows correlations between the HANK simulated truth and our model estimates (*DDFM*). Column *Sample X* shows correlations between the HANK simulated truth and a linearly-interpolated Sample $X \in \{C, A, D\}$. As in the figure above, conditional moments are averages by *Percentile Group* (Bottom 50%, Middle 50-90%, and Top 10%) of the distribution of consumption, income, and wealth and all time series are relative to the economy-wide average. Since Sample *C* does not contain wealth information and Sample *D* does not contain consumption information, correlations of conditional moments involving these variables cannot be computed. Sample *B*, which contains only income, is omitted for brevity. Correlations are averages over the ten simulated economies.

proximates well the marginal distributions as well as the correlational structure across variables.

Figure 8 provides the results of this exercise visually (for economy 1), plotting the evolution of consumption, income, and wealth by groups, relative to the economy-wide average. The figure compares the *DDFM* (solid red) estimates to the HANK simulated truth (dotted blue). For comparison, we plot *naive* estimates using Sample A in a light-pink line; simply a linear-interpolation of the black triangles which mark the observations in Sample A. This also visualizes the sparseness of the data that enters our estimation. In general, the movements across the DDFM and the HANK simulated truth are in sync, denoted by the high correlation and around the same magnitudes, denoted by the high R^2 across panels.[25] Using the naive linear-interpolation would miss key distributional move-

25. The wealth quantile function is steeper in the upper decile than the degree-11 Legendre ex-

Figure 8: Model Estimates (in red) vs. HANK Simulated Truth (in blue)



Notes: The figure reports estimates for consumption, income, and wealth relative to their economy-wide average. Estimates are shown for three groups: the bottom 50%, the next 40%, and the top 10%. Each panel compares model estimates with the HANK simulated truth. Model estimates (*DDFM*) from incomplete samples are shown as a solid red line; *HANK Simulated Truth* is shown as a dotted blue line. The black triangles labeled *HANK Data Used* correspond to the sample observations employed in estimating the red line (calculated based on the polynomial quantile estimator). *Corr. to Truth* reports the correlation between the red and blue lines. R^2 *to Truth* is the coefficient of determination from regressing detrended HANK simulated truth on detrended model estimates.

ments, especially for consumption and income.

pansion can resolve and the fat tail leads to a small sample bias of average wealth. Both together a -2 to -3% downward bias in the Top10 mean. Half of this is sampling, half the order of the polynomial being somewhat too small.

5 Distribution Dynamics over the Business Cycle

The focus of our chapter is to provide a method for synthesizing dynamic distributional data from micro and macro data inputs. In the previous section, we have established the reliability of that procedure. Before concluding, we show the usefulness of synthetic data in two application examples. The first example is still closely related to the dynamic distributional factor model that builds the backbone of our method. We ask what role aggregate shocks play in driving the dynamics of distributions. We find that aggregates explain most of the movements in the distribution in line with Bayer, Born, and Luetticke (2024) finding for an estimated HANK model. Second, we show how the method contributes to ongoing debates on interpreting and understanding macroeconomic dynamics. We document in our synthetic data characteristic differences across recessions in terms of their consumption distribution impact. This speaks to the emerging literature studying how consumption distributions and consumption risk move over the business cycle (see e.g., Berger, Bocola, and Dovis 2023; Bilbiie et al. 2025; Patterson 2023).

5.1 What drives the distribution dynamics?

Since the dynamic factor model that we estimate is a linear model, it allows us to answer straightforwardly the question whether the dynamics of the distribution is driven primarily by specific distributional shocks, e.g., skills becoming more diverse, asset returns more dispersed, intertemporal preferences more different over time, or whether aggregate shocks are driving the distribution. Bayer, Born, and Luetticke (2024) provides an analysis of this question based on selected distributional information and a structural HANK model. They find that aggregate shocks explain well the actual evolution of wealth inequality even at lower frequencies. In a more reduced form manner, but with the same message, Kuhn, Schularick, and Steins (2020) argue that price changes are essential to understand wealth dynamics.

Here, we revisit this question within our much less structural DDFM. For this purpose, we compute a forecast error variance decomposition (FEVD) for the distributional factors for the five-year horizon. Because reduced-form innovations in the state equation are correlated (i.e., Ω is non-diagonal), the FEVD is based on an identification that orthogonalizes shocks. Specifically, we apply a *blocked* Cholesky factorization with aggregates ordered first. This distinguishes between aggregate shocks as a whole and distributional shocks as a whole but makes no

Table 3: FEVD on the Distributional Factors, percent contribution of shocks

Distributional Factor No.	1	2	3	4	5	6	7	8
Aggregate shocks	99	99	70	89	93	86	86	77
Distributional shocks	1	1	30	11	7	14	14	23

Notes: Table reports the forecast error variance decomposition (FEVD) for the eight estimated distributional factors. Identification is via a blocked cholesky with aggregates ordered first. The columns show the percentage of variation in each factor explained by distributional shocks versus aggregate shocks. This is over a five-year horizon.

assumption on their causal ordering within group. Table 3 reports the results of this exercise. We find that aggregate shocks explain at least 70 percent of the variation for every factor. For six out of eight factors, the contribution exceeds 85 percent.

Table 4: FEVD of group means and population shares. Percent contribution

	Conditional means of ...					
	Consumption		*Income*		*Wealth*	
Group	D.-shocks	A.-shocks	D.-shocks	A.-shocks	D.-shocks	A.-shocks
Bottom 20%	19	81	17	83	20	80
Bottom 20 – 40%	18	82	10	90	6	94
Middle 40 – 80%	20	80	21	79	15	85
Top 20%	20	80	10	90	6	94
	Population shares with ...					
	Low Consumption		*High Consumption*			
Low Y, Low W	42	58	44	56		
Low Y, High W	40	60	42	58		
High Y, Low W	42	58	40	60		
High Y, High W	45	55	42	58		

Notes: Table reports the forecast error variance decomposition (FEVD, five year horizon) of Table 3 mapped to observables consumption, income, and wealth. The top panel looks at consumption, income, and wealth group means across four household groups given by the bottom 20%, the next 20%, the next 40% and the top 20% of the distribution of the respective variable (consumption (C), income (Y), and wealth (W)). The bottom panel reports the corresponding FEVD for the population shares that have any combination of above/below median consumption, income, and wealth. Identification is via a blocked cholesky with aggregates ordered first. Entries show the share of variation explained by distributional shocks versus aggregate shocks.

While these results for the factors are informative, it is not clear what this means for moments of the data that one would typically look at, like consumption of the poor or the rich. As these moments are linear in the factors, we can directly apply the FEVD.[26] Table 4 reports the contribution of aggregate shocks for these better interpretable moments, decomposing the marginal distributions of consumption, income, and wealth for four household groups: the bottom 20%,

26. We refer the reader to Appendix E for the transformations from factors to observables.

the bottom 20 – 40%, the bottom 40 – 80%, and the top 20% (top panel). We find that, typically, aggregate shocks explain 80 percent of the variation in consumption and even more for income and wealth. Looking at changes of the copula, here represented as the change in the share of four population groups within the joint distribution, for example, all observations with below median consumption, income, and wealth ("Low Consumption, Low Y, Low W"). We now find that distributional shocks are more important, explaining around 40 percent of the fluctuations. At the same time, this means that aggregate shocks still generate the majority of fluctuations.

5.2 Consumption Dynamics over the last three recessions

The relative distribution of consumption losses in recessions has received considerable attention in the macroeconomic literature (e.g., Coibion et al. 2017; Cloyne, Ferreira, and Surico 2020; Bilbiie et al. 2025; Holm, Paul, and Tischbirek 2021; Fagereng, Holm, and Natvik 2021). The motivation is partly normative, in the sense of identifying groups that benefit or are adversely affected, and partly theoretical because a greater cyclicality of incomes of high-MPC households can destabilize the economy (Bilbiie 2020). Using our synthetic data, we trace consumption dynamics along both the marginal and joint distributions of income and wealth across the last three U.S. recessions. We find that their distributional consequences are state-dependent: wealth losses dominated the dot-com bust, balance-sheet distress the financial crisis, and income disruption COVID-19. In each case, the most affected households are identified by their position in the joint distribution—not by income or wealth alone—challenging research strategies that rely on marginal distributions and unconditional cyclical elasticities.

To begin, for the three recessions, we first consider the consumption dynamics of households along the marginal distributions of income and wealth, creating four groups: (1) the bottom 20%, (2) the 20% to 40%, (3) 40% to 80%, and (4) the top 20%.[27] These are the first two rows of Figure 9. For the final row, we then consider the joint distribution and estimate consumption dynamics for four key household groups. Group (i) consists of liquidity-constrained hand-to-mouth households, defined as the bottom 20% income *and* bottom 20% wealth; group (ii) is a population of asset-rich, low-income households which offers some resemblance to the wealthy hand-to-mouth, defined here as the bottom 40% of income but top 30% of wealth; group (iii) is the middle class, presumably pru-

27. The income measure we use includes transfers, but is not net of taxes.

dent, patient "buffer-stock" households, located between the 40% to 80% in both income and wealth; and group (iv) comprises the richest households in terms of income and wealth, largely unconstrained, presumably Permanent Income Hypothesis households, residing in the top 20% in both income and wealth.[28]

For all groups, absolute consumption falls in a recession and rises in a recovery; therefore, we compare the business cycle dynamics of consumption for the different income and wealth groups, in each period t, *relative* to the household-wide consumption average at period t and then index these relative consumption dynamics to the quarter preceding the recession (the peak).[29] In this sense, our goal is to understand how a group's consumption shifts relative to the aggregate over a business cycle. The observable deviations from the aggregate represent the net impact of the distributional channels at work, which in turn amplify or dampen the corresponding macroeconomic shocks. Figure 9 summarizes how these deviations are distributed across households depending on the nature of each recession.

Dot-com recession (Panels a, d, g). This recession resembles a financial wealth shock as income differences played little role: The consumption dynamics are uniform along the income distribution (Panel a), whereas along wealth (Panel d), low-wealth households *gained* relative to average—benefiting from aggressive rate cuts—while the wealthy lost ground. The asset-rich, low-income group (purple, Panel g) suffered heavily, as their equity wealth collapsed.

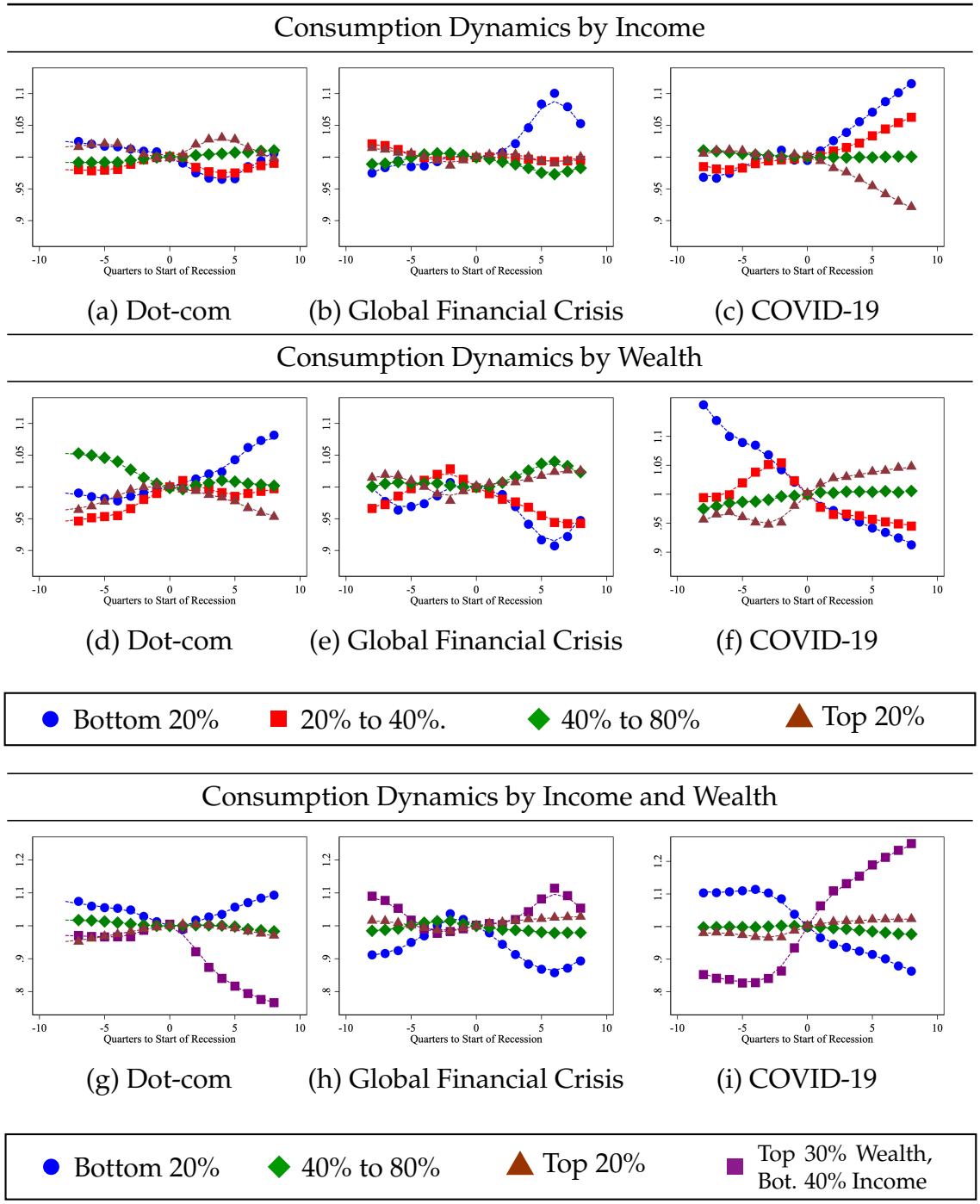
Global Financial Crisis (Panels b, e, h). This recession looks like a balance-sheet recession where now low-wealth, indebted households (Panel e) experienced the sharpest consumption declines—around 10% (evidence for the balance-sheet channel from Mian, Rao, and Sufi (2013)). Low-income households fared relatively well (Panel b), supported by fiscal transfers. The wealthy remained stable or gained.

28. For our analysis, groups are defined by contemporaneous ranks, so membership can change from quarter to quarter. While this implies some *churning* relative to a true panel, it remains informative because it tracks the currently vulnerable population relevant for real-time policy and because compositional changes within our broad bins are unlikely to overturn the qualitative patterns (residual churning is concentrated near bin cutoffs) (cp. Kuhn, Schularick, and Steins (2020)).

This interpretation is consistent with evidence that mobility is limited: positive shocks required for upward mobility are rare at the bottom, while large negative shocks are infrequent at the top (Guvenen et al. 2021); hand-to-mouth status is persistent (Aguiar, Bils, and Boar 2025); wealth positions are especially sticky due to persistent heterogeneity in returns (Fagereng et al. 2020); and earnings mobility is largely confined to adjacent deciles (Ettmeier, Hyun Kim, and Schorfheide 2024). Using coarse groups mitigates these concerns, though they become more salient when conditioning on multiple dimensions.

29. We construct symmetric 3-quarter moving averages and use this moving average for the normalization to the pre-recession quarter.

Figure 9: Comparison of Consumption Dynamics during Recessions



Notes: Relative consumption dynamics during recessions along the income and wealth distribution. Consumption dynamics of each income/wealth group are shown relative to average household consumption. These relative consumption time series for each group are indexed to the peak. The vertical axis shows changes of consumption over time relative to the change of average consumption over time. The horizontal axis shows the time relative to the start of the recession. The recessions are the dot-com recession in 2001Q1, the Global Financial Crisis in 2007Q4, and the COVID-19 recession in 2019Q4.

COVID-19 recession (Panels c, f, i). During this last recession, income becomes the dominant dimension: relative consumption is almost perfectly inversely ordered by income (Panel c). The bottom 20% gain roughly 10% relative to average (due to large government transfers). By wealth (Panel f), the pattern differs: low-wealth households lose persistently with no recovery two years out. The asset-rich, low-income group (purple, Panel i) outperforms all others by over 20%, benefiting from transfers, asset-price rebounds, and insulation from labor-market shocks.

The key finding is *state-dependence*: the channels through which recessions affect households—financial, balance-sheet, or labor-market—vary with the nature of the shock. During the dot-com bust, wealth mattered; during the Global Financial Crisis, debt mattered; during COVID-19, income and fiscal transfers mattered. Neither income nor wealth alone is a sufficient statistic for understanding consumption dynamics. This state-contingency cautions against research strategies that rely on uniform unconditional cyclical elasticities. It underscores that stabilization policy needs to diagnose the primary channel through which a crisis operates and take a holistic view of the sum of all stabilization measures.

6 Conclusion

In this chapter, we present a new method to derive synthetic distributional consumption, income, and wealth data. The method contributes to the modern theory of macroeconomic dynamics that has the joint distribution of consumption, income, and wealth as a key determinant of aggregate dynamics. Our method closes a gap as it provides a method for studying the empirical distributional dynamics as a counterpart to the existing theory at business-cycle frequency over time. We have shown that the method can incorporate information from various microdata sources regardless of their frequency and coverage of variables. By forecasting out of sample, we show that our method can generate joint distributional information at high frequency with good precision. Using a HANK model as a laboratory, we validated that the synthetic data from our model closely follow the "true" distributional dynamics when supplied with micro data samples as rich and frequent as in the U.S.

To illustrate the method's usefulness, we further traced consumption dynamics across the last three U.S. recessions along the income and wealth distribution. The application suggests that distributional consequences of recessions are state-dependent, with different channels dominating in different episodes—financial, balance-sheet, or labor-market. Although these findings are only suggestive,

they demonstrate how synthetic data can inform debates about the role of heterogeneity in macroeconomic dynamics. Our method also shows that most business cycle fluctuations of the distribution of consumption, income, and wealth are driven by aggregate shocks.

Beyond what we present in this chapter, our methodology provides a flexible tool for researchers: it can readily accommodate distributions based on, e.g., liquid and illiquid assets, financial income, or secured and unsecured debt, and even beyond households (e.g., firms and banks) and, in principle, be of higher frequency. Another promising direction is to apply the method to short panels where the joint distribution spans current and lagged values of the same variable. This would allow researchers to estimate high-frequency dynamics of household-level autocorrelation structures as in Almuzara et al. (2025), but incorporating a richer set of variables. We leave these extensions for future work.

References

Aguiar, Mark, Mark Bils, and Corina Boar. 2025. "Who are the Hand-to-Mouth?" *Review of Economic Studies* 92 (3): 1293–1340.

Almuzara, Martn, Manuel Arellano, Richard W Blundell, and Stéphane Bonhomme. 2025. "Nonlinear micro income processes with macro shocks." *FRB of New York Staff Report*, no. 1162.

Alvaredo, Facundo, Anthony Atkinson, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. 2016. "Distributional National Accounts (DINA) guidelines: Concepts and methods used in WID. world."

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

Auclert, Adrien. 2019. "Monetary policy and the redistribution channel." *American Economic Review* 109 (6): 2333–2367.

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub. 2021. "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models." *Econometrica* 89 (5): 2375–2408.

Bai, Jushan, and Kunpeng Li. 2016. "Maximum likelihood estimation and inference for approximate factor models of high dimension." *Review of Economics and Statistics.*

Bakam, Yves I Ngounou, and Denys Pommeret. 2023. "Nonparametric estimation of copulas and copula densities by orthogonal projections." *Econometrics and Statistics.*

Balakrishnan, Sivaraman, Martin J Wainwright, and Bin Yu. 2017. "Statistical guarantees for the EM algorithm: From population to sample-based analysis."

Babura, Marta, and Michele Modugno. 2014. "Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data." *Journal of applied econometrics* 29 (1): 133–160.

Barigozzi, Matteo, and Matteo Luciani. 2019. "Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the EM algorithm." *arXiv preprint arXiv:1910.03821.*

Barnard, John, Robert McCulloch, and Xiao-Li Meng. 2000. "Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage." *Statistica Sinica,* 1281–1311.

Bartscher, Alina K, Moritz Schularick, Moritz Kuhn, and Paul Wachtel. 2022. "Monetary policy and racial inequality." *Brookings Papers on Economic Activity* 2022 (1): 1–63.

Batty, Michael, Jesse Bricker, Joseph Briggs, Sarah Friedman, Danielle Nemschoff, Eric Nielsen, Kamila Sommer, and Alice Henriques Volz. 2020. "The Distributional Financial Accounts of the United States."

Baumeister, Christiane, Pascal Frank, Florian Huber, and Gary Koop. 2025. "Oil, Inflation Expectations, and Household Characteristics: A Nonlinear Heterogeneous Agent VAR Approach." Working paper, University of Notre Dame.

Bayer, Christian, Benjamin Born, and Ralph Luetticke. 2024. "Shocks, frictions, and inequality in US business cycles." *American Economic Review* 114 (5): 1211–1247.

Bayer, Christian, Ralph Luetticke, Lien Pham-Dao, and Volker Tjaden. 2019. "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk." *Econometrica* 87 (1): 255–290.

Berger, David, Luigi Bocola, and Alessandro Dovis. 2023. "Imperfect risk sharing and the business cycle." *The Quarterly Journal of Economics* 138 (3): 1765–1815.

Bhandari, Anmol, David Evans, Mikhail Golosov, and Thomas J Sargent. 2021. "Inequality, Business Cycles, and Monetary-Fiscal Policy." *Econometrica* 89 (6): 2559–2599.

Bilbiie, Florin O. 2020. "The new Keynesian cross." *Journal of Monetary Economics.*

Bilbiie, Florin O, Sigurd Molster Galaasen, RS Gürkayna, Mathis Mæhlum, and Krisztina Molnar. 2025. "Hanksson."

Blanchet, Thomas, Emmanuel Saez, and Gabriel Zucman. 2022. *Real-time inequality.* Technical report. National Bureau of Economic Research.

Boehl, Gregor. 2024. "DIME MCMC: A Swiss Army Knife for Bayesian Inference." *Journal of Econometrics.*

Breitung, Jörg, and Sandra Eickmeier. 2006. "Dynamic factor models." *Allgemeines Statistisches Archiv* 90 (1): 27–42.

Chang, Minsu, Xiaohong Chen, and Frank Schorfheide. 2024. "Heterogeneity and aggregate fluctuations." *Journal of Political Economy.*

Chang, Minsu, and Frank Schorfheide. 2024. *On the Effects of Monetary Policy Shocks on Income and Consumption Heterogeneity.* Technical report 32166. NBER.

Chang, Yoosoon, Chang Sik Kim, and Joon Y Park. 2016. "Nonstationarity in time series of state densities." *Journal of Econometrics* 192 (1): 152–167.

Chen, Weilong, Meng Joo Er, and Shiqian Wu. 2005. "PCA and LDA in DCT domain." *Pattern Recognition Letters* 26 (15): 2474–2482.

Chodorow-Reich, Gabriel, Plamen T Nenov, and Alp Simsek. 2021. "Stock market wealth and the real economy: A local labor market approach." *American Economic Review.*

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

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia. 2017. "Innocent Bystanders? Monetary policy and inequality." *Journal of Monetary Economics.*

Di Maggio, Marco, Amir Kermani, and Kaveh Majlesi. 2020. "Stock market returns and consumption." *The Journal of Finance* 75 (6): 3175–3219.

Diebold, Francis X, and Canlin Li. 2006. "Forecasting the term structure of government bond yields." *Journal of Econometrics* 130 (2): 337–364.

Doz, Catherine, Domenico Giannone, and Lucrezia Reichlin. 2012. "A quasi–maximum likelihood approach for large, approximate dynamic factor models." *Review of economics and statistics* 94 (4): 1014–1024.

Durbin, James, and Siem Jan Koopman. 2012. *Time series analysis by state space methods.* Vol. 38. OUP Oxford.

Ettmeier, Stephanie, Chi Hyun Kim, and Frank Schorfheide. 2024. *Measuring the Effects of Aggregate Shocks on Cross-sectional Distributions: Functional vs. Panel Approach.* Technical report.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2020. "Heterogeneity and persistence in returns to wealth." *Econometrica* 88 (1): 115–170.

Fagereng, Andreas, Martin B Holm, and Gisle J Natvik. 2021. "MPC heterogeneity and household balance sheets." *American Economic Journal: Macroeconomics* 13 (4): 1–54.

Flood, Sarah, Miriam King, Renae Rogers, Steven Ruggles, J. Robert Warren, and Michael Westberry. 2023. *IPUMS CPS: Version 11.0 [dataset].* Minneapolis, MN.

Gouskova, Elena, Patricia Andreski, and Robert F Schoeni. 2010. *Comparing estimates of family income in the PSID and the March CPS, 1968-2007.* Survey Research Center, Institute for Social Research.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2021. "What do data on millions of US workers reveal about lifecycle earnings dynamics?" *Econometrica.*

Harvey, Andrew, and Chia-Hui Chung. 2000. "Estimating the underlying change in unemployment in the UK." *Journal of the Royal Statistical Society: Series A.*

Holm, Martin Blomhoff, Pascal Paul, and Andreas Tischbirek. 2021. "The transmission of monetary policy under the microscope." *Journal of Political Economy.*

Insolera, Nora E, Beth A Simmert, and David S Johnson. 2021. "An overview of data comparisons between psid and other us household surveys." *Technical Series Paper,* 21–02.

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

Kim, Yong-Seong, and Frank P Stafford. 2000. "The quality of PSID income data in the 1990s and beyond." *Technical Series Paper.*

Kneip, Alois, and Klaus J Utikal. 2001. "Inference for density families using functional principal component analysis." *Journal of the American Statistical Association.*

Koop, Gary, Stuart McIntyre, James Mitchell, and Ping Wu. 2026. *Incorporating Micro Data into Macro Models Using Pseudo VARs.* Working Paper 26-04. Federal Reserve Bank of Cleveland.

Kuhn, Moritz, Moritz Schularick, and Ulrike I Steins. 2020. "Income and wealth inequality in America, 1949–2016." *Journal of Political Economy* 128 (9): 3469–3519.

Lewandowski, Daniel, Dorota Kurowicka, and Harry Joe. 2009. "Generating random correlation matrices based on vines and extended onion method." *Journal of multivariate analysis* 100 (9): 1989–2001.

McCracken, Michael W., and Serena Ng. 2021. "FRED-QD: A Quarterly Database for Macroeconomic Research." *Review* 103 (1): 1–44.

McKay, Alisdair, and Christian K Wolf. 2023. "Monetary policy and inequality." *Journal of Economic Perspectives* 37 (1): 121–144.

Meeks, Roland, and Francesca Monti. 2023. "Heterogeneous beliefs and the Phillips curve." *Journal of Monetary Economics* 139:41–54.

Mian, Atif, Kamalesh Rao, and Amir Sufi. 2013. "Household balance sheets, consumption, and the economic slump." *The Quarterly Journal of Economics* 128 (4): 1687–1726.

Mian, Atif R, Ludwig Straub, and Amir Sufi. 2020. *The Saving Glut of the Rich.* Technical report. National Bureau of Economic Research.

Patterson, Christina. 2023. "The matching multiplier and the amplification of recessions." *American Economic Review* 113 (4): 982–1012.

Pfeffer, Fabian T, Robert F Schoeni, Arthur Kennickell, and Patricia Andreski. 2016. "Measuring wealth and wealth inequality: Comparing two US surveys." *Journal of economic and social measurement* 41 (2): 103–120.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income inequality in the United States, 1913–1998." *The Quarterly journal of economics* 118 (1): 1–41.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. "Distributional national accounts: methods and estimates for the United States." *The Quarterly Journal of Economics* 133 (2): 553–609.

Saez, Emmanuel, and Gabriel Zucman. 2016. "Wealth inequality in the United States since 1913: Evidence from capitalized income tax data." *The Quarterly Journal of Economics* 131 (2): 519–578.

Schorfheide, Frank, and Dongho Song. 2015. "Real-time forecasting with a mixed-frequency VAR." *Journal of Business & Economic Statistics* 33 (3): 366–380.

Sklar, Abe. 1973. "Random variables, joint distribution functions, and copulas." *Kybernetika* 9 (6): 449–460.

Smith, Matthew, Owen Zidar, and Eric Zwick. 2023. "Top wealth in America: New estimates under heterogeneous returns." *The Quarterly Journal of Economics.*

Stock, James H, and Mark W Watson. 2002. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business & Economic Statistics* 20 (2): 147–162.

Tsay, Ruey S. 2016. "Some methods for analyzing big dependent data." *Journal of Business & Economic Statistics* 34 (4): 673–688.

Vermeulen, Philip. 2018. "How fat is the top tail of the wealth distribution?" *Review of Income and Wealth* 64 (2): 357–387.

Wu, CF Jeff. 1983. "On the convergence properties of the EM algorithm." *The Annals of statistics*, 95–103.

Online Appendix

A Economic Fit

The series estimator fits a single set of coefficients $\theta_t^j = \left(\xi^j_{m,o_1,t}, \ldots, \kappa^j_{(o_1,\ldots,o_d),t}\right)$ by projecting the entire distribution of the variable(s) of interest onto an orthogonal polynomial basis; it is therefore a *global* method in the sense that one parameter set governs all quantile/copula regions simultaneously. If distributions were exactly composed of finitely many orthogonal polynomials, the global estimator would be exact. In practice, however, the truncation we impose may systematically over- or underestimate certain segments, potentially smoothing over economically relevant features.[30] Also, even without truncation, we might fit polynomials to regions with zero mass in the measure—generating, potentially, spurious variation. Both concerns are naturally important in our setting, where our task is to recover the within and across time variation in the entire distribution, *Economic Fit*, not just segments best approximated by our estimator.

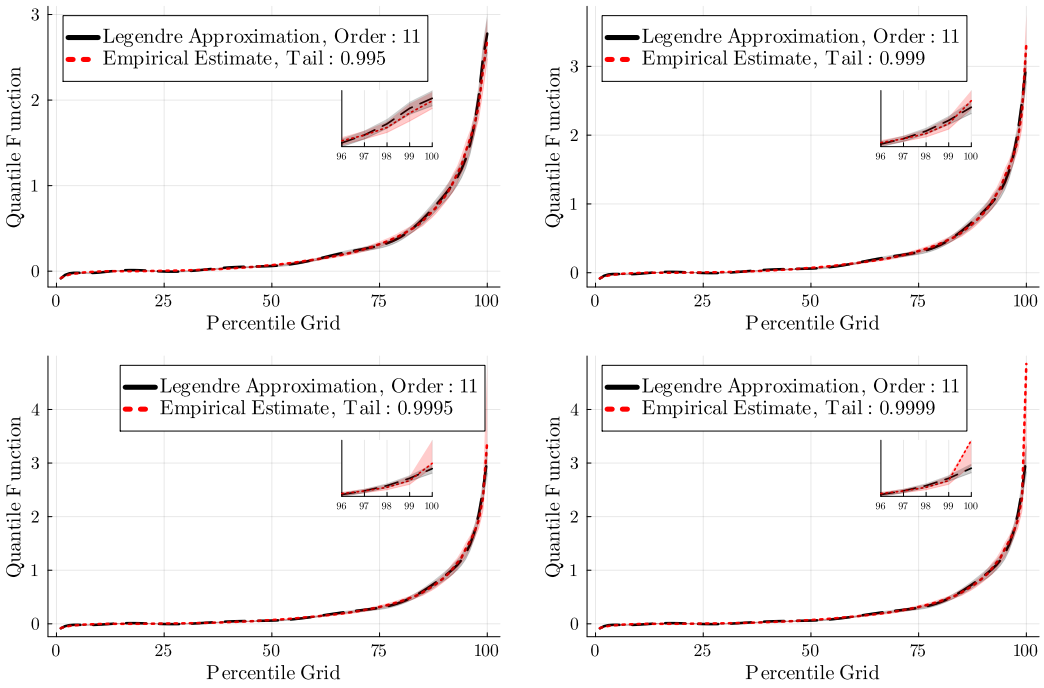
Two natural questions arise: (1) where does our estimator systematically fall short and (2) can these weaknesses be accommodated. For (2), this means (a) can the current (global) polynomial weights be adjusted to better capture specific regions (without compromising other regions), such as the upper tail of the wealth distribution or do they already carry enough flexibility to capture these specific regions (and the others)? and (b) can the estimator accommodate distributions with point-zero mass? In this spirit, the appendix addresses two points. First, it evaluates the degree to which we miss local variations. Second, it shows how an economist with a theoretical prior about which parts of the distribution to emphasize (e.g., the top of the wealth distribution) could adjust the coefficients accordingly. As a byproduct, for a given estimator, it also provides detailed evidence of variation within the distribution via observing weight changes within the distribution. More on this latter point later.

30. Our projection forces parts of the distribution onto the squareintegrable function space $\mathcal{L}^2$, even though those parts themselves do *not* belong to $\mathcal{L}^2$. Good examples are the upper tails of the consumption, income, and wealth, which are often modeled by Pareto laws. Vermeulen (2018) shows that wealth displays extremely heavy tailsso heavy that its variance is undefinedbecause the Pareto "alpha" parameters fall below 2. Likewise, Atkinson, Piketty, and Saez (2011) document that the U.S. incometail Pareto beta rose from 1.82 to 3.42 between 1976 and 2007, implying a shift in the corresponding alpha from 2.22 down to 1.41. As Toda and Walsh (2015) emphasize, analyses involving fat-tailed data require great care: when higherorder moments fail to exist, the Central Limit Theorem breaks down and conventional confidence intervalsfor instance, those on the top10 percent wealth sharelose validity. We are grateful to a referee for highlighting this issue.

A.1 Performance of the Estimator at the Tails

To investigate the first question, we begin with Figure 10, which compares two estimators: (1) the polynomial approximation as in the paper and (2) the empirical percentile function, shown with 99% interval bands. Each panel repeats the same percentile curve but varies its uppertail cutoff, using final grid points $p \in \{0.995, 0.999, 0.9995, 0.9999\}$. This addresses, in one panel, where (1) falls short, with particular focus on the tails, motivated by the footnote. As such, an inset for each panel is provided, zooming into the top 5% of the respective distribution. For wealth, we observe nearcoincidence of the two estimators up to about $p = 0.9995$. Only beyond the 99.95th percentile—the top onetwentieth of one percent—does the series estimate begin to pull away from the empirical percentile function (itself subject to sampling error). Hence, any inference inside the top 0.05 percent of the distribution should be treated with particular caution.[31]

Figure 10: Wealth Tails



Notes: Figure shows four panels plotting the percentile function for consumption, evaluated on a grid $G = \{0.01, \ldots, 0.99,\ p\}$. Consumption, as defined, is scaled by its respective aggregate and transformed via an inverse hyperbolic sine function. The four panels differ in the plotted estimate of the tail, denoted by p, which appears in the legend. The inset of each panel zooms in on the tail, covering the last five percentiles. No estimator is evaluated for $p = 1.0$.

31. Results on consumption and income are readily available and can be provided upon request.

A.2 Local vs. Global Estimation

To answer the second question, we introduce a *local* estimator that computes a set of coefficients within a moving kernel window. By sliding the window along a percentile grid G, we obtain $|G|$ coefficient vectors, each tailored to the local variation within that window. The local estimator is flexible and well-suited for capturing local structure.[32] By comparing the global estimators single set of coefficients with the collection of locally estimated coefficients, we can assess where and how they differ—especially around the tails of the distribution. Below, we describe the comparison procedure in detail and show that, for the economic data analyzed in this paper, our global estimator performs comparably to the local approach with minimal loss.

Data Let $w_i \in \mathbb{R}$ denote the economic variable of interest for household i as described in the main body of the text. The variable, as before, is first scaled by the corresponding nationalaccounts aggregate (quarterly, perhousehold level) and stabilized with an inverse hyperbolic sine transform. Results presented here rely on the 2019 wave of the PSID and are presented for the different measures consumption, income, and wealth. For this reason, we drop all time indices t and measure indices m.

Global Legendreseries estimator To have the appendix section self-contained, we provide again our estimator. Truncating the polynomial order O to 11, the shifted basis of orthogonal Legendre polynomials, evaluated on the empirical cumulative distribution u_i, is $Q_o(u_i)$ for $O = 0, \dots, 11$. Explicitly, $u_i = \frac{s_i}{\sum_i s_i}$ for s_i the survey weight corresponding to household i after sorting measure m. Projecting w_i on the basis yields coefficients

$$\hat{\xi}_o := N^{-1} \sum_i w_i Q_o(u_i) \qquad \text{for } o = 0, \dots, 11.$$

The estimated quantile function with these coefficients is

$$\hat{\Xi}^{-1}_{glob}(p) \;=\; \sum_{o=0}^{11} \hat{\xi}_o \, Q_o(p), \qquad p \in [0,1], \tag{27}$$

32. Naturally, the estimator will perform poorly outside the window, but we make clear that when evaluting the percentile function at a certain point, that we use the corresponding set of coefficients for that point i.e., the optimal coefficients in that window in a least squares sense.

where we use *glob* to denote that this is a global estimator and p some grid point in $[0, 1]$ for which we can evaluate the percentile function $\hat{\Xi}^{-1}_{glob}(p)$.

Local kernel Legendre estimator For the local estimation, let $p_g \in G$ be some probability grid point and $G = \{0.01, 0.02, \ldots, 0.995\}$. Let $K_h(u)$ denote a Gaussian kernel with bandwidth h[33]

$$K_h(u) \;=\; \frac{1}{h\sqrt{2\pi}}\,\exp\Big\{-\tfrac{1}{2}\big(u/h\big)^2\Big\}.$$

For a given survey weight of the PSID, s_i, we can generate the local weight $s_i^{(g)} = s_i\,K_h(u_i - p_g)$. Let $W_g = \mathrm{diag}(s_1^{(g)}, \ldots, s_n^{(g)})$ and $\Phi_{io} = Q_o(u_i)$. Solving the weighted leastsquares system

$$(\Phi^\top W_g \Phi)\,\boldsymbol{\xi}^{local}(p_g) \;=\; \Phi^\top W_g \mathbf{y}$$

produces the *local* coefficient vector $\boldsymbol{\xi}^{local}(p_g) = (\xi_0(p_g), \ldots, \xi_{11}(p_g))^\top$, which returns anew for every p_g. The estimated *local* quantile function then uses, for each grid point evaluation, a new set of coefficients $\xi_o^{local}(p_g)$, generating

$$\hat{\Xi}^{-1}_{local}(p_g) = \sum_{o=0}^{11} \xi_o^{local}(p_g) Q_o(p_g).$$[34]

Thus, we have the two extremes: the global approach of one set of coefficients to estimate the percentile function and the local approach of many sets of coefficients—localized throughout the entire distribution—to estimate the percentile function.

To see how the coefficients change across the distribution, independently of overall scale, each coefficient vector (from the different estimators) is transformed to unit ℓ^1 length:

$$\tilde{\xi}_o^{local}(p_g) \;=\; \frac{|\xi_o^{local}(p_g)|}{\sum_{o=0}^{11} |\xi_o^{local}(p_g)|}, \qquad \tilde{\xi}_o^{glob} \;=\; \frac{|\widehat{\xi}_o^{glob}|}{\sum_{o=0}^{11} |\xi_o^{glob}(p_g)|}.$$

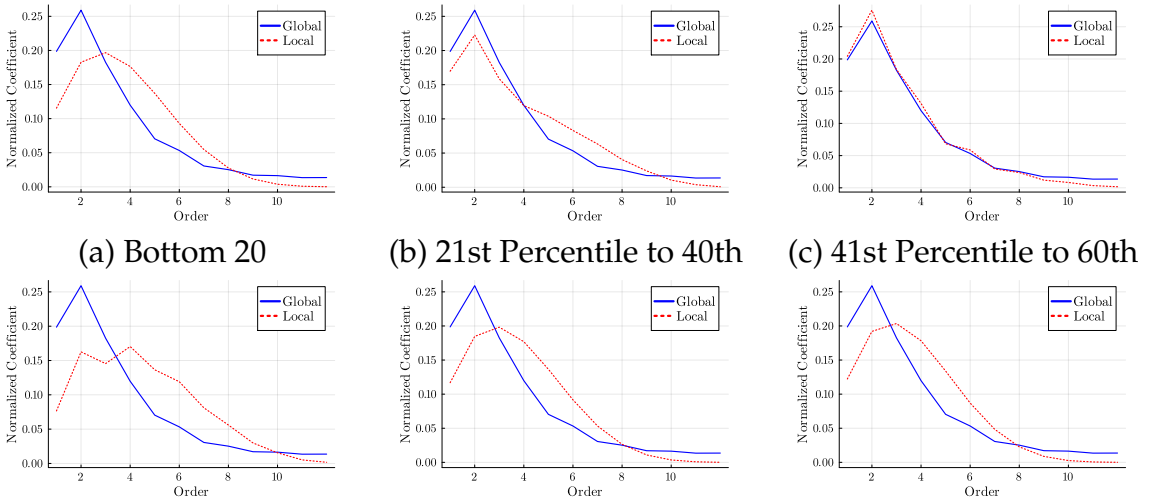
One can then average over the vector of coefficients (over intervals of G) to assess how these weights fluctuate for, for example, the bottom of the wealth distribution versus the top. Figure 11 shows precisely this (results are analogous for consumption and income). In short, our estimator assigns weights identically

33. Results below are unchanged for various smaller bandwidths $h \in \{.005, .01, .05, .10\}$

34. Alluding to our comment before, one can alternatively estimate $\hat{\Xi}^{-1}_{local}(p) = \sum_{o=0}^{11} \xi_o^{local}(p^*) Q_o(p)$, where $\xi_o^{local}(p^*)$ are the weights associated for a selected grid point p^* e.g., the 99th percentile.

to a local estimator that focuses on the 40th to 60th percentile. In every other case, it tends to assign more weight to the polynomials of order 4 to 8 in exchange for less weight on the first 3 polynomials. One might therefore conclude that we should instead use a different set of coefficients. Figure 12 provides evidence to the contrary.

Figure 11: Local vs. Global Estimator: Coefficients

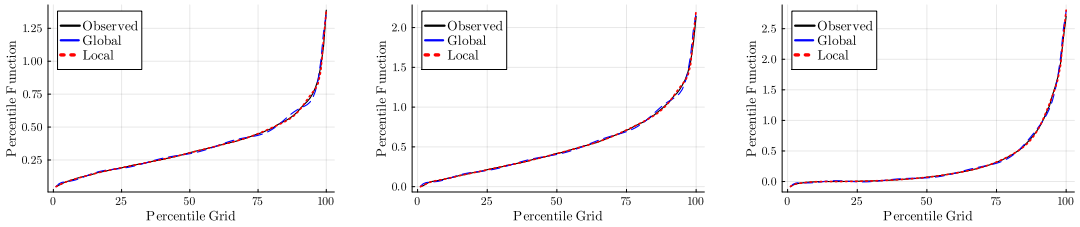


(a) Bottom 20 (b) 21st Percentile to 40th (c) 41st Percentile to 60th

(d) 61st Percentile to 80th (e) 81st Percentile to 100th (f) 91st Percentile to 100th

Notes: Figure shows panels comparing the global estimator to the local estimator for different segments of the wealth distribution. Each panel compares the contribution of each weight across the two estimators. Weights for the global estimator are the same across panels. Weights for the local estimator are calculated by averaging over the set of coefficients in that segment e.g., the Bottom 20 has 20 vectors of coefficients to average over.

Figure 12: Local vs. Global Estimator: Percentile Function



(a) Consumption (b) Income (c) Wealth

Notes: Figure shows three panels of percentile functions, comparing the global estimator to the local estimator, along with the empirical percentile function estimator. Percentile functions are evaluated over a set of grid points $G = \{0.01, \ldots, 0.99,\ 0.995\}$. Data is from the PSID 2019 wave.

A.3 Treatment of non-Differentiability of the Marginal

Some margins (e.g., earnings or financial assets) exhibit an atom at zero in addition to a continuous distribution over $\mathbb{R}_+$. Let Z_{mt} denote marginal m at time t and

$$\pi_{mt} := P_t(Z_{mt} = 0).$$

Write the CDF as a mixture

$$\Xi^{Z}_{mt}(z) = \pi_{mt}\mathbf{1}\{z \geq 0\} + (1 - \pi_{mt})\,\Xi^{Z,c}_{mt}(z), \qquad \Xi^{Z,c}_{mt}(z) := P_t(Z_{mt} \leq z \mid Z_{mt} > 0).$$

The unconditional quantile function is

$$\Xi^{Z,-1}_{mt}(u) = \begin{cases} 0, & u \leq \pi_{mt}, \\ \Xi^{Z,c,-1}_{mt}\Big(\dfrac{u - \pi_{mt}}{1 - \pi_{mt}}\Big), & u > \pi_{mt}, \end{cases} \tag{28}$$

i.e., the atom generates a flat segment of length π_{mt} and the positive part is rescaled to mass $1 - \pi_{mt}$.

We work with a strictly increasing transform $X_{mt} = T(Z_{mt})$ with $T(0) = 0$ (e.g. asinh), so (28) holds verbatim for X_{mt}; quantiles on the original scale follow as $\Xi^{Z,-1}_{mt}(u) = T^{-1}(\Xi^{X,-1}_{mt}(u))$.

Empirically, we estimate the mixed distribution in two steps: (i) estimate the point mass

$$\hat{\pi}_{mt} = \frac{\sum_i s_{it}\mathbf{1}\{Z_{mit} = 0\}}{\sum_i s_{it}},$$

and (ii) using only $Z_{mit} > 0$, estimate the conditional quantile function of $X_{mit} = T(Z_{mit})$ via the Legendre expansion

$$\Xi^{X,c,-1}_{mt}(u) \approx \sum_{o=0}^{O} \xi^{c,X}_{mot} Q_o(u).$$

We treat $\hat{\pi}_{mt}$ as an additional scalar coefficient alongside the Legendre coefficients and include it in the subsequent factor extraction and state-space analysis.

B Estimation of Factor Structures and Marginal Data Densities

A key input to our procedure is the mapping from factors to observables, defined by the projection matrix Γ. To estimate Γ, one method would be based on complete data alone. Complete here will mean there is no missing data in the spatial sense i.e., the data reports consumption, income, and wealth. For our case, this means using exclusively the PSID data to estimate Γ based on the PCA of $\tilde{\boldsymbol{\theta}}^{PSID}$. We consider different scenarios where we retain a different numbers of factors each time. Specifically, for this approach, we estimate the model with 3, 6, 7, and 8 distributional factors, obtained from the 12 PSID waves; however, this approach runs the risk of not fully representing the entire subspace that characterizes the evolution of the joint distributions because of the infrequent releases of the PSID data.

The literature has suggested alternatives for taking into account *incomplete* data in factor models. Here, this would mean surveys that only collect data on some, but not all marginals e.g., only income and wealth as in the SCF. In particular, we explore one alternative estimator for Γ: we consider the Tall-Wide algorithm of Bai and Ng (2021). We compare the quality of the model estimates under the alternative methods using marginal data densities for model comparison.

B.1 Tall-Wide Algorithm of Bai and Ng (2021)

An alternative approach to incorporating an additional block of data to the complete data is provided by Bai and Ng (2021). The paper tackles this problem by identifying two blocks of data within some larger T by N matrix— a tall block (data observed for all periods) and a wide block (time periods for which entire distribution is observed). For our setting, this means we add a tall block to our estimation of Γ, while the wide block would be what we call the complete data.[35] The tall block we add incorporates data on lower dimensional copulas of (income, consumption) and (income, wealth), as well as additional data on the marginals. This would be variation coming from both the CEX and SCF.

This alternative estimator for Γ, since it relies on more data, may improve model fit. For our application, we extract the factors that explain 80% of the

35. It is important to note that, given some data, such an estimator can consistently estimate the common component without making any assumptions on the nature of missingness. We refer the reader to the paper for more details on how the exact procedure is performed.

Table 5: Model Comparison

Model	Harmonic Mean	Bridge
	TW Projection Matrix	
12 distributional factors, 11 agg. factors	−81571.59	−85685.21
10 distributional factors, 11 agg. factors	−80555.15	−83940.35
	Standard Projection Matrix	
3 distributional factors, 3 agg. factors	−84379.20	−86262.30
6 distributional factors, 11 agg. factors	−72985.78	−77007.84
7 distributional factors, 11 agg. factors	−63227.01	−69535.44
8 distributional factors, 3 agg. factors	−54253.22	−56898.68
8 distributional factors, 11 agg. factors	−54137.45	−60434.82
8 distributional factors, 15 agg. factors	−54220.29	−61024.63

Notes: The table reports the log of the marginal data densities (MDD) across different model specifications. The MDD is estimated using two estimators: (1) Harmonic Mean and (2) a Bridge sampler. TW Projection Matrix are models estimated using the Tall-Wide algorithm of Bai and Ng (2021). Standard Projection Matrix is from a PCA estimator. Higher values means most efficient.

summable variation, as well as 85%. This corresponds to 10 and 12 factors; however, as mentioned before, it is unclear the degree of variation shared among this set of factors relative to the original set of factors. Furthermore, keeping more factors from this estimator runs the risk of over-parameterizing the model. Table 5 presents the marginal likelihoods of these *TW* projection matrices. Results suggest these models have marginally inferior fit than models that only rely on complete data.

B.2 Marginal Data Densities and Optimal Factor Structures

We estimate various versions of the state space model for different factor loadings Γ, that differ both in the number of factors and the estimation of Γ itself, and different numbers of aggregate factors. For each estimated model, we calculate the marginal data density (MDD) in order to discriminate between these alternatives. The set of models considered rely on the same data, but differ in the size of the parameter space. The MDD will effectively internalize these two features, selecting the model for which we can expect the best forecasting performance, while penalizing models through a lower density due to larger parameter spaces.

The marginal data densities $p(\tilde{\tilde{\boldsymbol{\theta}}})$ are estimated using Geweke's modified har-

monic mean estimator. Although standard, there may be concerns that, given the dimensionality of the model, such an estimator may not be appropriately approximated by a Gaussian distribution. Second, the estimator may not be numerically stable, as it requires the inversion of matrices—in this case, large ones. In this sense, we also estimate the density using a bridge sampler (Meng and Wong 1996), which is numerically stable (does not require any inversions) and may better internalize the shape of the posterior.

Table 5 covers the span of proposed models that were ultimately assessed, as well as the marginal data densities over the different estimators. First, we find that the MDD is most elastic to the number of distributional factors, as opposed to the number of aggregate factors. The model with the projection matrix from the complete PSID data, tagged *Standard Projector Matrix*, using the maximum number of 8 factors achieves the highest MDD for models with 11 aggregate factors.[36] Using the factor loadings Γ from the Bai and Ng (2021) TW-algorithm also achieves a lower MDD and thus a worse model fit.

C Minnesota Prior

Given the size of the model, we estimate it in a Bayesian framework and regularize the state equation parameters $(A, B, C, \text{diag}(D), \Omega)$ using a Minnesota prior.

Stacking the coefficients as $\theta := (\text{vec}(A)', \text{vec}(B)', \text{vec}(C)', \text{vec}(D)')'$ and setting

$$\theta \sim \mathcal{N}(\mu_{\text{Minn}}, V_{\text{Minn}}),$$

with mean

$$A_{ij} = \begin{cases} \kappa_2, & i = j \text{ (own first lag of distributional factors)}, \\ 0, & i \neq j, \end{cases}$$

$$B = \mathbf{0}, \quad C = \mathbf{0}, \quad D_{ii} = \kappa_3,$$

and diagonal variance

$$V_{\text{Minn},ii} = \begin{cases} \kappa_0 / l^2, & \text{own lags}, \\ (\kappa_0 \kappa_1 / l^2) \cdot (\hat{\sigma}_{ii}^2 / \hat{\sigma}_{jj}^2), & \text{cross-lags}. \end{cases}$$

36. Decreasing the number of aggregate factors increases the MDD, but these models are unfortunately incomparable as they have different measurement vectors. For example, the model with 15 aggregate factors has $10 \times T$ fewer contributions to the likelihood vs. the model with 25 aggregate factors.

we have our Minnesota prior. Here κ_2 and κ_3 control the prior persistence of the distributional and aggregate laws of motion. We work with unconstrained persistence parameters and map them to $(-0.99, 0.99)$ via a scaled inverse-logit; the unconstrained hyperpriors are diffuse, $\kappa_2, \kappa_3 \sim \mathcal{N}(0, 5)$. We fix the overall tightness at $\kappa_0 = 0.05$ motivated by Giannone, Lenza, and Primiceri (2015); κ_1 governs cross-lag shrinkage and is given a wide Normal hyperprior on its unconstrained representation, inducing an approximately uniform prior over $[0.2, 0.99]$ after transformation.

Prior for Ω. For the innovation covariance Ω, we place weakly informative priors on the diagonal elements (with variances governed by κ_4 and κ_5) and an LKJ prior on the correlation matrix. We parameterize the LKJ shape as

$$\eta = 1 + \log(\exp(\kappa_6) + 1),$$

so that $\eta \geq 1$, with $\kappa_4, \kappa_5 \sim \mathcal{N}(0, 1)$ and $\kappa_6 \sim \mathcal{N}(2, 1.5)$. Larger η shrinks correlations toward zero unless strongly supported by the data. We estimate $\{\kappa_i\}_{i=1}^{6}$ jointly with ψ_{par} (except κ_0, fixed at 0.05).

D Details on MCMC

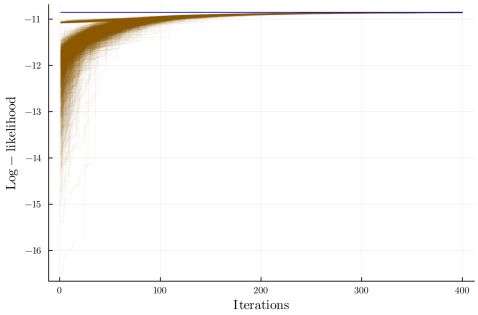
To estimate and sample from the posterior distribution, we employ the DIME sampler from Boehl (2024). The sampler is particularly advantageous for potentially complex, high-dimensional posterior distributions with ex-ante unknown properties.

We run an ensemble of $4n$ chains (for n the size of the parameter vector). To speed up adaptation, 10% of chains are initialized at a tentative mode from a separate optimization procedure; the remainder are initialized from the priors. The ensemble runs for 400 iterations, and we keep the last 25% of draws as posterior samples.[37] The sampler uses a single tuning parameter χ that governs the mixture between the local and global transition kernel; we set $\chi = 0.1$.

Figure 13 shows the traces of the (scaled) log-likelihood for all chains and indicates convergence.[38] Additional diagnostics are reported in the figure notes.

37. For the baseline $|\psi| \approx 400$, this implies $4(400) \times 400 = 640{,}000$ *efficient* runs.
38. Traces of all estimated models can be provided upon request.

Figure 13: Converging Chains



Notes: Figure shows the evolution of the ensemble of chains in terms of (scaled) log-likelihood; the last 25% of draws of each chain are retained. Chains initialized at the tentative mode start at higher likelihoods and generate the "cliff" at initialization. The sampler also reports the log-weight on the history of the proposal distribution and the standard deviation of likelihoods. Early on, log-weights are positive, indicating adaptation; after convergence they are close to zero (around 10^{-7}). Standard deviations are about 1% of the mode log-likelihood. Acceptance rates are in the 20%–40% range, consistent with effective exploration. Refer to the text for model specifications.

E Reconstructing Distributional Data

Following the procedure below, one can obtain the high-frequency distributional data. Given the estimated coefficients $\left(\hat{\xi}_{mot}, \hat{\kappa}_{o_1 \cdots o_d, t}\right)$ and the associated polynomials, the user only needs to specify $\mathbf{u} \in [0,1]^d$, the domain of integration. In our application, we integrate to construct deciles. In practice, the integration bounds exclude the exact endpoints 0 and 1, which are replaced by 10^{-6} and 0.9995, respectively. This is in accordance with our analysis in Appendix A and properties of our estimator.

Procedure:

$$\hat{\tilde{\theta}}_t = \hat{\Gamma}^{MF}\, \hat{F}_t \qquad \text{(Project factors)}$$

$$\hat{\theta}^j_{nt} = \sigma^j_{\zeta(n)} \times \hat{\tilde{\theta}}_{\zeta(n)t} + \mu^j_n + g(t)_n \qquad \text{(Unstandardize and add trend)}$$

$$\hat{\theta}^j_t = (\hat{\xi}^j_{m,o_1,t}, \ldots, \hat{\kappa}^j_{(o_1,\ldots,o_d),t}) \qquad \text{(Decomposition)}$$

$$\hat{\Xi}^{-1}_{jmt}(u_m) = \sum_{o=0}^{O} \hat{\xi}_{mot}\, Q_o(u_m) \qquad \text{(Reconstruct quantile function)}$$

$$d\hat{C}_{jt}(u_1, \ldots, u_d) = \sum_{o_1,\ldots,o_d=0}^{O} \hat{\kappa}_{o_1 \cdots o_d, t} \prod_{m=1}^{d} Q_{o_m}(u_m) \qquad \text{(Reconstruct copula density)}$$

Table 6: Description of Equation Components

Symbol	Description
$\hat{F}_t$	Estimated high-frequency factors
$\hat{\Gamma}^{MF}$	Factor loadings mapping factors to distributional observables
$\hat{\tilde{\theta}}_t$	Projected (standardized) factor representation
$\sigma^j_{\zeta(n)}$	Standard deviation term, varies by object
μ^j_n	Mean adjustment for coefficient n
$g(t)_n$	Non-parametric trend at time t
$\hat{\theta}^j_{nt}$	Unstandardized, trend-adjusted coefficients
$\hat{\xi}^j_{m,o,t}$	Marginal polynomial coefficients
$\hat{\kappa}^j_{(o_1,\ldots,o_d),t}$	Copula polynomial coefficients
$Q_o(u_m)$	Polynomial basis function of order o, O=11
$\hat{\Xi}^{-1}_{jmt}(u_m)$	Reconstructed quantile function (inverse CDF) for margin m
$d\hat{C}_{jt}(u_1,\ldots,u_d)$	Reconstructed copula density (dependence structure)

Notes: Table summarizes the role of each symbol in the above equations, moving from factor projections to reconstruction of marginal quantiles and copula density.

F Validation of Hyperparameter Choices

To evaluate our hyperparameter choices on the Bayesian estimation priors for the measurement error variances, we compare the series resulting from the Kalman smoother after estimation with the actual point estimates and their confidence bounds from the survey data.

Intuitively, if the prior mean for the measurement error variance is too low, it will force the estimator to exactly match each survey estimate of the distribution, despite the fact that each survey estimate is itself subject to measurement error. Thus, we should expect the smoother estimate to fall within the confidence bounds of each sample estimate at most with the corresponding confidence level of the bounds. The fact that the confidence level is an upper bound reflects that the estimated measurement error captures not only the sampling uncertainty that the confidence bounds capture, but also conceptual differences.

Choosing narrow measurement errors would overstate precision and potentially limit comovement with aggregates. As a result, it would drive parameter estimates for B, which captures this comovement, toward zero. Another reason not to be conservative with the measurement errors is that allowing for measurement error also accounts for the fact that we combine data from different sources

to produce a consensus estimate. These different data sources, despite their individual detrending, may produce some temporarily divergent estimates of the distributions. Without sufficient measurement error, the consensus estimate is then forced to oscillate between these different distribution estimates over short time intervals, rather than capturing their co-movements.

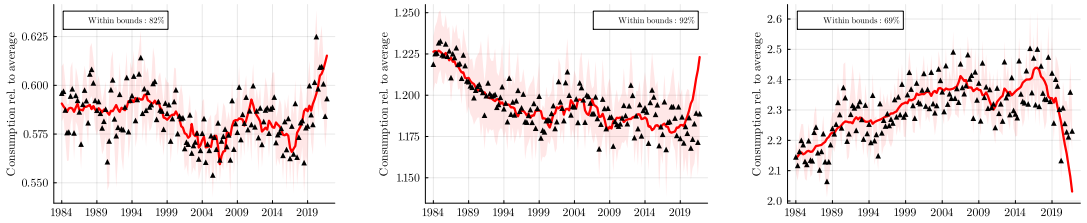
On the other hand, if the prior mean for the measurement error variance is too high, the estimator will treat the data as uninformative, and the smoother will miss the survey estimates more often and to a much greater extent than implied by its confidence bounds. We validate the choice of hyperparameters graphically for income and wealth in the SCF data and provide comprehensive summary statistics across all datasets and estimates. This visual inspection helps confirm that our smoothed series respect the noise inherent in the available microdata.

Figure 14 shows average consumption (top row), income (middle row) and wealth (bottom row) for the bottom 50 percent (first column), the next 40 percent (second column), and the top 10 percent (last column) of the respective distributions. It shows the point estimates from the surveys (black triangles), along with their 95% confidence limits, and the results from the Kalman smoother based on our estimates of the parameters of Equation (20). Overall, the smoothed estimates fall inside their respective confidence bounds in 85 out of the 108 observations (from (c) to (h)).

Table 7 provides a comprehensive summary of this validation approach. For all quantile functions and copulas, we report for each dataset and survey year how often the respective smoother estimate is within the confidence limits. Again, we use a confidence level of 95%. For the quantile functions, we find overall a modest difference (one to three percent) between the confidence level and the fraction of smoothed estimates that fall outside the confidence bounds. Only for the SIPP and to some extent the CPS, our estimator suggests a significant measurement error beyond sampling uncertainty reflecting differences in sample design and income/wealth measurement.

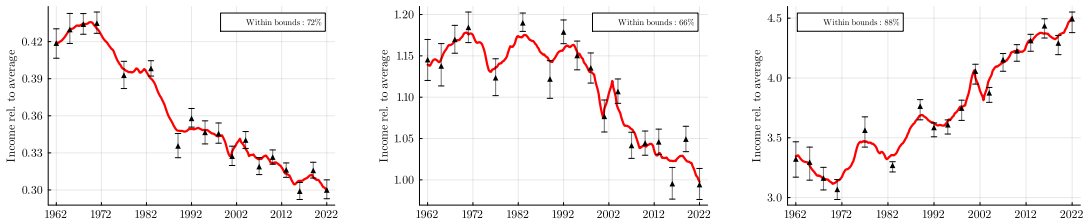
Figure 14: Comparison of Smoothed Distributional Data and Direct Survey Estimates

Consumption by Consumption



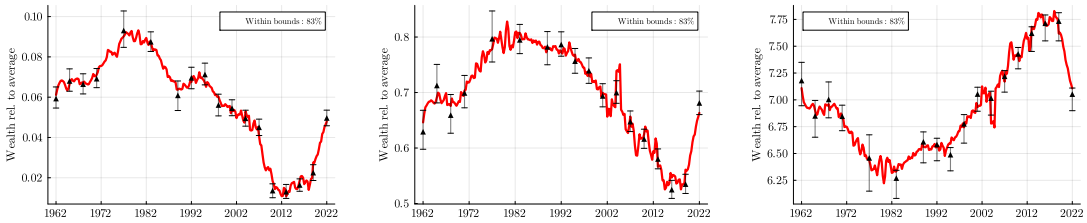
(a) Bottom 50 Percent (b) 50-90 Percent (c) Top 10 Percent

Income by Income



(c) Bottom 50 Percent (d) 50-90 Percent (e) Top 10 Percent

Wealth by Wealth



(f) Bottom 50 Percent (g) 50-90 Percent (h) Top 10 Percent

Model ▲ Data

Notes: Figure shows the average consumption, income, and wealth for the bottom 50 percent, 50-90 percent, and top 10 percent of households of the respective distribution. Dots show the estimates from the individual survey waves together with 95% confidence bounds. The solid red line shows the baseline estimate from the Kalman smoother at the posterior mode. Consumption shows CEX data and reconstruction. Income and wealth show SCF data and reconstruction. The legend reports for each Panel the share of smoothed estimates within the confidence bounds of the survey waves.

Table 7: Deviations of Smoothed Estimates and Microdata: Fraction within Confidence Bounds

Measure	CEX	CPS	SCF	SIPP	PSID	Overall
Consumption quantiles	77%	—%	—%	—%	100%	79%
Income quantiles	89%	93%	71%	46%	98%	77%
Wealth quantiles	—%	—%	77%	54%	99%	64%
Copula densities	98%	—%	94%	91%	98%	96%

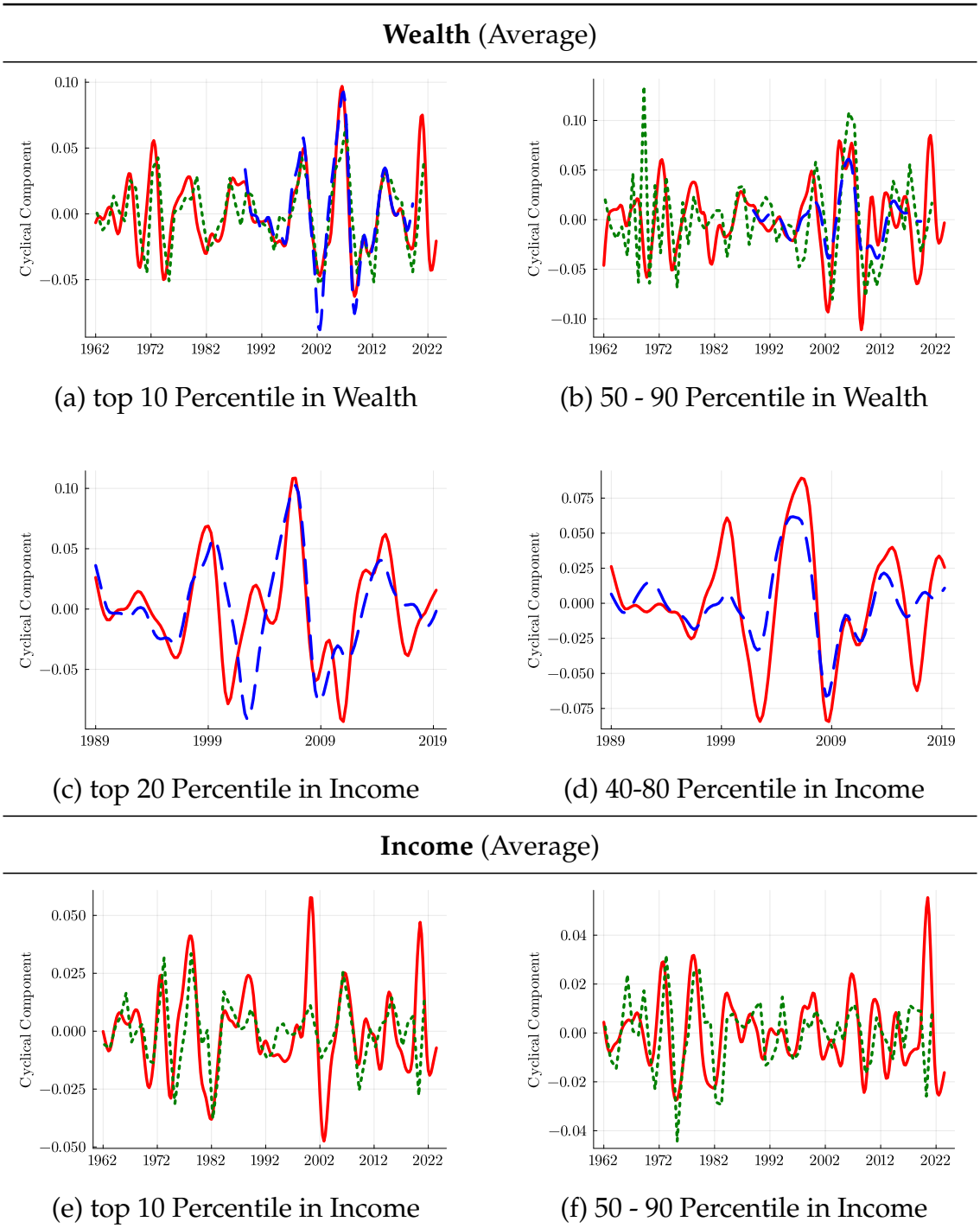
Notes: The table reports, by microdata and object, the fraction of estimates from the Kalman smoother at the posterior mode that fall within the 95% bootstrapped confidence intervals for the respective microdata. Quantile and copula estimates are defined on a decile grid.

G Comparison with External Estimates

The preceding sections established that our method predicts well out of sample and that the estimated distributions lie within sampling variability. We now assess whether the resulting high-frequency income and wealth dynamics align with external benchmarks. Since true distributional dynamics are unobservable, we compare our cyclical estimates to the *Distributional Financial Accounts (DFA)* (Batty et al. 2020) and the *World Inequality Database (WID)* (Piketty, Saez, and Zucman 2018). The DFA produces quarterly wealth distributions from the SCF using a different estimation approach, providing an "SCF flavor" benchmark. The WID provides annual estimates from administrative tax data without a time-series model, offering a robustness check for our assumption of stable and approximately linear distributional dynamics and their link to macro factors. To assess dependence dynamics (copula), we additionally compare our implied wealth-by-income series to the DFA.

Figure 15 reports the cyclical components (in logs). Panels (a)–(b) compare wealth by wealth group, focusing on the top 10% and the next 40% since the bottom half has wealth close to zero (Kuhn and Ríos-Rull 2016). Panels (c)–(d) compare wealth by income groups (informative about the joint distribution). Panels (e)–(f) compare income by income group (available in the WID). Across all cases, our estimates comove closely with DFA and WID and fall within the range of the external series.

Figure 15: Cyclical Component of Distributional Data vs. External Sources



Notes: Figure shows the cyclical component of (log) average wealth of (a) the wealthiest 10 percent, (b) the next wealthiest 40 percent, (c) the 20 percent income-richest households, and (d) the next 40 percent income-richest households. Bottom row shows the cyclical component of (log) average income of (e) the income-richest 10 percent and (f) the next income-richest 40 percent. Red lines show cyclical components from baseline model at quarterly frequency. Dotted green line show annual data from the *World Inequality Database* (WID). Dashed blue lines show quarterly data from the *Distributional Financial Accounts* (DFA). Cyclical components are obtained by an HP-filter with smoothing parameter $\lambda = 6$ for annual data and $\lambda = 1600$ for quarterly data.

H Data

The construction of these estimates relies on a great deal of data. An advantage with our method, however, is that it can incorporate these different microdata and their various differences in generating consensus estimates of the distributional data. Below, we describe the data, all expressed in 2019 dollars, and explain the mappings across data to ensure measures are at some base comparability (Curtin, Juster, and Morgan 1989; Czajka, Jacobson, and Cody 2003; Pfeffer et al. 2016). See Table 1 for information on their availability.

H.1 SIPP

The SIPP panel is a nationally representative, individual-level survey known for providing high-frequency dynamics on employment, earnings, wealth, household composition and program participation in the U.S.. For the data cleaning, the data is aggregated to the household-level, at quarterly frequency. Due to structural breaks induced by changing survey designs, we treat the SIPP as if it were three separate surveys (until 1983Q3-1995Q4, 1996Q1-2012Q4, 2013Q1-2022Q4). This is to minimize spurious fluctuations (e.g., seam bias) that obscure actual economic phenomena.

Income. For the 2014 releases and onward, we use the `THTOTINC` variable for income. For data releases prior, we sum over (1) earnings (`ws1_am, ws1_am`) (2) property/investment income (`tpprpinc`) (3) unemployment (`tuc1amt, tuc2amt, tuc3amt`) and (4) transfers (`tptrninc, tpscininc, twicamt, tfs_am, tssi_amt`) to construct household income.

Wealth. For the 2014 releases and onward, we use the `THNETWORTH` variable for wealth. For data releases prior, wealth is defined as total assets (`hhtwlth`) net total liabilities (`hhusdbt, hhscdbt`).
Note: The SIPP has undergone several major redesigns, and ignoring these seams, even with our prescribed data treatment, would lead to spurious swings in measured inequality—an issue documented by, among others, Moore (2008) and Czajka, Jacobson, and Cody (2003) and confirmed in our own diagnostics. To deal with this issue, we treat the SIPP as if it were three different surveys: (1) SIPP before 1996 (2) SIPP 1996-2012 and (3) SIPP 2013-2023. We found that around these seams, estimates tend to exhibit large jumps, which was at odds with their aggregate counterpart and general business cycle conditions. Ad-

ditional observations were dropped as well ($2003Q4, 1999Q4, 1988Q1, 1985Q4$), since they exhibited odd movements as well, related to the seams noted, as well as older seams. Fortunately, the estimation framework allows each of the three SIPP subsurveys to carry its own measurementerror process, so we retain almost all available information while preventing seam effects from distorting the results.

H.2 SCF+

The Survey of Consumer Finances (SCF), since its inception in 1983, is seen as the data gold mine for household information on income and wealth; however, due to the research excavations of Kuhn, Schularick, and Steins (2020), we are able to combine these triennial cross-sections with historical waves of the SCF; hence the name SCF+. Kuhn, Schularick, and Steins (2020) mention "... the SCF+ is the first dataset that makes it possible to study the joint distributions of income and wealth over the long run". Thus, it goes without saying how requisite this is for our study. The SCF+ is also augmented with the Forbes 400, from Forbes, for the years 2021 to 2024. For the years 1985 to 2020, we use the per capita dataset of Fernholz and Hagler (2023), whose observations originate from families of the Forbes 400.[39] Below we describe the measurements we used from the SCF+.

Income. Our definition of income follows Kuhn, Schularick, and Steins (2020), which consists of the following components: (1) labor income (i.e., earnings) (2) income from public transfers (3) income from professional practice and self-employment (4) income from rents (5) dividend income and (6) business/farm income. A different taxonomy that illustrates these components are taxable and transfer income.

Assets. Total assets include (1) liquid assets such as a household's checking and savings account, CDs, call/money market accounts, short-term government bonds, and mutual funds (2) illiquid assets such as housing and other real estate minus debt on that properties respectively, automobiles (3) defined-contribution retirement plans (4) the cash value of life insurance (5) stocks and (6) business equity.

39. Accessing the data via the usual link (e.g., `http://www.forbes.com/ajax/list/data?year=2005&uri=forbes-400&type=person`) returns Forbes 400 data, but unfortunately incomplete from the 1990s until the late 2000s—with only around 200 of the 400 observations there. Using Fernholz and Hagler (2023), with the same procedure imposed by Bricker, Hansen, and Volz (2019), generates the same result as in Bricker, Hansen, and Volz (2019) (increases the top wealth share by 1.5%), but is more complete (has almost all 400 every year).

Debt. We define debt of a household as the sum of personal (mostly unsecured) debt and housing (mortgage) debt. Housing debt includes debt from all properties and any loans made against the housing e.g., through HELOCs. Personal debt includes car loans, education loans, any loans from relatives, credit card debt, medical debt and legal debt.

Wealth. Wealth is total assets net total debt of a household.

H.3 PSID

The Panel Study of Income Dynamics complements the SCF+ extraordinarily well, as they take our estimations beyond more than half a century. In comparison to the post-1983 SCF, a deeper analysis of their similarity can be found in Pfeffer et al. (2016).

Income. The PSID has collected family income annually from 1968 to 1996 and then biennially from 1997 to 2021. Its measure of income is the sum of taxable income, transfers, and social security for the reference person, the spouse/partner (if any) and other members of the family.[40]

Assets. Data collection on household wealth took place in 1984, 1989, 1994, and then every wave beginning in 1999. The data on assets is split into liquid and illiquid assets. Although minor, the definition of liquid assets will vary between datasets, so careful attention is warranted here. Liquid assets for the PSID include checking and savings accounts, short-term instruments such as money-market accounts, certificates of deposit, and treasury bills. Illiquid assets include business equity, financial assets held in mutual funds, stocks, bond funds, investment funds; real assets held in real estate, vehicles like motor homes, boats, trailers, and cars; and retirement wealth in private annuities or IRAs.

Debt. For the PSID, we achieve the same debt split: personal and mortgage debt. This includes all kinds of real-estate debt, and unsecured debt such as credit card debt, student loans, medical debt, legal debt, and loans from relatives.

Wealth. Wealth is total assets net total debt of a household.

Consumption. Studying papers such as Skinner (1987), Cutler et al. (1991), Flavin

40. In the PSID, a family is a group of people living together who are economically interdependent.

and Yamashita (2002), Attanasio, Hurst, and Pistaferri (2014), and Attanasio and Pistaferri (2014), we define consumption as the sum of these expenditures: food, rent (for renters), housing rental equivalence (for home-owners), utilities, health, public transport, education, and childcare. We set the housing rental equivalence to be 6% of the home market value reported by households in the PSID. Consumption data is only available from 1999 in a biennial interval.

H.4 CPS

We use the Community Population Survey (CPS) Annual Social and Economic Supplement (ASEC). The sample is designed primarily to produce estimates of the labor force characteristics and runs from 1962 to 2022. Similar to the SIPP, we treat the CPS as a combination of two separate surveys to avoid fluctuations driven by survey design: 1967Q4-1993Q4 and 1994Q4-2021Q4.

Income. Income data are collected as part of the ASEC for the months of February, March and April as a supplement to the regular CPS monthly labor force interviews. The ASEC asks each person in the sample who is 15 years old and over about the amount of income received from a list of sources in the previous calendar year. We treat these observations as being observed in quarter four of the previous calendar year. For details on top-coding, see https://cps.ipums.org/cps/topcodes_tables.shtml.

H.5 CEX

The Consumption Expenditure Survey (CEX) is the most comprehensive household survey in the U.S. for recording the consumption habits of households. The CEX has two components: the interview survey (IS) and the diary survey (DS). The interview survey has sufficiently rich data on what we need, so we only use data from this component. Within this component, there are several files, each of which pertain to a topic, from which we can extract information. The following table breaks down each category of consumption, defining which `UCCs` belong to which category and which file it can be found in. All of these categories will combine to make the consumption variable. The table will also define wealth concepts of the CEX we use in our study. Since each household consumption record is with respect to a `UCC`, we find this presentation most apropos.

Item	UCCs / FMLI label	File
	Consumption	
Food	190904, 790220, 190901, 190902, 190903, 790410, 790430, 200900, 790330, 790420, 800700, 790230, 790240	MTBI
Rent	210110, 800710	MTBI
Utilities	250111, 250112, 250113, 250114, 250211, 250212, 250213, 250214, 250221, 250222, 250223, 250224, 250901, 250902, 250903, 250904, 250911, 250912, 250913, 250914, 260111, 260112, 260113, 260114, 260211, 260212, 260213, 260214, 270211, 270212, 270213, 270214, 270310, 270411, 270412, 270413, 270414,270101, 270102, 270104, 270105, 270310, 270311, 690116, 270901, 270902, 270903, 270904	MTBI
Health	570110, 570111, 570210, 570220, 570230, 560110, 560210, 560310, 560330, 560400, 340906, 540000, 550110, 550320, 550330, 550340, 570901, 570903, 570240, 580111, 580112, 580113, 580114, 580311, 580312, 580901, 580903, 580904, 580905, 580906, 580400, 580907	MTBI
Public Transport	520531, 520532, 530311, 530312, 530501, 530902, 530210, 530411, 530412, 520511, 520512, 520521, 520522, 520542, 520902, 520903, 520904, 520905, 520906, 520907, 530110, 530901, 520110, 520310	MTBI
Education	210310, 370903, 390901, 660110, 660210, 660310, 660900, 670110, 670210, 670901, 670902, 800802, 800804, 690111, 690112, 660410, 660902, 670410, 670903, 690114, 690310	MTBI
Child care	340210, 340211, 340212, 670310, 660901	MTBI
Rental Equivalence	910050, 800721 (market value of home), SIMHOUSX, RENTEQVX	FMLI, MTBI

Gas & Vehicle Repairs	470111, 470112, 470113, 470220, 470211, 470212, 480110, 480212, 480213, 480214, 490110, 490211, 490212, 490221, 490231, 490232, 490311, 490312, 490313, 490314, 490318, 490319, 490411, 490412, 490413, 490501, 490502, 490900, 520410, 480215, 620113	MTBI
	Other Concepts	
Housing Debt	QBLNCM1X, QBLNCM2X, QBLNCM3X, QBLNCM1G, PRINAMTX	MOR
Personal Debt	6001, 6002 (1990–2013), 5400, 5500, 5600, CREDITX, STUDNTX, OTHLONX, CREDITX1, CREDITX5, QBALNM1X	MTBI, ITBI, FMLI, FN2
Liquid Assets	SAVACCTX, CKBKACTX, USBNDX, 920010, 920020, 920030, 5100, LIQUIDX	FMLI, ITBI
Financial Assets	5800, 920040, STOCKX, SECESTX, OTHASTX	FMLI, ITBI
Income	FINCBTAX	FMLI

Notes: Table shows, by item, the identifiers necessary to construct each component of consumption, income and wealth for the CEX. The location of these identifiers can be found under the *File* column.

H.6 Aggregates

Together with the microdata, we specify a model component that captures the various aggregate shocks that potentially buffet the joint distribution of consumption, income, and wealth. This is represented in the state equation of the state-space model. The aggregate data we rely on to extract this information comes from the FRED-QD (McCracken and Ng 2021). This has various macrodata on industrial production, employment, housing, inventories, prices, earnings, productivity, household expectations, household balance sheets, interest rates, credit, etc. You can find more information on https://research.stlouisfed.org/econ/mccracken/fred-databases/.

Before performing the PCA on the aggregates, we are careful to check each series for non-stationarity. Recent literature has placed emphasis on the identifiability of orthogonal factors in high-dimensional settings, in particular for

macroeconomic aggregates, and finds non-stationarity to be the culprit of spurious variation (Onatski and Wang 2021; Hamilton and Xi 2022). Running the PCA on the non-stationary data will erroneously find that a large set of aggregates is confined to just a few factors. Taking note, we first remove any variation due to seasonality using the X13-ARIMA and closely follow the transformations (to induce stationarity) proposed by McCracken and Ng (2021). The resulting series satisfy an Augmented-Dickey-Fuller Test with a significance level of $\alpha = 0.05$ and are visually inspected for abnormalities.

The set of now stationary aggregates are concatenated with four of its lags to form a data matrix of quintuple the size and then column-wise standardized. A PCA on this block of data is performed and 11 orthogonal factors are kept. The number of factors chosen is based on Freyaldenhoven (2022). The baseline model estimation includes these 11 factors as inputs Y_t. More on the selection of factors is available upon request.

I Out of Sample Correlations

Table 9: Out of Sample Performance

Condition	Bottom			Middle			Top		
	C	I	W	C	I	W	C	I	W
Excluding Housing Cycle Wealth									
Entire Series	-	0.98	0.96	-	0.98	0.96	-	0.98	0.95
Specific Timeframe	-	0.98	0.94	-	0.96	0.97	-	0.96	0.97
Excluding Last 4 Years									
Entire Series	-	0.95	0.84	-	0.96	0.88	-	0.97	0.85
Specific Timeframe	-	0.65	-0.9	-	0.92	0.00	-	0.90	-0.34
Every 4 Years									
Entire Series	0.93	0.98	-	0.97	0.96	-	0.98	0.97	-
Specific Timeframe	0.94	0.99	-	0.98	0.96	-	0.95	0.97	-

Notes: Table reports correlations between the baseline cyclical estimates and the missing-data models, split by panel. Entire Series reports correlations for all estimates in the estimation timeframe. Specific Timeframe reports correlations of estimates in periods where data was intentionally left out of the estimation of the specific missing-data model. **Bottom**, **Middle**, and **Top** are the bottom 50, next 40, and top 10 of the respective distribution, denoted by **C, I,** and **W**. **C** is for consumption, **I** is for income, and **W** is for wealth. For the first three panels, correlations are made between SCF model estimates (no consumption). The final panel presents correlations from CEX model estimates (no wealth). Specific models (in header) are discussed in Section 4.

Appendix References

Atkinson, Anthony B, Thomas Piketty, and Emmanuel Saez. 2011. "Top incomes in the long run of history." *Journal of economic literature* 49 (1): 3–71.

Attanasio, Orazio, Erik Hurst, and Luigi Pistaferri. 2014. "The evolution of income, consumption, and leisure inequality in the United States, 1980–2010." In *Improving the measurement of consumer expenditures*, 100–140. University of Chicago Press.

Attanasio, Orazio, and Luigi Pistaferri. 2014. "Consumption inequality over the last half century: some evidence using the new PSID consumption measure." *American Economic Review* 104 (5): 122–126.

Bai, Jushan, and Serena Ng. 2021. "Matrix completion, counterfactuals, and factor analysis of missing data." *Journal of the American Statistical Association* 116 (536): 1746–1763.

Bricker, Jesse, Peter Hansen, and Alice Henriques Volz. 2019. "Wealth concentration in the US after augmenting the upper tail of the survey of consumer finances." *Economics Letters* 184:108659.

Curtin, Richard T, Thomas Juster, and James N Morgan. 1989. "Survey estimates of wealth: An assessment of quality." In *The measurement of saving, investment, and wealth*, 473–552. University of Chicago Press, 1989.

Cutler, David M, Lawrence F Katz, David Card, and Robert E Hall. 1991. "Macroeconomic performance and the disadvantaged." *Brookings papers on economic activity* 1991 (2): 1–74.

Czajka, John L, Jonathan E Jacobson, and Scott Cody. 2003. "Survey estimates of wealth: A comparative analysis and review of the Survey of Income and Program Participation." *Soc. Sec. Bull.* 65:63.

Fernholz, Ricardo T, and Kara Hagler. 2023. "Rising inequality and declining mobility in the Forbes 400." *Economics Letters* 230:111235.

Flavin, Marjorie, and Takashi Yamashita. 2002. "Owner-occupied housing and the composition of the household portfolio." *American Economic Review* 92 (1): 345–362.

Freyaldenhoven, Simon. 2022. "Factor models with local factorsdetermining the number of relevant factors." *Journal of Econometrics* 229 (1): 80–102.

Giannone, Domenico, Michele Lenza, and Giorgio E Primiceri. 2015. "Prior selection for vector autoregressions." *Review of Economics and Statistics* 97 (2): 436–451.

Hamilton, James D, and Jin Xi. 2022. "Principal Component Analysis for Nonstationary Series."

Kuhn, Moritz, and José-Víctor Ríos-Rull. 2016. "2013 Update on the US earnings, income, and wealth distributional facts: A View from Macroeconomics." *Federal Reserve Bank of Minneapolis Quarterly Review* 37 (1): 2–73.

Meng, Xiao-Li, and Wing Hung Wong. 1996. "Simulating ratios of normalizing constants via a simple identity: a theoretical exploration." *Statistica Sinica*, 831–860.

Moore, Jeffrey C. 2008. "Seam bias in the 2004 SIPP panel: Much improved, but much bias still remains." *US Census Bureau Statistical Research Division Survey Methodology Research Report Series* 3:2008.

Onatski, Alexei, and Chen Wang. 2021. "Spurious factor analysis." *Econometrica* 89 (2): 591–614.

Skinner, Jonathan. 1987. "A superior measure of consumption from the panel study of income dynamics." *Economics Letters* 23 (2): 213–216.

Toda, Alexis Akira, and Kieran Walsh. 2015. "The double power law in consumption and implications for testing Euler equations." *Journal of Political Economy* 123 (5): 1177–1200.

Chapter 2

Oil Supply Shocks & Inequality

1 Introduction

The impact of oil markets on prices and output has long been a central question in macroeconomics, as evidenced by the seminal work of Hamilton (1983). Geopolitical tensions between the U.S. and Iran have renewed interest in the question, with the disruption of the Strait of Hormuz, through which roughly 20% of the world's traded oil passes. This raises the specter of a repeat inflationary episode in the U.S. and globally and has policy makers debating whether they should *look through* these shocks or act proactively. The empirical literature on oil shocks has shown that such movements in oil markets have widespread macroeconomic effects, from Hamilton (1983) through Kilian (2009), Baumeister and Hamilton (2019), and Känzig (2021), but has remained comparatively silent on how these oil-market disturbances propagate to households. Through the lens of a VAR, this chapter will show oil supply shocks shape U.S. inequality, and *which type of* oil supply shock matters for who bears the cost.

The present study focuses on two oil supply shocks whose transmission may pass through to households: announcements / news concerning future oil supply and the realized, physical oil supply disruption itself. Känzig (2021) (KZ) identifies oil supply news shocks through high-frequency futures price surprises in narrow windows around OPEC announcements. Along the lines of Beaudry and Portier (2006), Känzig (2021) shows how such news shift the expectations of market participants and drive business cycle fluctuations. Baumeister and Hamilton (2019) (BH) would identify the physical supply shock, identified through informative Bayesian priors on the contemporaneous elasticities of oil supply and demand. A physical supply disruption drives nations to deplete oil inventories to compensate for the sudden shortfall, affecting input costs, industrial production, and eventually the real economy. Both shock series

enter the VAR (separately) as internal instruments (Plagborg-Møller and Wolf, 2021).

With that, the chapter documents the distributional consequences of both oil supply shock identifications side-by-side. The VAR embeds a parsimonious representation of the joint distribution of household income, consumption, and wealth alongside standard oil-market and macroeconomic variables. In practice, the distributional block will be a set of principal components derived from the Legendre-polynomial coefficients underlying the three-dimensional joint distribution of Bayer, Calderon, and Kuhn (2025). As a consequence of the treatment, I can back out, by linearity, any functional of the joint distribution: group-specific consumption means, robust Gini coefficients, interquantile ratios, joint income–wealth cells, etc. To accommodate the potentially large system, I adopt a Bayesian approach and elect for the asymmetric conjugate prior of Chan (2022), which allows for equation-specific shrinkage—critical to distinguish near-unit-root macroeconomic variables from the stationary distributional factors and the shock instrument itself.

Beyond the impulse responses themselves, I ask *through which channels* oil shocks affect macroeconomic aggregates. Following the causal mediation framework of Dufour and Wang (2024), I decompose each impulse response into the contributions of individual transmission channels e.g., from real activity, consumer prices, oil market variables, and the distributional block itself. The decomposition can trace, for example, how much of the inflation response is mediated through the price of oil or if monetary policy plays any role in the observed aggregate dynamics (Kilian and Lewis, 2011). Effectively, the decomposition shows over the impulse response horizon when a channel activates, how it propagates, and when it dissipates. This is different than a forecast error variance decomposition, which only informs on the variance contribution of a single shock on each variable. The level decomposition framework of Dufour and Wang (2024), given a single shock, shows the level contribution of *every* channel to *every* impulse response function.

With this richer model in place, I can revisit many conclusions about oil supply shocks and oil shocks in general. First, as most of the literature has already claimed, there is undoubtedly a causal link between oil prices and output: supply-driven oil price increases are contractionary. In response to oil supply shortfalls as defined by Baumeister and Hamilton (2019), real economic activity contracts, but *with a delay*, strengthening the main finding of Kilian (2008b). In contrast to Kilian (2008b), I find precise, persistent positive dynamic effects of headline CPI; however, the *core* CPI response under BH is credibly zero throughout the horizon—supporting Kilian's result that headline inflation is 'flat', shedding light on the importance of relative prices, and is consistent with Hooker (2002), which tested this pass-through from oil prices to inflation under different non-linearities and structural breaks.[1]

[1] I find that CPI owner-occupiers' equivalent rent places considerable downward pressure on CPI inflation, which may be driving the flatness. Recent literature has raised concerns precisely on the measurement of this

Under KZ, by contrast, all prices rise, including core CPI: The news shock is the regime in which oil-shock inflation has bite. I confirm the main findings of Känzig (2021) that oil supply news are indeed inflationary (all prices increase), contractionary, and swift. In the medium-run, for the same oil price impact response, physical disruptions have larger effects on the economy, however, than do oil supply news shocks. Despite the asymmetry in realized inflation dynamics, however, one-year-ahead inflation expectations—measured both for households (Michigan) and professional forecasters (SPF)—respond *identically* to the two shocks, a finding with direct implications for monetary-policy rules that condition on expected inflation.

Oil inventories also tell an interesting story. Under a KZ shock, the increase in inventories that Känzig (2021) interprets as announcement-driven hoarding only materializes after roughly two years, once oil prices ease back toward steady-state values. The channel decomposition clarifies why: the KZ shock channel *does* contribute positive pressure on inventories throughout—hoarding intent at the channel level—but it is overwhelmed in the short run by downward contributions from the inflation channel and the inventory channel itself, the latter a characterization of inventory stickiness. The oil-production channel, by contrast, contributes essentially nothing (Figure 23 shows neither OPEC nor non-OPEC production moves upon a KZ shock), so the inventory drawdown is a pure price-response, not a substitution for missing supply. Under a BH shock, the inventory response is consistently negative, in line with depletion of reserves upon a production shortfall. Despite the different signs after year two, inventory responses to both shocks share the same shape—pointing to a common price mechanism: the storage-arbitrage response of Pindyck (2001) and Working (1949), where inventory holders decrease (increase) stocks when spot prices are up (down), irrespective of whether supply is already disrupted (BH) or expected to tighten in the future (KZ).

As for the role of the distribution in the IRFs, I show in Appendix A that distributional factors predict both shocks; thus including the distribution is not merely cosmetic—it gets the aggregate IRFs *right*. Conceptually, the distributional block captures the demand-side information emphasized by Kilian and Lewis (2011)—heterogeneity in expenditure shares and household balance sheets—central to whether these shocks turn out inflationary. The most concrete demonstration of this is the monetary policy response: in the aggregates-only specification, oil supply shocks induce a tightening response of the policy rate; once the household distribution is included, this tightening disappears, consistent with the demand-side channels emphasized by Kilian and Lewis (2011). This has a complementary interpretation in the German context (Broer, Kramer, and Mitman, 2025), where a monetary tightening *does occur* under KZ and accounts for a substantial fraction of employment losses among low-income households. The choice of specification thus carries direct distributional stakes via a monetary channel.

component of CPI, for which consists of one-fourth the total costs considered in its computation (Chodorow-Reich et al., 2025).

To understand the underlying mechanisms for these results, I decompose the level and the variance of the impulse responses. The variance decomposition results reiterate the point that expectational shifts from oil supply news are more inflationary than realized supply disruptions *despite* similar oil-price responses. Oil supply news explains roughly two to three times the CPI variance share than do physical disruptions, even though the two shocks explain roughly comparable shares of oil-price variance (40% and 30%, respectively, in the medium run). The asymmetry lies in the price-pass-through channel and not in oil-price magnitude. Under BH, the variance contribution to oil prices and inflation is smaller and hump-shaped, implying a build up. Under KZ, they jump strongly on impact and gradually decline. Both shocks drive distributional dynamics similarly, with their FEVD share growing to about 10% over the horizon, consistent with a feedback role for the distribution.

Decomposing the level, I find that under the realized BH shock, the oil market and macroeconomy's responses are mediated primarily by the shock itself and by oil production—a cost-pull picture in which the physical shortfall (eventually) propagates through real activity. Under the KZ news shock, the response is mediated instead through oil prices and CPI inflation. The contribution of CPI inflation grows before fading slowly, in line with the gradual repricing of staggered price-setters (Calvo, 1983) and the slow updating of inflation expectations (Coibion and Gorodnichenko, 2015). Under KZ, opposite of BH, oil production plays essentially no role, as well as the shock itself, which may be due to OPEC production *and* non-OPEC production virtually not changing over the entire five-year trajectory (Figure 23). As for the U.S. household distribution, it contributes little to the *global* oil-market block, but extending the macro block to more domestic aggregates (Appendix J) reveals it as a medium-run amplifier of certain aggregates. Thus, although the two shocks would produce comparable aggregate IRFs, it would be through *different* mediators—a finding that would motivate different structural models of oil price transmission and naturally different policies.

These different underlying mechanisms will then matter for the distributional outcomes. While both shocks will have strong implications for inequality, the consumption and income dimensions go in opposite directions; wealth inequality, by contrast, moves in the same direction under both shocks, though with very different timing and profile. A sudden physical disruption will raise the Gini across every dimension in the joint distribution; increase the top-10% income share; decrease the bottom-50% consumption share—with asset-rich, income-rich households bearing the cost (in consumption terms) after the one-year lag required for production shortfalls to bind real activity. By contrast, the *asset-rich, income-poor* cells of the joint distribution—low-income households with substantial wealth—systematically benefit in consumption terms at intermediate horizons, as their wealth buffer absorbs the recessionary impulse: in the realized-supply-shock scenario, these are the only winners.

The Känzig (2021) supply news shock finds the opposite at medium horizons: a compressed consumption Gini, a fall in the top-10% income share, and an increase in the bottom-50% consumption share. The mechanism is different: news shocks are priced in opposite directions on impact depending on wealth, with financial-market participants—concentrated at the top of the wealth distribution—repricing expected supply scarcity first, while transfer-dependent households at the bottom are initially shielded by the higher pass-through of energy prices to nominal benefits. The asset-holders in the upper half of the wealth distribution contract consumption preemptively on this repricing—with one exception: at the very top of *joint* income and wealth, capital-income gains from the oil-related repricing dominate the precautionary motive, producing a consumption increase on impact that slowly fades over the horizon, making this group the KZ-side winner. Furthermore, low income and low wealth households decrease their consumption under both shocks in the medium-run, though results are imprecise.

Most relevant literature. This chapter connects several strands of the macroeconomic literature. I provide a detailed discussion in Section 2. Briefly, my work builds on the empirical oil shock literature initiated by Hamilton (1983) and developed by Kilian (2008b) and Kilian (2009), Baumeister and Hamilton (2019), and Känzig (2021); the nascent literature on oil shocks and inequality (Berisha et al., 2021; Drossidis, Mumtaz, and Theophilopoulou, 2024; Lei, Ludwig, and Ma, 2025; Broer, Kramer, and Mitman, 2025; Del Canto et al., 2023); the functional VAR methodology for distributional data (Chang, Chen, and Schorfheide, 2024; Chang and Schorfheide, 2024; Lenza and Savoia, 2024; Ettmeier, 2023; Bjørnland, Chang, and Cross, 2023); and the growing body of work that models household heterogeneity (Kaplan, Moll, and Violante, 2018; Auclert, 2019; Bayer, Born, and Luetticke, 2020).

To the best of my knowledge, this chapter makes three novel contributions. First, I am the first to compare how different types of oil disturbances — physical versus informational — propagate through the macroeconomy and the household sector. Second, I bring the recent econometric advances (Kilian and Zhou, 2023; Baumeister, 2025; Chan, 2022; Plagborg-Møller and Wolf, 2021; De Graeve and Westermark, 2025) to discipline the empirical framework, and apply the decomposition framework of Dufour and Wang (2024) to trace the shock's transmission through individual channels. Finally, within the same framework, I document how oil supply shocks shape the joint distribution of household income, consumption, and wealth — characterizing not just aggregate responses but who bears the cost across income, wealth, and consumption.

Outline. Section 2 reviews more broadly the related literature. Section 3 describes the data. Section 4 presents the empirical framework. Section 5 and Section 6 report the main results

for the macroeconomic aggregates and distribution respectively, comparing always the results from realized supply shocks and supply news shocks. Appendix F collects the robustness specifications. Section 7 concludes.

2 Literature

I build on and contribute to several literatures. I organize the discussion around four themes: the macroeconomic effects of oil shocks, identification in these oil-market systems, the distributional consequences of oil and energy price movements, and functional data methods for distributional macroeconomics.

Oil shocks and the macroeconomy. Oil prices have long been viewed as a key driver of macroeconomic fluctuations. Hamilton (1983) argued for a systematic relation between oil prices and output, noting oil price increases preceded seven of eight postwar U.S. recessions, motivating a large literature on the causal role of energy markets. This led to a central debate on whether oil price movements (which precede the recessions above) are driven primarily by supply or demand shocks. Using a structural VAR, Kilian (2009) decomposes oil price changes into supply, aggregate demand, and oil-specific demand shocks, showing that their macroeconomic effects differ sharply: demand-driven price increases tend to be expansionary, while supply-driven increases are contractionary. In contrast, Kilian (2008a) finds that exogenous supply disruptions have relatively modest effects on output, challenging earlier views that emphasized supply-side channels.

Identification. Early oil-VAR literature (Kilian, 2009; Kilian and Murphy, 2014a) identify oil shocks with timing zero-restrictions on contemporaneous responses (production and activity do not respond to each other within the month) plus sign restrictions on the remaining structural elasticities (supply slopes up, demand slopes down). Baumeister and Hamilton (2019) highlight that this approach is *set-identifying*, not point-identifying: sign restrictions bound structural parameters to an orthant but do not pin them down, and reported "point estimates" are typically arbitrary draws from an under-identified posterior. Because structural impulse responses are non-linear functions of the contemporaneous elasticities, posterior uncertainty in those elasticities propagates non-linearly into the IRFs, and the same data can support widely different conclusions depending on which point in the identified set one selects.

As such, the field has moved toward alternatives: structural Bayesian VARs with informative priors over the contemporaneous elasticities (Baumeister and Hamilton, 2019), external and internal instruments (Stock and Watson, 2018; Mertens and Ravn, 2013; Känzig, 2021;

Plagborg-Møller and Wolf, 2021) that use exogenous variation outside the VAR's information set to identify a column of the impact matrix by construction, and heteroskedasticity-based schemes. I follow the internal-instrument route of Plagborg-Møller and Wolf (2021), augmenting the VAR with the proxy ordered first via a Cholesky decomposition. The proxy point-identifies the column of the impact matrix corresponding to the supply shock under standard relevance and exogeneity conditions, sidestepping the under-identification debate.

Oil shocks, inequality, and the distribution. Compared to the vast literature on aggregate effects, remarkably little is known about the distributional consequences of oil shocks. The existing evidence is sparse and largely confined to reduced-form associations. Berisha et al. (2021) document a positive relationship between oil prices and income inequality (as measured by the Gini coefficient) across U.S. states, using panel data methods. Tan and Uprasen (2023) extend this analysis to the ASEAN countries, finding that rising oil prices tend to increase income inequality in oil-importing countries while reducing it in oil-exporting ones. Del Canto et al. (2023) study the distributional effects of inflationary shocks—a broader category that includes but is not limited to oil—using a feasible-set approach. They find that oil shocks are regressive, with asymmetric effects along the income distribution: households without a college degree require approximately $800 to afford their pre-shock consumption basket, while middle-aged college-educated households gain $833 from the same shock. Drossidis, Mumtaz, and Theophilopoulou (2024) use a factor-augmented VAR with external instruments to study the effects of oil supply news shocks on the U.S. income distribution. They document large negative effects at both tails of the income distribution—low-income households lose through wages and proprietors' income, while high-income households lose through corporate profits and interest income. Their analysis, however, is confined to income deciles and does not consider consumption or wealth. Baumeister et al. (2025) use a nonlinear heterogeneous agent VAR to study how households with different characteristics adjust their inflation expectations in response to oil supply shocks, combining a macro VAR block with a micro panel, allowing for sign and size asymmetries via Bayesian Additive Regression Trees.

Closest to my work are Broer, Kramer, and Mitman (2025) and Lei, Ludwig, and Ma (2025). Broer, Kramer, and Mitman (2025) use 45 years of German administrative data and the Känzig (2021) oil supply news shock series to document that oil shocks reduce earnings at the bottom of the distribution (by 0.75 percentage points two years after a 10% oil-price rise) and barely affect the top, driven primarily by separation-probability increases at the low end. They decompose the income response into a direct supply effect and an indirect effect from the systematic monetary-policy reaction, using the counterfactual-policy framework of Caravello, McKay, and Wolf (2024), and find that the monetary tightening accounts for a substantial part

of the employment losses among the poor. Lei, Ludwig, and Ma (2025) study the United States using an IV-SVAR combined with an incomplete markets model and find that oil price increases raise income and wealth inequality persistently, driven by stagflation that erodes the savings of low-wealth households and widens gaps as interest rates recover faster than wages. They also document a strong asymmetry, with oil price increases raising inequality more than decreases reduce it.

This chapter departs from all of these contributions in three respects. First, I study the *joint* distribution of income, consumption, and wealth—not any single margin. The joint distribution reveals mechanisms that marginal distributions cannot: in particular, the consumption insurance provided by flow income and wealth buffers. Second, I compare the distributional incidence of two fundamentally different oil disturbances—the realized supply shock of Baumeister and Hamilton (2019) and the supply news shock of Känzig (2021)—showing that whether the shock is physical or informational matters for *who* bears the cost. Third, I decompose each impulse response into the contributions of individual transmission channels via the causal-mediation framework of Dufour and Wang (2024), applied within a Bayesian VAR with the asymmetric conjugate prior of Chan (2022). This within-model decomposition identifies not just which variable absorbs the shock but the specific channels through which the response is mediated and how each channel's contribution evolves over the horizon—revealing that the BH/KZ wedge in distributional outcomes is fundamentally a wedge in transmission channels, not in aggregate footprints, a finding that variance decompositions or marginal-distribution analyses would miss.

Functional data methods for distributional macroeconomics. My empirical approach builds on recent work incorporating distributional data into VARs while addressing the key challenge of dimensionality. A full joint distribution over income, consumption, and wealth is too large to include directly, yet scalar summaries like Gini coefficients discard most information. Following Chang, Chen, and Schorfheide (2024) and Bayer, Calderon, and Kuhn (2025), I represent distributions as time-varying densities projected onto a finite set of basis functions whose coefficients evolve in a VAR, enabling standard impulse response analysis. I adopt the related framework of Bayer, Calderon, and Kuhn (2025), using Legendre polynomials and a copula-based representation to capture dependence across income, consumption, and wealth. This approach connects to a growing literature applying functional VARs to heterogeneous data (Ettmeier, 2023; Lenza and Savoia, 2024; Bjørnland, Chang, and Cross, 2023), and extends it by modeling a multivariate (three-dimensional) distribution to study the transmission of oil shocks.

Econometric methodology. I estimate the VAR using the asymmetric conjugate prior of Chan (2022), which allows equation-specific shrinkage—critical in my setting where the oil shock equation requires different regularization than the distributional factor equations. Hyperparameters are selected via marginal likelihood following Giannone, Lenza, and Primiceri (2015). I specify the VAR in log levels with priors for the long run (Giannone, Lenza, and Primiceri, 2019) to accommodate mixed persistence without differencing. I use 16 quarterly lags in the baseline, motivated by recent work showing that long-lag Bayesian VARs can reduce both bias and variance of impulse responses under shrinkage (De Graeve and Westermark, 2025). Ludwig (2024) establishes that iterated VAR impulse responses converge to direct local projection estimates as the lag order grows, and Antolín-Díaz and Surico (2025) demonstrate the practical importance of long lags for identifying the long-run effects of government spending. González-Casásus and Schorfheide (2025) develop an impulse response function criterion for jointly selecting the lag order and shrinkage.

3 Data

The analysis spans several data sources, but will primarily rely on five: (1) quarterly functional time-series data on a U.S. household distribution from Bayer, Calderon, and Kuhn (2025), (2) oil market variables from Känzig (2021), (3) macroeconomic aggregates from McCracken and Ng (2021), (4) the structural oil shock series from Känzig (2021) and (5) from Baumeister and Hamilton (2019). Depending on the estimation, the sample period will cover the mid 1970s to 2024, approximately 160-200 quarterly observations. Similarly, in documenting the robustness of the results, variables considered in the model may change. Details on these data and also data treatment are explained in the following sections. Appendix B provides an overview of all data used in this chapter, divided into broad categories.

3.1 Distribution of Consumption, Income, and Wealth

Description. To represent the distributional component of the model, I use the quarterly functional time-series data on the joint distribution of U.S. household consumption, income, and wealth constructed by Bayer, Calderon, and Kuhn (2025). Their methodology synthesizes microdata at the household-level with macroeconomic indicators of higher-frequency (McCracken and Ng, 2021) in a Bayesian state-space framework to produce a smooth and high-frequency representation of the joint distribution. The procedure relies on Sklar (1959): any joint distribution $\Xi_t(\boldsymbol{w})$ over $\boldsymbol{w} \in \mathbb{R}^d$ ($d = 3$ for consumption, income, wealth) can be decomposed into its marginal cumulative distribution functions $\Xi_{m,t}$ for $m \in \{c, y, w\}$ and a copula $C_t : [0,1]^d \to [0,1]$ that encodes the dependence structure independently of the marginals.

Representing the joint distribution then amounts to estimating the marginal CDFs and a copula.

Achieving a finite dimensional representation. These distributional objects are infinite-dimensional, making their inclusion in the model infeasible. To handle the infinite dimensionality of these objects, while also maintaining a functional approach, the marginal *quantile functions* $\Xi^{-1}_{m,t}$ and the copula *density* dC_t are projected onto an orthonormal basis of shifted Legendre polynomials $\{Q_o\}_{o\geq 0}$ on $[0,1]$:

$$\Xi^{-1}_{m,t}(u_m) = \sum_{o=0}^{\infty} \xi_{m,o,t}\, Q_o(u_m), \qquad u_m \in [0,1], \tag{1}$$

$$dC_t(u_1,\ldots,u_d) = \sum_{(o_1,\ldots,o_d)\in\mathbb{N}_0^d} \kappa_{(o_1,\ldots,o_d),t} \prod_{m=1}^{d} Q_{o_m}(u_m). \tag{2}$$

By orthonormality of the basis, the polynomial coefficients are inner products between the Legendre polynomials and the corresponding distributional object, be it $dC_t(u_1,\ldots,u_d)$ or $\Xi^{-1}_{m,t}(u_m)$. By infinite dimensionality of the distributional objects, the inner product is defined in the $\mathcal{L}^2$ space, which is the expected value of their product. This implies that the corresponding sample analogue is just a sample average over ranked microdata, shown in Equations 3 and 4:

$$\hat{\xi}_{m,o,t} = N_t^{-1} \sum_i w_{m,i,t}\, Q_o(u_{m,i,t}) \qquad \text{for the quantile function} \tag{3}$$

$$\hat{\kappa}_{(o_1,\ldots,o_d),t} = N_t^{-1} \sum_i \prod_m Q_{o_m}(u_{m,i,t}) \qquad \text{for the copula} \tag{4}$$

$$\boldsymbol{\xi}_{m,t} = \left(\xi_{m,0,t},\ldots,\xi_{m,O,t}\right)', \tag{5}$$

$$\boldsymbol{\kappa}_t = \left[\kappa_{(o_1,\ldots,o_d),t}\right]_{(o_1,\ldots,o_d)\in\{0,\ldots,O\}^d}, \tag{6}$$

$$\boldsymbol{\theta}_t = \left[\, \boldsymbol{\xi}'_{1,t},\ \ldots,\ \boldsymbol{\xi}'_{d,t},\ \ \mathrm{vec}(\boldsymbol{\kappa}_t)' \,\right]' \in \mathbb{R}^N \tag{7}$$

where $u_{m,i,t}$ is the data rank, derived from the cumulative sample weight at observation i after sorting the dimension m in the data and $w_{m,i,t}$ the data corresponding to the rank: the (treated) quantile for dimension m. Truncating the sums of Equation 1 and 2 at order O provides an approximation of the distributional objects. This truncation leaves us with a finite number coefficients 5 and 6 to be estimated, achieving the finite representation required to estimate the model. Once vertically concatenated, they form a finite vector $\boldsymbol{\theta}_t \in \mathbb{R}^N$, with $N = (O+1)^d + d\cdot(O+1)$ entries (quantile coefficients across d dimensions plus copula coefficients). These

polynomial coefficients $\boldsymbol{\theta}_t$ carry the business-cycle fluctuations of the joint distribution; the basis functions Q_o are known and invariant over time. For the interested reader, Appendix C restates the decomposition in greater detail.

Representation in BVAR. What will be important is how these data are represented in the Bayesian VAR. At the moment, to estimate the model with the coefficients $\boldsymbol{\theta}_t$ directly, although finite, would result in far too many equations. To reduce the dimensionality of the system, I exploit the correlational structure of $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T]'$ and extract the principal components of the coefficient matrix via PCA. From Bayer, Calderon, and Kuhn (2025), I know the data generation process of the joint distribution(s) estimated here and thus also know the first 8 principal components, $\boldsymbol{F}_t = (F_{1t}, \dots, F_{8t})'$, capture *all* the business cycle variation in the consumption, income, and wealth distribution. It is precisely these factors that enter the baseline Bayesian VAR. The treatment of the distributional data thereafter ensures impulse reponses $\boldsymbol{\Psi}_h^F$ still retain its properties as a distribution.[2] Appendix C explains how to then recover distributional moments of interest from these factor responses $\boldsymbol{\Psi}_h^F$, be it consumption of different household groups, inequality indices, and other conditional moments.

Robustness. I compare results where only the marginals of the distribution are included in the VAR. This amounts to only retaining coefficients $\boldsymbol{\xi}_{m,t}$ and in terms of variation, would be in line with the pre-dominant specification(s) found in the empirical inequality literature, allowing for a direct comparison. See Appendix D for more results.

Operationalization of concepts. The distributional factor estimates $\boldsymbol{F_t}$ rely on U.S. micro-data, U.S. macro data, a state-equation dictating the law of motion for the distribution, and a measurement equation which optimally weights the contribution of each data source to each estimate at each point in time, and which also reconciles differences in how each underlying micro-source operationalizes consumption, income, and wealth. In that sense, the operationalization of each concept — consumption, income, wealth — is a convex combination of these data sources and what the model believes is the best estimate for the unobserved distributional object.

Transforming these factors to observables, however, requires de-standardizing the coefficients with means and standard deviations estimated from a specific micro-dataset. This choice of de-standardization dataset enters only through the steady-state mean and unconditional

[2] Specifically, the reconstruction from factor estimates to distributional data preserves the properties of the underlying distributional objects across the impulse-response horizon $h \in 0, \dots, H$: the reconstructed marginal quantile functions $\hat{\Xi}_{m,h}^{-1}$ are weakly monotone in $u \in [0,1]$ for each m; the reconstructed copula density $d\hat{C}_h$ is non-negative on the trimmed cube $[\varepsilon, 1-\varepsilon]^d$; and the copula integrates to unity, $\int [0,1]^d d\hat{C}_h = 1$. Uniform marginals follow from the orthonormality of the Legendre basis. Monotonicity and non-negativity are verified ex post on a G^d tensor grid; deviations from the strict constraints are below 10^{-6} and treated as numerical noise.

volatility of each variable; the dynamics of the responses I report below are unaffected, since the measurement equation has already reconciled the operationalization differences across the underlying micro-sources. Since I am interested in the consumption, income, and wealth distribution, I use means and standard deviations from the PSID. To nonetheless provide some idea of the underlying variation constituting these estimates, I provide a description of the operationalization of concepts used in the PSID.

In the PSID, I measure **consumption** as containing food, rent (for renters) or housing rental equivalence (6% of home market value, for homeowners), utilities, health care, public transportation, education, childcare, and gas/vehicle repairs. **Income** contains labor earnings, public transfers, professional/self-employment income, rental income, dividends, and business/farm income, but is not net of taxes. **Wealth** is total assets minus total debt. Assets comprise liquid assets (checking, savings, CDs, money market, T-bills) plus illiquid assets (housing net of property debt, real estate, automobiles, retirement plans, life insurance, stocks, mutual funds, business equity). Debt comprises housing debt (mortgages, home equity lines of credit) plus personal debt (car loans, student loans, credit cards, medical, legal).

3.2 Oil Market and Macroeconomic Variables

Description. Alongside the distributional factors, the baseline VAR includes six oil market and macroeconomic variables: the real price of crude oil (West Texas Intermediate deflated by CPI-U), world crude oil production, a global industrial production index (OECD plus six major economies), U.S. industrial production, the consumer price index, and a proxy for OECD crude oil inventories. This block follows the standard specification in the oil shock literature (Kilian, 2009; Baumeister and Hamilton, 2019) to capture the global oil market and following Känzig (2021), I augment the global oil market structure with U.S. industrial production and U.S. consumer prices to capture domestic real activity and the inflation channel. Again, this defines the baseline of the model. I perform a battery of robustness checks to address concerns related to the above selection of variables. They are discussed below.

As my baseline inventory measure, I use the OECD log-inventory series constructed by cumulating Baumeister and Hamilton (2019)'s monthly Δi_t data into a quarterly log-inventory series. The underlying source is the same Kilian and Murphy (2014a) OECD series, also used by Hamilton (2009), Kilian and Murphy (2014b), and Känzig (2021), but the deseasoning and aggregation pipeline differs and is, by the argument of BH (Section III), less contaminated by measurement error, accounting for attenuation of its own estimates and general contamination of the system. I anchor the cumulated series at the level (of the noisy inventories variable) in 1975Q1 so that levels are comparable. BH's replication archive ends in 2019Q4; I extend the series through 2024Q2 by replicating their construction on current EIA inputs (U.S. crude,

OECD petroleum, U.S. petroleum stocks, global production). The replication agrees with BH's published Δi_t to a correlation of 0.997 over the 2000–2019 overlap window; details and the data-vintage caveat are in Appendix E. The results with the "noisy" series are provided for comparison in Appendix F.1.

Representation in BVAR. All oil market and macroeconomic variables enter in log levels rather than first differences and scaled by 100 for interpretability. First-differencing imposes that all structural shocks have permanent effects on the level of every variable, since a shock to Δy, the first difference, implies a level-shift forever. This would be a (unecessary) restriction at odds with economic theory, which holds that some shocks (e.g., monetary policy) have only transitory effects, while others (e.g., technology shocks) do not; especially since properly modeling the VAR in levels does not affect inference.[3] More on how my Bayesian setup accommodates these variables in Section 4. No further data treatment is taken.

Robustness. Here I consider robustness (and comparisons) using alternative oil or macro specifications, all detailed alongside their data sources in Table 3 of Appendix B.

Forward-looking information. Motivated by Mori and Peersman (2024), I address the potential predictability of both shocks by augmenting the baseline with a set of financial variables to capture forward-looking information. Their assessment on the predictability of the shocks relied on monthly data, so I computed my own Granger-causality tests for the quarterly data I use here (see Appendix A). The results show that the excess bond premium (EBP) of Gilchrist and Zakrajšek (2012) predicts the KZ shocks and the interest rates (term spread and short-term rates) predict the BH shocks. I have a robustness check that augments the baseline with precisely these channels. For completeness, I also report the Mori-Peersman block (VIX, S&P500, 1Y interest rate, EBP) that confounds monthly estimates, see Appendix F.2.

FRED-QD common factors. In a more brute force approach, I replace or augment the macro variables with principal component factors extracted from the FRED-QD quarterly database (McCracken and Ng, 2021)—a large panel of 113 macroeconomic series. I exclude from the PCA input panel any series that *already* appears as a named VAR observable; see Table 3, so the factor loadings cannot say *double-count* regressors. I use the levels factors $\boldsymbol{F}_t^{\text{lev}}$ in the spirit of Bai (2004), extracted from the log-level series. Factor count is selected following Freyaldenhoven (2022). Appendix A shows indeed these factors also predict the instrument. Does this

[3] Sims, Stock, and Watson (1990) show that OLS estimation of a VAR in levels produces consistent parameter estimates and asymptotically valid impulse responses regardless of the integration order of the variables, provided the model includes an intercept, so that pre-testing for unit roots is unnecessary. Toda and Yamamoto (1995) shows hypothesis testing on potentially mixed order systems as ones handled in Sims, Stock, and Watson (1990) is made possible via lag-augmentation, where the additional lags pick up the nonstationary nuisance parameters.

predictable variation overlap with the financial information set above? To see, I run a regression of each financial variable on the level factors. Appendix A show the level factors explain 92–99 % of variation in the rate-and-equity Granger-causing block above, 78% the term spread, but around 20% for EBP and 14% for VIX. Full impulse responses augmenting the baseline VAR with the levels factors are reported in Appendix F.7.

Oil-price measure. I replace WTI with two alternatives: the U.S. refiners' acquisition cost (RAC) of imported crude and the real-log Brent front-month price; construction details in Appendix B. This follows concerns raised by Kilian and Zhou (2023), where they argue WTI (and similar *U.S.* domestic prices) shouldn't proxy the *global* oil price post-1974 because U.S. price regulation until the early 1980s and 2010–2015 transportation bottlenecks broke arbitrage with world markets and that one should use the U.S. refiners' acquisition *cost* of imports (or Brent, post-mid-1980s); however Känzig (2021) shows in Appendix A.18-19 that results do not change when using the real refiner acquisition cost as oil price indicator, so I adopt the same baseline measure. I have run a robustness comparing impulse responses under these alternative measures and the qualitative conclusions are unchanged; figures available on request.

Capturing global demand. The baseline VAR uses world industrial production from Baumeister and Hamilton (2019) (OECD plus six major economies). Following Känzig (2021), as a robustness, I substitute in Kilian's (2009) Real Economic Activity (REA) index—a dry-cargo single-voyage shipping rate—motivated by the short-run identifying assumption that vessel capacity is fixed, so rate fluctuations track demand rather than supply (Kilian, 2019; Funashima, 2020). Two caveats apply: Kaufmann (2011) documents no statistically measurable relation between the shipping index and oil consumption over 1968–2008, weakening its interpretation as an oil-demand proxy; and Kolodzeij and Kaufmann (2014) flag that pairing the shipping index with RAC, as would be implied by Kilian (2009) and Kilian and Zhou (2023), as the price measure produces mechanical regression artifacts because transportation costs appear on both sides of the relevant equations. Neither concern appears in the baseline (World IP and real WTI); REA and RAC enter only as separate robustness checks on the activity and price measures, respectively.

Inventory measure. The baseline VAR uses the BH-cumulated inventory series described above; in a robustness (more of a comparison) I revert to the series of Kilian and Murphy (2014a) and Känzig (2021), which Baumeister and Hamilton (2019)'s Section III flags as carrying substantial measurement error. Their level correlation is 0.997, yet their $\Delta \log$ correlation is only 0.27, indicating that approximately 73 % of quarter-on-quarter variance is non-shared "construction noise." Reverting to the noisier series gives an explicit reading of how much of the

baseline result is sensitive to the measurement-error treatment. Comparison of results are in Appendix F.1; the more detailed comparison of the two series is in Appendix E.

Production aggregation. The baseline VAR uses world crude oil production. Following Kolodzeij and Kaufmann (2014), who argue that OPEC and non-OPEC producers operate under different output-setting criteria, I report a robustness that decomposes world oil production into OPEC and non-OPEC components. This is an interesting exercise, especially for the Känzig oil supply news shock because it reveals how non-OPEC nations react to these news, but also whether OPEC follows through on their announcements. See Appendix F.3.

All robustness exercises are estimated with the same prior and identification strategy as the baseline.

3.3 Oil Shock Series

To provide a thorough understanding of how oil supply disruptions propagate through the macroeconomy and household distributions, I provide a baseline estimation covering two empirically identified structural oil supply shock series.

Realized structural supply shock. The first shock is the realized structural oil supply shock of Baumeister and Hamilton (2019), identified through informative Bayesian priors on the structural impact matrix of a four-variable oil market VAR.[4] This shock captures *sudden* physical supply disruptions—*actual* shortfalls in oil production.

Oil supply news shock. The second shock are oil futures price surprises of Känzig (2021), constructed from a composite measure of WTI crude oil futures price changes in a narrow window around OPEC announcements, spanning the first-year term structure. This shock captures news regarding the future supply and operates through the *anticipation* channel. The announcements on oil production may or may not materialize later on—a key difference from Baumeister and Hamilton (2019).

The two identifications therefore isolate different stages of the same process. The KZ shock measures the market's *initial* response to OPEC's stated intentions—the repricing of

[4]The BH identification rests on prior beliefs about the contemporaneous structural elasticities of oil supply and demand—in particular, a tight prior centered on a small short-run price elasticity of demand ($\alpha_{pq} \approx -0.1$), disciplined by the micro-evidence on gasoline and oil demand at the pump (Hamilton, 2009). BH (their Section IV.D) argue that the wider, more "agnostic" priors used by Kilian and Murphy (2014a) are informative *against* the established micro evidence and effectively put non-trivial mass on values inconsistent with that literature. My use of their shock therefore inherits this prior structure: the supply-vs-demand decomposition that produces the realized supply shock is identified up to BH's prior over the four contemporaneous elasticities. The Känzig (2021) shock, identified by 30-minute announcement-window futures surprises, sidesteps this dependency by isolating supply news without taking a stand on demand elasticities, and serves as my comparison shock.

future supply expectations on announcement day. The BH shock measures the *realized* supply outcome—what actually happened to production, regardless of what was announced. The wedge between the two shocks reflects the difference between OPEC announcements and realized production.[5]

3.4 Sub-samples

I re-estimate the baseline VAR on three alternative subsamples that correspond to changes in the macro and oil-market regime. (1) The *post-1982* sample, which drops periods before 1982Q1, corresponds to an institutional regime change in OPEC: formal production quotas were introduced in March 1982, so the post-1982 sample is the period in which OPEC announcements correspond to quota commitments—the institutional setup the Känzig (2021) identification implicitly assumes. As Nakov and Pescatori (2010) emphasize, the period around 1981 was a time of dramatic change in world oil markets, in domestic energy production and consumption, and in U.S. monetary policy and the inflation environment; Hooker (2002) documents that the oil-price-to-inflation pass-through itself weakens substantially after this regime break, so the subsample doubles as a check that the inflation responses I report are not artifacts of pre-Volcker passthrough. (2) The *pre-Covid* sample truncates at 2019Q4, dropping the pandemic and its inflation aftermath. (3) The *pre-shale* sample truncates at 2010Q4, before the U.S. tight-oil boom materially raised the medium-run elasticity of global oil supply. Subsample IRFs are reported in Appendix F.4.

4 Bayesian Framework and Identification

This section presents the empirical framework for generating causal effects of oil supply shocks on macroeconomic outcomes and household distributions. This consists of a VAR that embeds both distributional and macroeconomic variables, a prior on the parameter space to discipline the model responses, and an identification strategy based on internal instruments. Each of these are discussed in turn.

[5] Realized production at the world level mixes OPEC compliance with non-OPEC production; Appendix F.3 reports a robustness that disentangles the OPEC and non-OPEC components, which directly provides some evidence of commitment (does the OPEC follow through), as well as whether non-OPEC offsets in any way. Baumeister (2023) show that OPEC compliance varies substantially over the sample period considered here.

4.1 Bayesian VAR

I embed the distributional factors alongside macroeconomic aggregates and oil market variables in a reduced-form VAR:

$$\boldsymbol{Y}_t = \boldsymbol{c} + \boldsymbol{A}_1 \boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{A}_p \boldsymbol{Y}_{t-p} + \boldsymbol{u}_t, \qquad \boldsymbol{u}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{8}$$

where $\boldsymbol{Y}_t = [z_t, \boldsymbol{M}_t', \boldsymbol{F}_t']'$ collects the oil supply instrument $z_t \in \mathbb{R}$, macroeconomic and oil market variables $\boldsymbol{M}_t \in \mathbb{R}^m$ following Känzig (2021), distributional factors $\boldsymbol{F}_t \in \mathbb{R}^d$ from Bayer, Calderon, and Kuhn (2025), and the reduced-form residual u_t . For the baseline, $\boldsymbol{M}_t$ includes the real price of crude oil (WTI deflated by CPI-U), world crude oil production, a global industrial production index (OECD plus six major economies), U.S. industrial production, the consumer price index, and the OECD crude oil inventories of Baumeister and Hamilton (2019). The number of lags is denoted by p and is equal to 16. With 15 variables and 16 lags, each equation contains 241 regressors (including constant). To address the resulting high dimensionality, I adopt a Bayesian framework.

Asymmetric conjugate prior. The standard Bayesian estimator of (8) stacks the reduced-form coefficient matrix $\boldsymbol{A} = [\boldsymbol{c}, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_p]$ and places a conjugate Normal–Inverse-Wishart (NIW) prior:

$$\operatorname{vec}(\boldsymbol{A}') \,\big|\, \boldsymbol{\Sigma} \;\sim\; \mathcal{N}\big(\operatorname{vec}(\boldsymbol{m}'),\; \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}\big), \qquad \boldsymbol{\Sigma} \;\sim\; \mathcal{IW}(\boldsymbol{\Psi}, \nu). \tag{9}$$

where $\boldsymbol{m}$ is the prior mean coefficient matrix (Minnesota-style: zero on every entry except the own first lag of each variable, set to a persistence parameter δ shared across all equations); $\boldsymbol{\Omega}$ is a positive-definite matrix encoding lag-decay shrinkage on the regressors; and $\boldsymbol{\Sigma}$ is the covariance matrix of the reduced-form residuals $\boldsymbol{u}_t$, on which one can place an Inverse-Wishart prior with scale matrix $\boldsymbol{\Psi}$ ($n \times n$, positive definite) and degrees of freedom $\nu > n - 1$.

The Kronecker structure $\boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}$ ties the prior covariance *across* equations: the same lag-decay matrix $\boldsymbol{\Omega}$ governs every equation's coefficients, scaled only by the contemporaneous covariance $\boldsymbol{\Sigma}$. In a Minnesota parameterization this collapses to a single scalar *own*-lag tightness κ_1, a single *cross*-lag tightness κ_2, and a *single* persistence parameter δ shared by every variable. This is restrictive: the shock instrument z_t should be a priori unpredictable ($\delta = 0$, very tight κ_1); macroeconomic aggregates are near-unit-root ($\delta \approx 1$, looser κ_1); distributional factors are stationary, smooth, and numerous, and benefit from aggressive cross-equation shrinkage ($\kappa_2 \to 0$). One $(\kappa_1, \kappa_2, \delta)$ triple cannot accommodate all three.

Triangular reparameterization. Chan (2022) replaces the symmetric NIW with a per-equation Normal–Inverse-Gamma prior by reparameterizing (8) into recursive structural form. Decom-

pose the reduced-form innovation covariance as

$$\boldsymbol{\Sigma} \;=\; \boldsymbol{A}_0^{-1}\,\boldsymbol{D}\,\boldsymbol{A}_0^{-\top}, \qquad \boldsymbol{A}_0 \;=\; \boldsymbol{I} - \boldsymbol{L}, \tag{10}$$

where $\boldsymbol{L}$ is strictly lower triangular (zero diagonal, free entries a_{ij} for $i > j$) and $\boldsymbol{D} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. Pre-multiplying (8) by $\boldsymbol{A}_0$ transforms the system to

$$\boldsymbol{A}_0\,\boldsymbol{Y}_t \;=\; \boldsymbol{A}_0\,\boldsymbol{c} \;+\; \boldsymbol{A}_0\,\boldsymbol{A}_1\,\boldsymbol{Y}_{t-1} \;+\; \cdots \;+\; \boldsymbol{A}_0\,\boldsymbol{A}_p\,\boldsymbol{Y}_{t-p} \;+\; \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \;\sim\; \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}), \tag{11}$$

with structural innovations $\boldsymbol{\varepsilon}_t = \boldsymbol{A}_0\,\boldsymbol{u}_t$ that are mutually orthogonal by construction. Reading off equation i gives the recursive form

$$y_{i,t} \;=\; \underbrace{-\sum_{j<i} a_{ij}\, y_{j,t}}_{\text{contemporaneous block}} \;+\; \boldsymbol{x}_t'\,\boldsymbol{\beta}_i \;+\; \varepsilon_{i,t}, \qquad \varepsilon_{i,t} \;\sim\; \mathcal{N}(0, \sigma_i^2), \tag{12}$$

where $\boldsymbol{x}_t = [\boldsymbol{Y}_{t-1}', \ldots, \boldsymbol{Y}_{t-p}', 1]'$ collects the lagged regressors and the constant. Conditional on the contemporaneous coefficients $\{a_{ij}\}_{j<i}$, equation i is a single-equation linear regression with Gaussian errors; the structural innovations $\varepsilon_{i,t}$ are independent across equations, so equation-by-equation priors are now coherent. This is the step that breaks the Kronecker constraint of (9).

Stacking the unknowns of equation i as $\boldsymbol{\theta}_i = [a_{i,1}, \ldots, a_{i,i-1}, \boldsymbol{\beta}_i']'$ (contemporaneous coefficients first, then lag coefficients and constant), Chan (2022) places independent Normal–Inverse-Gamma (NIG) priors:

$$\boldsymbol{\theta}_i \,\big|\, \sigma_i^2 \;\sim\; \mathcal{N}\big(\boldsymbol{m}_i,\; \sigma_i^2\,\boldsymbol{V}_i\big), \qquad \sigma_i^2 \;\sim\; \mathcal{IG}\Big(\tfrac{\nu_i}{2},\; \tfrac{S_i}{2}\Big), \qquad i = 1, \ldots, n. \tag{13}$$

The hyperparameters (ν_i, S_i) govern the marginal Inverse-Gamma prior on the structural-shock variance: $\mathbb{E}[\sigma_i^2] = S_i/(\nu_i - 2)$ for $\nu_i > 2$. I set $\nu_i = 3$ and $S_i = \hat{\sigma}_i^2$ so that $\mathbb{E}[\sigma_i^2] = \hat{\sigma}_i^2$, again anchored on the AR(p) residual variance. The Inverse-Gamma is the univariate analog of the Inverse-Wishart, conjugate to the equation-i Gaussian likelihood, so the posterior of $(\boldsymbol{\theta}_i, \sigma_i^2)$ is again Normal–Inverse-Gamma and draws are available in closed form.

The prior mean $\boldsymbol{m}_i$ is zero on every entry except the coefficient on the own first lag of variable i, which is set to δ_i (the equation-specific persistence prior). The prior variance is diagonal with the Minnesota structure of Litterman (1980) and Doan, Litterman, and Sims

(1984):

$$V_i^{(j,\ell)} = \begin{cases} \kappa_{1i} \,/\, (\ell^{2\lambda_3}\, \hat{\sigma}_i^2) & \text{own lag } (j = i), \\ \kappa_{2i} \,/\, (\ell^{2\lambda_3}\, \hat{\sigma}_j^2) & \text{cross lag } (j \neq i), \\ \kappa_{2i} \,/\, \hat{\sigma}_j^2 & \text{contemporaneous coefficient } a_{ij}, \\ \kappa_4 & \text{intercept}, \end{cases} \tag{14}$$

where $\hat{\sigma}_j^2$ is the residual variance from a univariate AR(p) for variable j, ℓ indexes the lag order, λ_3 governs lag decay, and κ_4 is a fixed loose prior on the constant. The crucial difference from (9) is that each equation now has its own $(\kappa_{1i}, \kappa_{2i}, \delta_i)$.

Equation-specific hyperparameter selection. Conjugacy of (13) delivers a closed-form equation-level marginal likelihood. I exploit this to select $(\kappa_{1i}, \kappa_{2i}, \delta_i)$ *per equation* by grid search:

$$(\kappa_{1i}^*, \kappa_{2i}^*, \delta_i^*) = \arg \max_{(\kappa_1, \kappa_2, \delta) \in \mathcal{G}} \log p\big(\{y_{i,t}\}_{t=1}^T \,\big|\, \kappa_1, \kappa_2, \delta;\ \mathcal{H}_{-i}\big), \tag{15}$$

where $\mathcal{H}_{-i}$ holds the system-wide hyperparameters fixed at their current values. The persistence grid is restricted as

$$\delta_i \in \begin{cases} \{0\} & \text{if } i \leq n_s \quad \text{(shock equations: dogmatic white-noise prior)}, \\ [\,0,\ 1\,] & \text{otherwise}, \end{cases} \tag{16}$$

where n_s is the number of shock instruments. The shock equation is pinned at $\delta_i = 0$ to encode the assumption that z_t is unpredictable from its own past; every other equation selects δ_i^* from the data, so I(0) factors typically receive small δ_i^*, near-I(1) macro variables receive $\delta_i^* \approx 1$, and stationary-but-persistent variables select intermediate values. The shared hyperparameters—λ_3 and the long-run prior tightness (Section 4.1)—are optimized at the system level using the marginal-likelihood approach of Giannone, Lenza, and Primiceri (2015).

The end result: every equation receives its own shrinkage $(\kappa_{1i}^*, \kappa_{2i}^*)$ and its own persistence prior δ_i^*, all chosen by the data, in contrast to the single $(\kappa_1, \kappa_2, \delta)$ imposed by the symmetric NIW prior in (9). The oil shock equation requires little shrinkage; the distributional factor equations, smooth and numerous, receive aggressive cross-equation regularization. Posterior draws are independent across equations, conditional on $(\kappa_{1i}^*, \kappa_{2i}^*, \delta_i^*)$, and available in closed form.

Long-run priors. All oil market and macroeconomic variables enter in log levels. To discipline long-run behavior without imposing unit roots, I augment the likelihood with the *prior for the long run* (PLR) of Giannone, Lenza, and Primiceri (2019), which regularizes the long-run multiplier matrix $\boldsymbol{A}(1) = \boldsymbol{I} - \boldsymbol{A}_1 - \cdots - \boldsymbol{A}_p$ and lets the data determine which linear combinations of variables share common stochastic trends, rather than hard-coding cointegration rank ex ante. The construction is data-adaptive (ADF-based stationarity classification + auto-detected cointegrating pairs), and the long-run tightness is optimized jointly with the Minnesota hyperparameters via marginal likelihood. Full implementation details are in Appendix G.

Lag length. I set $p = 16$ quarterly lags. Kilian and Zhou (2023) and Hamilton and Herrera (2004) show longer lags are needed to allow for the building and contraction of oil market cycles. This choice is also motivated by recent evidence that, under Bayesian shrinkage, long-lag VARs can simultaneously reduce both the bias and variance of impulse response estimates (De Graeve and Westermark, 2025). Relatedly, Ludwig (2024) shows that iterated VAR impulse responses converge to direct local projection estimates as the lag order grows with the horizon, implying that long-lag VARs inherit the misspecification robustness of local projections while retaining the efficiency of the parametric model. Following comments from Baumeister (2025), which shares the view of González-Casásus and Schorfheide (2025), I do not select the lag order via AIC or BIC, which target forecast loss and can yield specifications that are suboptimal for structural inference. Instead, I wish to speak on short- to medium-run dynamics and thus, p is set to reflect that. In Appendix H, I show how impulse responses change from varying the lag order p, which will take values $\in \{2, 4, 8, 12, 20\}$ around the baseline $p = 16$.

Posterior simulation. Following Chan (2022), the posterior of each equation is available in closed form as a NIG distribution, and draws are independent. I draw 5,000 independent coefficient vectors with a stable companion matrix (eigenvalues within the unit circle), discarding explosive draws, and compute structural impulse responses for each. Credible bands are constructed as pointwise 68% posterior intervals; 90% bands and posterior point summaries other than the median (mean and mode) are available on request and yield qualitatively identical conclusions.

4.2 Identification

To identify oil supply shocks, I use the internal instrument approach of Plagborg-Møller and Wolf (2021). The shock series z_t is included directly in the VAR as the first variable, ordered before the real oil price and all other endogenous variables.

Relevance and exogeneity in a Bayesian context. Identification is then based on treating each external proxy separately as an instrument, z_t, that satisfies the relevance and exogeneity condition regarding its underlying structural shock (Stock and Watson (2012) and Mertens and Ravn (2013))[6]:

$$cov(z_t, \varepsilon_{1t}) = \alpha \neq 0 \tag{17}$$

$$cov(z_t, \varepsilon_{-t}) = 0, \tag{18}$$

where ε_{1t} denotes the first shock in the system and ε_{-t} the remaining $n-1$ shocks. Condition (17) is the relevance condition of the instrument and is testable. Condition (18) is the exogeneity of the instrument. Because ε_{-t} is unobserved, this condition is not *directly* testable, but its observable implications can be examined—for example, orthogonality of z_t to pre-determined macro variables and to alternative external instruments for non-target shocks. I present such diagnostics in Appendix A for both shocks.

Under the Bayesian internal instrument approach employed here, relevance governs the informativeness of the likelihood about the structural impulse responses. A weak instrument implies a flat likelihood in the direction of the shock of interest, so that the posterior is dominated by the prior: credible bands remain wide and centered on prior-implied dynamics rather than being updated by the data. I provide evidence that this concern does not apply in my setting: the FEVD in Section 5 will show that the Baumeister and Hamilton (2019) shock explains a non-trivial share of the forecast error variance of all variables, and under the oil supply news instrument, the impulse responses largely align with those obtained in Känzig (2021), for which relevance is well-established.

Fundamentalness. A further consideration is the fundamentalness of the instrument relative to the VAR's information set. A shock is fundamental if it can be recovered from current and past observables in the system; non-fundamentalness arises when agents possess information not spanned by the econometrician's VAR. Miranda-Agrippino and Ricco (2023) show that, under non-fundamentalness, the external SVAR-IV approach produces biased impulse responses, and Bruns and Lütkepohl (2025) formally establish that the external and internal approaches are not equivalent precisely in this case.

This is what makes the internal instrument approach of Plagborg-Møller and Wolf (2021) attractive. It resolves the point mechanically by augmenting the VAR with z_t directly: the system's information set is enlarged to span the instrument, and fundamentalness holds by construction within the augmented system. The inclusion of the instrument, as opposed to

[6] Sufficient lags absorb the lag-exogeneity component by whitening the residuals; the lead-exogeneity component is an assumption about the construction of the proxy, which in my case follows from KZ's OPEC-announcement timing / BH's sign-restriction design.

the external-instrument approach, does affect inference. It may produce wider credible bands, since the model propagates the instrument's full parameter uncertainty—its own AR dynamics, its loadings on the other variables, and the associated rotation of the structural impact matrix — rather than only the first-stage impact uncertainty as with the external instrument approach (see Caldara and Herbst, 2019).[7] In this sense my inference is conservative. This pattern of small bands in external vs. large bands in internal is visible in Degasperi, Figure C.4, panel (a).

Structural impulse responses. Consider the mapping between the reduced form errors u_t and the structural shocks ε_t of Equation (8):

$$u_t = \boldsymbol{B}_0 \varepsilon_t \tag{19}$$

where $\text{cov}(u_t) = \boldsymbol{\Sigma} = \boldsymbol{B}_0 \boldsymbol{B}_0'$. I identify *oil supply* (Baumeister and Hamilton, 2019) and *oil supply news* (Känzig, 2021) shocks based on the internal instrument approach motivated by Plagborg-Møller and Wolf (2021) and order the analogous instrument z_t first in the following Cholesky ordering:

$$\boldsymbol{Y}_t = [z_t, \boldsymbol{M}_t', \boldsymbol{F}_t']'. \tag{20}$$

Ordering the instrument first ensures that the first structural shock, ε_{1t}, is the innovation to z_t that is orthogonal to lagged information in the system. The first column of $\boldsymbol{B}_0$ gives the impact responses of all variables to this shock. The remaining $n-1$ shocks are indeterminate up to an orthogonal rotation (Stock and Watson, 2018). As shown by Plagborg-Møller and Wolf (2021), this approach is robust to non-invertibility, abstracts from a necessary first stage regression outside the VAR and instrument relevance is reflected in the width of the posterior credible bands. The shock is normalized to produce a 10 percent increase in the real price of oil on impact.

The Wold moving average coefficients Ψ_h are computed recursively from the estimated VAR coefficients:

$$\Psi_0 = I, \qquad \Psi_h = \sum_{\ell=1}^{\min(p,h)} A_\ell \, \Psi_{h-\ell}, \quad h \geq 1. \tag{21}$$

The structural impulse response of variable i at horizon h to the identified oil shock is then

$$\Theta_{i,h} = e_i' \Psi_h B_0 \, e_1, \tag{22}$$

[7] This widening is more pronounced when the instrument is weak, because a fully Bayesian treatment internalizes the weak-signal case as part of the posterior, whereas the external approach takes the first-stage estimate as a *point* and builds bands conditional on it, so the instrument's own uncertainty does not propagate.

where e_i and e_1 are selector vectors. Since the oil shock is ordered first, the first column of B_0 gives the on-impact responses, and Ψ_h propagates these forward through the VAR dynamics.

5 Macroeconomic Effects of Oil Supply Shocks

This section presents the first part of the empirical results, with the focus on macroeconomic aggregates. The section opens with the aggregate impulse responses.[8] I then present forecast error variance decompositions, to provide some evidence on the relevance of the oil supply shocks on the uncertainty of the variables in my system. Next, I use the framework of Dufour and Wang (2024) to decompose the *level* of the IRFs and uncover the different channels driving the observed responses using the framework of Dufour and Wang (2024). Throughout the results section, I compare the Baumeister and Hamilton (2019) physical oil supply shocks with the Känzig (2021) oil supply news shock, to distinguish the two economic channels: sudden physical disruptions versus expectational shifts in oil supply.

To complement these aggregate responses, Appendix I runs a *rotating* Factor-Augmented Vector Autoregression (FAVAR). This amounts to running many Bayesian VARs, where one observable is *rotated* out for each Bayesian VAR, with the idea of obtaining evidence on the transmission of both shocks across the many parts of the macroeconomy. This strategy has been similarly adopted by Känzig (2021) and Lei, Ludwig, and Ma (2025), among others, and a quick way to obtain *a sense* of the transmission mechanisms one can draw from these shocks.[9]

5.1 Aggregate Impulse Responses

Figure 1 reports the impulse responses of key macroeconomic and oil market variables to the two shocks. To allow for a clean comparison, both shocks are normalized to produce a 10 percent increase in the real price of oil on impact.[10] Given they are otherwise identical systems, the normalization allows us to even the field and ask how an oil price hike transmits in both settings. All impulse responses are estimated to horizon $h = 20$.

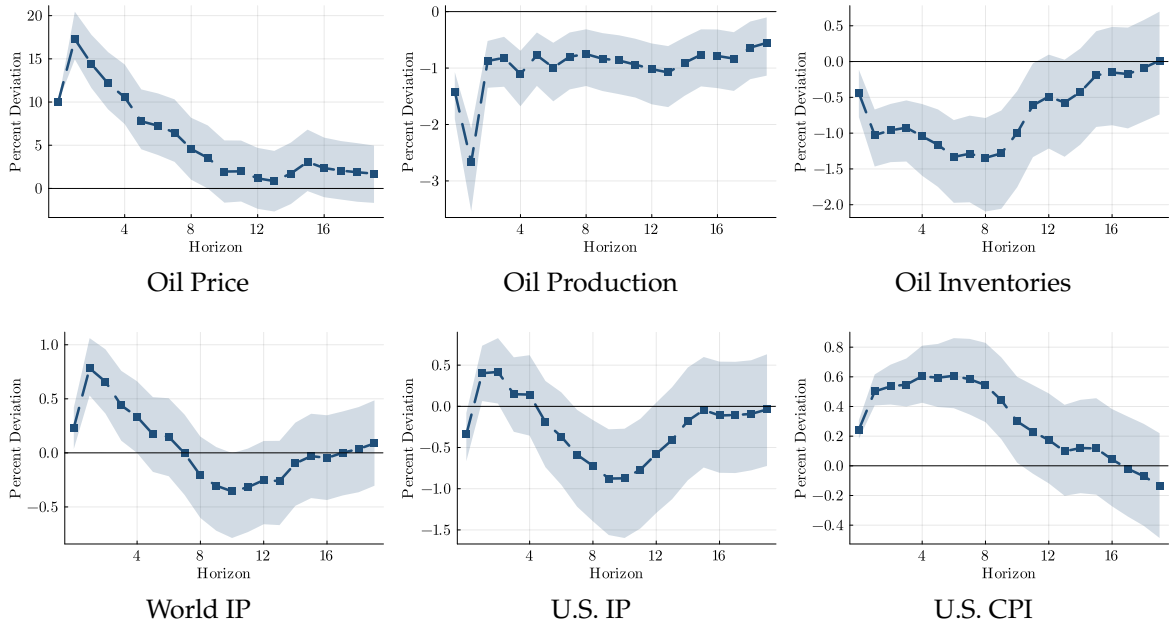
[8] Appendix F.5 provides a comparison between my VAR augmented with distributional factors and those without as typically examined in the literature.

[9] I generate IRFs for variables on real economic activity, labor markets, income, fiscal policy, consumer prices, consumption, interest rates, financial conditions, credit supply, household debt, and others, for the U.S. economy. For each variable, I plot two IRFs: the observable IRF with aggregate factors identified with the McCracken and Ng (2021) FRED-QD and another where distributional factors are added on top.). These responses were only meant to inform my hypothesis of the main results, which *do* model the oil structure, which I have argued is important for identification. The rotating FAVARs *do not* have the oil variables. Nevertheless, the sensitivity I raised thus far on modeling the oil variables concerned only oil inventories. The decomposition I show later on provides a formal identification of these mechanisms within a single model.

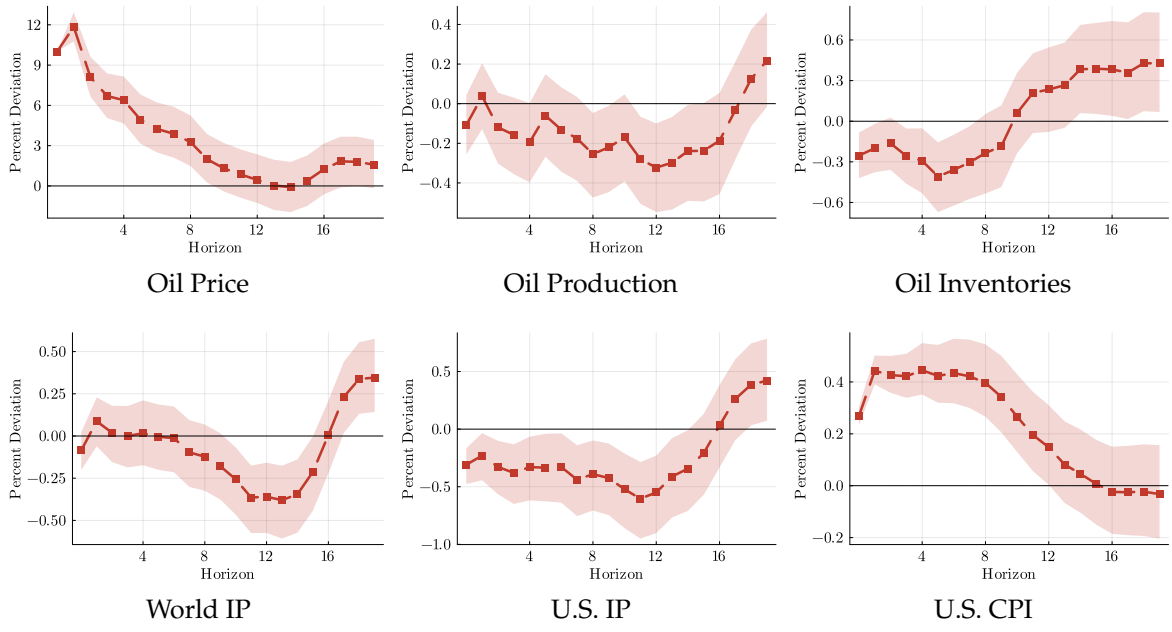
[10] The variable normalized on (the oil price) responds contemporaneously under both shocks, so the contemporaneous unit-effect normalization is appropriate (Stock and Watson, 2018, §2.1.3); the variables that respond with a delay under KZ (production, inventories) are correctly traced as IRFs from this normalization, not used as the normalizing object.

Figure 1: Aggregate impulse responses to oil supply shocks

Panel A: Baumeister & Hamilton (2019) realized supply shock



Panel B: Känzig (2021) supply news shock



Notes: Impulse responses to oil supply shocks normalized to produce a 10% increase in the real price of oil on impact. Solid lines: posterior median. Shaded areas: 68% credible sets. Horizon in quarters.

First, Panel A of Figure 1 shows the classical oil supply shock: a sudden production shortfall, coupled with a depletion of oil inventories. The model finds sudden disruptions in the availability of oil *increases* general prices on impact; which remain elevated for about two years. Industrial production, both for the U.S. and abroad, show a rather delayed effect, with U.S. industrial production decreasing near 1% two years later. The initial rise in World and U.S. IP seems to be mechanical. The geography of World IP is OECD + 6 non-member economies, many of which are oil-producing nations that benefit from the OPEC production cuts, taking prices as given. You can see from U.S. IP that they are a contributor of the rise. Eventually, demand side factors do drive production downward. I have a robustness specification in Appendix F.6 which replaces World IP with Kilian's (2009a) measure of real economic activity (REA), deflated dry-cargo shipping rate. There, it is free of this geographic artifact, showing a much more muted rise and quicker decline.[11]

To continue, Panel B of Figure 1 shows the macroeconomic responses under an oil supply news shock. For prices, the same pattern emerges: prices jump on impact and remain elevated for two years before gradually returning to their steady state levels. Interestingly, industrial production falls on impact and for three years, is negative. The response of World IP is quite muted relative to the BH shock and similarly decreases with a delay as in the BH shock, albeit without the strong decline in oil production. The IRFs suggest that the anticipation channel dominates the oil-price-driving extraction effect: in mere anticipation of a lower oil supply following OPEC's announcement, industrial firms reduce production. I confirm this in Figure 4 which shows the channel decomposition. Another interesting phenomenon of these news shock is the recovery: World IP, U.S. IP, and oil production observe a significant rise around year five, returning and exceeding pre-shock levels, as opposed to the BH shock. This feature is robust to several specifications.

Discussion on oil inventories. Under the oil supply news shock, I can only confirm the baseline response of oil inventories Känzig (2021) finds for the medium, but not for the short term. Whereas Figure 3 of KZ documents the characteristic precautionary "piling up" of inventories on impact—consistent with the interpretation that news shocks shift inventory demand in anticipation of future supply shortfalls—my estimates show inventories *declining* initially (about one-third of the BH response), with an accumulation emerging only after approximately two years. This result is present in virtually all my robustness checks and is not an artifact of the different inventories variable I use here (see Figure 17). Other checks show their result is sensitive to primarily truncation bias (Figure 34) and similarly, omitted variable bias (Figure 35 and

[11] Kaufmann (2011) documents that the dry-bulk shipping index — the basis of the Kilian REA — has no statistically measurable relation to oil consumption over 1968–2008, weakening its interpretation as a global oil-demand proxy. The baseline VAR uses worldip (OECD + 6 industrial production); REA enters only as a robustness on the activity measure.

32). Examining the KZ results, I find two drivers of the deviation: the absence of an oil-market structure in the VAR and additional conditioning variation that absorbs macroeconomic dynamics unrelated to oil markets.[12]

I do not interpret my findings as an absence of the speculative-demand mechanism, but rather for oil inventories, there may be evidence that other forces operate in the opposite direction and dominate at the announcement horizon. For example, *storage arbitrage* may exist, where the impact spot-price spike compresses the carry premium and pushes optimal inventories *down* (Pindyck, 2001). Second, *imperfect "cartel" discipline*, for which there is substantial evidence. OPEC members routinely produce above their assigned quotas: historical compliance rates range from 40 to 80 percent, varying with the price cycle and internal cohesion. This is why several suthors have stated that OPEC is not a cartel (Colgan, 2014; Ratti and Vespignani, 2015; Molchanov, 2003; Baumeister and Kilian, 2016). To the extent that OPEC production announcements affect oil prices (conditional on non-members beliefs), the probability of deviation necessarily compresses the revision in $\mathbb{E}_t[P_{t+1}]$ (from a change in expected supply) that would otherwise motivate hoarding; and lastly, perhaps the *post-shale supply elasticity* weakens the precautionary motive over the post-2010 part of my sample, which would be roughly 1/4 of the entire sample. This last channel is assessed Figure 26 and indeed shows less depletion of oil inventories in the first year. [13]

What conclusions can we draw? First, the realized BH shock and the supply news KZ shock produce remarkably similar oil-price dynamics and aggregate dynamics, but the timing of the general dynamics is worth noting: The *real*-side transmissions are fundamentally different. The BH shock operates through a *cost-pull* channel—actual production shortfalls raise input costs on impact (Appendix Figure 45 and 38), but real activity (Appendix Figure 36), labor markets (Appendix Figure 37) and the SP500 (Appendix Figure 41) contract with a one year lag required for those costs to bind, strengthening the result of Kilian (2008b). CPI energy and CPI transportation (Figure 38) do rise on impact and remain elevated for roughly two years, consistent with energy-cost pass-through into both upstream producer prices and the

[12]Känzig (2021) runs a robustness check with FRED-MD factors layered on his 6-variable VAR but does not report the result. Running the baseline (oil + distributional factors) with FRED-QD macroeconomic factors on top (Figure 32) shows a compromise: oil inventories decline on impact, but remain stable for two years (credible interval contains zero), then increase afterwards similarly to my IRF. I thus identify a sensitivity of the KZ result to truncation bias (Figure 34) and omitted variable bias (Figure 32).

[13]Signs of this result also appear in Känzig (2021). In fact, a closer look shows the early negative response and subsequent rise more closely resembles the inventories response to oil supply *surprises* in KZ's two-shock specification (Appendix Figure A.13), and aligns with his LP-IV results (Appendix Section A.2.2, Figure A.4), which shows the strong negative response when one adds more lags/more controls.[14] They also align with the Figure 3 oil inventories response of Mori and Peersman (2024), which as mentioned already, deal with a specific form of contamination in the instrument.[15] The result suggests the presence of other forces that matter for the anticipation channel in oil inventories, which actually relieves concerns that this is a special storage demand shock as stated by Kilian and Zhou (2023).

energy/transport components of consumer prices; but, CPI-core does not—the credible interval contains zero throughout the entire IRF horizon—again in line with Kilian (2008b), who find this similar profile but for CPI inflation, and with Hooker (2002), who documents that the pass-through from oil prices to inflation weakens substantially after the early-1980s regime break. In contrast, the corresponding KZ panels show that input costs rise, but real activity, labor markets, and the SP500 all respond on impact and the contractions are persistent. Furthermore, KZ is inflationary—all prices rise.

A striking related observation concerns *inflation expectations*. Using the University of Michigan household one-year-ahead inflation expectations and the SPF median one-year CPI forecast (Figure 44), the response of expected inflation to BH and KZ shocks is essentially identical. Both surveys—the household survey and the professional-forecaster survey—treat the two shocks the same, even though their underlying inflation dynamics differ: BH raises headline but not core, KZ raises all prices including core. Households and forecasters either do not distinguish between physical and informational oil disturbances, or they do not have the information set the VAR uses to decompose them. This has direct monetary-policy implications: a central bank that responds to inflation expectations sees the *same* signal regardless of shock type, even though the warranted policy response is plausibly different (more muted under BH given the absent core response; more active under KZ given the persistent broad pass-through).

What about monetary policy? The KZ inflation response induces a somewhat stronger monetary policy reaction on impact, but the effect is not credibly different from zero. Beyond impact, policy is largely unresponsive over the first two years and subsequently eases in response to recessionary pressures; the BH shock exhibits a similar profile. Taken together, these results suggest that monetary policy does not systematically tighten in response to oil supply shocks, even if they may be inflationary, consistent with Kilian and Lewis (2011), who emphasizes the importance of demand-side channels. Notably, excluding the household distribution yields the opposite pattern—an increase in policy rates—indicating that incorporating the distribution dampens the case for a tightening response.

What about the medium-run? As a result of the cost-pull channel, the BH shock generates larger effects despite the delay. Except for World IP, aggregate responses are larger in the medium run under a BH shock, indicating the presence of adverse general equilibrium effects kicking in. Rotating FAVAR results show this is a systematic response. The decomposition later on will reveal the key drivers of these effects.

And what about recovery? The year-five recovery and overshoot under KZ, absent under BH, is itself interesting: when expectations of supply scarcity prove only partially realized or even *not* realized (consistent with the OPEC non-compliance evidence in the inventories dis-

cussion above and also Figure 24), the implicit precautionary contraction unwinds—capital, inventory, and saving decisions made under the initial belief gradually reverse over the horizon and real activity rebounds toward—and in some cases beyond—its pre-shock level. Both findings echo Kilian (2009)'s seminal point that *not all oil price shocks are alike*—but here the two shocks fall on opposite sides of the same OPEC decision, depending on whether identification picks up the *announcement* or the *realization*.[16]

5.2 Forecast Error Variance Decompositions

Description. Figure 2a and Figure 2b present the forecast error variance decomposition (FEVD) of the identified shocks.[17] The BH shock explains approximately 60 percent of oil production variance near impact, gradually declining to below 40% over the long horizon; about 30 percent of oil price variance, and 10–20 percent of CPI variance. In the distributional block, it accounts for little on impact, but incrementally grows past 10% at horizon $h = 20$, a non-trivial share, suggesting that the joint distribution contains non-negligible information about oil shock transmission that grows gradually. Oil inventories would follow a similar profile. Under the current specification, the shock explains relatively little of industrial production variance, both in the U.S. and abroad—echoing Kilian (2009)'s finding that exogenous oil-supply disruptions contribute only modestly to real-activity fluctuations.[18]

In comparison, the KZ news shock explains the largest share of forecast variance for the two macro variables most associated with oil episodes: the oil price and CPI. About 45% of oil price variance is explained by the shock on impact, declining modestly to 40% by horizon $h = 20$, which is larger than the BH shock (30% stable). CPI variance follows a similar trajectory, jumping to 40%, but declines quicker. Unlike the BH shock, oil production starts near zero and rises to just under 10% — the shock doesn't move production immediately (it is news about future supply), but gradually explains more as anticipated production adjustments perhaps materialize. The distributional block's FEVD share grows similarly to under the BH shock—small on impact, 10% by horizon $h = 20$—confirming that the distribution carries non-trivial information about transmission under both identifications. Oil supply shocks drive less the uncertainty of the other variables, but nonetheless the contribution of the shocks are distinguishable and non-trivial, providing evidence of their relevance.

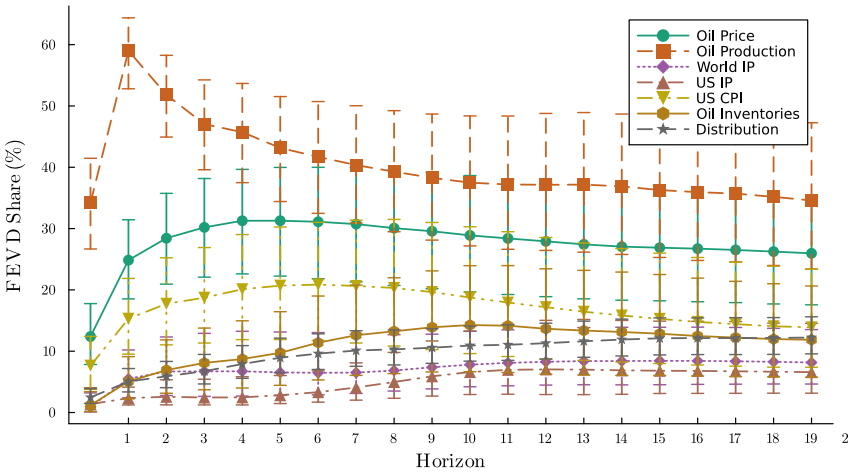
[16] The BH shock concerns *world* oil production, so, not exactly opposites of the same OPEC decision and indeed, Figure 24 shows both OPEC and non-OPEC oil production decline on impact, but *only* OPEC production is negative and different than zero across nearly all horizons—the change in non-OPEC oil production is indistinguishable from zero.

[17] Under Stock and Watson (2018), these are identified, since the IRFs are identified and the system is, by construction, invertible.

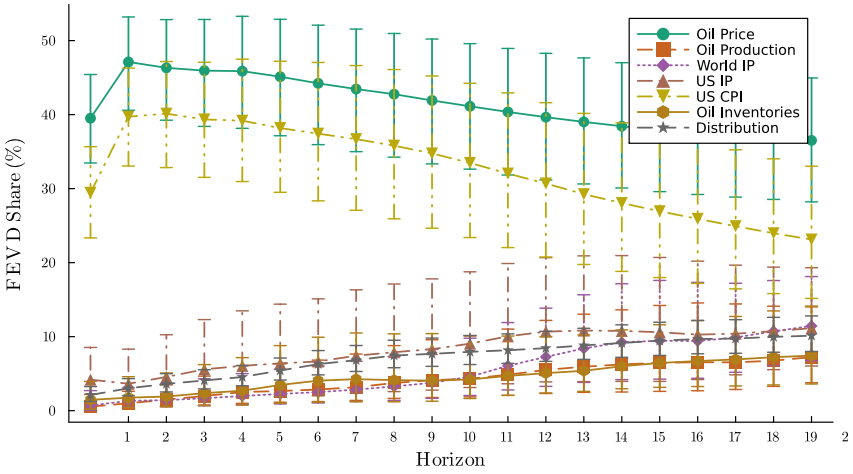
[18] This result is also robust to using the measure of Kilian (2009). The result is stored and available upon request.

Main finding. Expectational shifts from oil supply news are markedly more inflationary than realized supply disruptions *despite* the similar oil-price magnitude in Figure 1. The KZ shock explains roughly two to three times the CPI variance share than the BH shock does, even though the two shocks explain roughly comparable shares of oil-price variance (40% and 30%, respectively, in the medium run). The asymmetry lies in the price-pass-through channel and not in oil-price magnitude. Under BH, the contribution to oil prices and inflation is smaller and hump-shaped, implying a build up. Under KZ, they jump on impact and gradually decline. Regarding the distribution, both shocks drive distributional dynamics similarly, with their FEVD share growing to about 10% over the horizon, consistent with a feedback role for the distribution, which I further example through a within-model decomposition in the next section, Section 5.3. To my knowledge, this is the first FEVD of identified oil supply shocks for distributional aggregates.

Figure 2: Forecast error variance decomposition (FEVD): oil supply shocks



(a) Realized oil supply shock (Baumeister and Hamilton, 2019)



(b) Oil supply news shock (Känzig, 2021)

Notes: Share of forecast error variance explained by the identified shocks for each variable in the VAR, including the distributional block. Markers: posterior median. Error bars: 68% credible sets. Horizon in quarters.

5.3 What Drives Aggregate Responses?

The aggregate responses of CPI to both shocks are quantitatively similar; however they are the net effect of several channels and from solely the IRFs, it is not clear what drives the CPI response for each shock. To understand the underlying mechanisms, I apply the generalized impulse response (GIR) framework of Dufour and Renault (1998) and the causal mediation interpretation of Dufour and Wang (2024).[19] This methodology decomposes each impulse response into variable-level contributions, revealing not only *which* channels drive the response but also *when* each channel's influence materializes and how large this contribution is.[20,21]

The Dufour and Wang (2024) decomposition. The key insight is that the total impulse response $\Theta_{i,h}$ of variable i at horizon h can be written as a sum of *channel contributions*:

$$\Theta_{i,h} = \sum_{m=1}^{n_y} C^{(h)}_{m\to i}, \qquad C^{(h)}_{m\to i} = \sum_{n=0}^{h} c^{(n,h)}_{m\to i}, \tag{23}$$

where $C^{(h)}_{m\to i}$ is the total contribution of variable m to the response (the IRF) of variable i at horizon h. The impulse response above would then be the sum of $C^{(h)}_{m\to i}$ for all n_y variables, each a possible mediator/channel m. To compute the contribution of a specific channel, $C^{(h)}_{m\to i}$, requires estimating $c^{(n,h)}_{m\to i}$, which is the contribution made at some intermediate date n, for each n until we reach horizon h. One can then interpret (n, h) in superscript as the contribution from n to h. This contribution $c^{(n,h)}_{m\to i}$ can be decomposed into a product of two forces:

$$c^{(n,h)}_{m\to i} = \sum_{\ell=1}^{\min(n+1,\,p)} \underbrace{[\varphi_\ell^{(h-n)}]_{i,m}}_{\substack{\text{Propagation:}\\ \text{how } m \text{ at lag } \ell\\ \text{reaches } i \text{ over}\\ h-n \text{ periods}}} \times \underbrace{\Theta_{m,\,n+1-\ell}}_{\substack{\text{Activation:}\\ \text{how strongly}\\ \text{the shock moves}\\ m \text{ at horizon } n+1-\ell}} \tag{24}$$

For intuition on Equation 24, I provide the first two terms of $C^{(h)}_{m\to i}$, complemented with Figure 3, to better understand what exactly is being measured. The first term is $c^{(0,h)}_{m\to i}$, which corresponds to $n = 0$, the date of the shock's impact. The term quantifies the contribution of channel m that occurrs at $n = 0$ (if activated) and how this channel's response $\Theta_{m,0}$ propagates from $n = 0$ to $n = h$. So, the shock hits at $n = 0$, moves m on impact, m then propagates to

[19] This is *not* to be confused with the generalized impulse responses of Pesaran and Shin (1998), which only captures the causal effect of a structural shock, possibly in a nonlinear setting, but does not capture the *mediation* effects (the channels), the object of interest here.

[20] An important disclaimer: this is not a counterfactual exercise and more a descriptive accounting exercise i.e., an ex-post attribution of the responses to different channels, given the DGP; however, the decomposition is an important empirical step to identify an active channel which can in turn be used to discipline a structural model that then microfounds that channel. One can then do actual policy counterfactuals in the structural model.

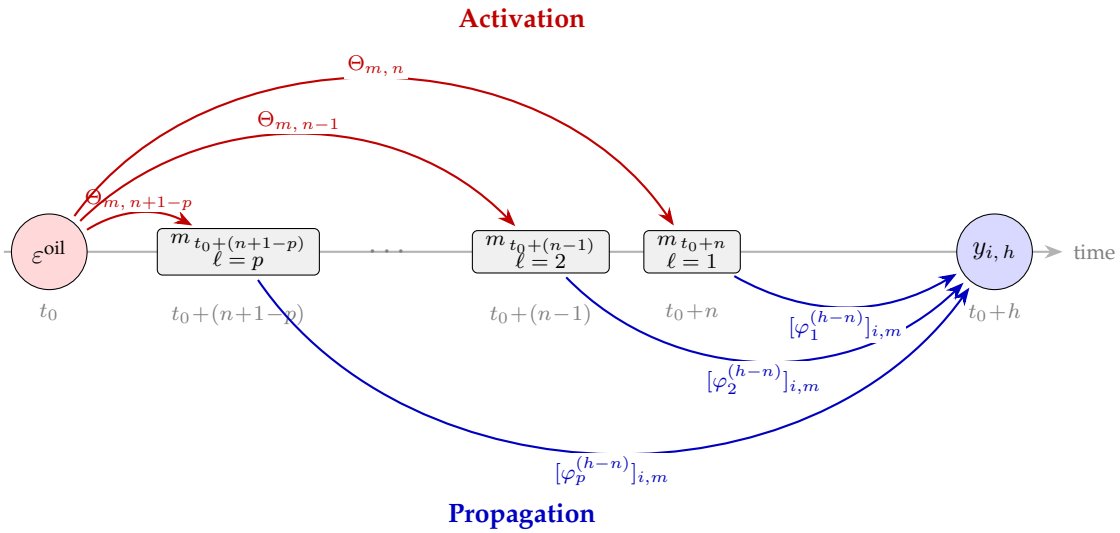
[21] The decomposition is order-invariant as the moments used to identify the contributions are also order-invariant.

variable i over h periods. For $n = 1$, we only need to consider the propagation from $n = 1$ to $n = h$, hence the $(h - 1)$, where the superscript on the φ terms can be interpreted as the length of propagation. Then it's about measuring how the impact $n = 0$ response contributed to this in-between horizon of $n = 1$ and $n = h$ and how the $n = 1$ response contributed to the same in-between horizon. The remaining terms for $n = 2$ to $n = h$ are assessed analogously.

$$c_{m\to i}^{(0,h)} = \underbrace{[\varphi_1^{(h)}]_{i,m}}_{\text{propagation over } h \text{ periods}} \times \underbrace{\Theta_{m,0}}_{\text{impact response of } m} \tag{25}$$

$$c_{m\to i}^{(1,h)} = \underbrace{[\varphi_1^{(h-1)}]_{i,m}\,\Theta_{m,1}}_{m \text{ at } n=1 \text{ propagated over } h-1 \text{ periods}} + \underbrace{[\varphi_2^{(h-1)}]_{i,m}\,\Theta_{m,0}}_{m \text{ at } n=0 \text{ entering as a 2-period lag}} . \tag{26}$$

Figure 3: Anatomy of a single channel contribution $c_{m\to i}^{(n,h)}$



Notes: Visualization of equation (24) for a single mediating variable m, evaluation horizon n, and outcome horizon h. The oil supply shock ε^{oil} fires at time t_0 and *activates* the channel variable m at each of the time points $t_0 + (n + 1 - \ell)$ for $\ell = 1, \ldots, \min(n + 1, p)$ (red arcs above the time axis). Each activated value of m then *propagates* forward to outcome i at time $t_0 + h$ (blue arcs below). Arc length is drawn proportional to time-distance: more-recent m values have long activation arcs and short propagation arcs, older m values the reverse, with shock-to-outcome spans summing to h in every case. The $n = 0$ case in equation (25) corresponds to keeping only the rightmost arc pair ($\ell = 1$); the $n = 1$ case in equation (26) adds the next pair ($\ell = 2$), and so on up to $\min(n + 1, p)$.

Measuring propagation. $\varphi_\ell^{(h)}$ measures the amount of propagation over h periods and acts as a dynamic multiplier that scales the activation into the outcome IRF. The subscript ℓ identifies which lag of channel m is being propagated, so at each intermediate date n, I have the total contribution from m aggregates $\varphi_\ell^{(h-n)}$ across each of its active lags $\ell = 1, \ldots, \min(n+1,\, p)$. These are the GIR coefficients of Dufour and Renault (1998), computed via the recursion $\varphi_\ell^{(h+1)} = \varphi_{\ell+1}^{(h)} + \Psi_h A_\ell$ with initial conditions $\varphi_\ell^{(1)} = A_\ell$. Element $[\varphi_\ell^{(h)}]_{i,m}$ captures how a perturbation

in variable m at lag ℓ propagates to variable i over h periods, accounting for all intermediate feedback through every other variable in the system.

The decomposition allows to examine many questions such as the role of inflation on the transmission of such shocks; the role of financial conditions; or the role of the distributional in shaping the aggregate responses themselves; all within a single model.

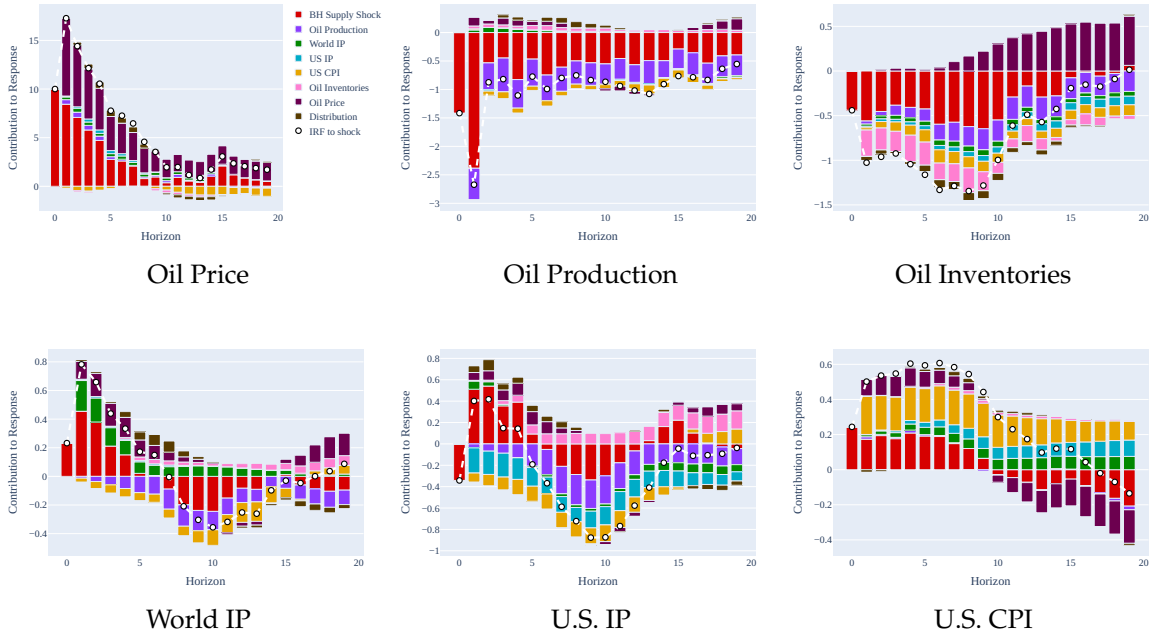
Why not two VARs? ? I investigate the impact of oil supply shocks through the lens of a *single* model, as opposed to running two VARs, where one is some baseline system and the other removes some channel of interest, keeping everything else the same. A naive econometrician would attribute the difference in aggregate impulse responses to the omitted block / the channel of interest; but the two-VAR gap conflates omitted-variable bias in the smaller model, genuine mediation through the distribution, and identification drift (impact vector not invariant to information set)—which cannot be separated in a two-VAR comparison. The GIR decomposition I adopt in this section is immune to all three issues: it fixes the model, fixes the identification, and analytically attributes the identified impulse responses above to each mediator through the Wold representation (and not through omission). This setup permits to make quantitative statements like "channel m accounts for $X\%$ of the macro response"—not possible with two VARs.

Results. Figure 4 presents the channel decomposition for the aggregate responses for both shocks. The stacked bars show the posterior median contribution of each channel at each horizon, where each channel is identified by a separate color and the color scheme is identical across panels, so the meaning of purple for example is the same across panels. The dashed white line and outlined white circles correspond to the median point estimates from Figure 1.

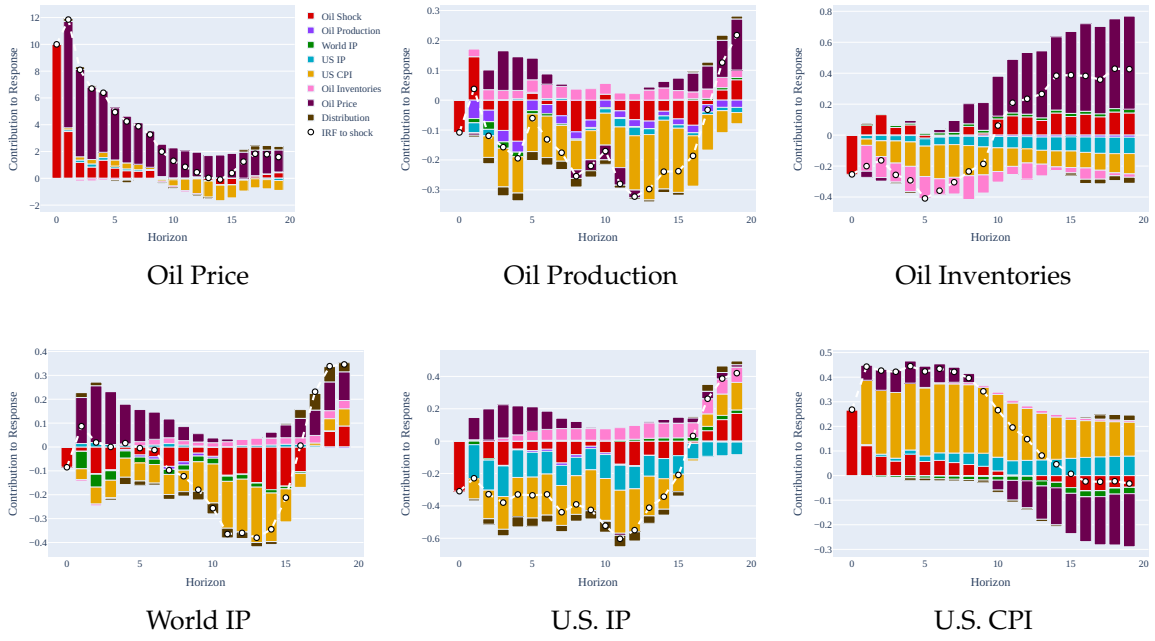
Before diving in on each specific panel, a bird's-eye view reveals which channels dominate. Under BH realized disruptions, red and purple — the shock itself and oil production — are the biggest contributors to the aggregate dynamics. Under KZ supply news, the picture shifts: dark purple and mostly yellow dominate, that is, oil price and CPI—similar to the FEVD picture. The decomposition makes the inflation *stickiness* of the news shock concrete: once the announcement has repriced both oil and consumer prices on impact, those two channels propagate themselves forward and account for the bulk of the aggregate response for the rest of the horizon, with the shock itself and oil production contributing essentially nothing beyond the first quarter. This self-propagating-price pattern is absent under BH, where the shock and production channels remain the dominant carriers throughout. I confirm Kilian (2009)'s conclusion that not all oil price shocks are alike: The two oil supply shocks differ in nature and thus motivate different macroeconomic models and different policy.

Figure 4: Channel Decomposition of Oil Supply Shocks

Panel A: Baumeister & Hamilton (2019) realized supply shock



Panel B: Känzig (2021) supply news shock



Notes: Generalized impulse response decomposition (Dufour and Wang, 2024) of the macro impulse response into contributions from the oil supply shock, aggregate macro channels, and the distributional channel. Stacked bars: posterior median contribution of each channel at each horizon (color scheme identical across panels). Dashed white line and outlined white circles: median point estimates from Figure 1. Panel A: Baumeister and Hamilton (2019) realized supply shock; Panel B: Känzig (2021) oil supply news shock.

The oil inventories plot in Panel A and Panel B shows the distinguishing features of the KZ shock. On impact, the shock — whether BH or KZ — is responsible for the entire response, reflected by the full red bar at $h = 0$ with no other channels. This is mechanical (the shock is ordered first), so it can be seen as an implementation check. From $h = 1$, new colors appear: the shock at $h = 0$ activated channels at $h = 1$ that will now propagate until $h = 20$.

From this point, the key similarities and differences emerge. The shocks are similar in the contribution of oil inventories and oil prices. In both cases, inventories exert downward pressure on themselves, which is in line with oil inventories being rigid, while the oil price channel becomes increasingly important after roughly two years and then propagates strongly. The magnitude of this price mechanism, however, differs across shocks. It is more muted under BH and more pronounced under KZ. Why? Under KZ, oil production (the IRF) is relatively unaffected, allowing supply to accommodate the increased demand associated with lower oil prices. Under BH, by contrast, production shortfalls are large and persistent, limiting this adjustment. This difference is visible in the decomposition: the production channel (purple) plays essentially no role in the inventories response under KZ, but is clearly present under BH.

The shocks' channel themselves (red) will differ in their implications for inventories, but also generally. The shocks contribution is larger and persistent under BH and much smaller under KZ, but operate in the same direction for all aggregates except one: oil inventories. The BH shock exerts a sizeable *downward* pressure on oil inventories throughout the horizon. In contrast, the KZ shock—beyond the impact response—generates *upward* pressure on inventories, particularly after two years. This pattern supports the interpretation of the KZ shock as directly driving inventory accumulation.

Examining the role of the U.S. distribution for the global oil market system reveals that it is a small one. The response of U.S. CPI is not driven by the distribution, and only a small fraction of U.S. IP is. With only two domestic variables — U.S. IP and U.S. CPI — there is also little scope for the U.S. distribution to matter, since the contribution of each channel will depend on the scope of relevant channels available. Appendix J extends the macro block to more domestic aggregates, where the distribution's contribution is larger (but still minor) and has the common feature of becoming active in the medium-run and as an amplifier.

6 The Distributional Effects of Oil Supply Shocks

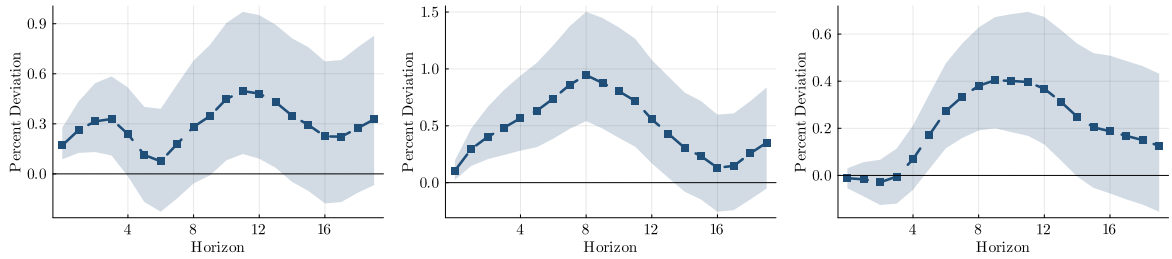
This section provides insights on how the two different shocks affect inequality and household-level consumption responses. As before, I compare the Baumeister and Hamilton (2019) realized oil supply shocks with the Känzig (2021) oil supply news shock, to distinguish the two economic channels: sudden physical disruptions versus expectational shifts in oil supply.

6.1 Aggregate Measures of Inequality

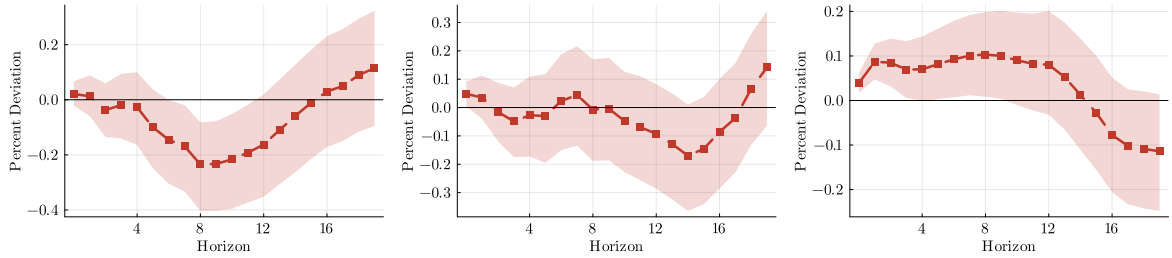
I examine how oil supply shocks affect inequality across the three dimensions of the joint distribution: consumption, income, and wealth. Figure 5 reports the accumulated stationary impulse responses of the Gini coefficient.[22] Figures 6, 7, and 8 report shares held by different groups. Appendix K reports the log 90/10 and 90/50 quantile ratios, and the standard deviation of logs for each dimension.

Figure 5: Inequality responses to oil supply shocks

Panel A: Baumeister & Hamilton (2019) realized supply shock



Panel B: Känzig (2021) supply news shock



Notes: Impulse responses of the Gini coefficient for income, consumption, and wealth. Units: percentage point deviation from steady state. Solid lines: posterior median. Shaded areas: 68% credible sets.

Ginis. Under the BH shock, all three Gini coefficients increase, but with distinct short-run dynamics. Consumption and income inequality rise on impact, whereas wealth inequality responds with a delay. In the medium-run, after year two, all Ginis elevate and peak before waning off in the longer horizon. These patterns for income and wealth align with Lei, Ludwig, and Ma (2025), while the consumption response is novel. Under the KZ shock, a different story emerges: consumption and income inequality are rather flat on impact and more so for income. Whereas a physical disruption widens inequality in the medium-run, the anticipation

[22] For all three dimensions I use the bounded Gini normalization of Raffinetti, Siletti, and Vernizzi (2015), which replaces the classical denominator $2(N-1)\bar{x}$ with $2(N-1)(T^+ + T^-)$, where T^+ and T^- are the weighted totals of positive and absolute-negative values. This guarantees $G \in [0, 1]$ even when some observations are negative—relevant for wealth, where leveraged households can carry negative net worth—and collapses to the standard Gini when all values are non-negative.

channel appears to compress it at medium horizons before recovering toward the steady state by the end of the horizon.[23] Based on results I show later (Figure 10), I find that this is because households in the upper half of the wealth distribution reduce consumption preemptively while transfer-dependent households at the bottom are initially shielded (Figure 43). This is consistent with households who participate in financial markets being more forward-looking and responsive to news. The wealth Gini is the only dimension where both shocks agree, increasing under both identifications; but still, their profiles are different. Under KZ, wealth inequality rises mildly and has a flat profile; under BH it is rather hump-shaped. Figures 6, 7, and 8 decompose the Gini dynamics into the underlying group shares and reveal what is moving them.

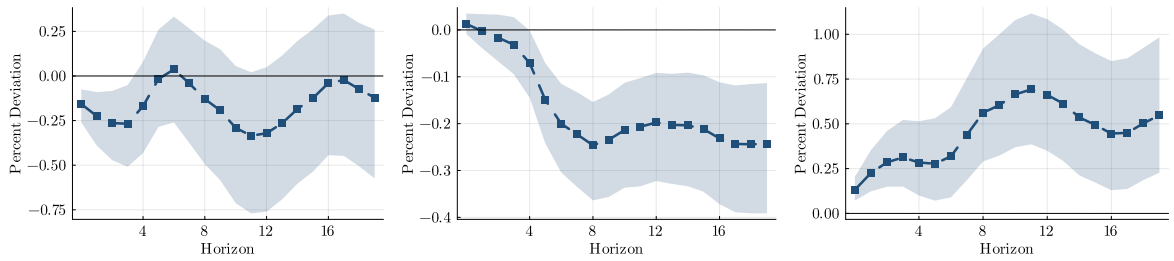
Consumption shares. Under the BH shock, the Bottom 50% share falls on impact and remains negative for three quarters, while the Top 10% rises by a comparable amount—consistent with the standard oil-redistribution channel (e.g., Lei, Ludwig, and Ma, 2025). After one year, however, the pattern shifts: the increase in the Gini is driven primarily by a persistent contraction in the Middle 50%-90% share, which bottoms out around quarter eight and dominates at medium horizons, regardless of the oil supply shock. Although the initial decline in the Bottom 50% is larger, its response is volatile, likely reflecting labor supply adjustments (intensive and extensive margins, Figure 37), transfer timing (Figure 43), and non-homothetic consumption (e.g., higher energy shares among low-income households; Edelstein and Kilian, 2009). Under the KZ shock, redistribution also occurs but with a delay and in the opposite direction: the Bottom 50% share rises relative to the Middle 50%-90%, rather than a contemporaneous increase in the Top 10% as under BH.

Income shares. Figure 7 shows the income-share decomposition. Under the realized BH shock, the Top 10% income share response is strictly monotonic, peaking at 1.5% at quarter 8, possibly coming from capital and corporate-profit gains (Figure 45). The shares of the Bottom 50% and Middle 50%-90% respond equally strongly, but in the opposite direction, with the Bottom 50% responding two times stronger than the Middle 50%-90%. Under the KZ shock, the pattern is different: changes in income inequality only materialize in the medium-run, remaining basically flat for most of the time horizon. An interesting finding however is that from quarter eight onward, the shape of the IRFs is similar—both shocks produce a hockey-shaped trajectory in each share, just at opposite levels (BH peaks positive, KZ troughs negative)—and it is mostly in the short-run that they meaningfully differ.

[23] Although not explicitly documented, this decrease in income inequality can be found, to some extent, in Figure A.3 of Lei, Ludwig, and Ma (2025).

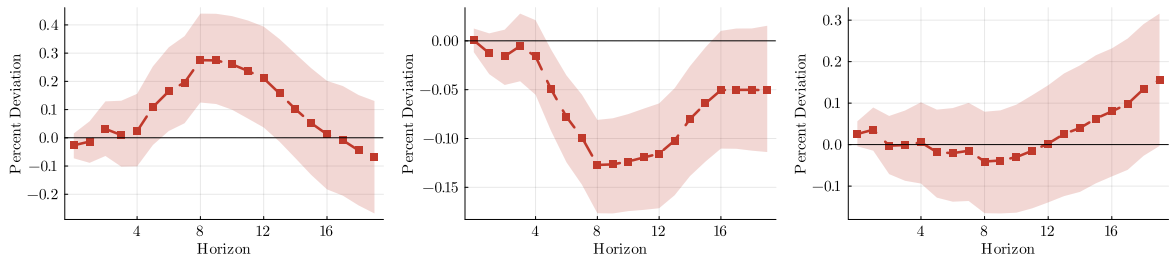
Figure 6: Consumption share responses to oil supply shocks

Panel A: Baumeister & Hamilton (2019) realized supply shock



Bottom 50% Middle 50%-90% Top 10%

Panel B: Känzig (2021) supply news shock

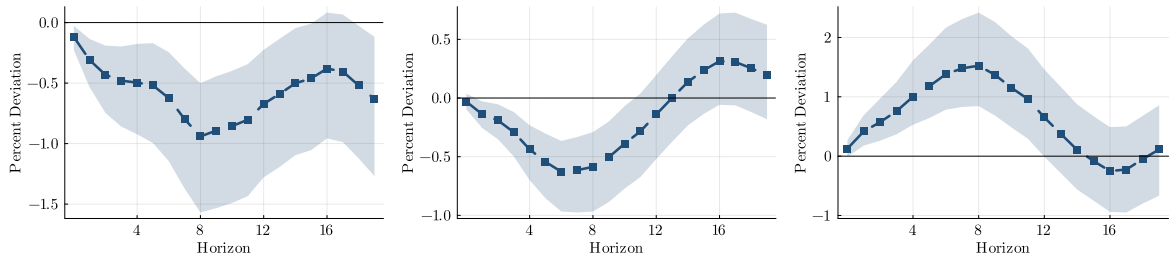


Bottom 50% Middle 50%-90% Top 10%

Notes: Impulse responses of the consumption share held by the Bottom 50% (0-50th percentile), Middle 50%-90% (50th-90th), and Top 10% (90th-100th) of the consumption distribution. Units: percentage point deviation from steady state. Solid lines: posterior median. Shaded areas: 68% credible sets.

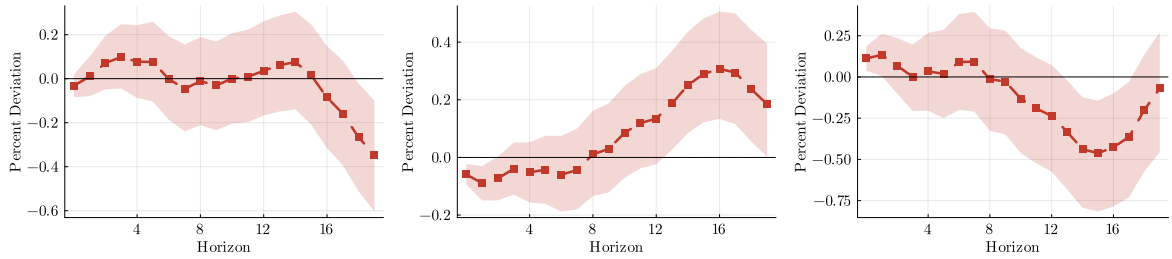
Figure 7: Income share responses to oil supply shocks

Panel A: Baumeister & Hamilton (2019) realized supply shock



Bottom 50% Middle 50%-90% Top 10%

Panel B: Känzig (2021) supply news shock



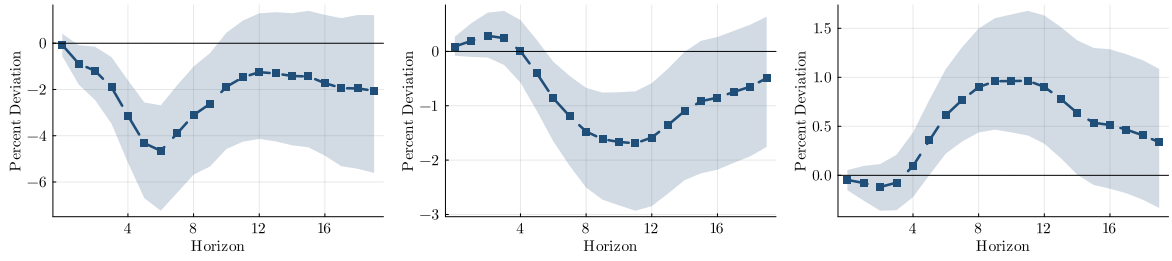
Bottom 50% Middle 50%-90% Top 10%

Notes: Impulse responses of the income share held by the Bottom 50%, Middle 50%-90%, and Top 10% of the income distribution. Units: percentage point deviation from steady state. Solid lines: posterior median. Shaded areas: 68% credible sets.

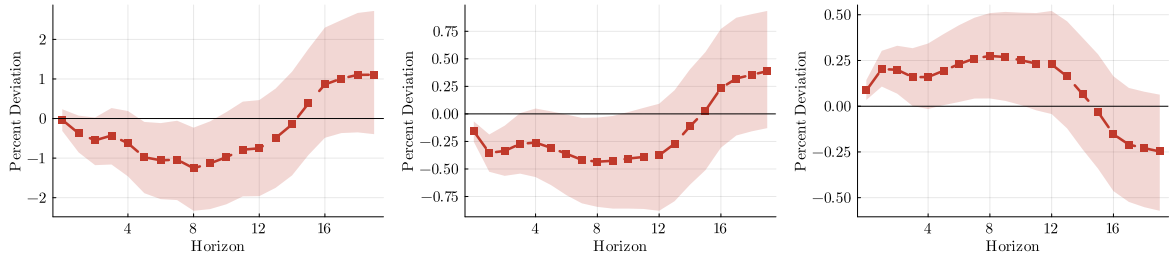
Wealth shares. Figure 8 shows the wealth-share decomposition. Under both identifications, I see a similar pattern: the Top 10% wealth share rises and the other groups fall. This is interesting because it suggests that the two shocks, in this dimension, possibly share the same underlying mechanism. Still, these responses will differ in their magnitudes and timing. Under BH, the Bottom 50% deplete their savings strongly within the first year, most likely attempting to curb the hike in energy prices. The Middle 50%-90% as well, but with a delay. The Top 10% also reacts with a delay, but ultimately peaks at 1%. All these effects are muted under a news shock and linear, suggesting that the news shock interacts with fewer channels, but these channels are persistent and correlated with the other variables in the system. Figure 4 shows precisely this.

Figure 8: Wealth share responses to oil supply shocks

Panel A: Baumeister & Hamilton (2019) realized supply shock



Panel B: Känzig (2021) supply news shock



Notes: Impulse responses of the wealth share held by the Bottom 50%, Middle 50%-90%, and Top 10% of the wealth distribution. Units: percentage point deviation from steady state. Solid lines: posterior median. Shaded areas: 68% credible sets.

6.2 Household-Level Responses

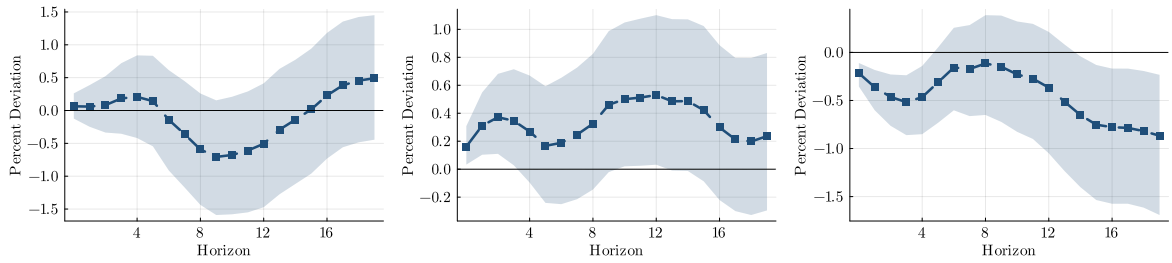
This section zooms in on consumption and provides insights on the sources of consumption inequality by showing household-level consumption impulse responses along the income and wealth distribution. This analysis uncovers heterogeneity that is hidden in the aggregate—offering a more nuanced and informative picture than the inequality measures alone. Figure 9 shows consumption by income and Figure 10 shows consumption by wealth, each for three

groups: Bottom 50%, Middle 50%-90%, and Top 10%. The section concludes with the full 5×5 income–wealth grid (Figures 11 and 12), which reports consumption responses for each (income, wealth) cell.

Consumption responses by income. Figure 9 reports the consumption impulse responses for households grouped into Bottom 50%, Middle 50%-90%, and Top 10% income groups.[24] Under the BH shock, the response of Bottom 50% income households is relatively flat the first five quarters before declining sharply to about -0.5% around quarter eight. Middle 50%-90% income households show a different, rather non-monotonic path, but generally increasing consumption over the time horizon. Top 10% income households display the opposite pattern of Middle 50%-90% income households: an immediate, but small decline that gradually zeros after the first year, but has a medium-run negative impact. Under the KZ news shock, there is again this flatness across all groups in the short-run, but responses then begin to mimic the BH shock.

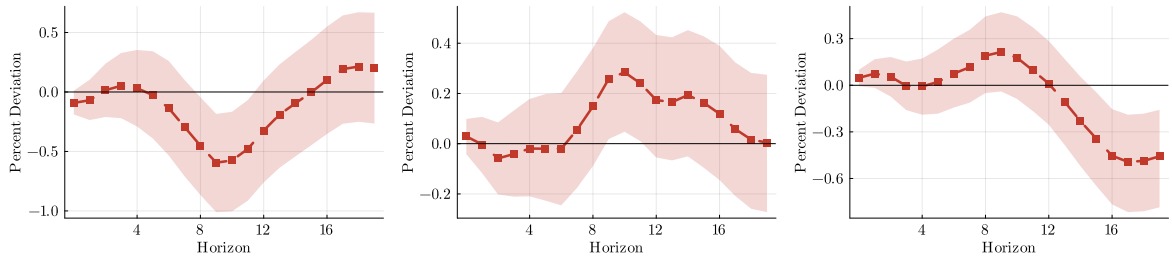
Figure 9: Consumption responses by position in the income distribution

Panel A: Baumeister & Hamilton (2019) realized supply shock



Bottom 50% Income | Middle 50%-90% Income | Top 10% Income

Panel B: Känzig (2021) supply news shock



Bottom 50% Income | Middle 50%-90% Income | Top 10% Income

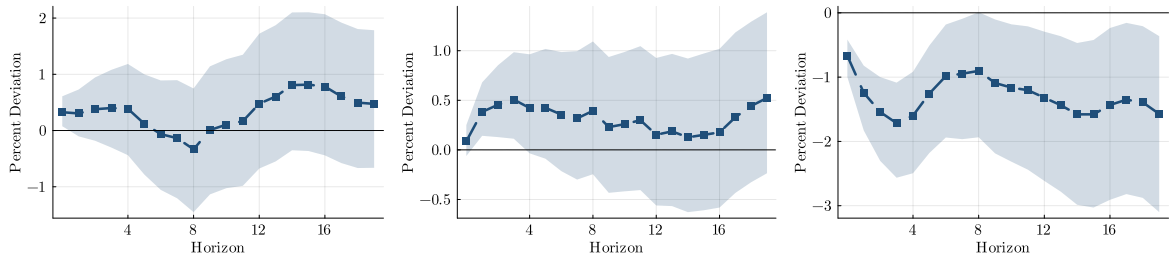
Notes: Impulse responses of mean consumption for households grouped into Bottom 50% (0–50%), Middle 50%-90% (50–90%), and Top 10% (90–100%) of the marginal income distribution. Solid lines: posterior median. Shaded areas: 68% credible sets. Horizon in quarters.

[24] Throughout this section I use five overlapping marginal cuts of the income (and analogously the wealth) distribution: Bottom 20% (0–20%), Bottom 50% (0–50%), Middle 50%-90% (50–90%), Top 20% (80–100%), and Top 10% (90–100%). The tail cuts (Bottom 20%, Top 10%) are nested inside the broader cuts (Bottom 50%, Top 20%) and capture the extremes of the marginal distribution; the remaining three span the bulk of the population. The body figures use the three broad cuts (Bottom 50%, Middle 50%-90%, Top 10%); the joint-distribution figures and the appendix tail cuts use all five.

Consumption responses by wealth. Figure 10 repeats the exercise along the wealth dimension. Under BH, Bottom 50% wealth households respond similar to Bottom 50% income households, but with larger uncertainty. The Middle 50%-90% group again increases consumption on impact, but uncertainty widens in the medium-run. The Top 10% wealth group decreases consumption by nearly 1% in the first year—a sharp result. These households do not return to steady state in the medium-run. Under KZ, Bottom 50% wealth households increase consumption in the short-run and in the medium-run. Middle 50%-90% wealth and Top 10% wealth households lose in the medium-run as in BH, but not in the short-run. This indicates the wealth channel is similar at the top, but different for the Bottom 90%. Appendix L extends both the income and wealth views to the tails of each margin (Bottom 20% and Top 10%), where the magnitudes are sharper.

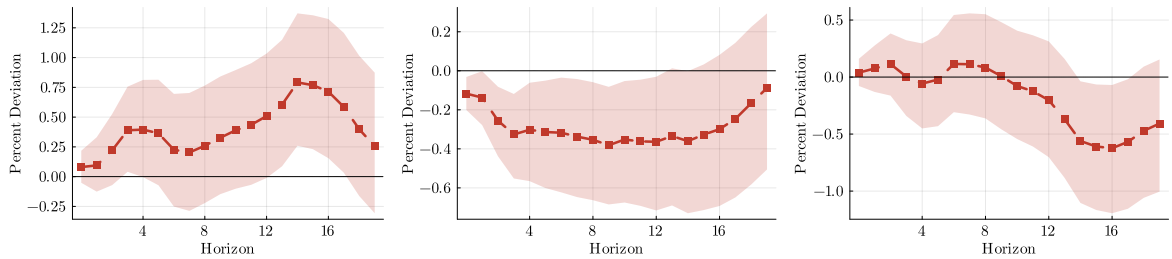
Figure 10: Consumption responses by position in the wealth distribution

Panel A: Baumeister & Hamilton (2019) realized supply shock



Bottom 50% Wealth | Middle 50%-90% Wealth | Top 10% Wealth

Panel B: Känzig (2021) supply news shock



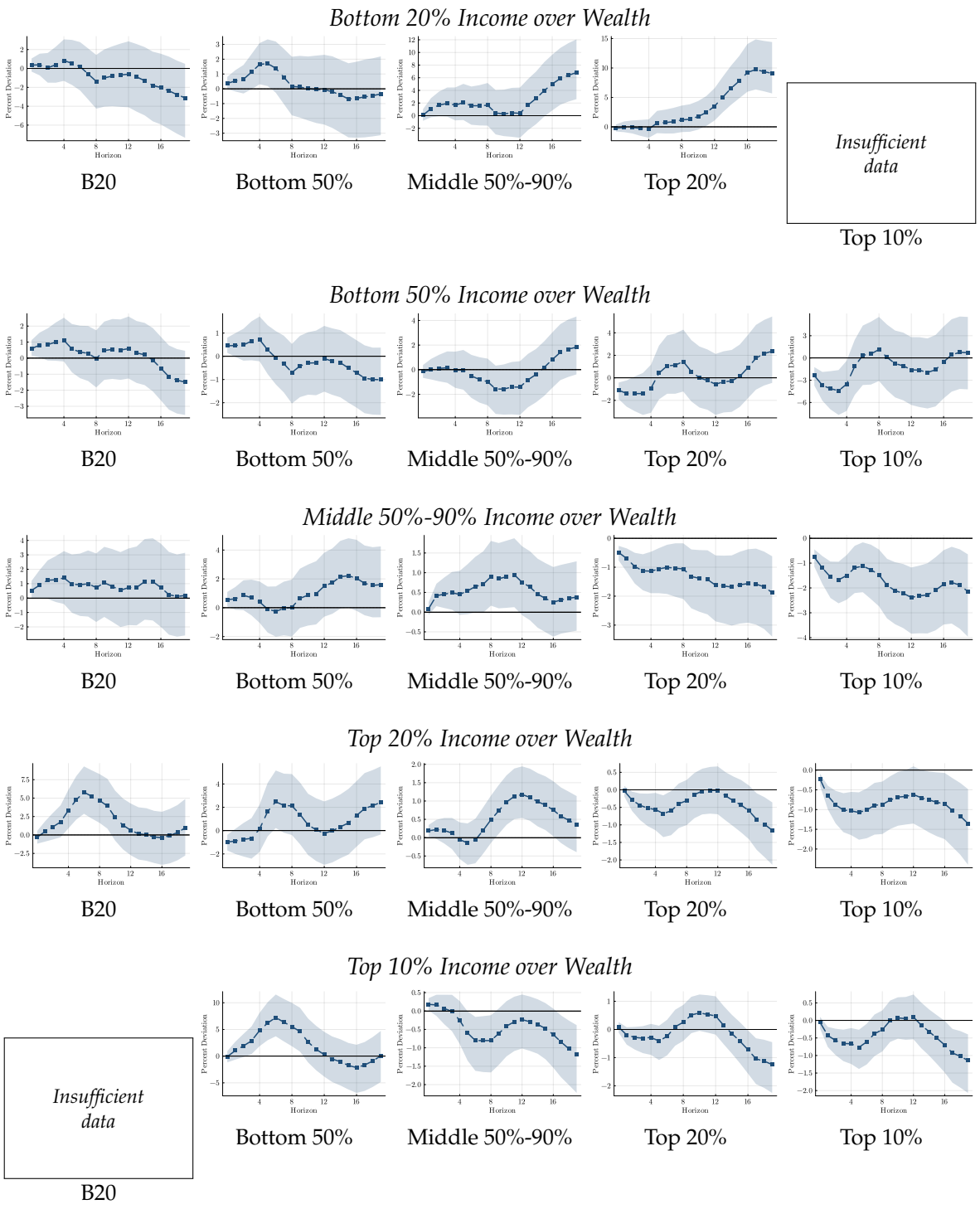
Bottom 50% Wealth | Middle 50%-90% Wealth | Top 10% Wealth

Notes: Impulse responses of mean consumption for households grouped into Bottom 50% (0–50%), Middle 50%-90% (50–90%), and Top 10% (90–100%) of the marginal wealth distribution. Solid lines: posterior median. Shaded areas: 68% credible sets. Horizon in quarters.

Consumption responses by income and wealth. Figures 11 and 12 report consumption impulse responses across the full 5×5 income–wealth grid for the BH and KZ shocks, respectively. Each row corresponds to an income group, similar to before, with two new groups: Bottom 20% and Top 20%; columns are the analogous wealth groups.[25]

[25] Two cells (Bottom 20% income/Top 10% wealth and Top 10% income/Bottom 20% wealth) are absent because the underlying micro data does not find a correlation between these groups.

Figure 11: Consumption responses by joint income–wealth cell — BH realized supply shock



Notes: Consumption impulse responses for households in each (Income, Wealth) cell of the joint distribution; Baumeister and Hamilton (2019) realized supply shock. Rows correspond to income groups (Bottom 20%, Bottom 50%, Middle 50%-90%, Top 20%, Top 10%; overlapping ranges, see footnote on first introduction); columns within each row correspond to the analogous wealth groups. Cells absent due to insufficient observations: Bottom 20% income/Top 10% wealth and Top 10% income/Bottom 20% wealth. Solid lines: posterior median. Shaded areas: 68% credible sets. Horizon in quarters.

Figure 12: Consumption responses by joint income–wealth cell — KZ oil supply news shock

Notes: Consumption impulse responses for households in each (Income, Wealth) cell of the joint distribution; Känzig (2021) oil supply news shock. Rows correspond to income groups (Bottom 20%, Bottom 50%, Middle 50%-90%, Top 20%, Top 10%; overlapping ranges, see footnote on first introduction); columns within each row correspond to the analogous wealth groups. Cells absent due to insufficient observations: Bottom 20% income/Top 10% wealth and Top 10% income/Bottom 20% wealth. Solid lines: posterior median. Shaded areas: 68% credible sets. Horizon in quarters.

Reading the BH grid (Figure 11) by blocks, four patterns stand out. (i) The *asset-rich, income-poor* cells in the upper-right 2×3 block (Bottom 20%, Bottom 50% income) $\times$ (Middle 50%-90%, Top 20%, Top 10% wealth) increase consumption in the medium run, often substantially: the wealth buffer at the bottom of the income distribution absorbs the recessionary impulse, and households in this block are the only ones that systematically *benefit* from the realized supply shock at intermediate horizons. (ii) The opposite block — the bottom right 3×3 (Middle 50%-90%, Top 20%, Top 10% income) $\times$ (Middle 50%-90%, Top 20%, Top 10% wealth) — contracts uniformly, with two exceptions: (Middle 50%-90% income, Middle 50%-90% wealth) and (Top 20% income, Middle 50%-90% wealth). (iii) The 2×2 of (Bottom 20%, Bottom 50% income) $\times$ (Bottom 20%, Bottom 50% wealth) display a small short-run rise followed by a medium-run decline — consistent with transfer-indexed nominal income protecting the bottom in the short run before the recessionary contraction binds—though the posterior contains zero. (iv) The (Middle 50%-90%, Top 20%, Top 10% income) $\times$ (Bottom 20%, Bottom 50% wealth) do not contract; they rise on impact and revert.

The KZ grid, Figure 12, preserves the upper-right "wealth-buffer wins" block (i) and block (iii), but block (ii) reads differently. Under KZ, Middle 50%-90% income households split: at Bottom 20% or Bottom 50% wealth they increase consumption gradually over the entire horizon, while at Middle 50%-90%, Top 20%, or Top 10% wealth they decrease in the short or medium run — a pattern consistent with asset-holders pricing the announcement and non-asset holders not. Top 20% income households at the bottom of the wealth distribution increase consumption in the medium run; at the top of the wealth distribution they rise marginally on impact before fading along zero. Top 10% income / Top 10% wealth households are the cleanest news responders: a stronger impact consumption increase consistent with capital-income gains as oil-related expectations reprice, slowly waning to zero. The contrast with BH is informative — realized supply contractions punish the asset-rich, income-rich; news shocks are priced in opposite directions on impact depending on wealth, with financial-market participants reacting first.

7 Conclusion

This chapter has documented that oil supply shocks have first-order effects on the joint distribution of U.S. household income, consumption, and wealth, and that the distributional incidence depends qualitatively on the type of shock identified. Realized supply disruptions and supply news shocks move many of the same aggregate variables in similar directions but redistribute in opposite directions: the Baumeister and Hamilton (2019) realized shock raises consumption inequality through a cost-pull channel that punishes asset-rich, income-rich house-

holds after a one-year lag, while the Känzig (2021) news shock compresses consumption inequality on impact through an expectations channel that operates first on financial-market participants. Across robustness checks—alternative oil-price measures, sub-sample windows, lag lengths, financial-control sets, and posterior central tendencies—this wedge survives in sign and shape. The analysis thus shares the view of Kilian and Lewis (2011) that policy responses should depend on the underlying causes of oil price shocks.

Beyond the substantive findings, the chapter's methodological contribution is the combination of the asymmetric conjugate Bayesian VAR of Chan (2022) with the within-model causal-mediation decomposition of Dufour and Wang (2024), applied to a system that embeds the joint income–consumption–wealth distribution of Bayer, Calderon, and Kuhn (2025)—a template that other applied work on shock transmission to heterogeneous households can adopt.

Three implications stand out. First, structural identification matters not only for the aggregate response, as Kilian (2009) and Känzig (2021) emphasize, but also for the redistributive incidence—different parts of the same OPEC decision (announcement versus realization) hit different parts of the distribution. Second, the joint distribution is not a passive add-on: dropping it from the VAR changes the aggregate IRFs themselves—most starkly, the monetary-policy response flips sign—so studying oil shocks at the aggregate level alone leads, in this dataset, to an omitted-variable problem. This is the empirical analogue of the demand-side-information argument of Kilian and Lewis (2011): heterogeneity in expenditure shares and household balance sheets carries demand-side variation that the small aggregate-only oil VARs of the literature systematically omit. Third, the wealth dimension is the one place where the two shocks agree directionally, suggesting that wealth inequality is shaped by oil shocks through balance-sheet revaluations that operate regardless of identification, while consumption and income inequality are shaped by the channel through which the shock reaches the household.

Several extensions come to mind that would complement the results presented here. First, the sufficient-statistics counterfactual framework of Caravello, McKay, and Wolf (2024) can be applied to the impulse responses to ask how the distributional incidence would look under alternative monetary regimes—a fixed-rate peg, strict inflation targeting, or a Taylor rule that internalizes distributional concerns—turning the present empirical wedge into a normative comparison of policy rules, building on Broer, Kramer, and Mitman (2025)'s application of the same framework to German labor-market microdata but with the joint income–consumption–wealth distribution as the object of incidence. Second, the mechanism documented here can perhaps be modeled in a heterogeneous-agent New Keynesian model with a realistic oil-market structure, replicating the flip inequality result I find, using also the distributional IRFs reported here as targets. Third, applying the framework to other supply-side

shocks should clarify whether the "physical-versus-informational" wedge is specific to oil or general to commodity markets where announcement effects coexist with realized supply changes.

A final caveat concerns non-linearities. My baseline is a linear Bayesian VAR, and the muted headline-inflation pass-through under the BH realized shock is a statement about the sample-average shock size. Baumeister et al. (2025) document substantial size and sign asymmetries in oil-shock transmission using Bayesian Additive Regression Trees, and earlier non-linear work (e.g., Hooker, 2002) shows that headline pass-through can re-emerge under unusually large or compounding episodes. A state-dependent or non-linear specification with a sufficiently large BH-type impulse—comparable to the 1973–74 or 1979–80 episodes—could plausibly find a more inflationary realized-supply response (and other responses) than the linear average I report.

References

[1] Juan Antolín-Díaz and Paolo Surico. "The Long-Run Effects of Government Spending". In: *American Economic Review* 115.7 (2025), pp. 2376–2413.

[2] Adrien Auclert. "Monetary policy and the redistribution channel". In: *American Economic Review* 109.6 (2019), pp. 2333–2367.

[3] Jushan Bai. "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data". In: *Journal of Econometrics* 122.1 (2004), pp. 137–183.

[4] Christiane Baumeister. "Measuring Market Expectations". In: *Handbook of Economic Expectations* (2023), pp. 413–442.

[5] Christiane Baumeister. *Discussion of" Local Projections or VARs? A Primer for Macroeconomists" by José Luis Montiel Olea, Mikkel Plagborg-Møller, Eric Qian, and Christian K*. Tech. rep. Wolf, Working paper, Prepared for the NBER Macroeconomics Annual, 2025.

[6] Christiane Baumeister and James D. Hamilton. "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks". In: *American Economic Review* 109.5 (2019), pp. 1873–1910.

[7] Christiane Baumeister and Lutz Kilian. "Forty years of oil price fluctuations: Why the price of oil may still surprise us". In: *Journal of Economic Perspectives* 30.1 (2016), pp. 139–160.

[8] Christiane Baumeister et al. *Oil, Inflation Expectations, and Household Characteristics: A Nonlinear Heterogeneous Agent VAR Approach*. Tech. rep. University of Notre Dame, 2025.

[9] Christian Bayer, Benjamin Born, and Ralph Luetticke. "Shocks, frictions, and inequality in US business cycles". In: (2020).

[10] Christian Bayer, Luis Calderon, and Moritz Kuhn. *Distributional Dynamics*. Tech. rep. 2025.

[11] Paul Beaudry and Franck Portier. "Stock prices, news, and economic fluctuations". In: *American economic review* 96.4 (2006), pp. 1293–1307.

[12] Edmond Berisha et al. "Income inequality and oil resources: Panel evidence from the United States". In: *Energy Policy* 159 (2021), p. 112603.

[13] Hilde C Bjørnland, Yoosoon Chang, and Jamie Cross. "Oil and the stock market revisited: A mixed functional var approach". In: *Available at SSRN* (2023).

[14] Tobias Broer, John Kramer, and Kurt Mitman. "The Distributional Effects of Oil Shocks". In: *IMF Economic Review* 73.3 (2025), pp. 851–889.

[15] Martin Bruns and Helmut Lütkepohl. "Comparing external and internal instruments for vector autoregressions". In: *Journal of Economic Dynamics and Control* 177 (2025), p. 105131.

[16] Dario Caldara and Edward Herbst. "Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs". In: *American Economic Journal: Macroeconomics* 11.1 (2019), 157–92.

[17] Guillermo A. Calvo. "Staggered prices in a utility-maximizing framework". In: *Journal of Monetary Economics* 12.3 (1983), pp. 383–398.

[18] Tomás E Caravello, Alisdair McKay, and Christian K Wolf. *Evaluating policy counterfactuals: A var-plus approach*. Tech. rep. National Bureau of Economic Research, 2024.

[19] Joshua C C Chan. "Asymmetric conjugate priors for large Bayesian VARs". In: *Quantitative Economics* 13.3 (2022), pp. 1145–1169.

[20] Minsu Chang, Xiaohong Chen, and Frank Schorfheide. "Heterogeneity and Aggregate Fluctuations". In: *Journal of Political Economy* 132.12 (2024), pp. 4021–4067.

[21] Minsu Chang and Frank Schorfheide. *On the Effects of Monetary Policy Shocks on Income and Consumption Heterogeneity*. Working Paper 32166. National Bureau of Economic Research, 2024.

[22] Gabriel Chodorow-Reich et al. "Homeownership and the Cost of Living". In: *Manuscript, Harvard University* (2025).

[23] Olivier Coibion and Yuriy Gorodnichenko. "Information rigidity and the expectations formation process: A simple framework and new facts". In: *American Economic Review* 105.8 (2015), pp. 2644–2678.

[24] Jeff D. Colgan. "The Emperor Has No Clothes: The Limits of OPEC in the Global Oil Market". In: *International Organization* 68.3 (2014), pp. 599–632.

[25] Ferre De Graeve and Andreas Westermark. *How long is long enough? Long-lag VARs*. Tech. rep. Sveriges Riksbank Working Paper Series, 2025.

[26] Riccardo Degasperi. *Identification of expectational shocks in the oil market using OPEC announcements*. University of Warwick, Department of Economics, 2025.

[27] Felipe N Del Canto et al. *Are inflationary shocks regressive? A feasible set approach*. Tech. rep. National Bureau of Economic Research, 2023.

[28] Thomas Doan, Robert Litterman, and Christopher Sims. "Forecasting and conditional projection using realistic prior distributions". In: *Econometric reviews* 3.1 (1984), pp. 1–100.

[29] Theo Drossidis, Haroon Mumtaz, and Angeliki Theophilopoulou. "The distributional effects of oil supply news shocks". In: *Economics Letters* 240.C (2024).

[30] Jean-Marie Dufour and Eric Renault. "Short run and long run causality in time series: theory". In: *Econometrica* (1998), pp. 1099–1125.

[31] Jean-Marie Dufour and Endong Wang. *Causal mechanism and mediation analysis for macroeconomics dynamics: a bridge of Granger and Sims causality*. Tech. rep. 2024.

[32] Paul Edelstein and Lutz Kilian. "How sensitive are consumer expenditures to retail energy prices?" In: *Journal of Monetary Economics* 56.6 (2009), pp. 766–779.

[33] Stephanie Ettmeier. *No Taxation Without Reallocation: The Distributional Effects of Tax Changes*. CRC TR 224 Discussion Paper Series crctr224_2023_436. University of Bonn and University of Mannheim, Germany, 2023.

[34] Simon Freyaldenhoven. "Factor models with local factors—determining the number of relevant factors". In: *Journal of Econometrics* 229.1 (2022), pp. 80–102.

[35] Yoshito Funashima. "Global Economic Activity Indexes Revisited". In: *Economics Letters* 193 (2020), p. 109269.

[36] Domenico Giannone, Michele Lenza, and Giorgio E Primiceri. "Prior selection for vector autoregressions". In: *Review of Economics and Statistics* 97.2 (2015), pp. 436–451.

[37] Domenico Giannone, Michele Lenza, and Giorgio E Primiceri. "Priors for the long run". In: *Journal of the American Statistical Association* 114.526 (2019), pp. 565–580.

[38] Simon Gilchrist and Egon Zakrajšek. "Credit spreads and business cycle fluctuations". In: *American economic review* 102.4 (2012), pp. 1692–1720.

[39] Oriol González-Casásus and Frank Schorfheide. *Misspecification-Robust Shrinkage and Selection for VAR Forecasts and IRFs*. Tech. rep. NBER Working Paper 33474, 2025.

[40] James D Hamilton. "Oil and the macroeconomy since World War II". In: *Journal of political economy* 91.2 (1983), pp. 228–248.

[41] James D Hamilton. *Causes and Consequences of the Oil Shock of 2007-08*. Tech. rep. National Bureau of Economic Research, 2009.

[42] James D Hamilton and Ana Maria Herrera. "Comment: oil shocks and aggregate macroeconomic behavior: the role of monetary policy". In: *Journal of Money, credit and Banking* (2004), pp. 265–286.

[43] Mark A Hooker. "Are oil shocks inflationary? Asymmetric and nonlinear specifications versus changes in regime". In: *Journal of money, credit and banking* (2002), pp. 540–561.

[44] Diego R Känzig. "The macroeconomic effects of oil supply news: Evidence from OPEC announcements". In: *American Economic Review* 111.4 (2021), pp. 1092–1125.

[45] Greg Kaplan, Benjamin Moll, and Giovanni L Violante. "Monetary policy according to HANK". In: *American Economic Review* 108.3 (2018), pp. 697–743.

[46] Robert K. Kaufmann. "The role of market fundamentals and speculation in recent price changes for crude oil". In: *Energy Policy* 39.1 (2011), pp. 105–115.

[47] Lutz Kilian. "Exogenous oil supply shocks: how big are they and how much do they matter for the US economy?" In: *The review of economics and statistics* 90.2 (2008), pp. 216–240.

[48] Lutz Kilian. "The economic effects of energy price shocks". In: *Journal of economic literature* 46.4 (2008), pp. 871–909.

[49] Lutz Kilian. "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market". In: *American economic review* 99.3 (2009), pp. 1053–1069.

[50] Lutz Kilian. "Measuring global real economic activity: Do recent critiques hold up to scrutiny?" In: *Economics Letters* 178 (2019), pp. 106–110.

[51] Lutz Kilian and Logan T Lewis. "Does the Fed respond to oil price shocks?" In: *The Economic Journal* 121.555 (2011), pp. 1047–1072.

[52] Lutz Kilian and Daniel P. Murphy. "The Role of Inventories and Speculative Trading in the Global Market for Crude Oil". In: *Journal of Applied Econometrics* 29.3 (2014), pp. 454–478.

[53] Lutz Kilian and Daniel P Murphy. "The role of inventories and speculative trading in the global market for crude oil". In: *Journal of Applied econometrics* 29.3 (2014), pp. 454–478.

[54] Lutz Kilian and Xiaoqing Zhou. "The econometrics of oil market VAR models". In: *Advances in Econometrics* 45 (2023), pp. 65–95.

[55] Marek Kolodzeij and Robert K. Kaufmann. "Oil demand shocks reconsidered: A cointegrated vector autoregression". In: *Energy Economics* 41 (2014), pp. 33–40.

[56] Xiaowen Lei, Julian F. Ludwig, and Xiaohan Ma. "Oil Prices and Inequality". In: *European Economic Review* (2025). Forthcoming.

[57] Michele Lenza and Ettore Savoia. *Do we need firm data to understand macroeconomic dynamics?* Working Paper Series 438. Sveriges Riksbank (Central Bank of Sweden), 2024.

[58] Robert B Litterman. "Techniques for forecasting with vector autoregressions". PhD thesis. Ph. D. thesis, University of Minnesota, 1980.

[59] Julian F. Ludwig. *Local Projections Are VAR Predictions of Increasing Order*. Tech. rep. SSRN Working Paper 4882149, 2024.

[60] James G. MacKinnon. "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests". In: *Journal of Business & Economic Statistics* 12.2 (1994), pp. 167–176.

[61] Michael W. McCracken and Serena Ng. "FRED-QD: A Quarterly Database for Macroeconomic Research". In: *Review* 103.1 (2021), pp. 1–44.

[62] Alisdair McKay and Christian K Wolf. "What Can Time-Series Regressions Tell Us About Policy Counterfactuals?" In: *Econometrica* 91.5 (2023), pp. 1695–1725.

[63] Karel Mertens and Morten O Ravn. "The dynamic effects of personal and corporate income tax changes in the United States". In: *American economic review* 103.4 (2013), pp. 1212–1247.

[64] Silvia Miranda-Agrippino and Giovanni Ricco. "Identification with external instruments in structural VARs". In: *Journal of Monetary Economics* 135 (2023), pp. 1–19.

[65] Pavel Molchanov. "A statistical analysis of OPEC quota violations". In: *Economics* (2003), pp. 1–31.

[66] Lorenzo Mori and Gert Peersman. *Estimating the macroeconomic effects of oil supply news*. Tech. rep. CESifo Working Paper, 2024.

[67] Anton Nakov and Andrea Pescatori. "Oil and the great moderation". In: *The Economic Journal* 120.543 (2010), pp. 131–156.

[68] M Hashem Pesaran and Yongcheol Shin. "Generalized impulse response analysis in linear multivariate models". In: *Economics Letters* 58.1 (1998), pp. 17–29.

[69] Robert S. Pindyck. "The dynamics of commodity spot and futures markets: A primer". In: *The Energy Journal* 22.3 (2001), pp. 1–29.

[70] Mikkel Plagborg-Møller and Christian K Wolf. "Local projections and VARs estimate the same impulse responses". In: *Econometrica* 89.2 (2021), pp. 955–980.

[71] Emanuela Raffinetti, Elena Siletti, and Achille Vernizzi. "On the Gini coefficient normalization when attributes with negative values are considered". In: *Statistical Methods & Applications* 24.3 (2015), pp. 507–521.

[72] Ronald A Ratti and Joaquin L Vespignani. "OPEC and non-OPEC oil production and the global economy". In: *Energy Economics* 50 (2015), pp. 364–378.

[73] Christopher A Sims, James H Stock, and Mark W Watson. "Inference in linear time series models with some unit roots". In: *Econometrica: Journal of the Econometric Society* (1990), pp. 113–144.

[74] Christopher A Sims and Tao Zha. "Bayesian methods for dynamic multivariate models". In: *International Economic Review* (1998), pp. 949–968.

[75] Abe Sklar. "Vol. 8 of Fonctions de Répartition à n Dimensions et Leurs Marges, 229–231". In: *Paris: Publications de l'Institut de statistique de l'Université de Paris* (1959).

[76] James H Stock and Mark W Watson. *Disentangling the Channels of the 2007-2009 Recession*. Tech. rep. National Bureau of Economic Research, 2012.

[77] James H Stock and Mark W Watson. "Identification and estimation of dynamic causal effects in macroeconomics using external instruments". In: *The Economic Journal* 128.610 (2018), pp. 917–948.

[78] Rainer Storn and Kenneth Price. "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces". In: *Journal of global optimization* 11 (1997), pp. 341–359.

[79] Yan Tan and Utai Uprasen. "Asymmetric effects of oil price shocks on income inequality in ASEAN countries". In: *Energy Economics* 126.C (2023).

[80] Hiro Y Toda and Taku Yamamoto. "Statistical inference in vector autoregressions with possibly integrated processes". In: *Journal of econometrics* 66.1-2 (1995), pp. 225–250.

[81] Halbert White. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity". In: *Econometrica* 48.4 (1980), pp. 817–838.

[82] Holbrook Working. "The Theory of Price of Storage". In: *American Economic Review* 39.6 (1949), pp. 1254–1262.

A Granger-Causality Tests

Mori and Peersman (2024) document that structural oil-market shocks identified in standard monthly SVARs are predictable by financial variables, in violation of the lead-lag exogeneity required for valid identification. I replicate their predictability test at quarterly frequency for both shocks I use. (The companion "recoverability" test on the oil-price residual is mechanically satisfied in my specification, since the shock instrument enters the VAR endogenously and is therefore a linear combination of the system's innovations by construction; it is informative only for the external/proxy-SVAR reading, which I do not adopt.)

Granger-causality test. For each candidate predictor x_t (a financial variable or common factor), I run the auxiliary regression

$$z_t = \alpha + \gamma_1 x_{t-1} + \cdots + \gamma_L x_{t-L} + e_t, \tag{27}$$

where z_t is the shock instrument and $L \in \{2, 4\}$. I test $H_0 \colon \gamma_1 = \cdots = \gamma_L = 0$ with a heteroskedasticity-robust Wald statistic using the HC0 sandwich estimator of White (1980) and report the p-value from an $F(L, T - L - 1)$ reference distribution. The $L = 2$ choice corresponds to a six-month horizon, matching the lag structure used by Mori and Peersman (2024) at monthly frequency (six lags); $L = 4$ extends to one year as a robustness check.

Results. Table 2 reports the predictability test for both the Känzig (2021) oil supply news shock and the Baumeister and Hamilton (2019) supply shock. No individual financial variable robustly predicts either shock instrument across both lag specifications, supporting the exogeneity of the instruments at quarterly frequency. For the KZ shock, the one exception is the excess bond premium, which is correlated with Factor 5; for BH, the term spread and Factor 6 are the exceptions. The Granger-causality patterns documented by Mori and Peersman (2024) at monthly frequency are therefore substantially attenuated at quarterly frequency, with the

residual predictive content concentrated in EBP (KZ) and the term spread (BH)—the channels that the `mp_ebp` robustness specification controls for directly.

Information overlap between the QD factor block and financial variables. A natural question is whether the baseline robustness specification `paper_kanzig_kqd_smfdd`, which augments the baseline VAR with the six McCracken and Ng (2021) levels factors $\boldsymbol{F}_t^{\text{lev}} = (qdl_{f1}, \ldots, qdl_{f6})$, implicitly addresses the financial-variable predictability documented in Table 2. To assess this I compute, for each financial variable x_t, the in-sample R^2 of a linear projection on the six levels factors:

$$x_t = \alpha + \boldsymbol{\beta}' \boldsymbol{F}_t^{\text{lev}} + \eta_t. \tag{28}$$

Table 1 reports these R^2 values over the 1975Q1–2024Q3 sample.

Table 1: R^2 from regressing each financial variable on the six McCracken and Ng (2021) levels factors

Financial variable	R^2
Federal funds rate	0.986
3M T-bill	0.968
1Y interest rate	0.963
10Y interest rate	0.952
BAA corporate yield	0.949
S&P 500	0.924
Term spread	0.788
USD exchange rate	0.677
Excess bond premium	0.220
VIX	0.138

Note: OLS R^2 from regressing each financial variable on a constant and the six FRED-QD levels factors $qdl_{f1}, \ldots, qdl_{f6}$ used in `paper_kanzig_kqd_smfdd`. Sample 1975Q1–2024Q3, $n = 199$ for all rows except S&P 500 ($n = 193$). The factor block is constructed with the variable-overlap exclusions of Section 4, so the regressors do not contain the dependent variables themselves.

The six levels factors span 92–99 % of the variation in interest rates, the S&P 500, and the BAA corporate yield, and roughly 70–80 % of the term spread and USD exchange rate. The two financial variables that the factor block does *not* span are the excess bond premium ($R^2 = 0.22$) and the VIX ($R^2 = 0.14$): both reflect higher-frequency credit-stress and volatility innovations that are orthogonal to the slow-moving levels factor space. Two implications follow.

First, including $\boldsymbol{F}_t^{\text{lev}}$ in the VAR implicitly controls for the financial-predictor block of Table 2 along the rate-and-equity dimension. The Granger-causality concerns of Mori and Peersman (2024) for monthly SVARs—which manifest in my quarterly results primarily for short and long Treasury yields, the federal funds rate, the term spread, and corporate yields—are addressed by the kqd specification through factor-block conditioning, without the need to add each financial variable as a separate VAR observable.

Second, the kqd and `mp_ebp` specifications are complementary rather than substitutes. The QD factor block absorbs the slow-moving level and term-structure information; the Mori and Peersman (2024) financial block adds the credit-stress (EBP) and volatility (VIX) dimensions that levels factors do not span. I report both specifications as robustness checks: `paper_kanzig_kqd_smfdd` guards against omission of the rate/equity portion of the financial information set, while `paper_kanzig_mp_ebp` guards against omission of the residual credit and volatility components. Impulse responses are stable across both specifications, indicating that the baseline conclusions are not driven by either component of the financial information set in isolation.

Table 2: Granger-causality tests — Shock instruments

	Känzig (2021)		Baumeister & Hamilton (2019)	
	VAR(2)	VAR(4)	VAR(2)	VAR(4)
Common factors (all)	**0.00*****	**0.00*****	**0.00*****	**0.00*****
F_1	1.00	0.67	0.86	0.96
F_2	0.58	0.88	0.97	0.89
F_3	0.92	0.94	**0.03****	**0.01*****
F_4	0.32	0.34	0.31	0.40
F_5	0.13	0.31	0.59	0.08*
F_6	0.31	0.52	0.10	0.12
F_7	0.53	0.61	0.35	0.38
Dist. factors (all)	0.06*	**0.04****	**0.02****	**0.00*****
D_1	0.44	0.84	0.41	0.16
D_2	0.39	**0.00*****	0.58	0.16
D_3	0.16	0.40	**0.02****	**0.02****
D_4	0.45	0.78	0.52	0.90
D_5	0.07*	**0.02****	0.06*	**0.02****
D_6	0.49	0.54	0.73	0.82
D_7	**0.05****	0.29	0.07*	**0.01****
D_8	0.10	0.41	0.46	0.33
S&P 500	0.95	0.68	0.12	0.34
VIX	0.45	0.77	0.83	0.82
1Y interest rate	0.83	0.70	0.75	0.07*
10Y interest rate	0.88	0.33	0.35	0.16
3M T-bill	0.64	0.71	0.81	**0.03****
BAA corporate yield	0.39	0.47	0.26	0.10
Federal funds rate	0.57	0.36	0.80	**0.02****
Excess bond premium	0.26	0.09*	0.25	0.33
Term spread	0.17	0.22	**0.01****	**0.01****
USD exchange rate	0.65	0.70	0.48	0.76

Note: p-values from robust (HC0) F-tests of the null that lagged variables do not Granger-cause the shock instrument. Since the instrument is external to the VAR, results are invariant to the VAR information set; only the baseline specification is reported. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Sample: 1975Q1–2023Q1. Common factors are the first seven principal components of the McCracken and Ng (2021) quarterly dataset. Distributional factors are the eight smoothed states of the Bayer–Calderon–Kuhn (2025) state-space model on PSID income, consumption, and copula coefficients. Term spread is the 10-year minus 3-month Treasury yield.

B Data

This appendix documents the macroeconomic, financial, and oil-market variables used in the baseline VAR and across the robustness specifications. All series are at quarterly frequency; monthly source data is aggregated appropriately. The full set is bundled in the canonical macro file `oil_macro_all.csv`, which is constructed from three sources: the Känzig (2021) replication archive (oil-market block), McCracken and Ng (2021) FRED-QD (macroeconomic and financial), and FRED monthly downloads for series not in FRED-QD (selected CPI components and the Brent futures price). Distributional variables and their reconstruction are documented in Appendix C.

Table 3: Macroeconomic, financial, and oil-market variables

Variable	Description	Source	Transform	Used in
Oil market block — baseline				
`poil`	Real WTI spot price	Känzig (2021)	log levels	all baseline configs
`oilprod`	Global crude oil production	Känzig (2021)	log levels	all baseline configs
`worldip`	OECD + 6 NME industrial production	Känzig (2021)	log levels	all baseline configs
`usip`	U.S. industrial production	FRED-QD (`INDPRO`)	log levels	all baseline configs
`log_cpi`	U.S. CPI, all items	FRED-QD (`CPIAUCSL`)	log levels	all baseline configs
`oilstocksM_BH`	OECD crude inventories, BH cumulated	Baumeister and Hamilton (2019) replication archive (extended via EIA inputs through 2024Q2; anchored at 1975Q1)	log levels	baseline VAR
Oil-market robustness				
`poil_rac`	Refiners' acquisition cost (imported), real	EIA	log levels	`paper_*_rac`
`poil_rac_comp`	Refiners' acquisition cost (composite), real	EIA	log levels	sensitivity (`poil_rac*`)
`poil_rac_dom`	Refiners' acquisition cost (domestic), real	EIA	log levels	sensitivity (`poil_rac*`)
`poil_brent`	Real Brent (deflated by CPI)	EIA `MCOILBRENTEU`	log levels (1987Q2+)	`paper_*_brent`
`oilstocksM`	OECD crude oil inventories, noisy series	Känzig (2021) (Kilian–Murphy 2014 construction)	log levels	`paper_*_oldinv`

Continued on next page

Table 3 – continued from previous page

Variable	Description	Source	Transform	Used in
`oilstocksL`	OECD crude inventories, level (no log)	derived from `oilstocksM`	raw level	sensitivity
`kilian_rea`	Kilian Index of Global Real Economic Activity (dry-cargo shipping rate)	Kilian (2009), Dallas Fed update	quarterly mean of monthly (already % deviation from trend)	`paper_*_rea`
`oilprod_opec`	OPEC crude (incl. lease condensate) production, Kanzig-anchored via EIA share	EIA INTL (`productId=57`, X-12-ARIMA SA) × Känzig (2021) world	log levels (1973Q1+, sample limited by `oilprod`)	`paper_*_opec`
`oilprod_nopec`	Non-OPEC crude production (residual: World − OPEC, identity preserved)	same as above	log levels	`paper_*_opec`
Real activity (FRED-QD)				
`gdp`	Real GDP	`GDPC1`	log levels	medium block
`disp_income`	Real disposable income	`DPIC96`	log levels	FAVAR
`consumption`	Real PCE	`PCECC96`	log levels	FAVAR
`investment`	Real gross private domestic invest.	`GPDIC1`	log levels	FAVAR
`pce_durables`	Real PCE durables	`PCDGx`	log levels	FAVAR
`pce_nondurables`	Real PCE nondurables	`PCNDx`	log levels	FAVAR
`pce_services`	Real PCE services	`PCESVx`	log levels	FAVAR
`unrate`	Civilian unemployment rate	`UNRATE`	level (%)	medium block, FAVAR
`payems`	Total nonfarm payrolls	`PAYEMS`	log levels	FAVAR
`manemp`	Manufacturing employment	`MANEMP`	log levels	FAVAR
`ahetpi`	Avg. hourly earnings, prod. workers	`AHETPIx`	log levels	FAVAR
`weekly_hours`	Avg. weekly hours, prod. workers	`CES0600000007`	level	FAVAR
`sentiment`	Michigan consumer sentiment	`UMCSENTx`	log levels	FAVAR
Prices (FRED-QD + FRED monthly)				
`cpi_core`	Core PCE price index	`PCEPILFE`	log levels	medium block
`cpi_oer`	CPI owners' equivalent rent	`CUSR0000SEHC`	log levels	FAVAR
`cpi_food`	CPI food	FRED `CPIUFDSL`	log levels	FAVAR

Continued on next page

Table 3 – continued from previous page

Variable	Description	Source	Transform	Used in
cpi_energy	CPI energy	FRED CPIENGSL	log levels	FAVAR
cpi_rent	CPI rent of primary residence	FRED CUSR0000SEHA	log levels	FAVAR
cpi_transp	CPI transportation	CPITRNSL	log levels	FAVAR
ppi	PPI all commodities	PPIACO	log levels	FAVAR
imp_prices	Import price index, BLS IPP	FRED IR, spliced pre-1982Q3 with BEA B021RG3Q086SBEA	log levels	medium block, FAVAR
Financial (FRED-QD)				
sp500	S&P 500 index	WolfAggs (CRSP)	log levels	MP+EBP, FAVAR
vix	CBOE VIX	VIXCLSx	log levels	MP+EBP, FAVAR
gs1	1-year Treasury yield	GS1	level (%)	MP+EBP, FAVAR
gs10	10-year Treasury yield	GS10	level (%)	FAVAR
tb3ms	3-month T-bill yield	TB3MS	level (%)	FAVAR
baa	BAA corporate yield	BAA	level (%)	FAVAR
mortgage30	30-year mortgage rate	MORTGAGE30US	level (%)	FAVAR
fedfunds	Federal funds rate	FEDFUNDS	level (%)	FAVAR
term_spread	10Y minus 3M Treasury spread	GS10TB3Mx	level (%)	FAVAR
ebp	Excess bond premium	Gilchrist and Zakrajšek (2012)	level	MP+EBP
exrate	Trade-weighted USD index	TWEXAFEGSMTHx	log levels	FAVAR
corp_profits	Corporate profits before tax	B020RE1Q156NBEA	level	FAVAR
house_prices	Case-Shiller 20-city HPI	SPCS20RSA	log levels	FAVAR
Macro aggregates (WolfAggs)				
inv	Real gross private domestic investment	WolfAggs (replication archive of McKay and Wolf, 2023, based on FRED GPDIC1)	log levels	extended-macro VAR
cons	Real personal consumption expenditures	WolfAggs (FRED PCECC96)	log levels	extended-macro VAR
wages	Real wages, production / non-supervisory workers	WolfAggs (FRED AHETPIx)	log levels	extended-macro VAR
unemp	Civilian unemployment rate	WolfAggs (FRED UNRATE)	level (%)	extended-macro VAR
infl	PCE-deflator inflation, q-on-q	WolfAggs (FRED PCECTPI)	level (%)	extended-macro VAR

Continued on next page

Table 3 – continued from previous page

Variable	Description	Source	Transform	Used in
`ffr`	Federal funds rate	WolfAggs (FRED `FEDFUNDS`)	level (%)	extended-macro VAR
Household credit (FRED-QD + FRED monthly)				
`consumer_credit`	Total consumer credit outstanding, real	FRED-QD `TOTALSLx`	log levels	extended-macro VAR, FAVAR
`consumer_loans`	Consumer loans at all commercial banks, real	FRED-QD `CONSUMERx`	log levels	FAVAR
`cc_debt`	Revolving consumer credit, real (≈ credit-card debt)	FRED-QD `REVOLSLx`	log levels	FAVAR
`nonrev_credit`	Non-revolving consumer credit, real (auto + student)	FRED-QD `NONREVSLx`	log levels	FAVAR
`dlq_consumer`	Delinquency rate on consumer loans	FRED `DRCLACBS`	level (%)	FAVAR
`dlq_cc`	Delinquency rate on credit-card loans	FRED `DRCCLACBS`	level (%)	FAVAR
`dlq_mortgage`	Delinquency rate on single-family residential mortgages	FRED `DRSFRMACBS`	level (%)	FAVAR
`chgoff_consumer`	Charge-off rate on consumer loans	FRED `CORCACBS`	level (%)	FAVAR
`chgoff_cc`	Charge-off rate on credit-card loans	FRED `CORCCACBS`	level (%)	FAVAR
Income composition (NIPA)				
`transfers`	Personal current transfer receipts (nominal $bn)	FRED `A063RC1`	log levels	FAVAR
`income_ex_transfers`	Real personal income excluding transfer receipts	FRED `W875RX1`	log levels	FAVAR
Inflation expectations				
`infl_exp_mich`	U. Michigan 1-year expected inflation (consumer survey)	FRED `MICH`	level (%)	FAVAR
`infl_exp_spf_cpi1y`	SPF median 1-year CPI inflation forecast (forecasters)	Philadelphia Fed SPF (`INFCPI1YR`)	level (%)	FAVAR
Common factors (PCA on FRED-QD)				
$F^{\text{lev}}_{1..6}$	Levels common factors (Bai 2004)	PCA on log-levels of FRED-QD ($n = 113 - 12$, post-exclusion); 6 factors	—	`paper_*_kqd_smfdd`
Shock instruments				

Continued on next page

Table 3 – continued from previous page

Variable	Description	Source	Transform	Used in
`oil_supply_news`	OPEC announcement futures surprise	Känzig (2021)	quarterly sum	baseline (KZ)
`bh_supply_shock`	Realized structural supply shock	Baumeister and Hamilton (2019)	—	baseline (B&H)
`paveldje_shock`	Demand-side news shock	Degasperi (2025)	—	inactive (commented)

Notes: All series at quarterly frequency, 1975Q1–2024Q3 unless otherwise noted. Monthly source data is aggregated by quarterly mean. "Log levels" means $\log(x_t)$ used directly without further differencing or detrending; the Bayesian framework handles mixed orders of integration without pre-testing (Section 4). The FRED-QD common factors are extracted with the variable-overlap exclusion list (`OILPRICEx`, `INDPRO`, `CPIAUCSL`, `CPILFESL`, `FEDFUNDS`, `GS1`, `GS10`, `VIXCLSx`, `S&P 500`, `UNRATE`, `GDPC1`, `GPDIC1`, `PCECC96`, `AHETPIx`, `PAYEMS`) so that any series used elsewhere in the VAR as an explicit observable is removed from the PCA input panel before extraction. Brent (`poil_brent`) and the BH-cumulated inventory series (`oilstocksM_BH`) are constructed in the build pipeline; details on the latter are in Appendix E. The Kilian REA (`kilian_rea`) is already in percent-deviation-from-trend units and enters the VAR without further log-transform.

C Reconstruction of Distributional Objects

This appendix details how impulse responses estimated in factor space are mapped back to distributional objects—quantile functions, copula densities, and household-level moments—following the methodology of Bayer, Calderon, and Kuhn (2025).

C.1 From Factor IRFs to Coefficient IRFs

The VAR is estimated on a small number of distributional factors $\boldsymbol{F}_{D,t} \in \mathbb{R}^K$, with $K = 8$ in the baseline specification. Let $\hat{\boldsymbol{\theta}}_t \in \mathbb{R}^N$ denote the raw Legendre polynomial coefficient vector, with $N = (O+1)^d + d \cdot (O+1)$ collecting copula and marginal quantile-function coefficients (cf. Section 3). The factors are obtained via principal components analysis on the *standardized* coefficients $\boldsymbol{\Sigma}_\theta^{-1/2}\big(\hat{\boldsymbol{\theta}}_t - \bar{\boldsymbol{\theta}}\big)$, so the implied factor model for the raw coefficients is

$$\hat{\boldsymbol{\theta}}_t \;=\; \bar{\boldsymbol{\theta}} \;+\; \boldsymbol{\Sigma}_\theta^{1/2}\big(\boldsymbol{\Lambda}\,\boldsymbol{F}_{D,t} + \boldsymbol{\gamma}\,\boldsymbol{f}_t\big), \tag{29}$$

where $\bar{\boldsymbol{\theta}}$ is the sample mean, $\boldsymbol{\Sigma}_\theta^{1/2}$ is the diagonal matrix of per-coefficient standard deviations, $\boldsymbol{\Lambda} \in \mathbb{R}^{N \times K}$ is the factor loading matrix, and $\boldsymbol{f}_t$ captures idiosyncratic variation orthogonal to the common factors.

Mixed-frequency measurement and the four-quarter projection. A subtle point governs the loading used for reconstruction. Bayer, Calderon, and Kuhn (2025) estimate the factor model in a mixed-frequency state-space framework: the latent factors $\boldsymbol{F}_t$ live at quarterly frequency, but the PSID coefficient observations are annual averages. The measurement equation for the PSID dataset takes the form

$$\tilde{\boldsymbol{\theta}}_t^{\text{PSID}} \;=\; \frac{1}{4}\,\boldsymbol{\Lambda}\left(\boldsymbol{F}_t + \boldsymbol{F}_{t-1} + \boldsymbol{F}_{t-2} + \boldsymbol{F}_{t-3}\right) \;+\; \boldsymbol{\nu}_t^{\text{PSID}}, \tag{30}$$

where $\boldsymbol{\nu}_t^{\text{PSID}}$ captures sampling and operationalisation noise. Equation (30) is the source of a reconstruction issue when the smoothed factors enter the VAR as observables. Applying $\boldsymbol{\Lambda}$ alone to a quarterly factor IRF produces an over-attenuated coefficient response, because each quarter of the IRF activates a single factor rather than the four-quarter sum the loadings $\boldsymbol{\Lambda}$ were estimated to fit. To recover the quarterly-frequency loading that maps a single quarter of factor variation into its associated coefficient impact, I fit by ordinary least squares

$$\hat{\boldsymbol{\Lambda}}_{4q} \;=\; \arg\min_{\boldsymbol{\Lambda}} \sum_t \left\| \hat{\boldsymbol{\theta}}_{t|T}^{\text{PSID}} \;-\; \boldsymbol{\Lambda}\,\tfrac{1}{4}\big(\hat{\boldsymbol{F}}_{t|T} + \hat{\boldsymbol{F}}_{t-1|T} + \hat{\boldsymbol{F}}_{t-2|T} + \hat{\boldsymbol{F}}_{t-3|T}\big) \right\|^2, \tag{31}$$

where $\hat{\boldsymbol{\theta}}_{t|T}^{\text{PSID}}$ and $\hat{\boldsymbol{F}}_{t|T}$ are the BCK smoother's posterior-mean reconstructions of the PSID coefficient and the latent quarterly factor at quarter t, respectively. The four-quarter rolling sum on the right-hand side enforces the BCK measurement equation (30); OLS thus recovers the loading $\hat{\boldsymbol{\Lambda}}_{4q}$ that closes the measurement-equation residual.

Mapping factor IRFs to coefficient IRFs. Quarterly factor responses $\boldsymbol{\Psi}_h^F$ are mapped to raw-coefficient responses by

$$\Delta\hat{\boldsymbol{\theta}}_h \;=\; \boldsymbol{\Sigma}_\theta^{1/2}\,\hat{\boldsymbol{\Lambda}}_{4q}\,\boldsymbol{\Psi}_h^F, \tag{32}$$

which undoes the per-coefficient standardization $\boldsymbol{\Sigma}_\theta^{1/2}$ from the PCA. The result $\Delta\hat{\boldsymbol{\theta}}_h$ is the IRF of the polynomial coefficients in raw units; the steady-state mean is not reintroduced.

Reconstructing nonlinear functionals. To evaluate distributional functionals ϕ (quantile values, copula densities, household-group means, inequality indices) at horizon h, the reconstruction is non-linear in the coefficient vector and therefore requires the level coefficient

$$\hat{\boldsymbol{\theta}}_h \;=\; \bar{\boldsymbol{\theta}} + \Delta\hat{\boldsymbol{\theta}}_h, \tag{33}$$

where $\bar{\boldsymbol{\theta}}$ is the time-invariant steady-state coefficient (including any trend component removed prior to estimation). The reconstructed objects are evaluated via (35)–(36) below at the level

$\hat{\boldsymbol{\theta}}_h$, and the IRF of any functional ϕ is reported as

$$\Psi_h^{\phi} \;=\; \phi(\hat{\boldsymbol{\theta}}_h) \;-\; \phi(\bar{\boldsymbol{\theta}}). \tag{34}$$

For functionals that are linear in $\boldsymbol{\theta}$ (the marginal quantile values evaluated pointwise via (35)), this reduces to direct evaluation at $\Delta\hat{\boldsymbol{\theta}}_h$. For copula-based functionals (household-group means, conditional moments, Ginis), the level coefficient must be used because monotonicity and non-negativity constraints couple the reconstruction nonlinearly across coefficients.

C.2 Reconstructing Quantile Functions and Copula Densities

The coefficient vector $\hat{\theta}_h$ partitions into copula coefficients $\kappa_h \in \mathbb{R}^{(O+1)^d}$ and marginal quantile-function coefficients $\xi_h^m \in \mathbb{R}^{O+1}$ for each dimension $m \in \{\text{consumption, income, wealth}\}$. These reconstruct the distributional objects via the Legendre polynomial basis $\{Q_o\}_{o=0}^{O}$:

Marginal quantile functions. For dimension m evaluated on a grid $u \in [0,1]$:

$$\hat{\Xi}_{m,h}^{-1}(u) = \sum_{o=0}^{O} \xi_{o,h}^{m}\, Q_o(u). \tag{35}$$

In practice, I evaluate this on a grid of $G = 20$ equally spaced quantile points (vigintiles), giving the income, consumption, or wealth level at each vigintile of the distribution.

Copula density. The joint dependence structure is reconstructed as:

$$d\hat{C}_h(u_1,u_2,u_3) = \sum_{o_1,o_2,o_3=0}^{O} \kappa_{(o_1,o_2,o_3),h} \prod_{m=1}^{3} Q_{o_m}(u_m). \tag{36}$$

This is evaluated on the $G^3 = 8{,}000$ grid points of the three-dimensional unit cube, giving the joint density at each combination of consumption, income, and wealth vigintiles.

C.3 Household Groups and Conditional Moments

A household group is defined by a range of quantiles in each dimension. For a group occupying quantile range $U_i^c \times U_i^y \times U_i^w$ (e.g., the bottom two income vigintiles crossed with all wealth vigintiles), I compute:

Mass (population share). The fraction of households in the cell:

$$\omega_{i,h} = \iiint_{(u_c,u_y,u_w)\in U_i^c\times U_i^y\times U_i^w} d\hat{C}_h(u_c, u_y, u_w). \tag{37}$$

Conditional mean consumption. The average consumption of households in the cell:

$$\bar{c}_{i,h} = \frac{1}{\omega_{i,h}} \iiint_{U_i^c\times U_i^y\times U_i^w} \hat{\Xi}_{c,h}^{-1}(u_c)\, d\hat{C}_h(u_c, u_y, u_w). \tag{38}$$

Income and wealth conditional means are defined analogously by replacing $\hat{\Xi}_{c,h}^{-1}$ with the corresponding marginal quantile function.

Numerical implementation. The continuous integrals (37)–(38) are discretized on the $G \times G \times G$ vigintile grid. Let $\mathcal{G}_i \subseteq \{1,\dots,G\}^3$ denote the grid indices belonging to household group i, and write $u_g = (g-1/2)/G$ for the midpoint of vigintile g. The discrete analogues of (37) and (38) are

$$\hat{\omega}_{i,h} = \frac{1}{G^3}\sum_{\mathbf{g}\in\mathcal{G}_i} d\hat{C}_h(u_{g_c}, u_{g_y}, u_{g_w}), \qquad \hat{\bar{c}}_{i,h} = \frac{1}{G^3\,\hat{\omega}_{i,h}}\sum_{\mathbf{g}\in\mathcal{G}_i} \hat{\Xi}_{c,h}^{-1}(u_{g_c})\, d\hat{C}_h(u_{g_c}, u_{g_y}, u_{g_w}), \tag{39}$$

with $\mathbf{g} = (g_c, g_y, g_w)$. The copula density is rescaled across the full grid so that $\sum_{\mathbf{g}} d\hat{C}_h(u_{g_c}, u_{g_y}, u_{g_w}) = G^3$, ensuring $\sum_i \hat{\omega}_{i,h} = 1$ when groups partition the unit cube.

C.4 Computation Summary

Table 4: Reconstruction parameters

Parameter	Value
Polynomial order (O)	11
Distribution dimensions (d)	3 (consumption, income, wealth)
Total coefficients ($N = (O+1)^d + d\cdot(O+1)$)	1,764
Copula coefficients ($(O+1)^d$)	1,728
Marginal coefficients ($d\cdot(O+1)$)	36
Distributional factors (K)	8
Variance explained by K factors	$> 95\%$
Evaluation grid (G)	20 (vigintiles)
Grid points for copula (G^3)	8,000
Gini evaluation grid (n)	200
Household groups	5 income $\times$ 5 wealth = 25 cells
Posterior draws	5,000

For each of the 5,000 posterior draws, the full reconstruction pipeline (factor IRF $\rightarrow$ coefficient IRF $\rightarrow$ quantile functions + copula density $\rightarrow$ household moments + inequality measures)

is executed at each horizon $h = 1, \ldots, 20$. The reported impulse responses are posterior medians with 68% credible sets computed pointwise across draws. The tensor contraction over the G^d grid is the computational bottleneck; the implementation exploits the separable structure of the Legendre basis—each factor $Q_{o_m}(u_m)$ depends on only one dimension—to reduce the $O\big(G^d(O+1)^d\big)$ naïve evaluation to $O\big(G^d(O+1)\big)$ operations per draw via successive 1D contractions along each dimension.

Formally, the reconstruction proceeds as follows. Given the estimated factors $\hat{\boldsymbol{F}}_t$, I first recover the standardized coefficients, then undo normalization and trend adjustments:

$$\hat{\tilde{\theta}}_t = \hat{\Lambda}\,\hat{\boldsymbol{F}} \qquad \text{(Project factors)} \tag{40}$$

$$\hat{\theta}_{nt} = \sigma_{\zeta(n)}\hat{\tilde{\theta}}_{\zeta(n)t} + \mu_n + g(t)_n \qquad \text{(Unstandardize and add trend)} \tag{41}$$

$$\hat{\theta}_t = (\hat{\xi}_{m,o_1,t}, \ldots, \hat{\kappa}_{(o_1,\ldots,o_d),t}) \qquad \text{(Decomposition)} \tag{42}$$

$$\tag{43}$$

See Table 5 for details on notation. To generate economically interpretable objects—quantile functions, copula densities, Gini coefficients, percentile shares, and other moments—these reconstructed objects are evaluated on $u \in [0,1]^d$, with boundary points excluded in practice (the grid uses $[10^{-6}, 0.9995]$ for numerical stability) and then integrated to obtain the distributional statistics reported in the body.

Table 5: Reconstruction of Distributional Objects: Notation

Symbol	Description
$\hat{\boldsymbol{F}}_t$	Estimated high-frequency factors
$\hat{\Lambda}$	Factor loadings mapping factors to coefficients
$\hat{\tilde{\theta}}_t$	Standardized coefficient representation
$\hat{\theta}_{nt}$	Final coefficient vector for coefficient n
$\zeta(n)$	denotes a distributional object—either a quantile function or a copula, which consists of n coefficients
$\sigma_{\zeta(n)}$	Standard deviation *per* dist. object
μ_n	Mean adjustment *per* coefficient
$g(t)_n$	Trend component per coefficient
$\hat{\xi}_{m,o,t}$	Marginal polynomial coefficients
$\hat{\kappa}_{(o_1,\ldots,o_d),t}$	Copula coefficients
$Q_o(u_m)$	Legendre polynomial basis function

Notes: The table summarizes the mapping from factor estimates to reconstructed distributional objects.

D Marginal-Only Specifications

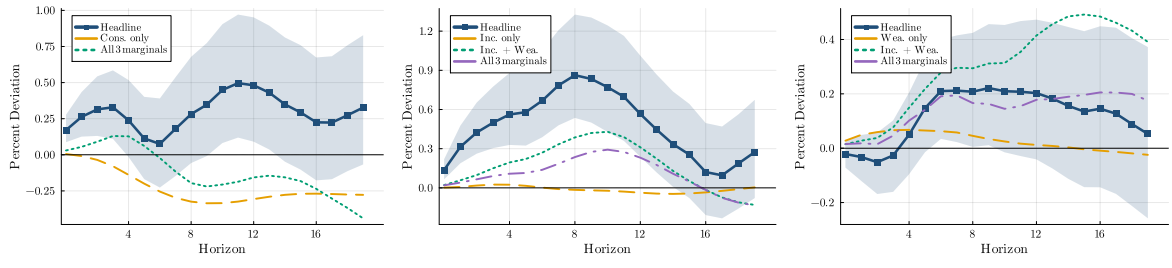
This appendix complements the baseline specification — which models the joint distribution of consumption, income, and wealth via the copula together with all three marginals — with five *marginal-only* specifications that selectively drop distributional information from the VAR. In each case, the macroeconomic block is unchanged; only the distributional factor block varies. The purpose is twofold. First, it benchmarks the baseline against the empirical strategy that dominates the literature: estimation on a single marginal dimension or a stack of marginals *without* the copula. Second, it isolates how much of the baseline inequality response is driven by the joint distribution rather than by any individual marginal.

Concretely, the five specifications are: (i) consumption marginal only (12 Legendre coefficients), (ii) income marginal only (12), (iii) wealth marginal only (12), (iv) income and wealth marginals stacked but no copula (24), and (v) all three marginals stacked but no copula (36). All five share the same baseline macro block, lag length, prior, and posterior draws; they differ only in which distributional coefficients enter the VAR and (mechanically) in the number of distributional principal components carried into the system, which is capped at three for marginal-only specifications.

Figures 13 and 14 report the Gini IRF for each dimension under the baseline (solid line, 68% credible band) and under each marginal-only specification that includes that dimension (dashed lines, posterior median only). The takeaway is direct: when the copula is dropped, the inequality response shifts in magnitude and sometimes in shape, and the shift is more pronounced under the realized supply shock than under the news shock. Modeling the joint matters most for the dimensions where wealth-income comovement carries the cyclical signal — consumption and wealth — and matters least for income, whose marginal response is largely picked up by the income coefficients alone.

Figure 13: Gini IRFs across marginal-only specifications

Panel A: Baumeister & Hamilton (2019) realized supply shock

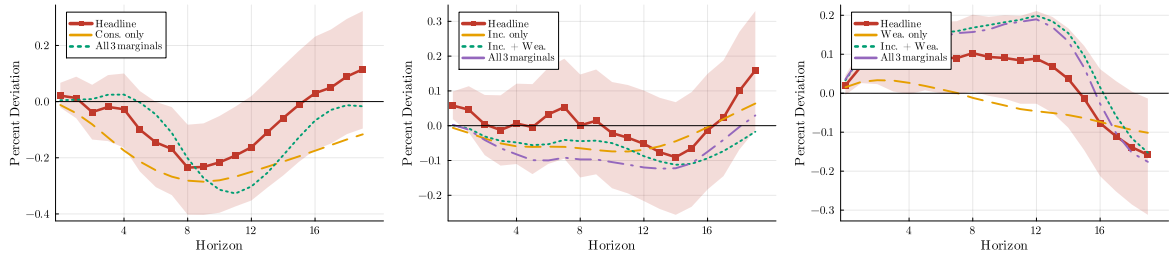


Consumption Gini | Income Gini | Wealth Gini

Notes: IRF of the Gini, percent deviation from steady state. Solid line + 68% credible band: baseline VAR (full distribution: copula + all three marginals). Dashed lines: marginal-only specifications that include the named dimension (consumption-only, income-only, wealth-only, income+wealth, all three marginals; see Section D for definitions). Macro block, lag length, prior, and posterior draws are identical across specifications.

Figure 14: Gini IRFs across marginal-only specifications

Panel B: Känzig (2021) supply news shock



Notes: IRF of the Gini, percent deviation from steady state. Solid line + 68% credible band: baseline VAR (full distribution: copula + all three marginals). Dashed lines: marginal-only specifications that include the named dimension. Macro block, lag length, prior, and posterior draws are identical across specifications.

E Inventory Measurement Error and Robustness

Baumeister and Hamilton (2019), Section III, document substantial measurement error in the OECD crude-oil inventory series of Kilian and Murphy (2014a) and propose a measurement-error-corrected reconstruction; details are in their paper. I adopt their construction for the baseline VAR: cumulating their monthly Δi_t series, anchored at `oilstocksM` in 1975Q1, yields the quarterly log-level series `oilstocksM_BH`, which I extend through 2024Q2 by replicating their construction on current EIA inputs and splicing at 2019M12. Figure 15 compares the result to the legacy `oilstocksM` of Känzig (2021): the level correlation is 0.997, but the $\Delta \log$ correlation is only 0.27 — the empirical content of BH's measurement-error equation, since a VAR identifies dynamics from changes.

Appendix F.1 reverts to the noisier legacy `oilstocksM` measure, holding the structural setup, prior, identification, and lag length fixed. The qualitative pattern is intact, ruling out a measurement-error explanation.

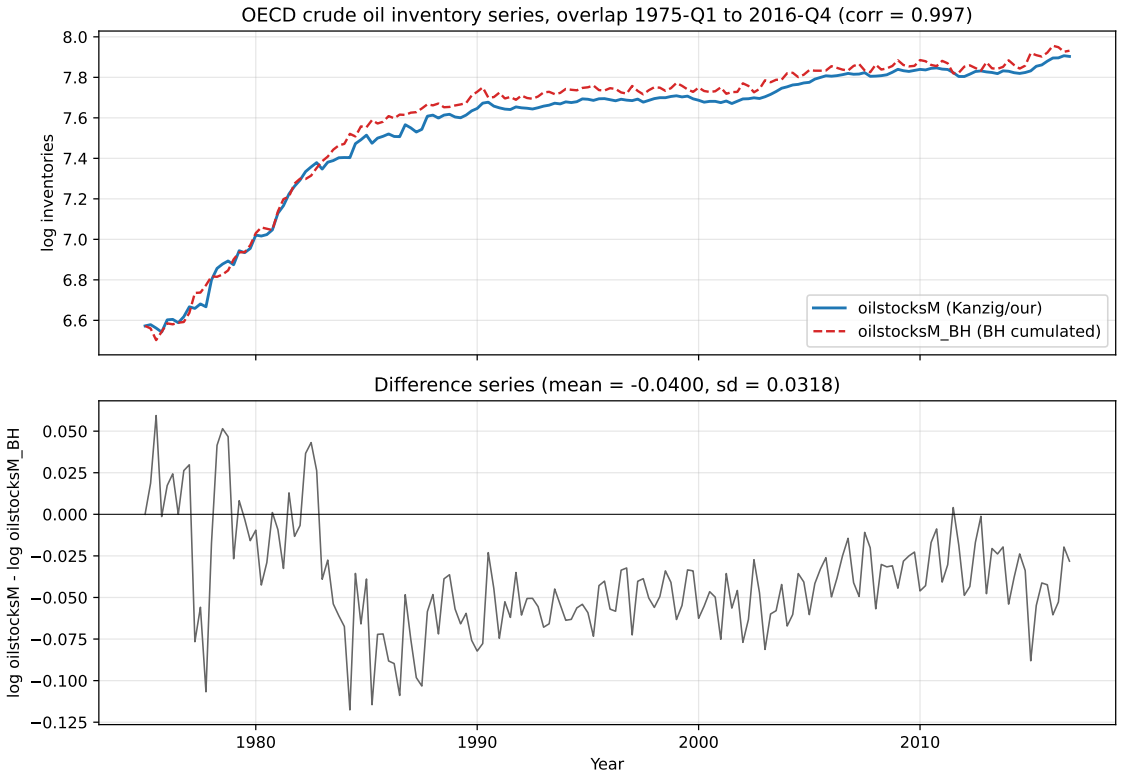


Figure 15: Comparison of two OECD crude-oil inventory series, 1975Q1–2016Q4

Notes: `oilstocksM`: legacy series from Känzig (2021). `oilstocksM_BH`: reconstructed by cumulating Baumeister and Hamilton (2019)'s monthly Δi_t data, anchored at 1975Q1. Top: log-level overlay. Bottom: difference series.

F Robustness

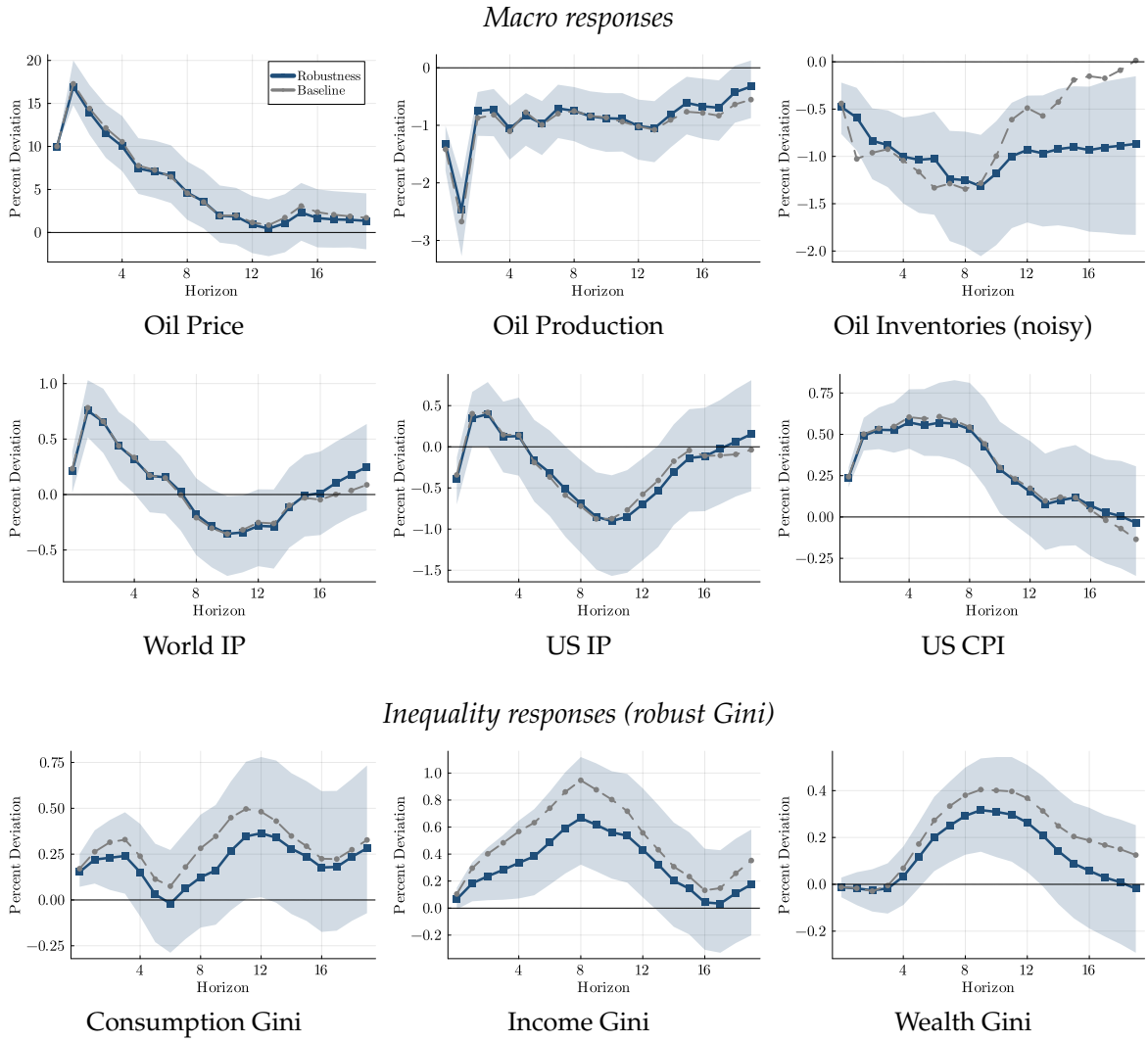
This appendix collects baseline-vs-robustness IRF comparisons for every robustness specification estimated in the paper. Each subsection corresponds to one robustness theme; within a subsection, each figure overlays the baseline IRF (solid + 68% credible band, shock-aware color) with the robustness IRF (dashed gray) for six macro variables and three robust-Gini measures of inequality.[26]

F.1 Legacy Inventory Series

I revert to the legacy `oilstocksM` inventory measure (Kanzig 2021's deseasonalised OECD series) instead of the baseline `oilstocksM_BH` (Baumeister & Hamilton 2019's measurement-error-corrected reconstruction); reverting to `oilstocksM` restores the full 1975Q1–2024Q3 sample (vs. 1975Q1–2016Q4 under the baseline).

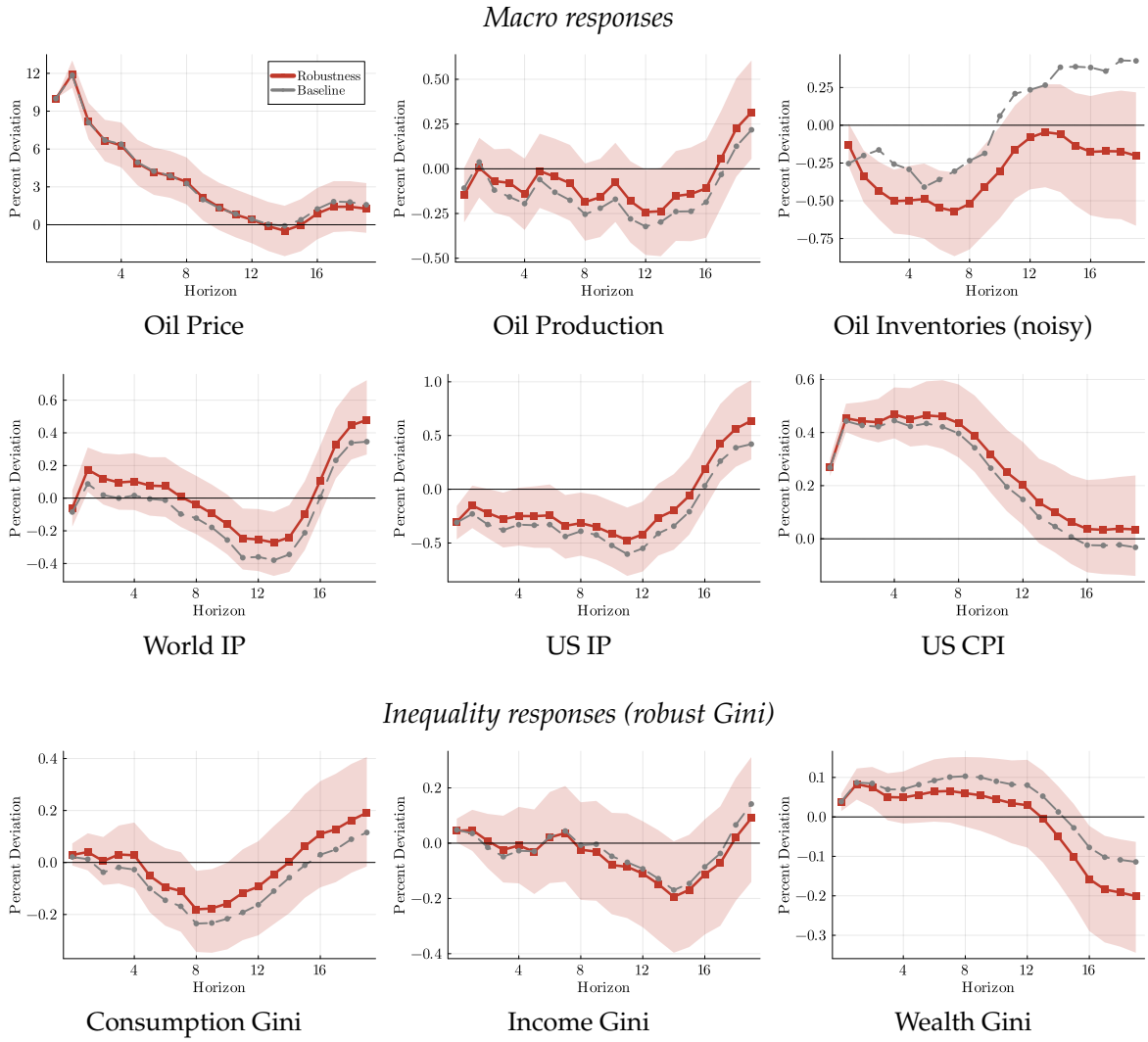
[26] Results are also unchanged under several additional checks not reported here for space: alternative oil-price measures (the U.S. refiners' acquisition cost of imported crude and the real-log Brent front-month price, addressing Kilian and Zhou 2023); the QD-levels factor specification with the distributional block dropped; posterior mean and posterior mode in place of the median (addressing the asymmetry concerns of Kilian and Zhou 2023); and 90% pointwise credible bands in place of the 68% bands. Figures available on request.

Figure 16: Legacy oilstocksM — Baumeister and Hamilton (2019) supply shock



Notes: Solid + 68% band: robustness (Legacy oilstocksM). Dashed gray: baseline. Macro responses are level IRFs.

Figure 17: Legacy oilstocksM — Känzig (2021) supply news shock

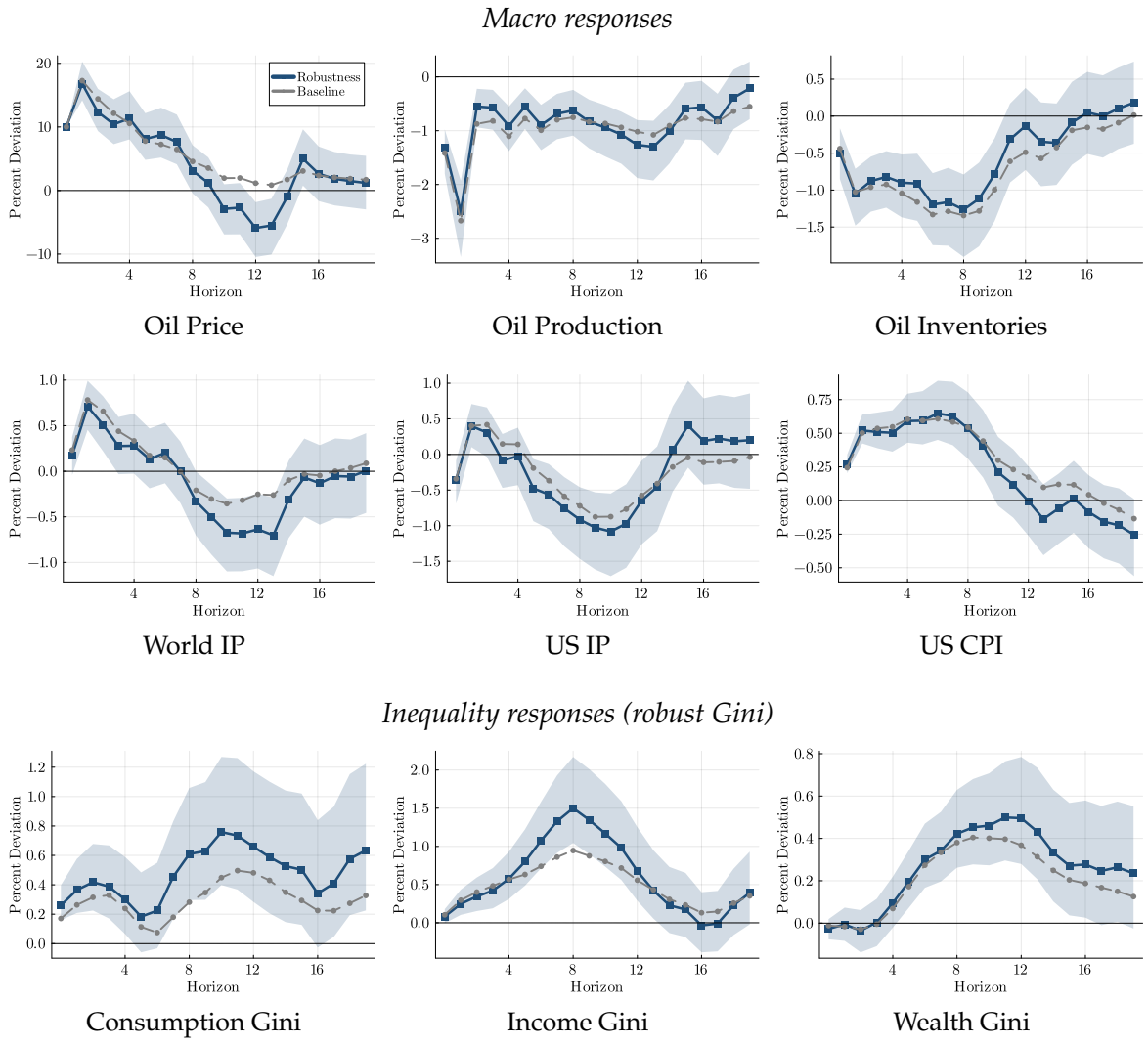


Notes: Solid + 68% band: robustness (Legacy oilstocksM). Dashed gray: baseline. Macro responses are level IRFs.

F.2 Mori-Peersman Financial Block

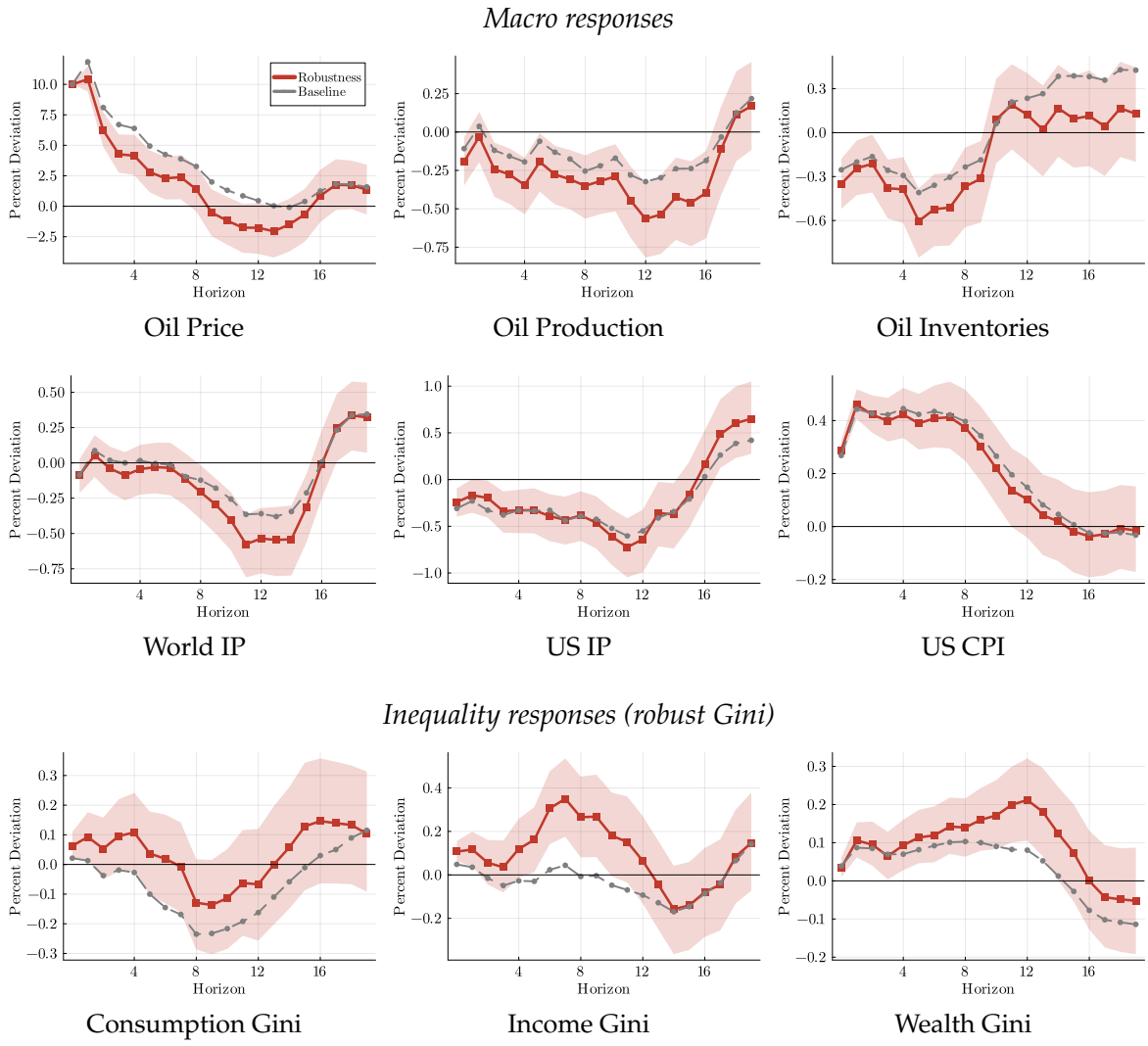
I augment the baseline VAR with the Mori-Peersman financial block in two variants. The bare *MP* block adds the 1-year Treasury yield, S&P 500, and VIX. The *MP+EBP* block additionally appends the Gilchrist-Zakrajšek excess bond premium (`ebp`). The motivation is the Mori and Peersman (2024) predictability concern: at quarterly frequency, my own Granger tests (Appendix A) point to the EBP as the predictor for the KZ shock, while interest rates predict the BH shock. The MP block is the literature-matched specification; MP+EBP additionally controls for the EBP channel that binds at quarterly frequency.

Figure 18: MP financial block (GS1, S&P 500, VIX) — Baumeister and Hamilton (2019) supply shock



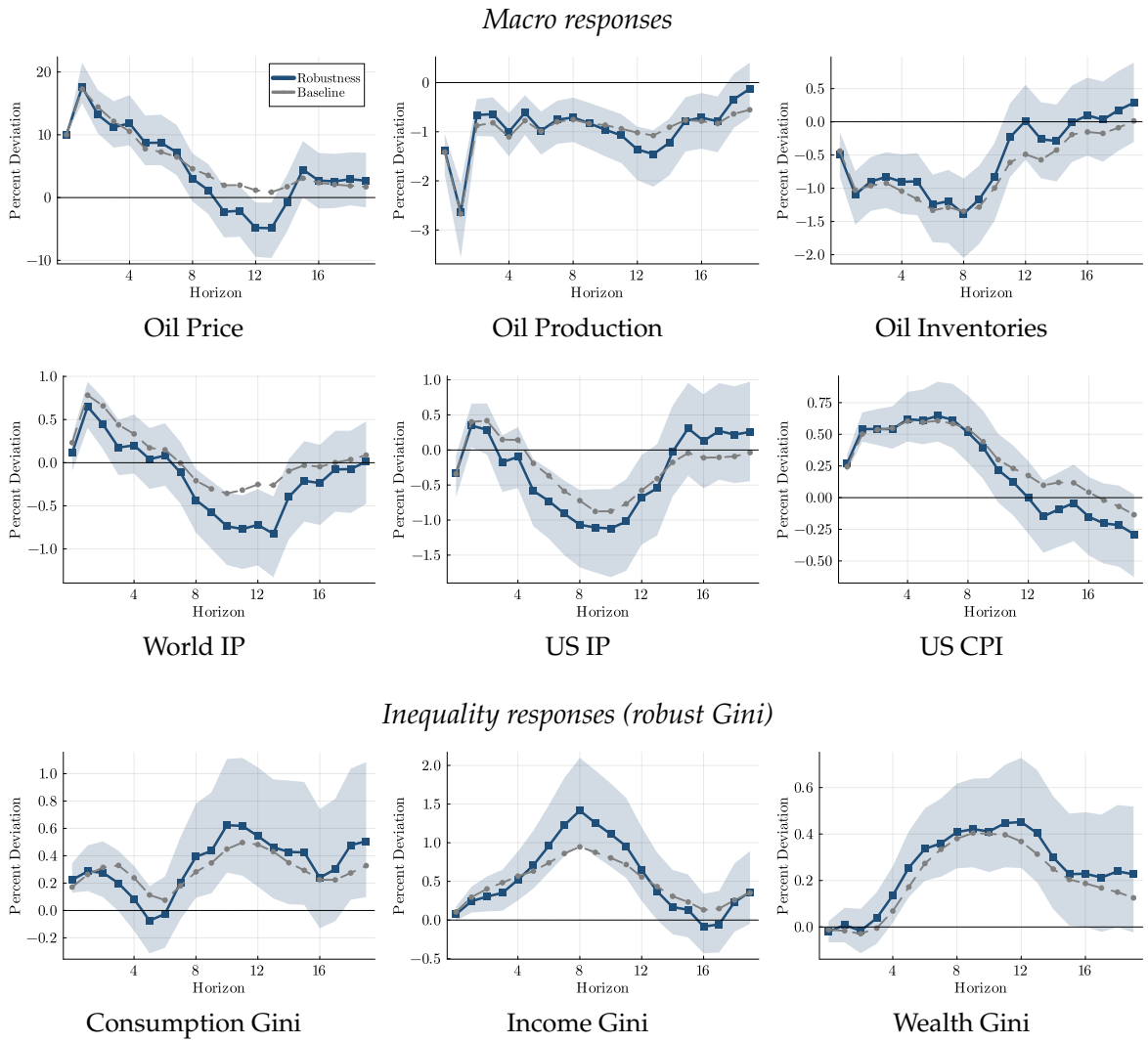
Notes: Solid + 68% band: robustness (MP financial block (GS1, S&P 500, VIX)). Dashed gray: baseline. Macro responses are level IRFs.

Figure 19: MP financial block (GS1, S&P 500, VIX) — Känzig (2021) supply news shock



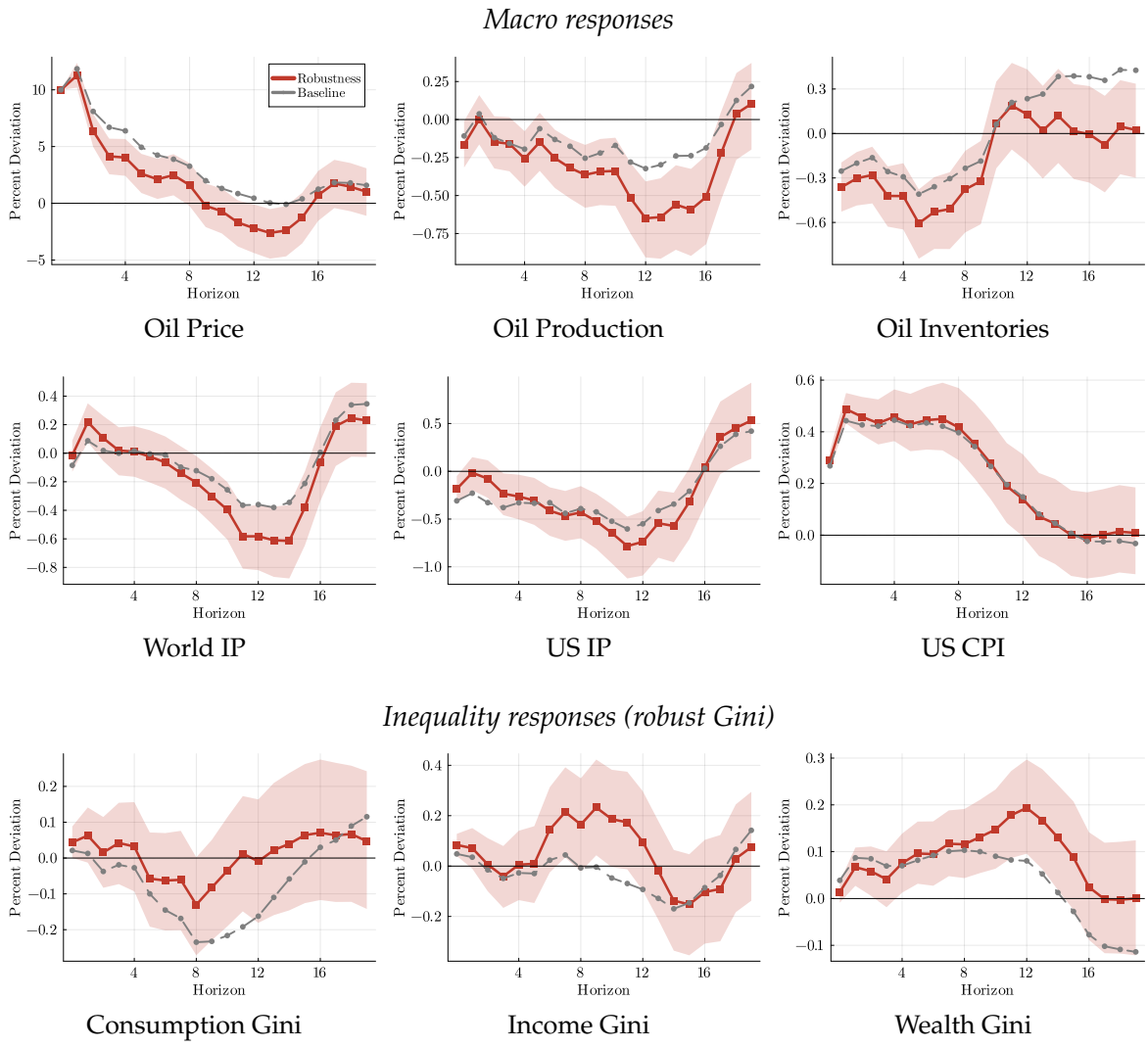
Notes: Solid + 68% band: robustness (MP financial block (GS1, S&P 500, VIX)). Dashed gray: baseline. Macro responses are level IRFs.

Figure 20: MP + EBP financial block — Baumeister and Hamilton (2019) supply shock



Notes: Solid + 68% band: robustness (MP + EBP financial block). Dashed gray: baseline. Macro responses are level IRFs.

Figure 21: MP + EBP financial block — Känzig (2021) supply news shock

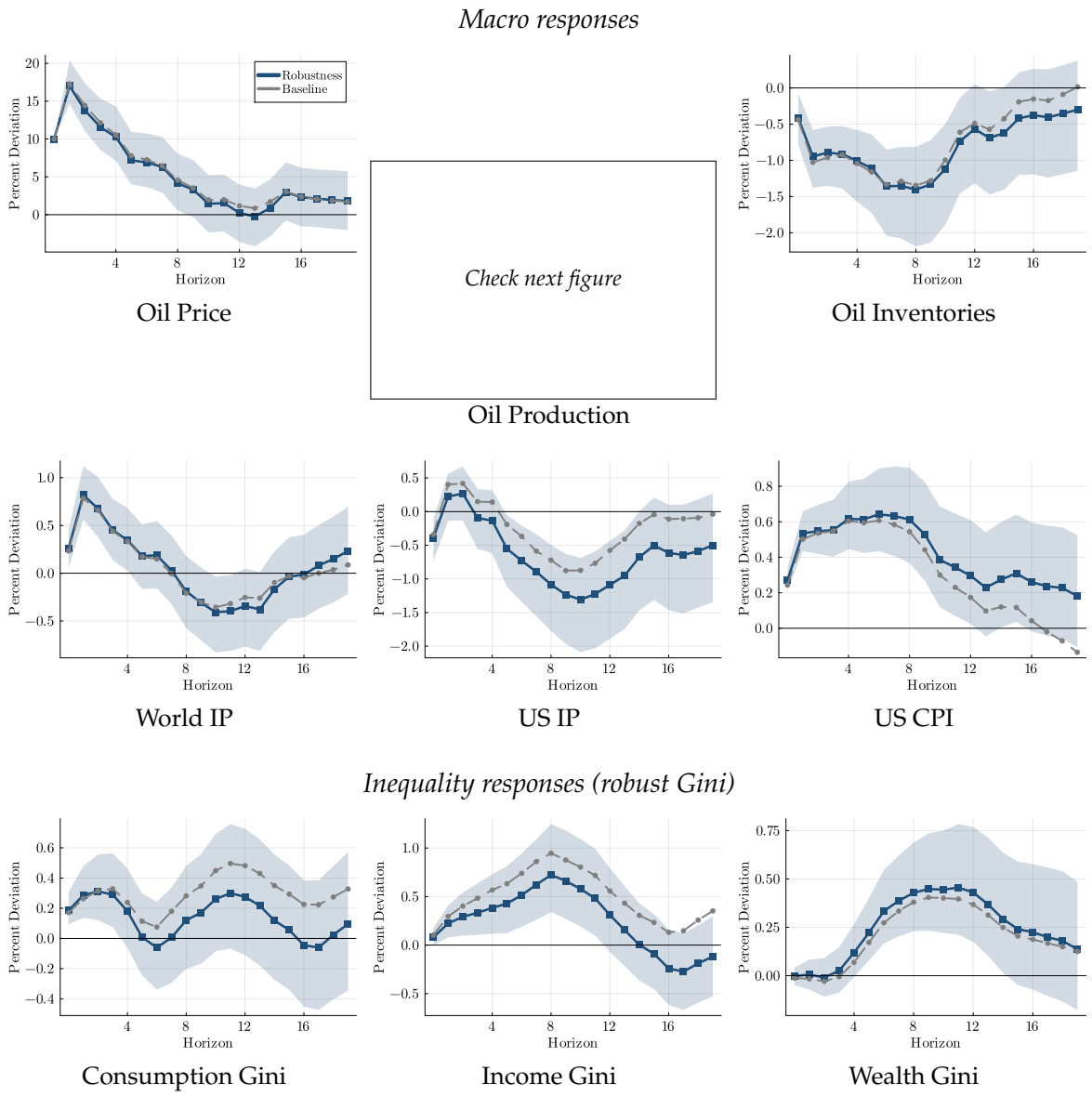


Notes: Solid + 68% band: robustness (MP + EBP financial block). Dashed gray: baseline. Macro responses are level IRFs.

F.3 OPEC / Non-OPEC Production Split

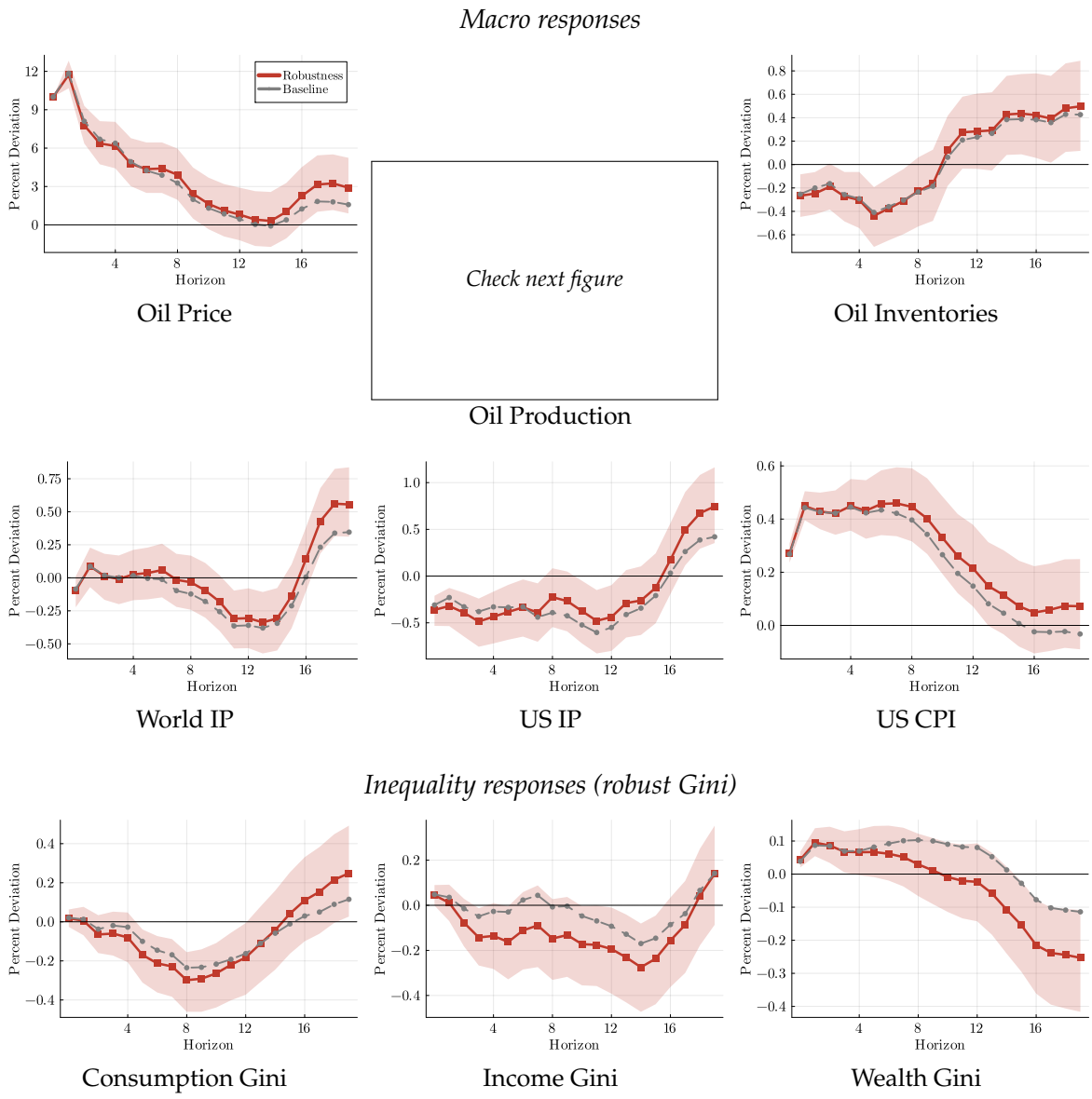
Following Kolodzeij and Kaufmann (2014), who argue that OPEC and non-OPEC producers operate under different output-setting criteria so that reductions in OPEC production tend to raise oil prices while non-OPEC movements act differently, I replace the aggregate `oilprod` variable with two components: `oilprod_opec` (OPEC production) and `oilprod_nopec` (Non-OPEC production). Series are constructed by applying EIA INTL OPEC and Non-OPEC shares (X-12-ARIMA seasonally adjusted) to the baseline `oilprod` level, so that $\exp(\texttt{oilprod\_opec}) + \exp(\texttt{oilprod\_nopec}) = \exp(\texttt{oilprod})$ by construction. This tests whether baseline supply-shock dynamics are sensitive to treating OPEC discipline separately from non-OPEC supply (US shale, North Sea, etc.). The `oilprod` panel below is absent because the OPEC specification does not contain an aggregate production variable; the OPEC and non-OPEC component IRFs are reported separately in Figure 24.

Figure 22: OPEC + Non-OPEC split — Baumeister and Hamilton (2019) supply shock



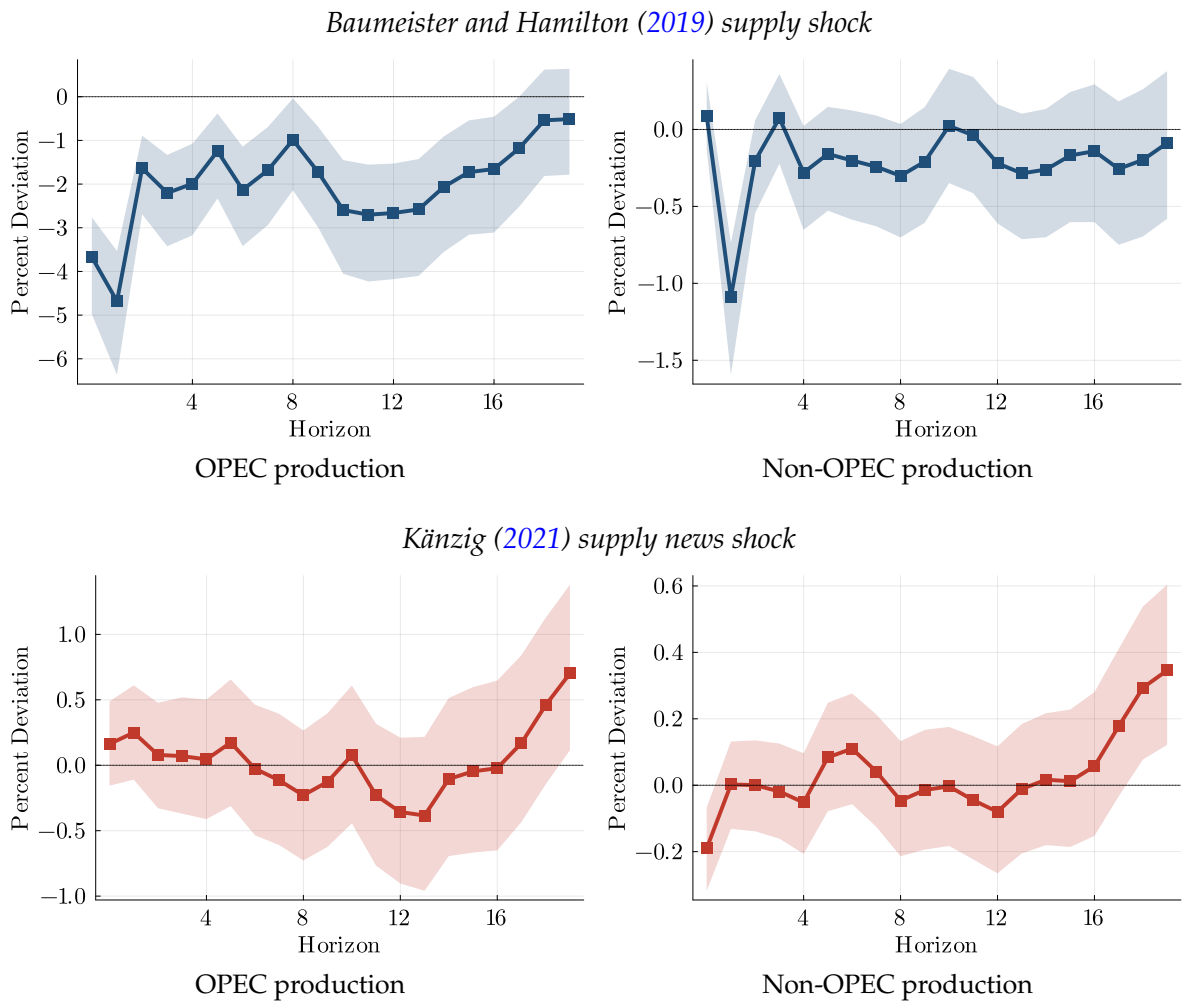
Notes: Solid + 68% band: robustness (OPEC + Non-OPEC split). Dashed gray: baseline. Macro responses are level IRFs.

Figure 23: OPEC + Non-OPEC split — Känzig (2021) supply news shock



Notes: Solid + 68% band: robustness (OPEC + Non-OPEC split). Dashed gray: baseline. Macro responses are level IRFs.

Figure 24: OPEC and Non-OPEC production responses (OPEC-split specification)

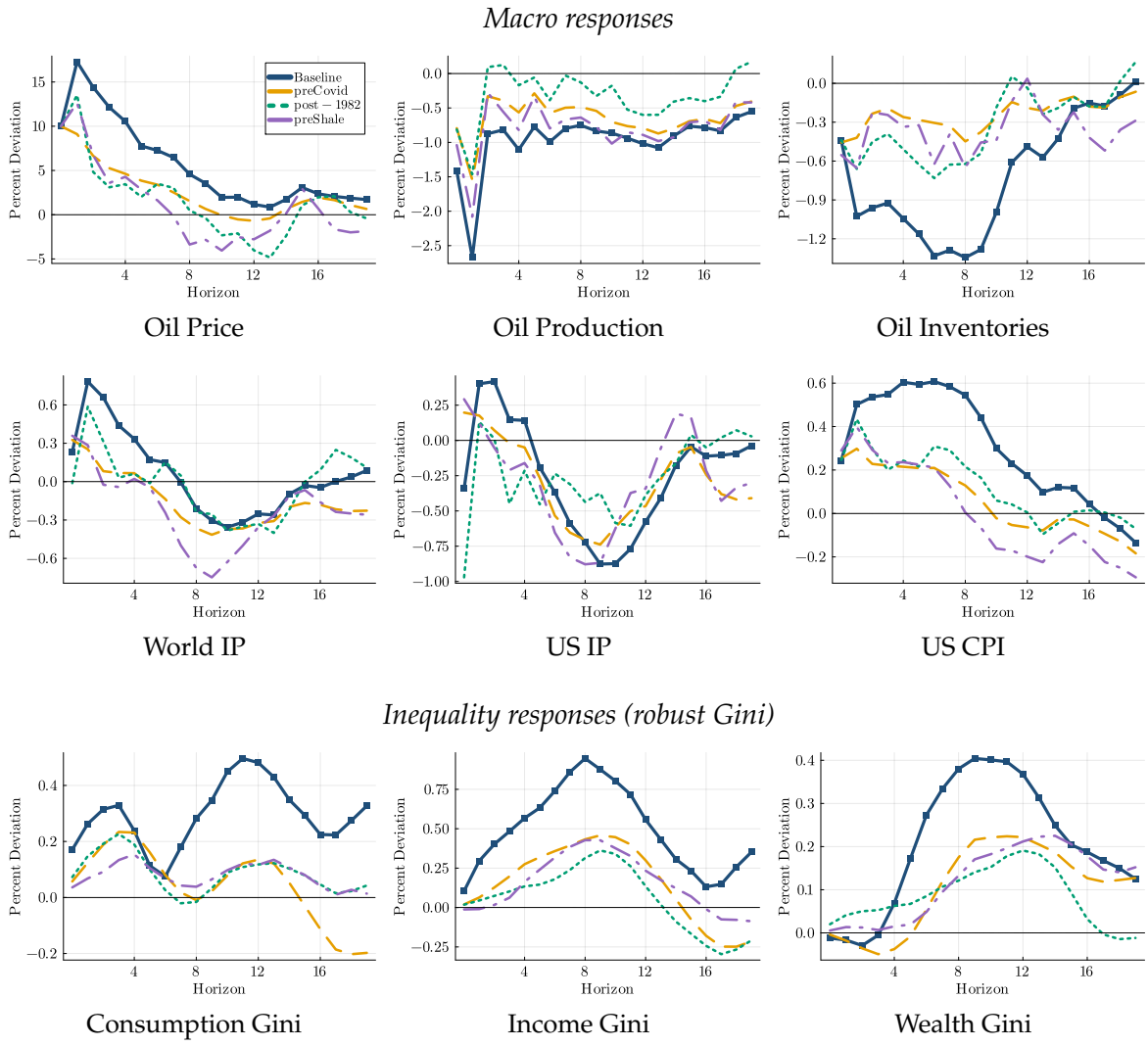


Notes: Solid line: posterior median; shaded band: 68% credible set. The OPEC-split specification replaces `oilprod` with `oilprod_opec` and `oilprod_nopec`; these IRFs have no baseline counterpart in the baseline VAR (where only the aggregate `oilprod` is identified) and are reported on their own.

F.4 Subsamples

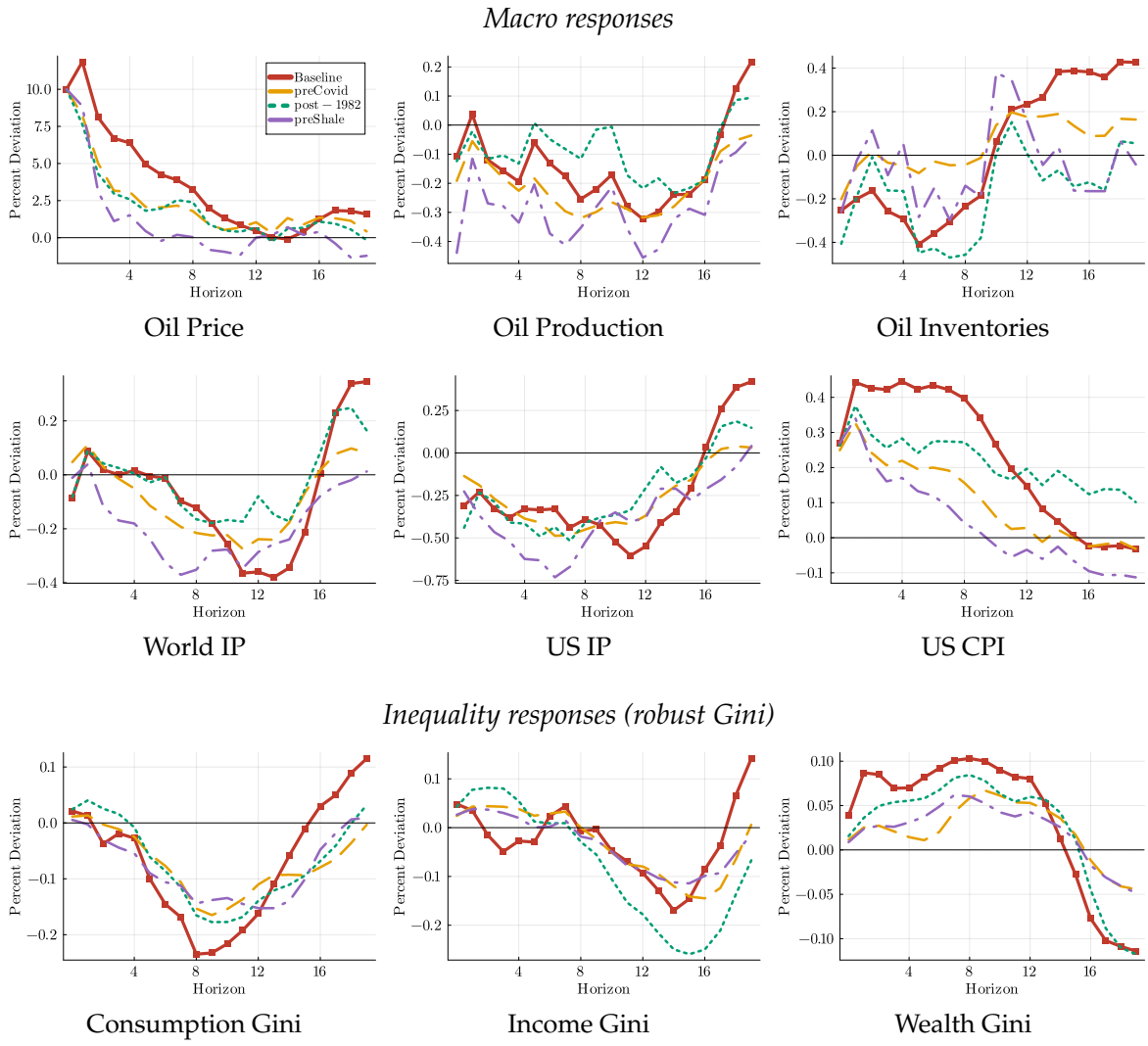
I re-estimate the baseline VAR on three subsamples: *preCovid* (truncated at 2019Q4 to exclude the pandemic and its aftermath), *post1982* (starting in 1982Q1 to drop the early Volcker disinflation period), and *preShale* (truncated at 2010Q4, before the U.S. shale-oil revolution materially raised the medium-run elasticity of supply). The post-1982 cut is also motivated by Nakov and Pescatori (2010), who note that the period around 1981 marks a regime break in world oil markets, in U.S. energy production and consumption, and in monetary policy and inflation; Hooker (2002) further documents that the pass-through from oil prices to inflation weakens substantially after this break, so the subsample doubles as a check that the inflation responses are not artifacts of pre-Volcker passthrough.

Figure 25: Subsample sweep — Baumeister and Hamilton (2019) supply shock



Notes: Each panel: baseline (solid, shock-aware color, full sample) + three dashed lines for *preCovid* ($\leq$ 2019Q4), *post-1982* ($\geq$ 1982Q1), and *preShale* ($\leq$ 2010Q4). No credible bands. Macro responses are level IRFs.

Figure 26: Subsample sweep — Känzig (2021) supply news shock

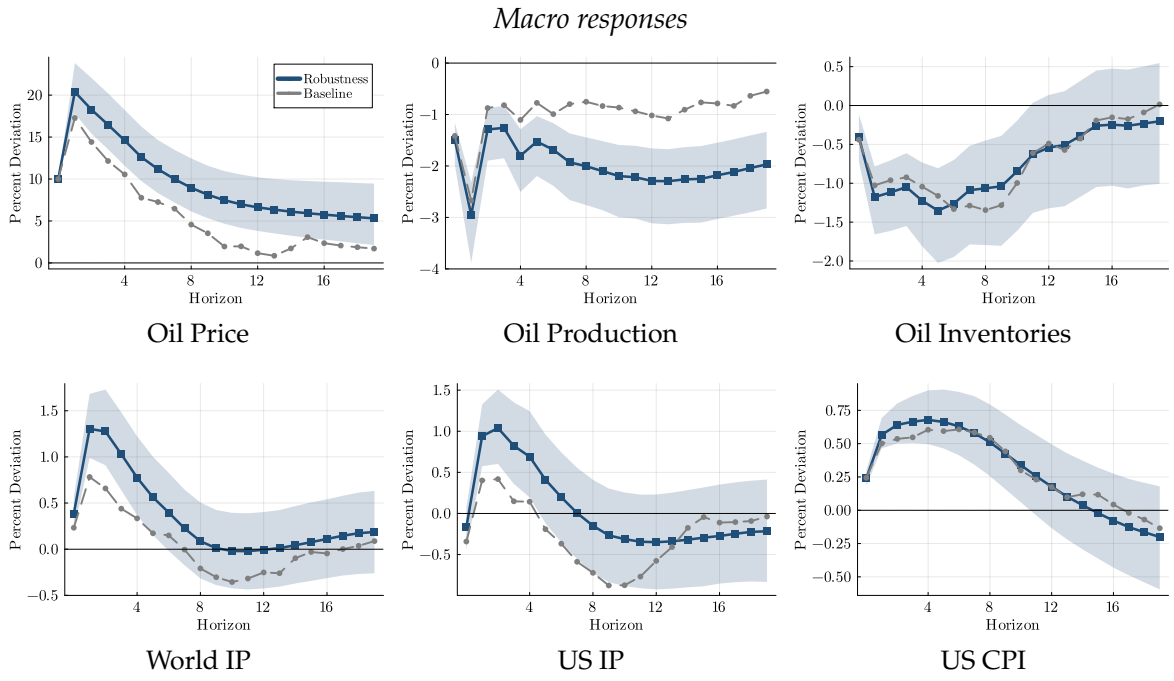


Notes: Each panel: baseline (solid, shock-aware color, full sample) + three dashed lines for *preCovid* ($\leq$ 2019Q4), *post-1982* ($\geq$ 1982Q1), and *preShale* ($\leq$ 2010Q4). No credible bands. Macro responses are level IRFs.

F.5 Aggregates-Only Specification

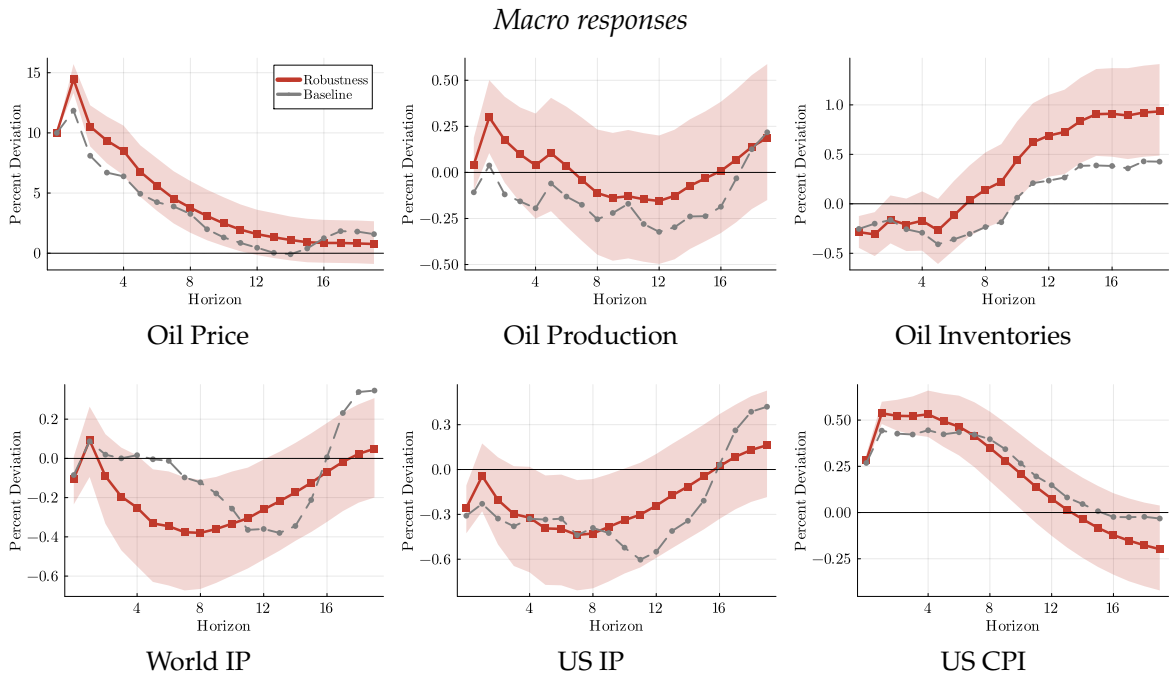
I drop the distributional block from the VAR. This is the standard small SVAR; the comparison reveals whether including the distribution changes the aggregate IRFs. (Inequality panels are absent because the model has no distributional measures.)

Figure 27: Aggregates only — Baumeister and Hamilton (2019) supply shock



Notes: Solid + 68% band: robustness (Aggregates only). Dashed gray: baseline. Macro responses are level IRFs.

Figure 28: Aggregates only — Känzig (2021) supply news shock

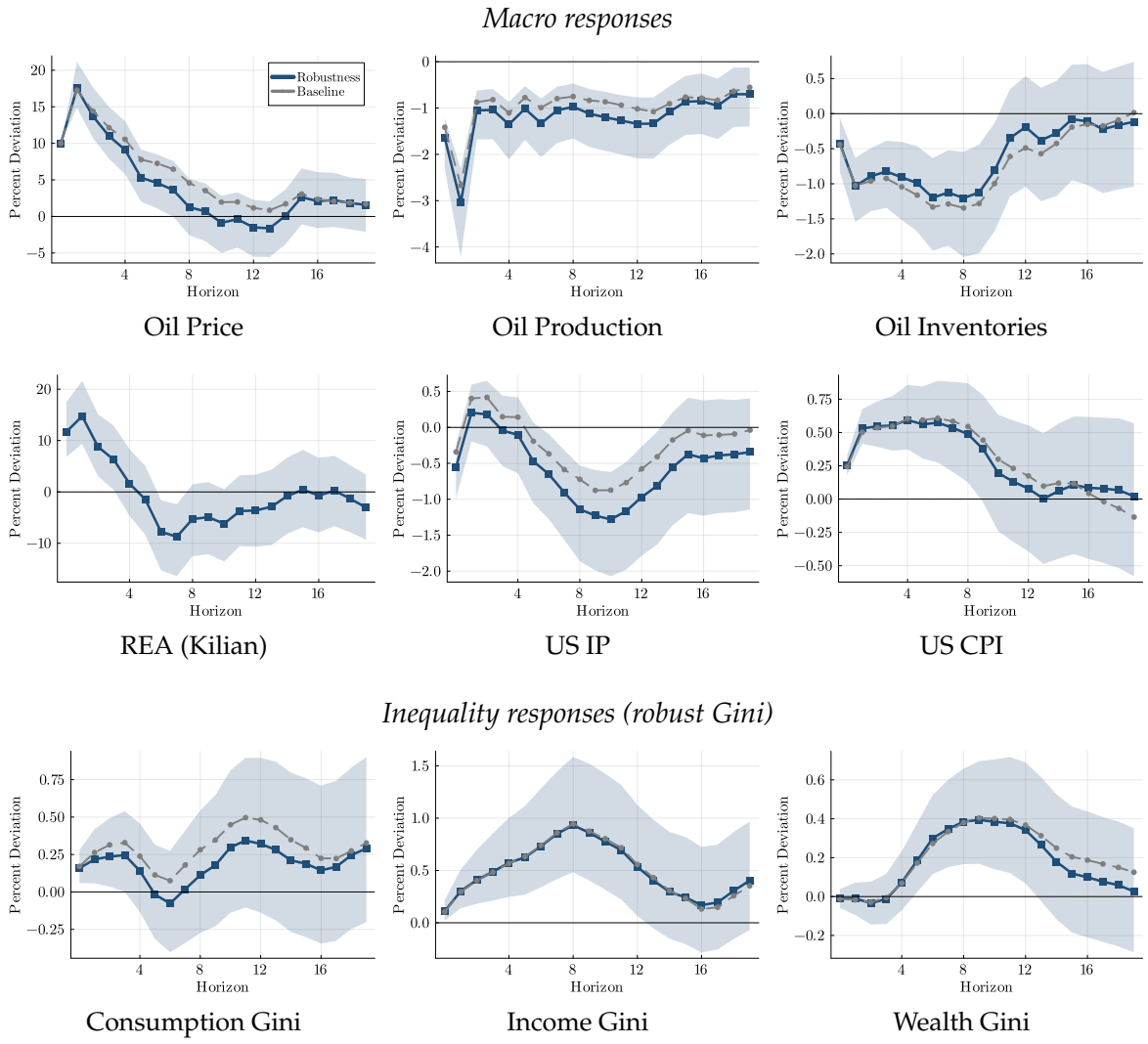


Notes: Solid + 68% band: robustness (Aggregates only). Dashed gray: baseline. Macro responses are level IRFs.

F.6 Kilian REA in Place of World IP

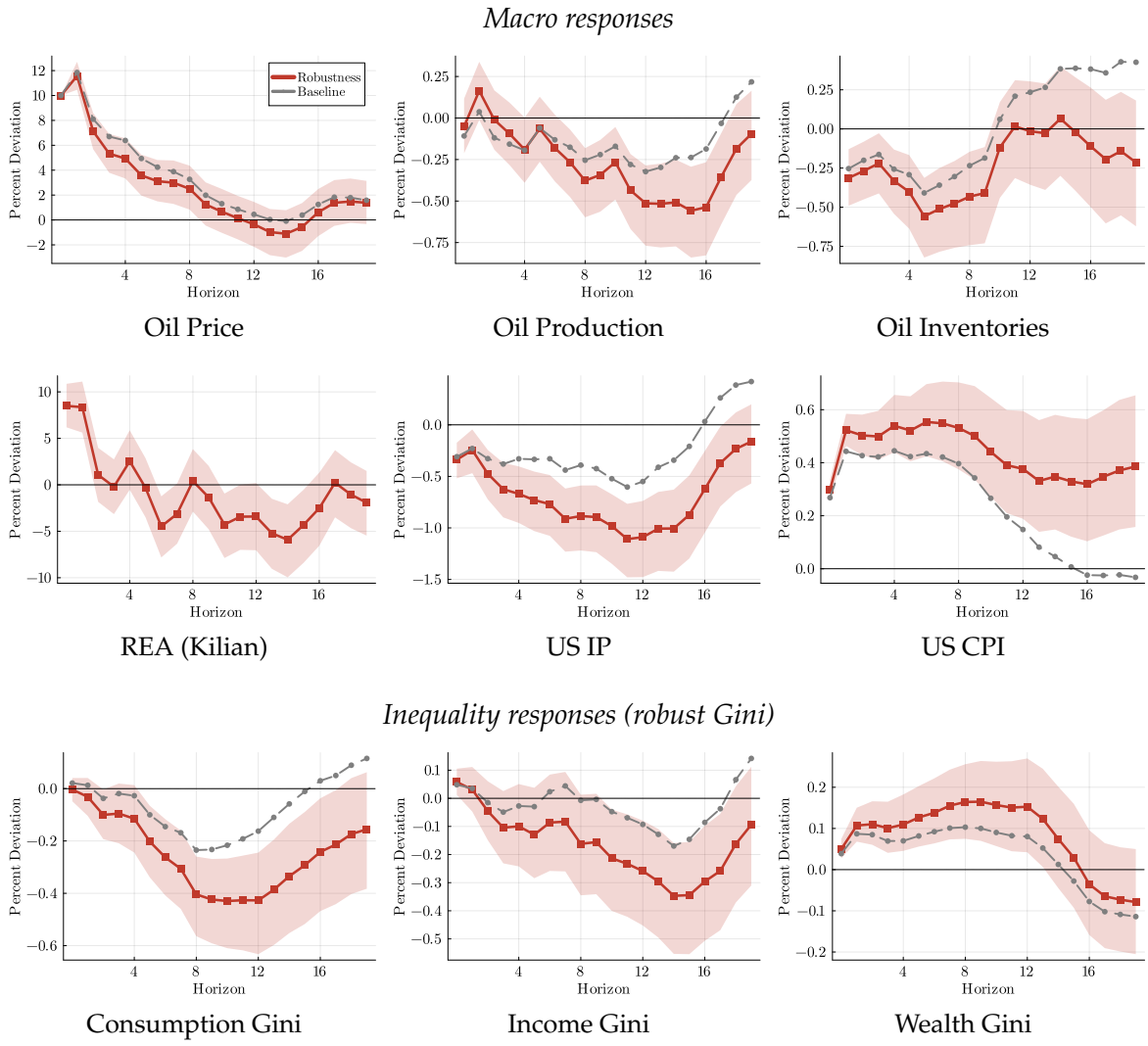
I replace world industrial production with Kilian's (2009a) Real Economic Activity index (deflated dry-cargo shipping rate, detrended). The data window is determined by the REA series.

Figure 29: Kilian REA — Baumeister and Hamilton (2019) supply shock



Notes: Solid + 68% band: robustness (Kilian REA). Dashed gray: baseline. *World IP / REA panel*: only the robustness REA series is shown because the two measures are on different scales. Macro responses are level IRFs.

Figure 30: Kilian REA — Känzig (2021) supply news shock

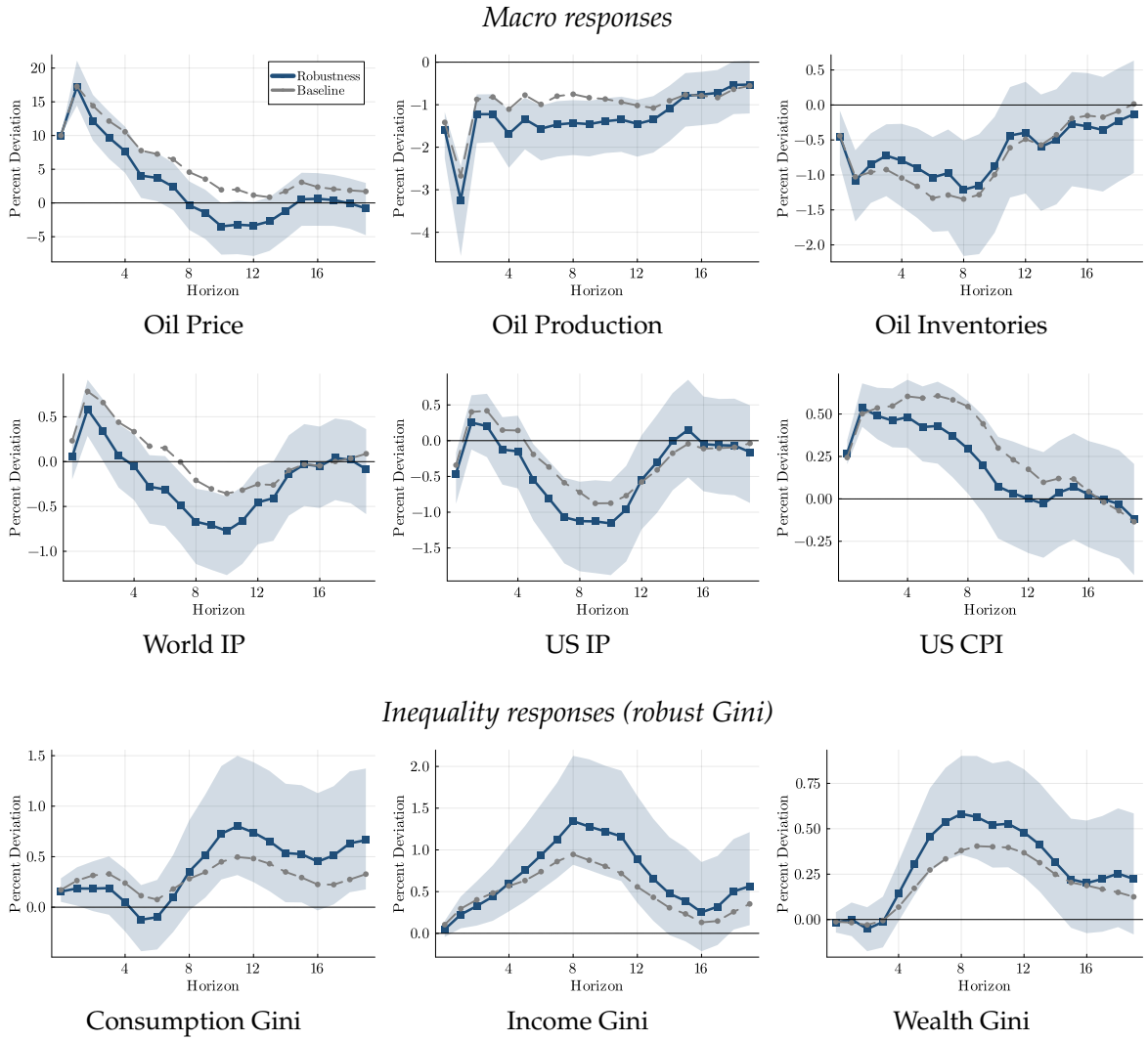


Notes: Solid + 68% band: robustness (Kilian REA). Dashed gray: baseline. *World IP / REA panel*: only the robustness REA series is shown because the two measures are on different scales. Macro responses are level IRFs.

F.7 QD-Levels Macro Factors

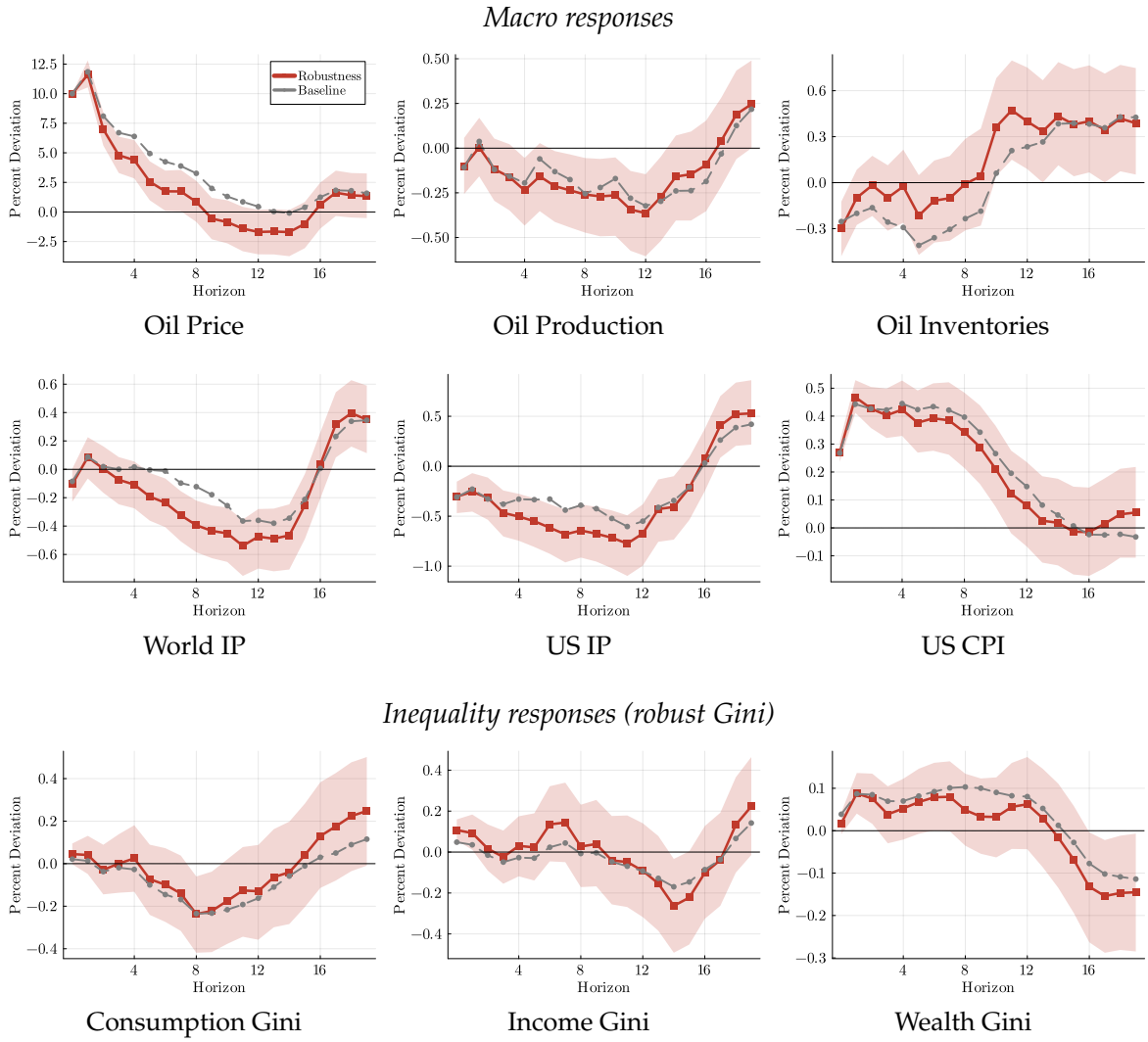
I re-estimate the baseline VAR augmenting the macro block with the first principal components of the FRED-QD log-level macro panel (Bai 2004 non-stationary PCA). This guards against the baseline result being an artefact of a six-variable macro block missing common factors.

Figure 31: QD-levels factors — Baumeister and Hamilton (2019) supply shock



Notes: Solid + 68% band: robustness (QD-levels factors). Dashed gray: baseline. Macro responses are level IRFs.

Figure 32: QD-levels factors — Känzig (2021) supply news shock



Notes: Solid + 68% band: robustness (QD-levels factors). Dashed gray: baseline. Macro responses are level IRFs.

G Prior for the Long Run: Construction and Diagnostic

This appendix details the construction of the prior for the long run (PLR) used in estimation and reports the auto-detection output for the baseline VARs (`paper_bh_smfdd` and `paper_kanzig_smfdd`: shock instrument + 6 macroeconomic variables + 8 smoothed distributional factors = 15 endogenous variables, 16 lags, 1975–2024 quarterly).

Dummy-observation construction. The PLR of Giannone, Lenza, and Primiceri (2019) regularizes the long-run multiplier matrix $\boldsymbol{A}(1) = \boldsymbol{I} - \boldsymbol{A}_1 - \cdots - \boldsymbol{A}_p$ by augmenting the likelihood with n artificial observations $(\boldsymbol{Y}_d, \boldsymbol{X}_d)$ prepended to the data. The dummy in row i encodes the prior belief that the long-run combination $\boldsymbol{H}_{i\cdot}\boldsymbol{Y}_t$ has mean $\bar{y}_{0,i}$, with strength inversely proportional to a tightness parameter ϕ_i:

$$\boldsymbol{Y}_{d,i,\cdot} \;=\; \frac{\boldsymbol{H}_{i\cdot}\,\bar{\boldsymbol{y}}_0}{\phi_i}\,[\boldsymbol{H}^{-1}]_{\cdot i}, \tag{44}$$

with $\phi_i \to 0$ collapsing the prior toward exact cointegration in direction $\boldsymbol{H}_{i\cdot}$ and $\phi_i \to \infty$ removing it.

Adaptation to the asymmetric conjugate prior. The PLR was originally derived for the symmetric Normal–Inverse-Wishart prior of (9). I extend the dummy-observation construction to the per-equation NIG marginal likelihood of (13) so that all priors operate on the same augmented data, and the joint marginal-likelihood optimization remains internally consistent. This resolves the implementation issue that had previously made PLR incompatible with the Chan (2022) per-equation prior structure I use in estimation.

Joint optimization. The PLR tightness vector $\boldsymbol{\phi}$ is optimized *jointly* with the Minnesota hyperparameters $(\kappa_1, \kappa_2, \lambda_3)$ at the system level by maximizing the closed-form marginal likelihood under (13) on the dummy-augmented data; the per-equation persistence parameters δ_i that would otherwise be system-level (and discarded once the per-equation grid in (15) fires) are held at loose defaults during this stage. The resulting structure—a Chan asymmetric conjugate base augmented with PLR dummies, with both stages selected jointly via marginal likelihood—accommodates the mixed persistence of my system (an $I(0)$ shock instrument, near-$I(1)$ macroeconomic aggregates, and $I(0)$ distributional factors) without requiring the researcher to take a stand on individual unit root tests, which tend to have low power (Sims and Zha, 1998).

Auto-detection of $\boldsymbol{H}$. I construct $\boldsymbol{H}$ data-adaptively: each variable is classified $I(0)$ or $I(1)$ via an augmented Dickey–Fuller test *with deterministic trend* (MacKinnon (1994) critical value -3.41 at 5%), and pairs of $I(1)$ variables whose first difference is itself stationary are encoded as cointegrating rows of $\boldsymbol{H}$ in the form $\boldsymbol{e}_i - \boldsymbol{e}_j$. The trend-included specification is essential: the no-trend ADF can spuriously reject the unit root for series with strong deterministic components by absorbing the trend into the intercept.

ADF classification with deterministic trend. Table 6 reports the trend-included ADF t-statistic for each of the 14 endogenous variables. The 5 % MacKinnon (1994) critical value is -3.41. Variables with $t < -3.41$ are classified $I(0)$; the rest are $I(1)$ candidates for the cointegration step.

Table 6: ADF unit-root tests (with deterministic trend), baseline VAR

Variable	t-stat	Class	Row interpretation in $\boldsymbol{H}$
`poil`	-2.03	$I(1)$	identity
`oilprod`	-3.08	$I(1)$	cointegrating: `oilprod` $-$ `oilstocksM_BH`
`worldip`	-2.20	$I(1)$	identity
`usip`	-1.44	$I(1)$	identity
`log_cpi`	-4.06	$I(0)$	identity (trend-stationary)
`oilstocksM_BH`	-2.57	$I(1)$	identity
`dist_f1`	-7.40	$I(0)$	identity (stationary)
`dist_f2`	-6.91	$I(0)$	identity (stationary)
`dist_f3`	-7.28	$I(0)$	identity (stationary)
`dist_f4`	-5.58	$I(0)$	identity (stationary)
`dist_f5`	-5.32	$I(0)$	identity (stationary)
`dist_f6`	-5.13	$I(0)$	identity (stationary)
`dist_f7`	-5.45	$I(0)$	identity (stationary)
`dist_f8`	-5.13	$I(0)$	identity (stationary)

Note: ADF specification: $\Delta y_t = \alpha + \delta\, t + \gamma\, y_{t-1} + \sum_{j=1}^{p^*} \beta_j\, \Delta y_{t-j} + \varepsilon_t$, with $p^* = \lfloor (T-1)^{1/3} \rfloor$ capped at 12. MacKinnon (1994) 5 % critical value -3.41. The trend term is essential for log-level macroeconomic series with non-zero drift; without it, the regression absorbs the deterministic trend into the intercept and over-rejects the unit root for variables with persistent upward levels (e.g., `oilstocksM_BH` returns $t = -4.11$ under the no-trend specification).

$\boldsymbol{H}$ matrix. The auto-built $\boldsymbol{H}$ is 15×15 with two cointegrating rows; the rest are identity. The leading detected pair is

$$\boldsymbol{e}_{\texttt{oilprod}} - \boldsymbol{e}_{\texttt{oilstocksM\_BH}},$$

which combines into $\log(\texttt{oilprod}) - \log(\texttt{oilstocksM\_BH}) = -\log(\text{days of cover})$, the negative log of the OECD inventory-to-production ratio (bounded by storage capacity and operational requirements).

Optimal tightness vector $\boldsymbol{\phi}$. Joint marginal-likelihood optimization over $(\kappa_1, \kappa_2, \lambda_3, \phi_1, \ldots, \phi_{15})$ at the system level is run separately for each shock identification, since the shock equation dif-

fers across the BH and KZ runs and the joint marginal likelihood depends on all 15 equations:

Table 7: Optimal PLR tightness ϕ_i^* under BH and KZ identifications, baseline VARs

Row of $\boldsymbol{H}$	BH ϕ_i^*	KZ ϕ_i^*
shock instrument	62.06	**14.11**
`poil`	90.75	79.65
`oilprod`	99.55	97.53
`worldip`	99.87	99.35
`usip`	98.08	99.94
`log_cpi`	99.93	99.17
`oilstocksM_BH`	98.88	99.37
`dist_f1`	75.14	49.56
`dist_f2`	54.37	25.38
`dist_f3`	32.94	89.05
`dist_f4`	42.05	83.98
`dist_f5`	**0.37**	43.86
`dist_f6`	12.89	47.17
`dist_f7`	27.72	31.57
`dist_f8`	93.05	75.03

Note: Search range $\phi_i \in [0.01, 100]$ on a log grid; optimization via differential evolution (Storn and Price, 1997) on the closed-form Chan marginal likelihood, dummy-augmented per (44). Larger ϕ_i corresponds to a looser long-run prior on row i of $\boldsymbol{H}$. Optimal Minnesota hyperparameters: $(\kappa_1, \kappa_2, \lambda_3) = (0.156, 0.022, 0.430)$ for BH and $(0.116, 0.017, 0.361)$ for KZ. Bold entries flag the rows where the PLR is most binding under each shock.

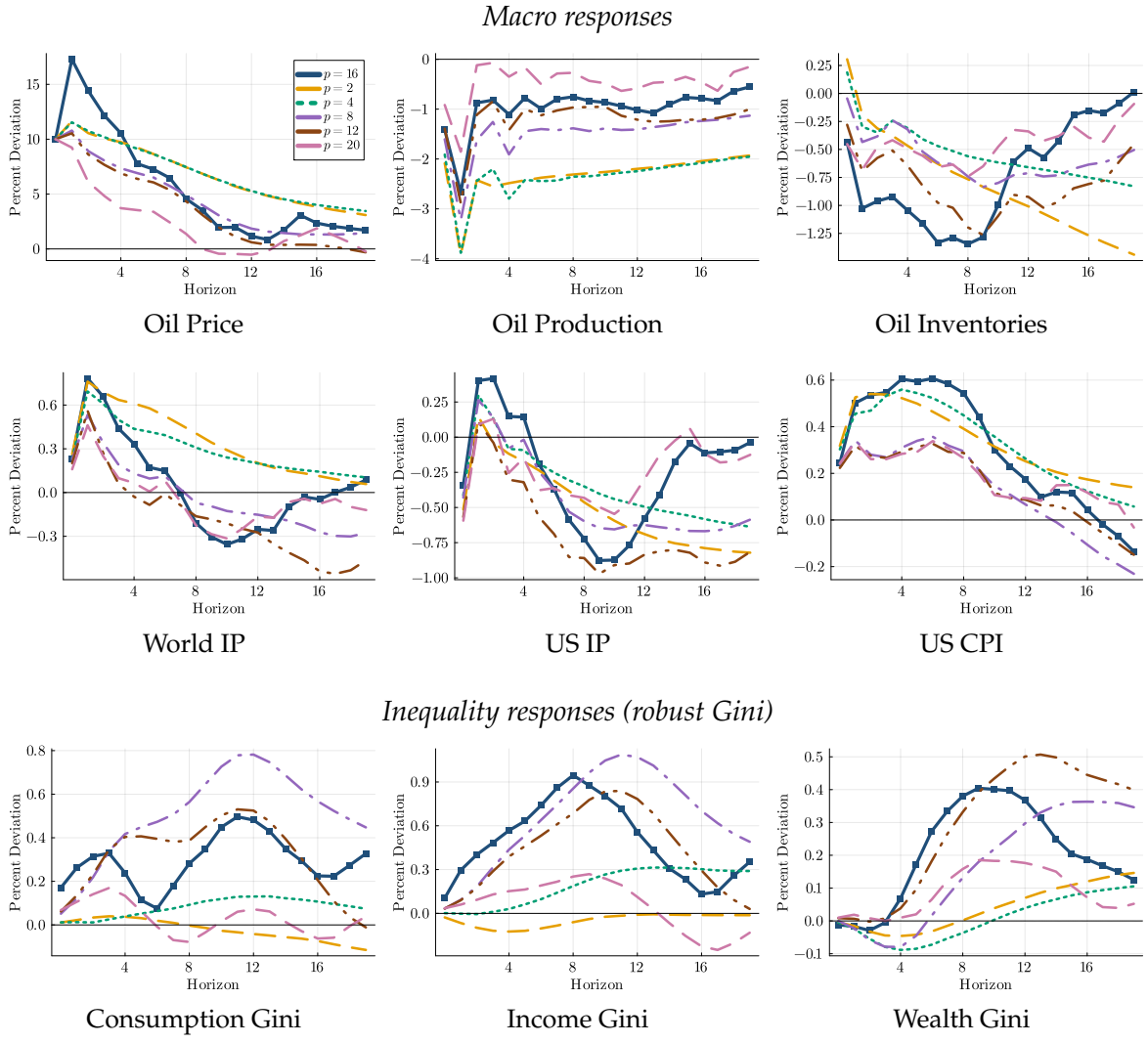
Interpretation. Three patterns stand out. First, the macro block is essentially non-binding under both shocks: `poil` sits in the loose-to-upper-bound range and the remaining five macro rows hover at the upper bound, indicating that the long-run prior contributes negligible additional shrinkage once the per-equation Chan persistence $\delta_i \approx 1$ is in force. The cointegrating row (`oilprod` $-$ `oilstocksM_BH`) receives a corner ϕ under both identifications: the auto-detected pair is interpretable as a stationary log-days-of-cover ratio (Working, 1949; Pindyck, 2001) but not statistically binding in this sample.[27] Second, the shock-instrument row is shock-specific by construction and the data prefer different long-run shrinkage on it: $\phi^*_{\text{KZ}} = 14.11$ is meaningfully tight, $\phi^*_{\text{BH}} = 62.06$ is moderate. Third, the distributional block is where the PLR does real work, and the binding row depends on the shock: $\phi^*_{D_5} = 0.37$ under BH (very tight); under KZ the dist factors take moderate values throughout, with $\phi^*_{D_2} = 25$ the tightest. For smooth, near-stationary latent factors with small steady-state means, the cointegration-style prior provides informative shrinkage that complements the per-equation Minnesota structure—exactly the asymmetry the Chan (2022) design philosophy is meant to deliver, and one that the data sharpen differently across shocks.

[27] Kilian and Murphy (2014a) report no cointegration on a pre-2009 monthly sample. I leave the auto-detected pair in the prior because the post-2009 quarterly regime makes the long-run ratio empirically more stable, and the PLR is non-binding in this row regardless.

H Robustness: Lag Order

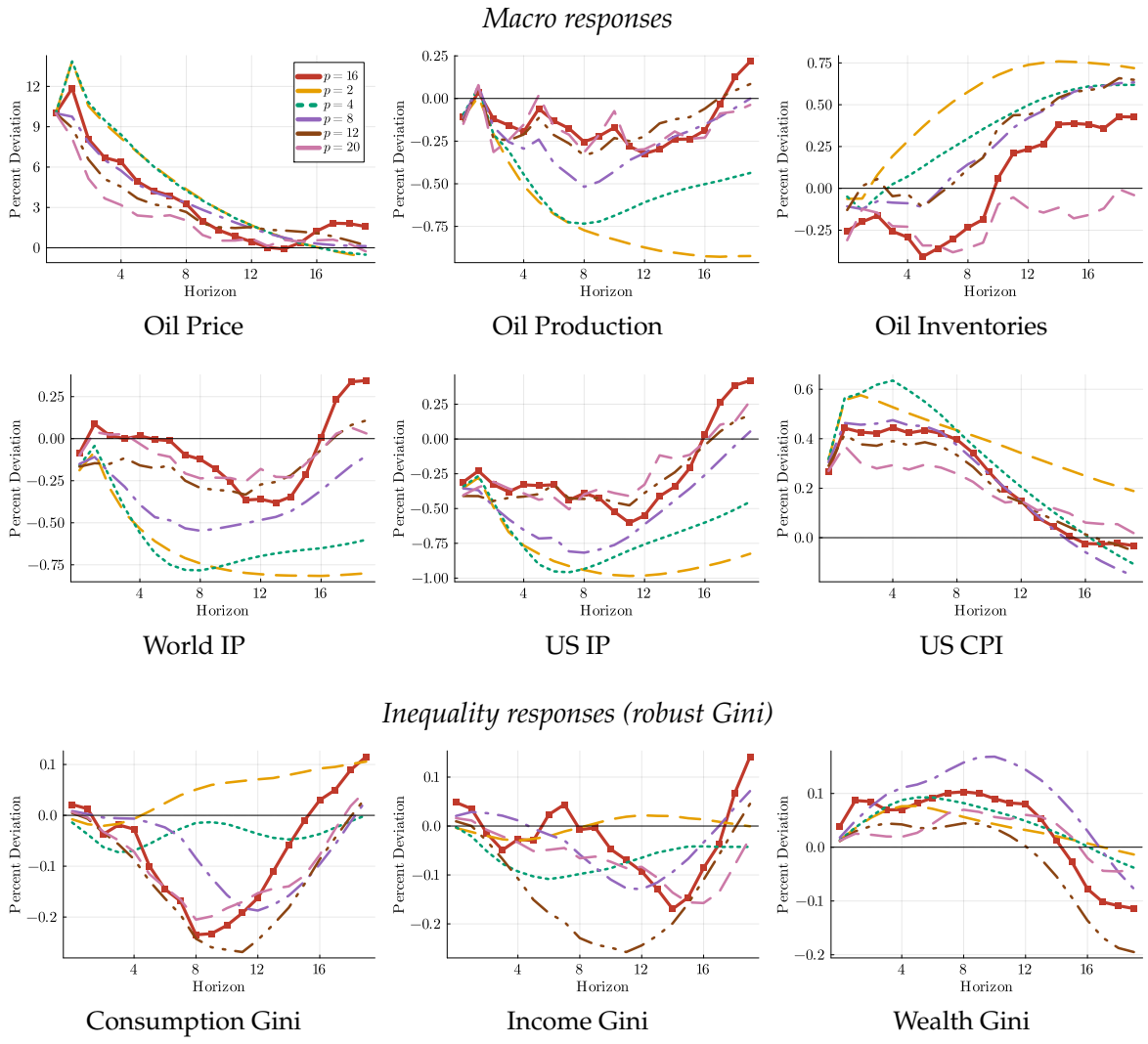
We vary the VAR lag length $p \in \{2, 4, 8, 12, 20\}$ around the baseline ($p = 16$). Shorter lags risk omitted dynamics; longer lags strain the data.

Figure 33: Lag length sweep — Baumeister and Hamilton (2019) supply shock



Notes: Each panel: baseline (solid, shock-aware color, $p = 16$) + five dashed lines for $p \in \{2, 4, 8, 12, 20\}$. No credible bands. Macro responses are level IRFs.

Figure 34: Lag length sweep — Känzig (2021) supply news shock



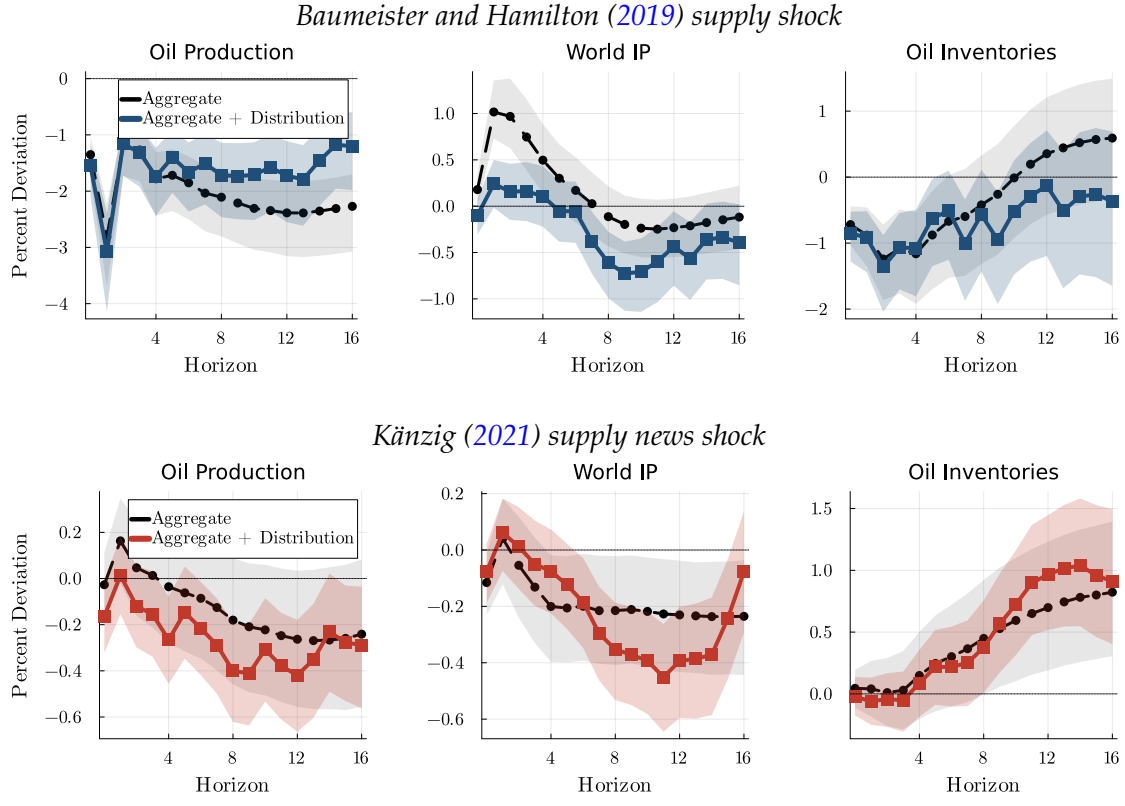
Notes: Each panel: baseline (solid, shock-aware color, $p = 16$) + five dashed lines for $p \in \{2, 4, 8, 12, 20\}$. No credible bands. Macro responses are level IRFs.

I IRFs from Rotating FAVAR

This appendix reports impulse responses from the rotating FAVAR exercise described in Section 4. Each FAVAR has the form $[\varepsilon^{\text{oil}}, \log p^{\text{oil}}, F_1, \ldots, F_7, x_t]$, where $F_1, \ldots, F_7$ are the seven stationary FRED-QD common factors (McCracken and Ng, 2021) and x_t is one of $|\mathcal{X}|$ macro and financial variables that enters the VAR ordered last. The shock is identified via the same internal-instrument scheme as the baseline VAR, and the lag length, prior, and posterior draws all match the baseline specification (Chan asymmetric conjugate prior, $p = 16$, 5,000 posterior draws, PLR enabled).

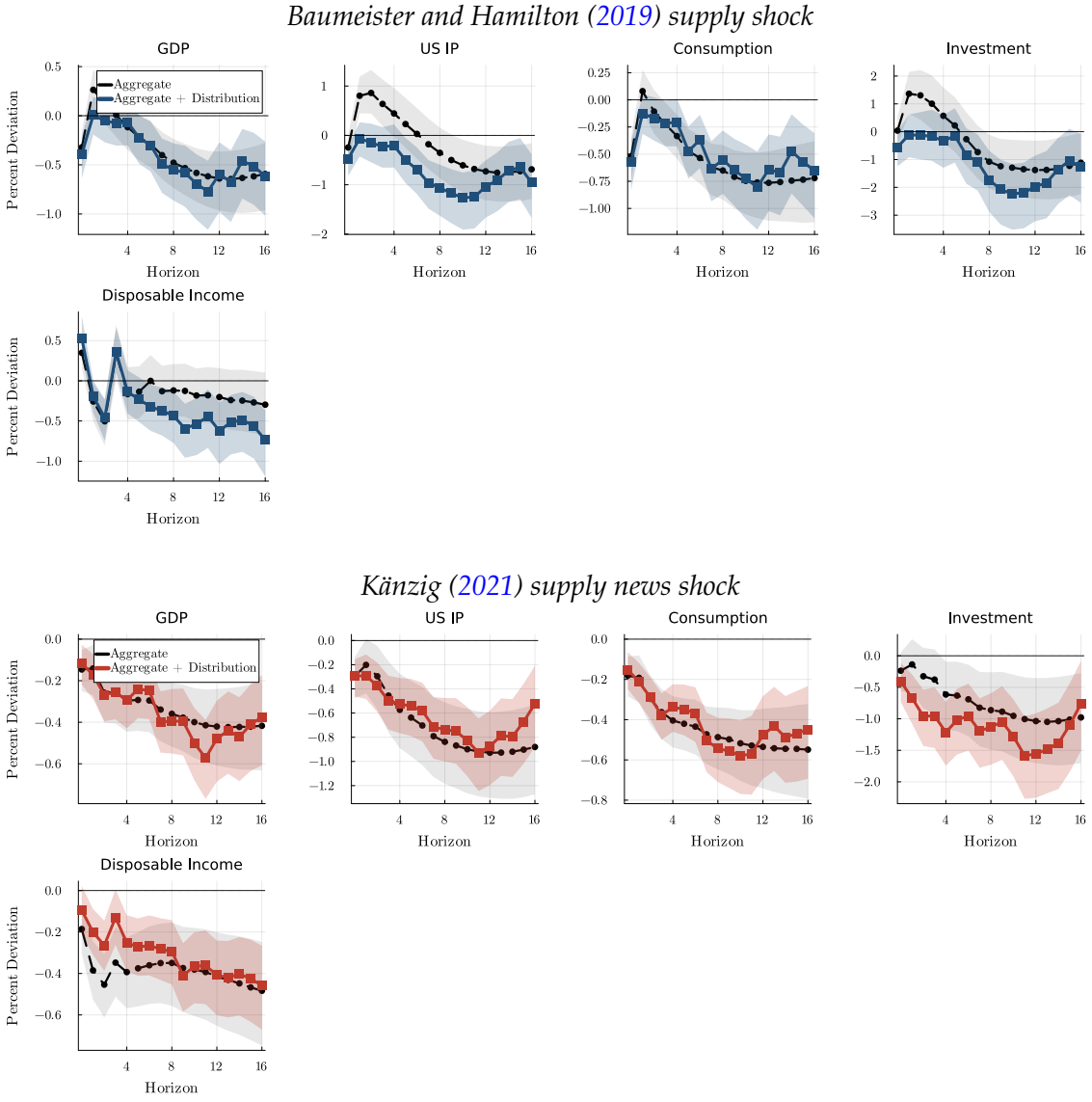
To gauge whether including the joint distribution alters the implied macro responses, each panel below overlays two IRFs for the same rotating variable: a solid line from the *aggregate-only* FAVAR (specification above), and a dashed line from a *distribution-augmented* FAVAR that appends the eight smoothed distributional factors $\boldsymbol{\xi}_t$ to the same aggregate block, $[\varepsilon^{\text{oil}}, \log p^{\text{oil}}, F_1, \ldots, F_7, \boldsymbol{\xi}_t, x$ Both lines come with 68% credible bands. The two shocks are color-coded: blue for the Baumeister and Hamilton (2019) realised supply shock and red for the Känzig (2021) supply news shock; within a figure, the dashed (distribution-augmented) line uses a darker shade of the same family. Variables are grouped into ten economic channels.

Figure 35: Rotating FAVAR — oil market



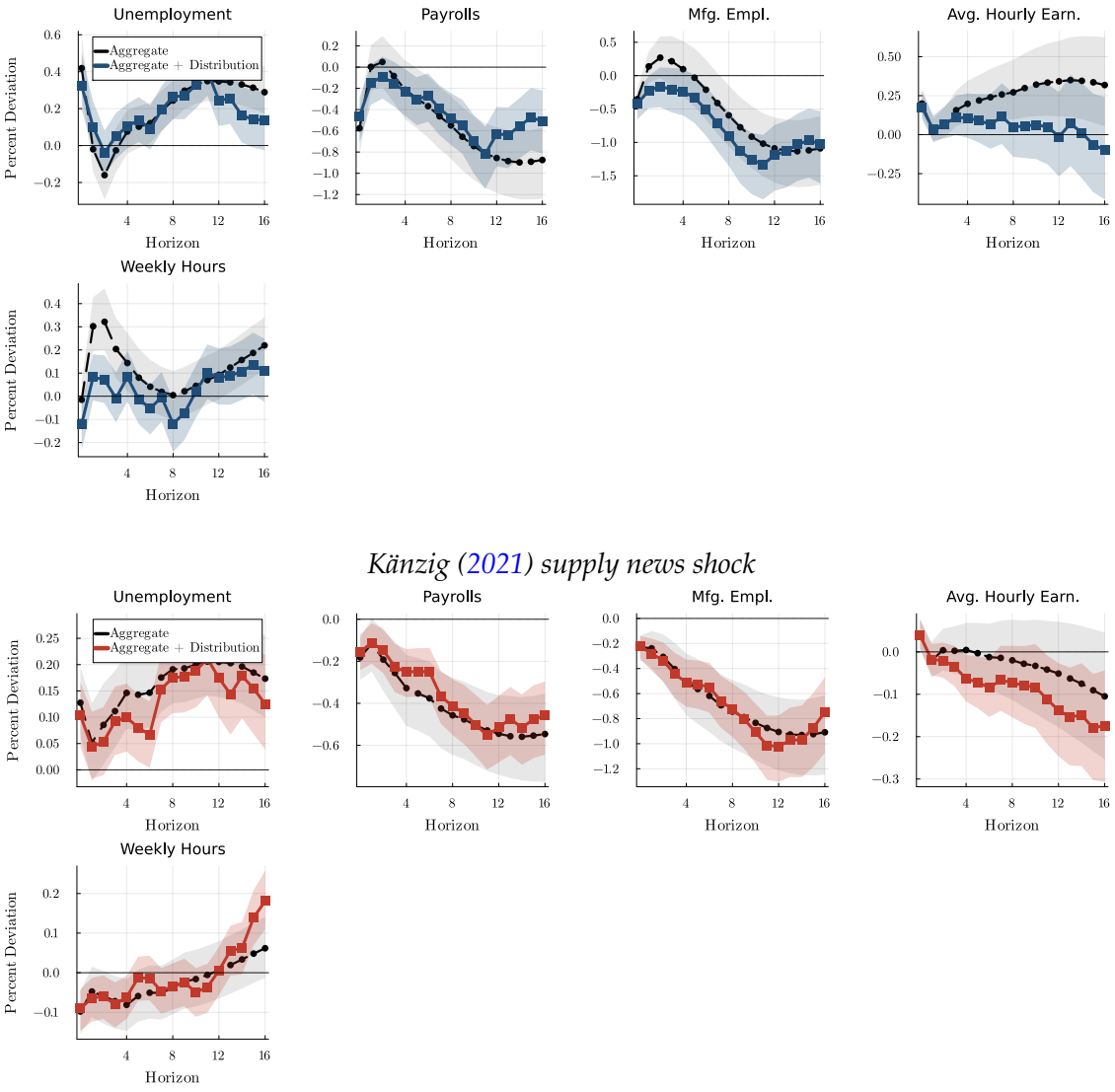
Notes: Each panel shows two IRFs for the named variable: solid line — aggregate-only FAVAR; dashed line — FAVAR augmented with the eight smoothed distributional factors $\boldsymbol{\xi}_t$. Shaded regions are 68% credible sets.

Figure 36: Rotating FAVAR — real economic activity



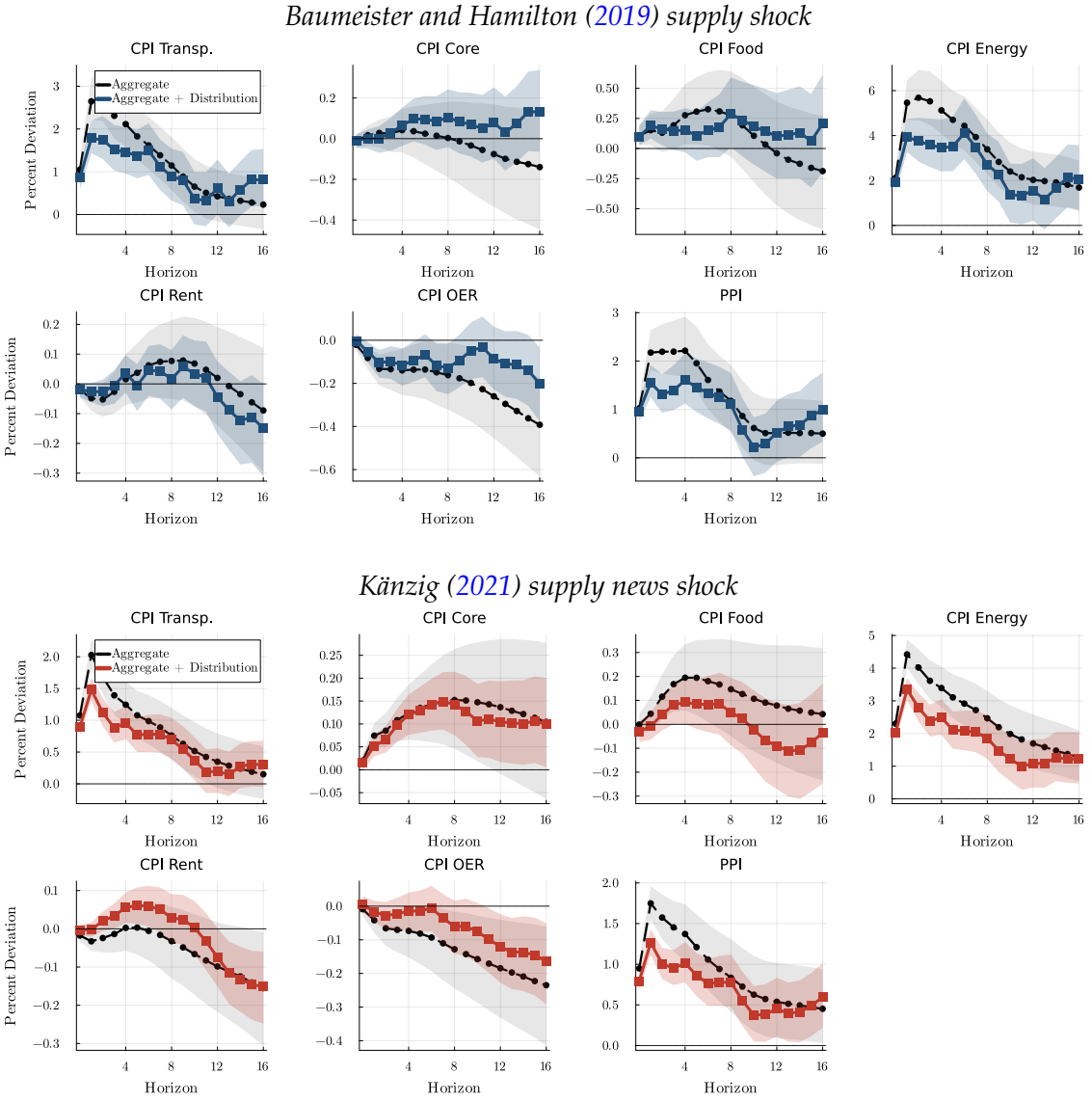
Notes: Solid: aggregate-only FAVAR. Dashed: FAVAR augmented with the eight smoothed distributional factors. Shaded regions are 68% credible sets.

Figure 37: Rotating FAVAR — labor market



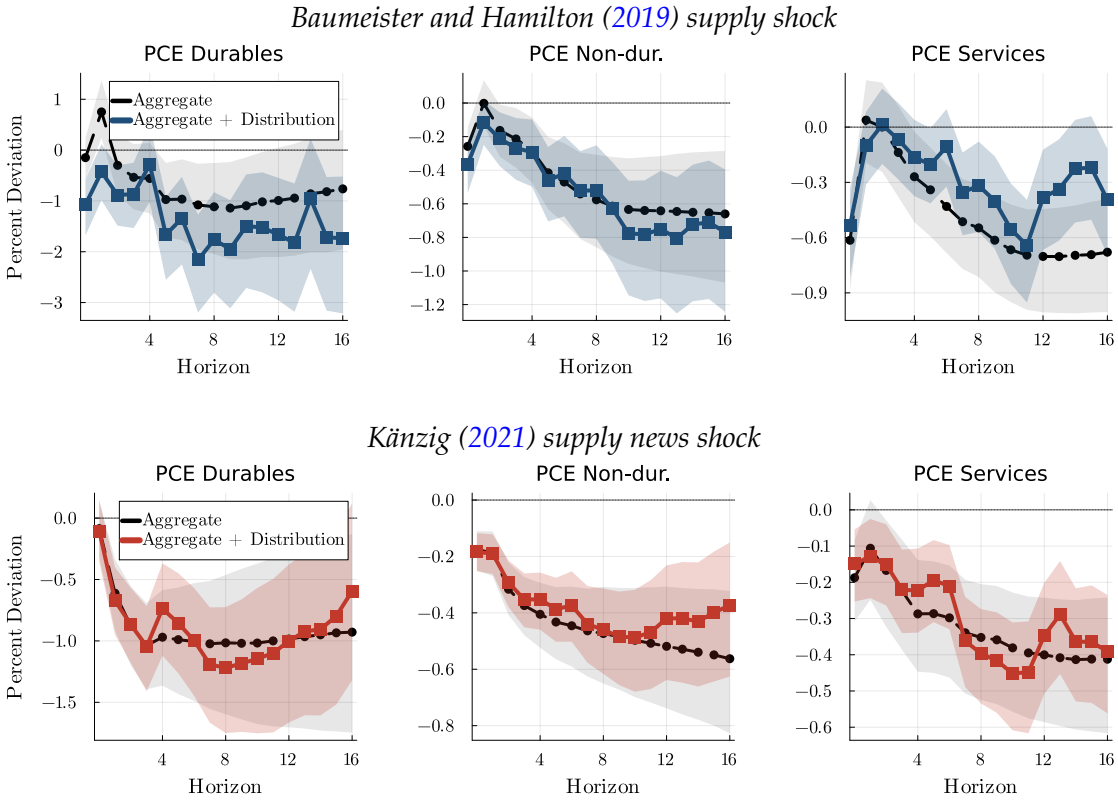
Notes: Solid: aggregate-only FAVAR. Dashed: FAVAR augmented with the eight smoothed distributional factors. Shaded regions are 68% credible sets.

Figure 38: Rotating FAVAR — consumer prices



Notes: Solid: aggregate-only FAVAR. Dashed: FAVAR augmented with the eight smoothed distributional factors. Disaggregated CPI components (food, energy, rent, transportation, OER). Shaded regions are 68% credible sets.

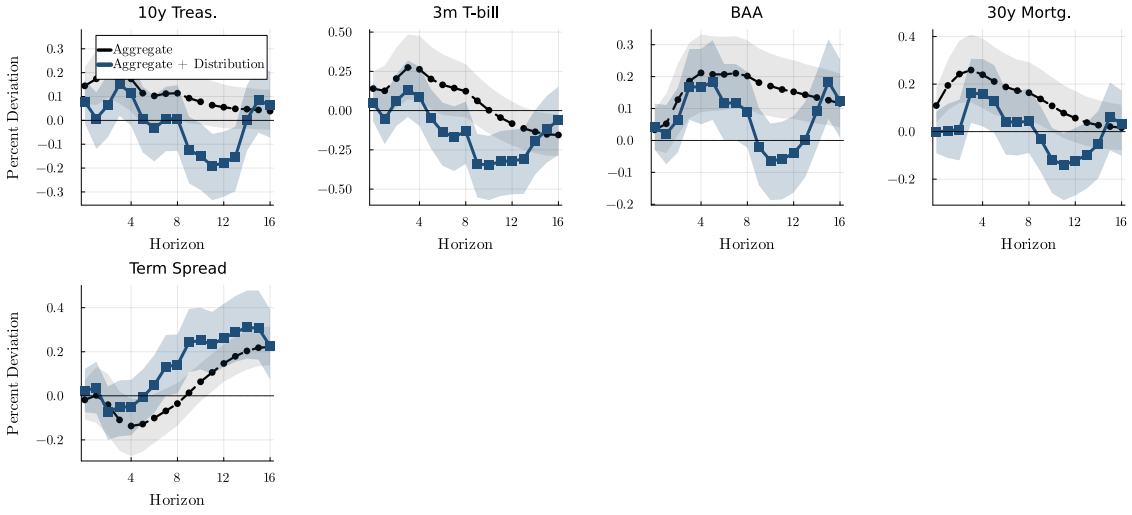
Figure 39: Rotating FAVAR — consumption



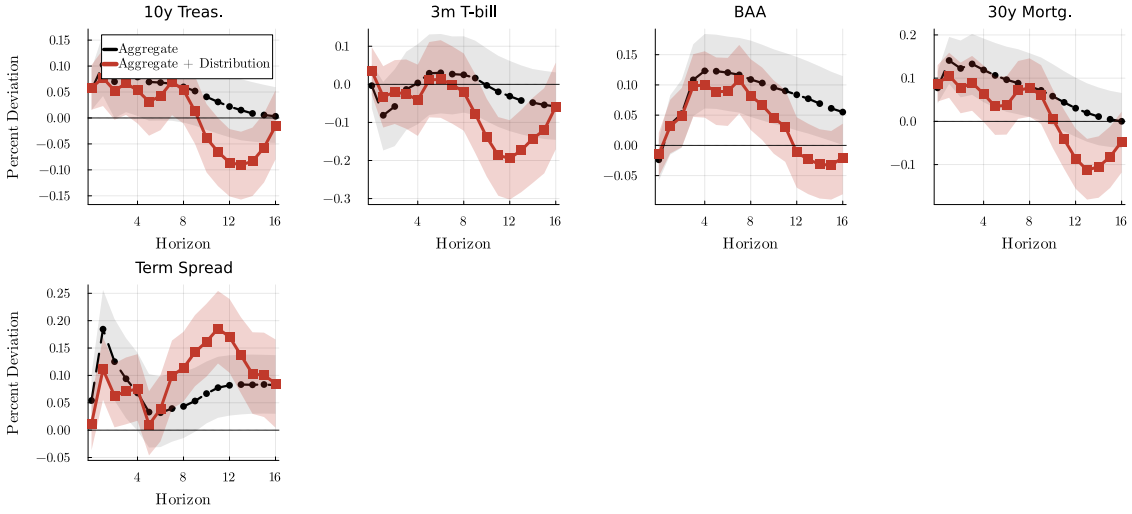
Notes: Solid: aggregate-only FAVAR. Dashed: FAVAR augmented with the eight smoothed distributional factors. PCE components (durables, non-durables, services). Shaded regions are 68% credible sets.

Figure 40: Rotating FAVAR — interest rates

Baumeister and Hamilton (2019) supply shock

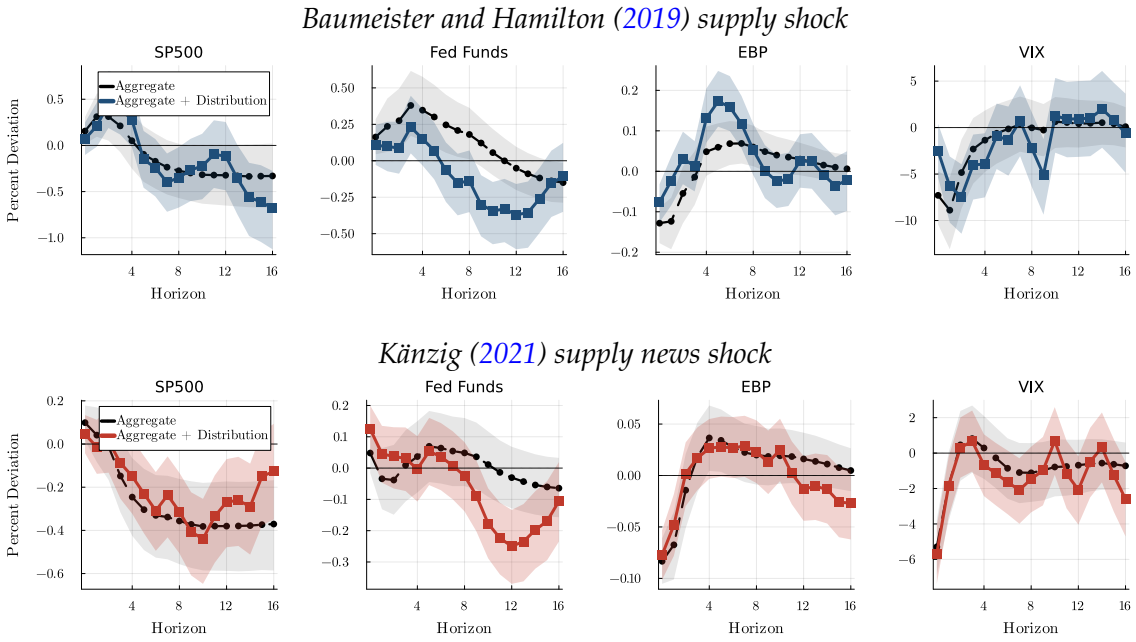


Känzig (2021) supply news shock



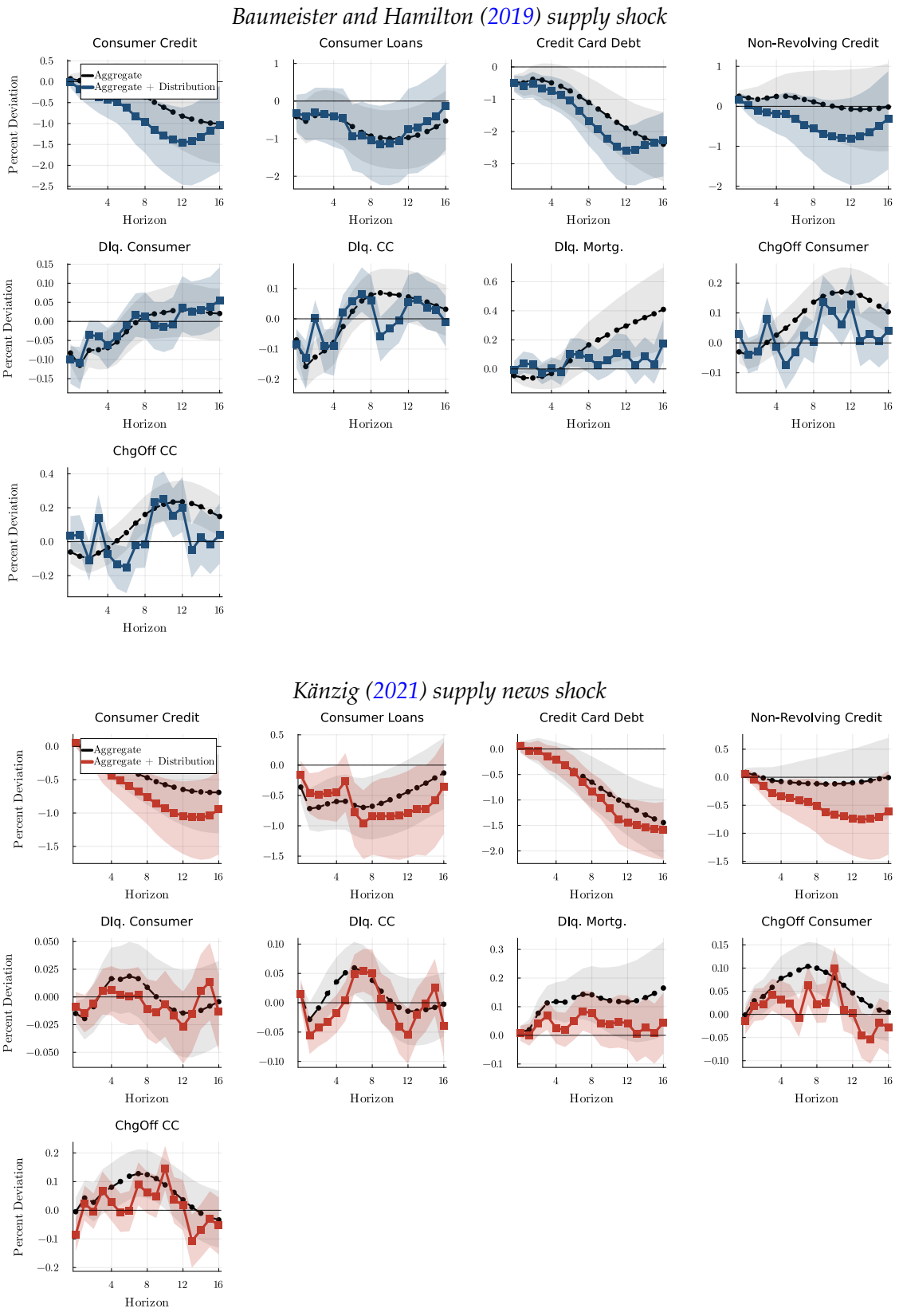
Notes: Solid: aggregate-only FAVAR. Dashed: FAVAR augmented with the eight smoothed distributional factors. Treasury, federal funds, and corporate yields. Shaded regions are 68% credible sets.

Figure 41: Rotating FAVAR — financial conditions



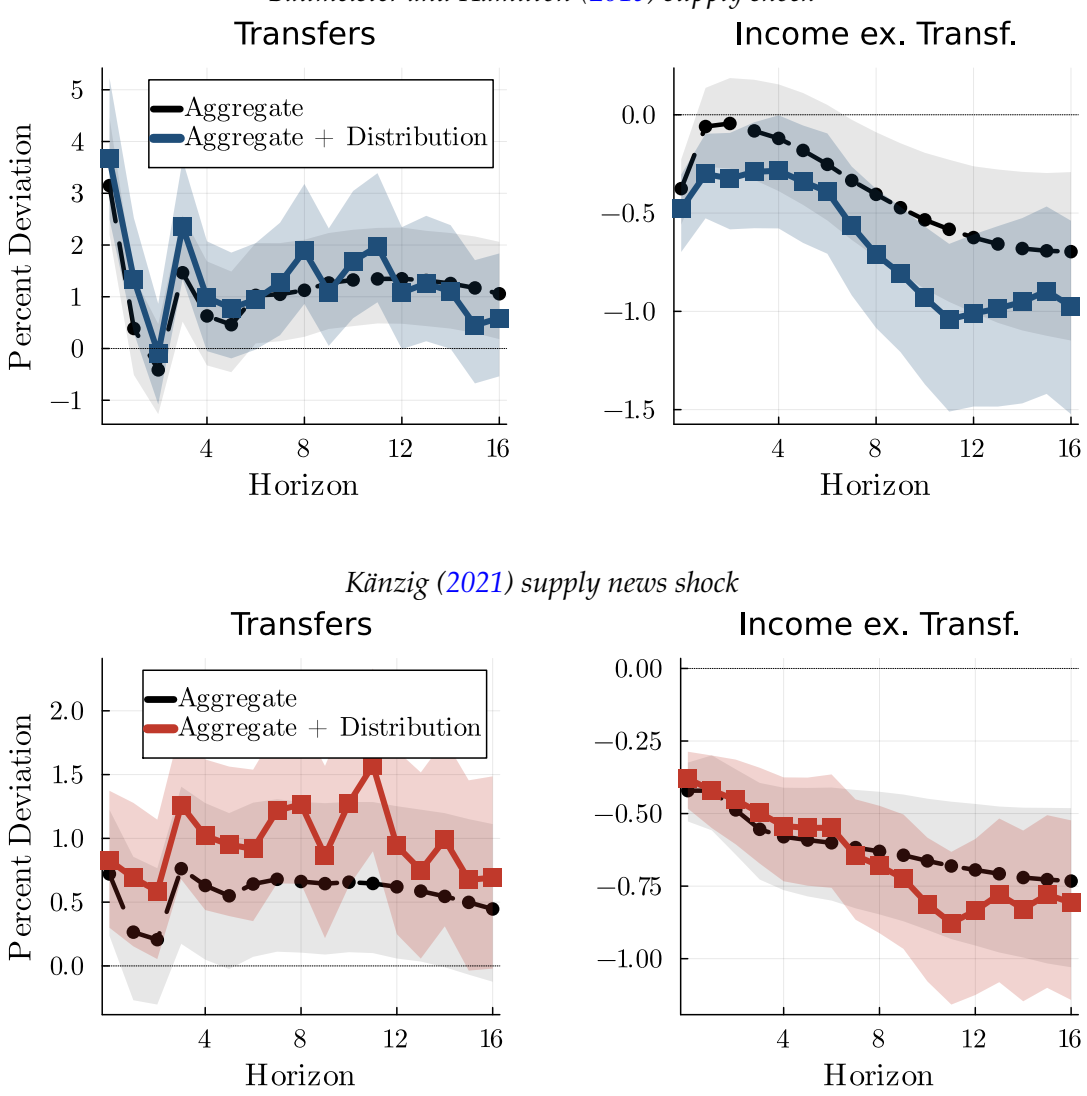
Notes: Solid: aggregate-only FAVAR. Dashed: FAVAR augmented with the eight smoothed distributional factors. S&P 500, VIX, Excess Bond Premium, federal funds rate. Shaded regions are 68% credible sets.

Figure 42: Rotating FAVAR — household credit supply and delinquency



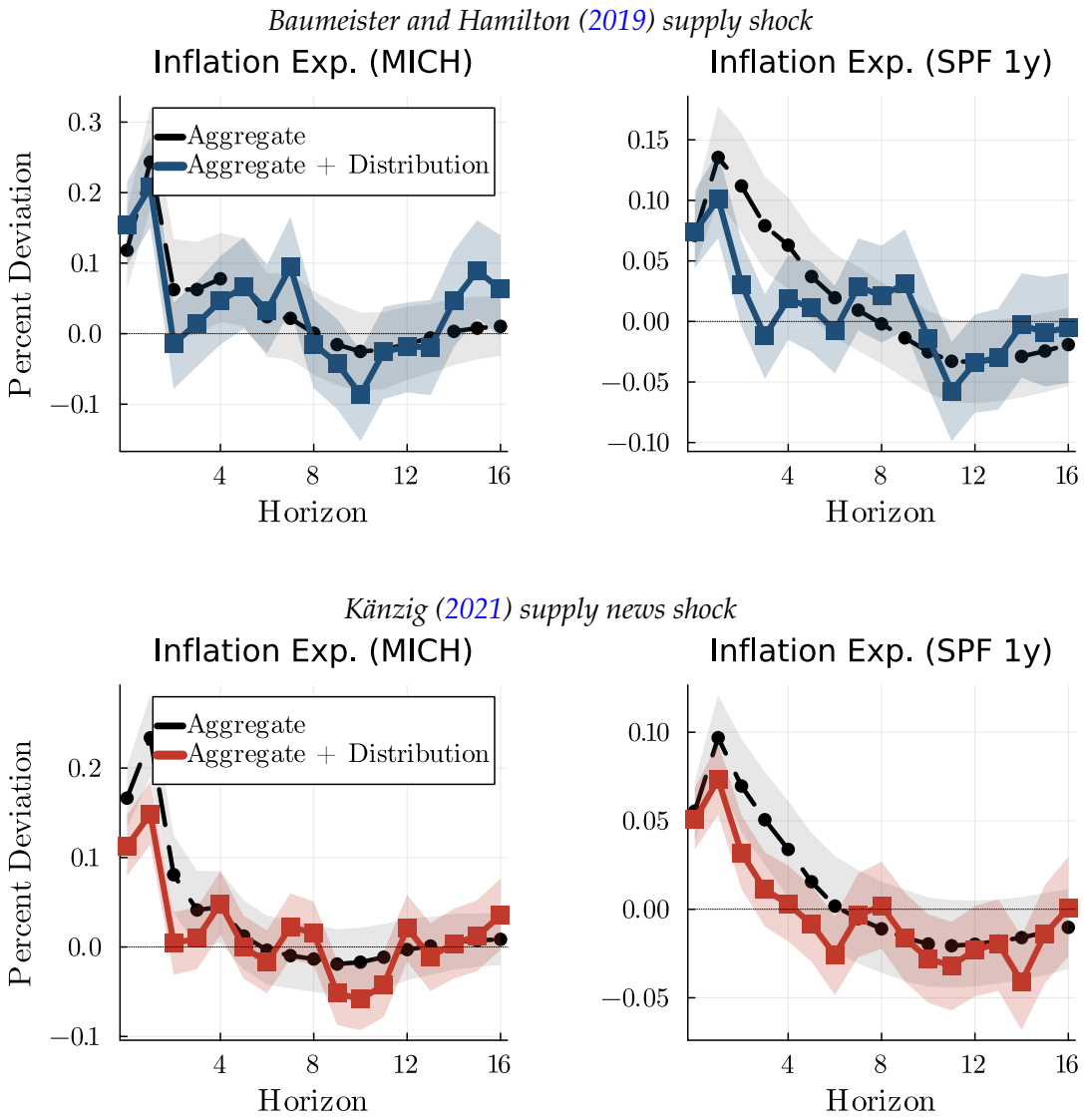
Notes: Solid: aggregate-only FAVAR. Dashed: FAVAR augmented with the eight smoothed distributional factors. Consumer credit and consumer-loan stocks; delinquency and charge-off rates on consumer loans, credit-card balances, and mortgages. Shaded regions are 68% credible sets.

Figure 43: Rotating FAVAR — income composition



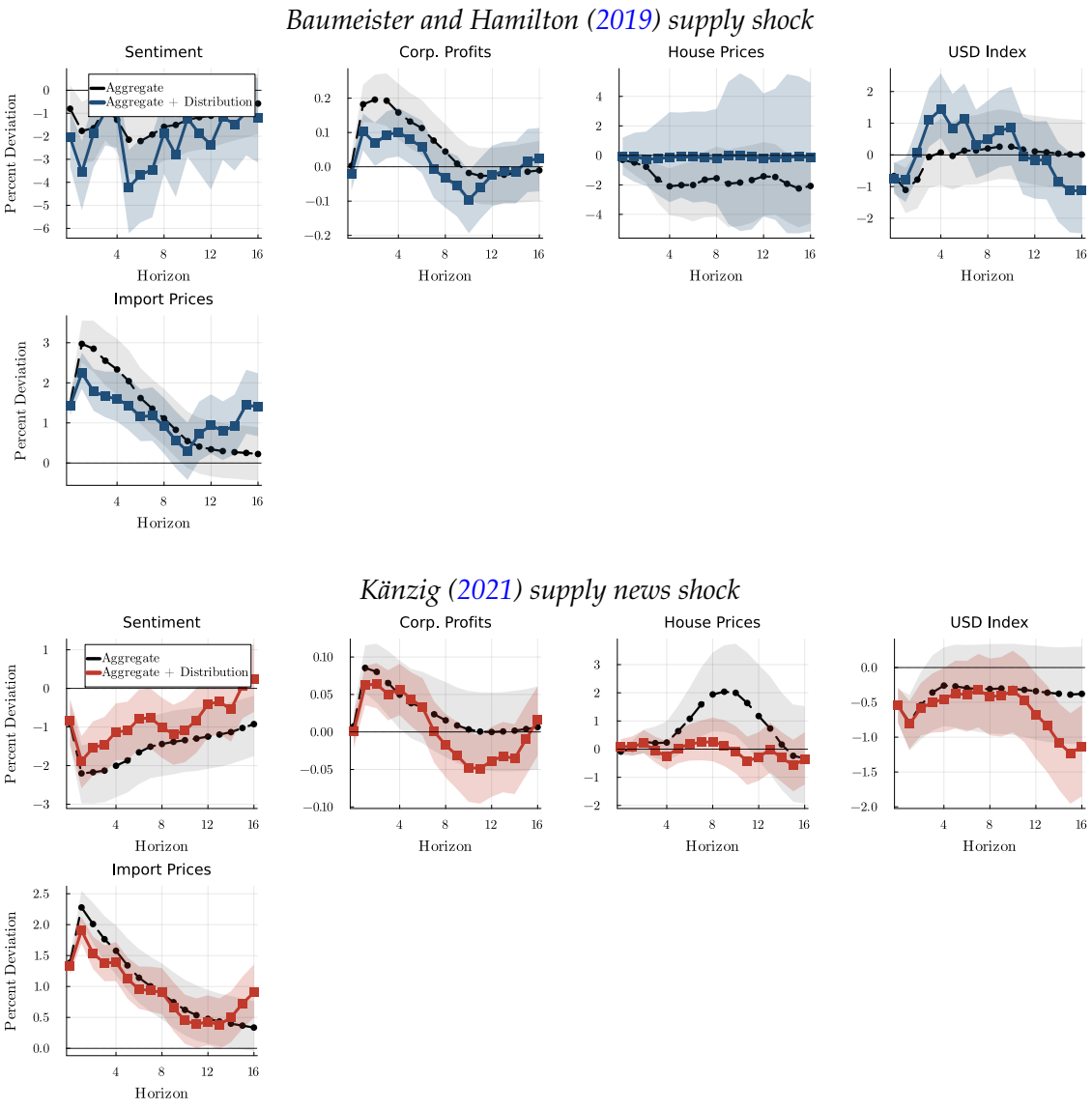
Notes: Solid: aggregate-only FAVAR. Dashed: FAVAR augmented with the eight smoothed distributional factors. Government transfers and personal income net of transfers. Shaded regions are 68% credible sets.

Figure 44: Rotating FAVAR — inflation expectations



Notes: Solid: aggregate-only FAVAR. Dashed: FAVAR augmented with the eight smoothed distributional factors. Inflation expectations: University of Michigan 1-year household survey and SPF 1-year CPI forecasters. Shaded regions are 68% credible sets.

Figure 45: Rotating FAVAR — other



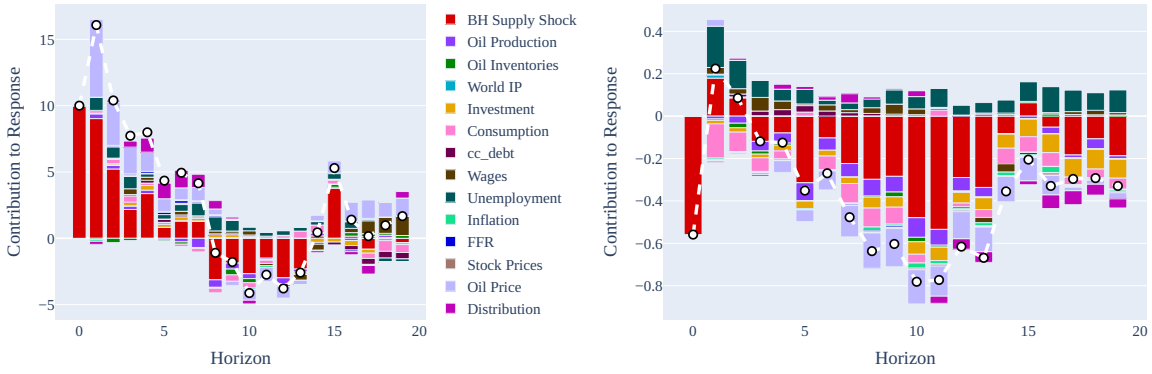
Notes: Solid: aggregate-only FAVAR. Dashed: FAVAR augmented with the eight smoothed distributional factors. Other macro and financial aggregates not classified above (e.g., trade-weighted dollar, corporate profits, house prices). Shaded regions are 68% credible sets.

J Channel Decomposition under the Extended Macro VAR

This appendix reports the impact-horizon ($n = 0$) channel decomposition of the oil-price and aggregate-consumption responses under the extended-macro specification, which augments the baseline VAR with a richer US block (`inv`, `cons`, `cc_debt`, `wages`, `unemp`, `infl`, `ffr`, `sp500`) plus the oil-market block. The decomposition follows the same Dufour–Wang construction used for the baseline (Section 5) but operates on the larger 12-variable macro block, so additional channels (credit, labor, monetary, financial) appear alongside the oil-market and inequality channels.

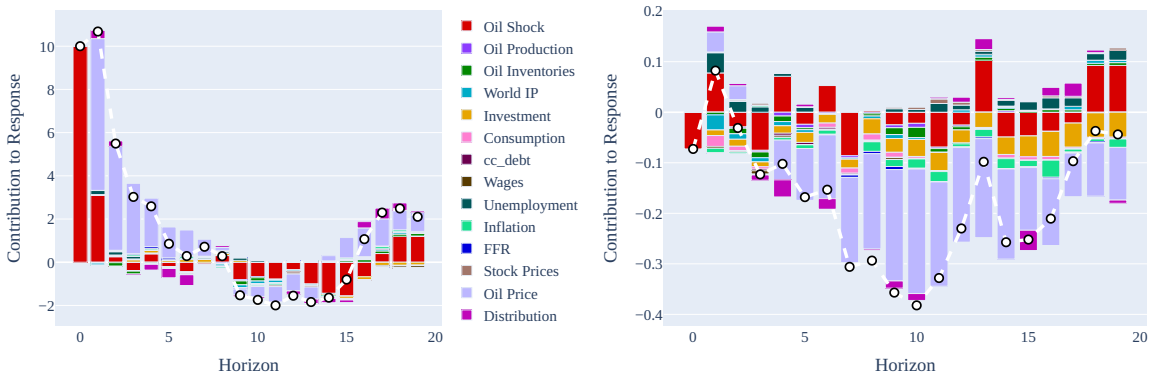
Figure 46: Channel decomposition at $n = 0$, extended-macro VAR — oil price (`poil`) and aggregate consumption (`cons`)

Panel A: Baumeister and Hamilton (2019) realized supply shock



Oil price (`poil`) Aggregate consumption (`cons`)

Panel B: Känzig (2021) oil supply news shock



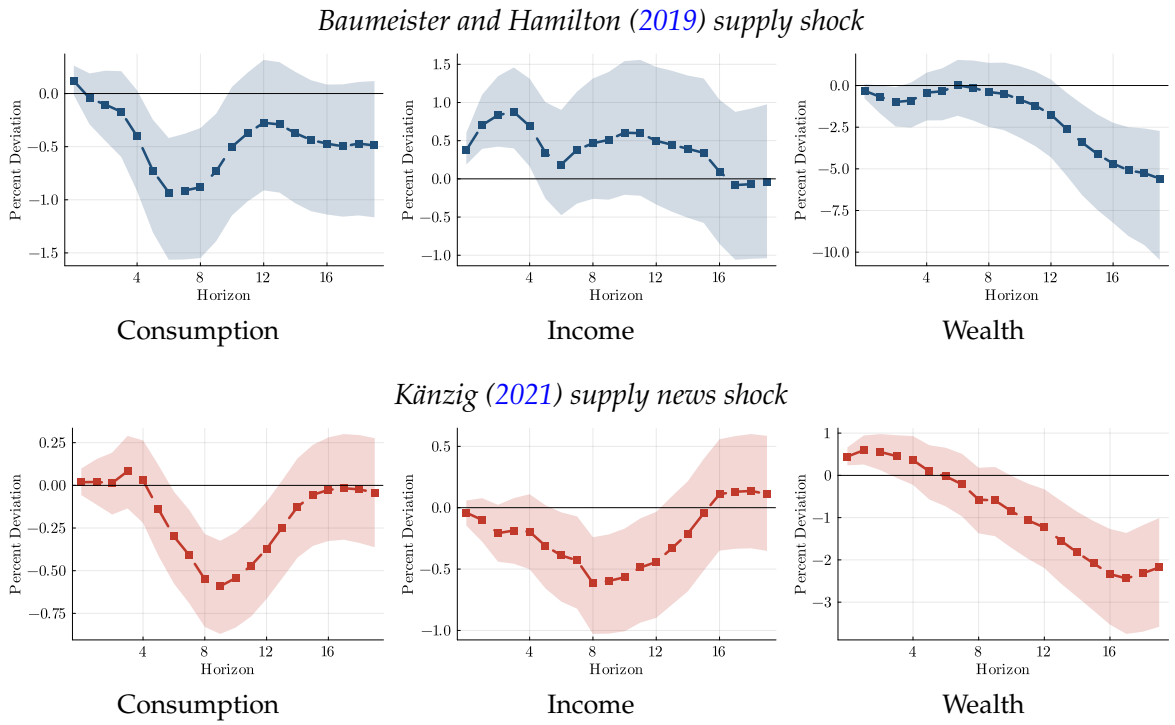
Oil price (`poil`) Aggregate consumption (`cons`)

Notes: Channel contributions $c_{m\to i}^{(h,0)}$ at impact ($n = 0$) under the extended-macro VAR. Each colored bar is the total contribution of channel m to the response of the outcome i at horizon h, evaluated from impact. Channels are grouped following the propagation classification in Section 4: oil market (oil price, oil production, inventories, world activity), real activity (investment, consumption, wages, unemployment), prices (PCE inflation), monetary/financial (federal funds rate, S&P 500), credit (`cc_debt`), and the eight smoothed distributional factors. Bars sum to the total impulse response at each horizon by construction.

K Additional Inequality Measures

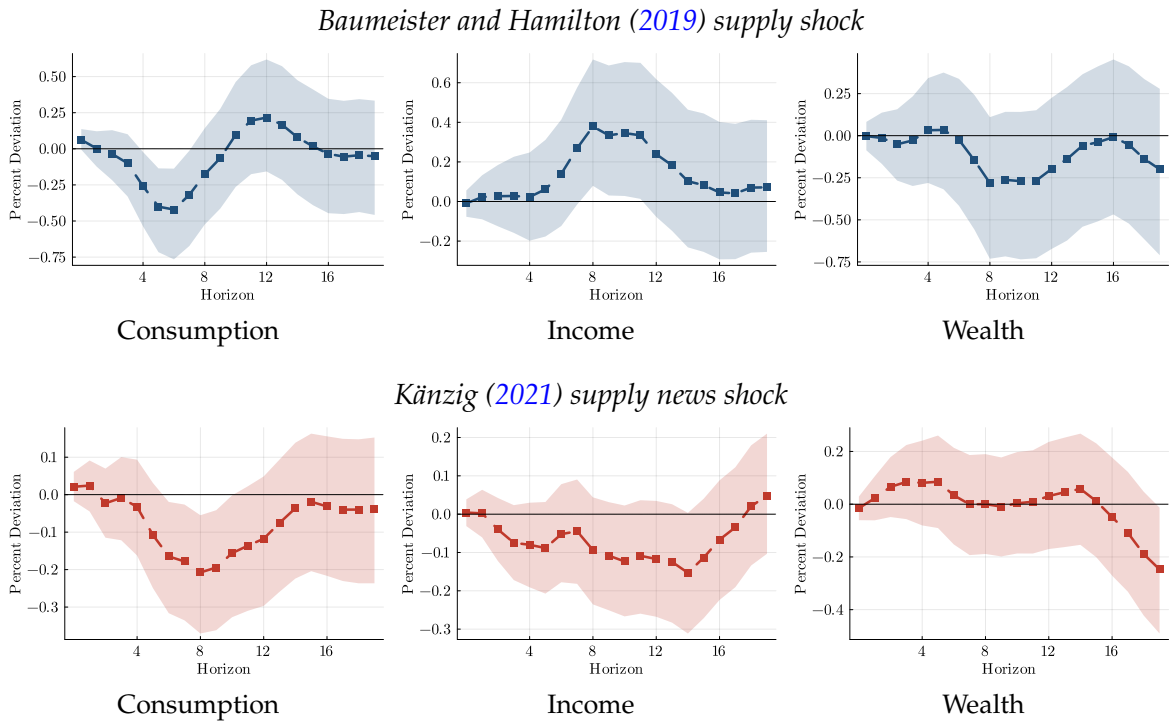
This appendix complements the Gini-coefficient responses reported in Section 6 with three additional summary measures of dispersion: the log 90/10 and 90/50 percentile ratios and the standard deviation of logs. Each is reported for the consumption, income, and wealth marginals and for both identification schemes. The percentile ratios summarize between-tail spread (a rise in 90/10 indicates a widening tail) while the standard deviation of logs summarizes overall dispersion.

Figure 47: Log 90/10 percentile ratio



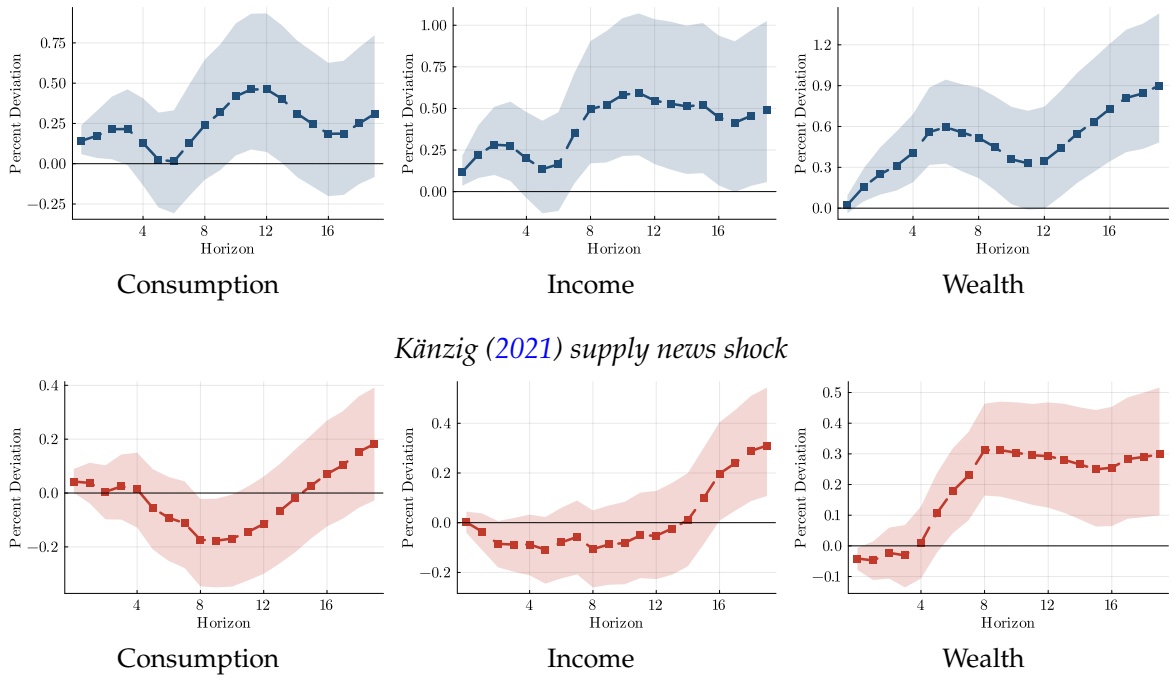
Notes: Impulse responses of the log 90/10 percentile ratio for consumption, income, and wealth. Units: percentage point deviation from steady state. Solid lines: posterior median. Shaded areas: 68% credible sets.

Figure 48: Log 90/50 percentile ratio



Notes: Impulse responses of the log 90/50 percentile ratio for consumption, income, and wealth. Units: percentage point deviation from steady state. Solid lines: posterior median. Shaded areas: 68% credible sets.

Figure 49: Standard deviation of logs



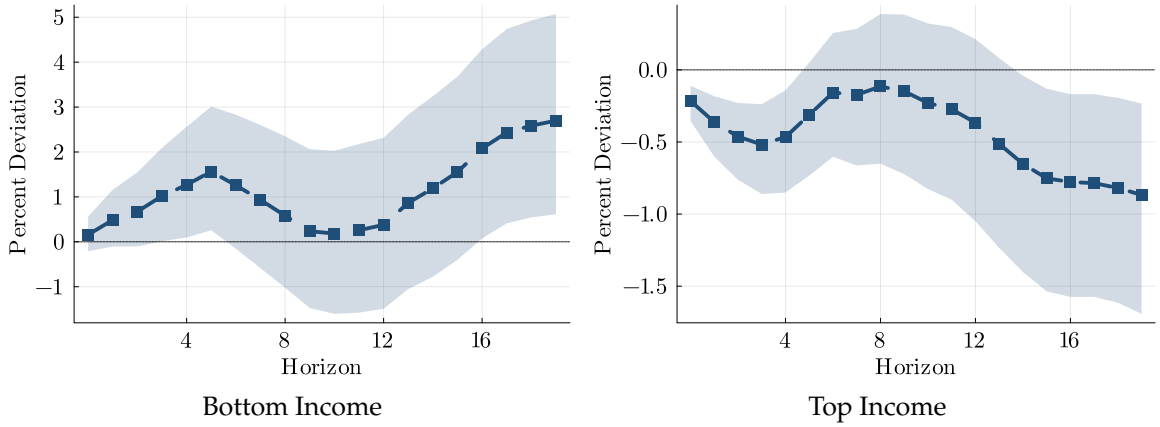
Notes: Impulse responses of the standard deviation of logs for consumption, income, and wealth. Units: percentage point deviation from steady state. Solid lines: posterior median. Shaded areas: 68% credible sets.

L Consumption Responses for the Distribution Tails

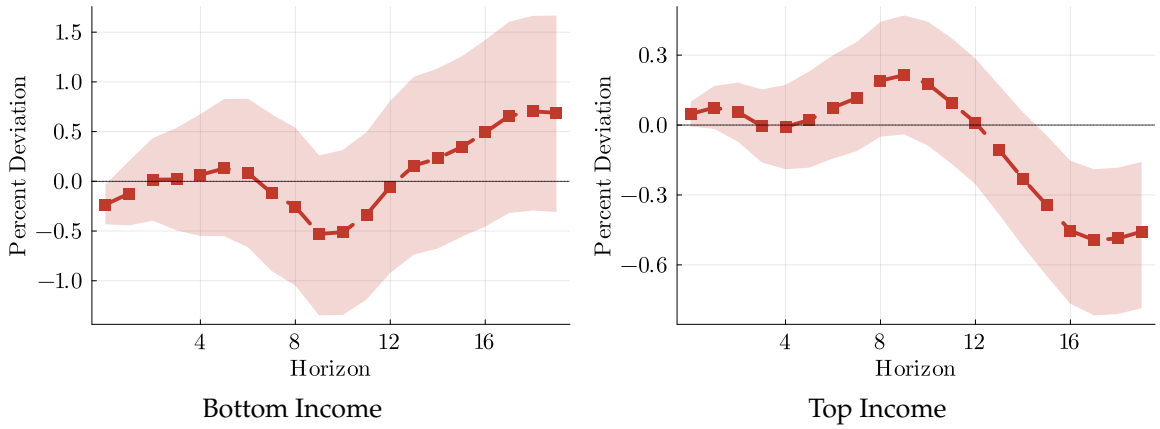
Figures 9 and 10 in the body report consumption responses by the Low (0–50%), Middle (50–90%), and Top (90–100%) groups of the income and wealth distributions. This appendix complements those plots with the responses of households at the *tails* of the marginal distributions: the Bottom (0–20%) and the Top (90–100%) of each margin. The qualitative picture extends naturally from the broad cuts, but the magnitudes are sharper at the tails, particularly for the bottom-income and top-wealth groups.

Figure 50: Consumption responses by extreme positions in the income distribution

Panel A: Baumeister and Hamilton (2019) realized supply shock



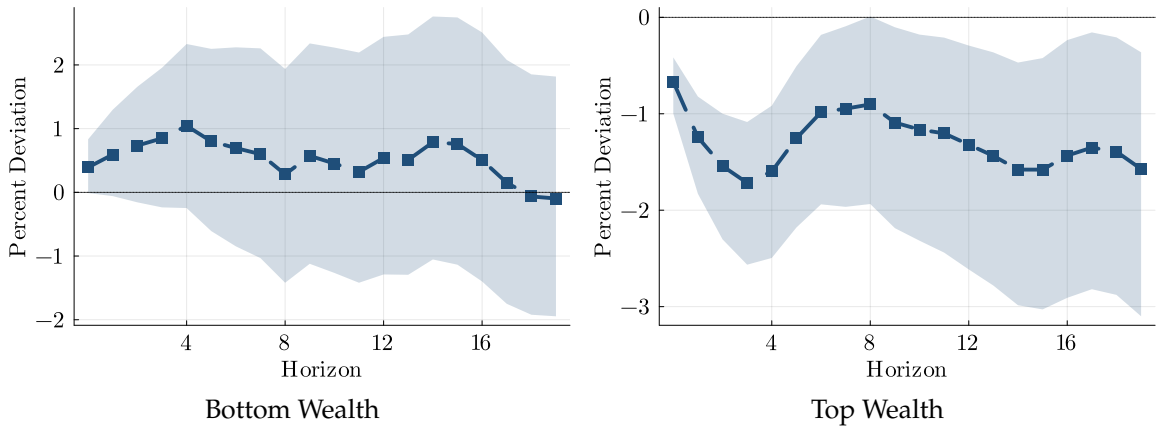
Panel B: Känzig (2021) oil supply news shock



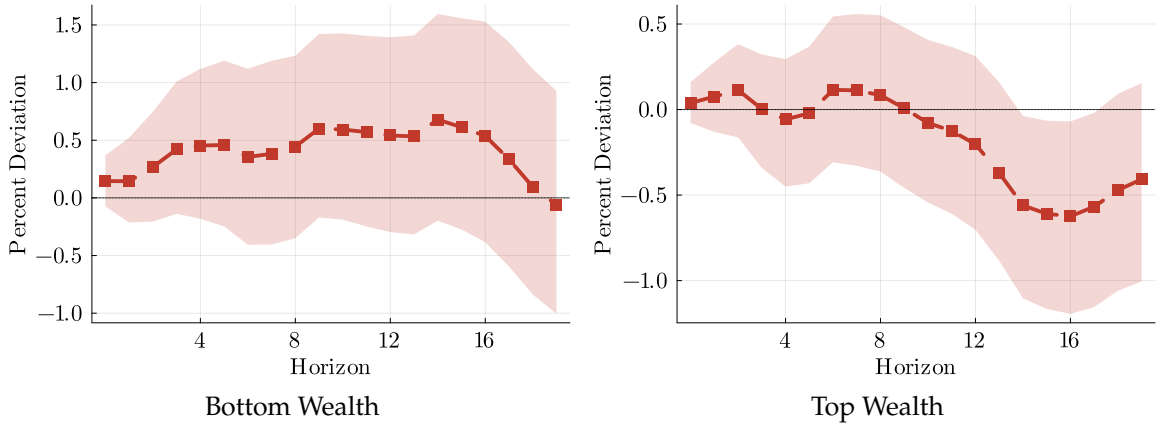
Notes: Impulse responses of mean consumption for households at the bottom and top of the marginal income distribution. Solid lines: posterior median. Shaded areas: 68% credible sets. Horizon in quarters.

Figure 51: Consumption responses by extreme positions in the wealth distribution

Panel A: Baumeister and Hamilton (2019) realized supply shock



Panel B: Känzig (2021) oil supply news shock



Notes: Impulse responses of mean consumption for households at the bottom and top of the marginal wealth distribution. Solid lines: posterior median. Shaded areas: 68% credible sets. Horizon in quarters.

Chapter 3

Oil Supply Shocks & Monetary Policy

1 Introduction

Oil supply shocks tend to be stagflationary: they increase the price of goods and services and depress real activity simultaneously. This was precisely the situation the central bank faced in the 1970s and history seems to be repeating itself. This presents a trade-off for the central bank: raise rates to deter realized or expected inflation, but at the risk of causing a recession or *look through* with the potential risk of inflation spiraling. Previous work has suggested the central bank's role is not innocuous for the transmission of these shocks and may even be responsible for the large macroeconomic outcomes historically observed (Bernanke, Gertler, and Watson, 1997). This view has been challenged in subsequent empirical and theoretical work (see e.g., Kilian and Lewis, 2011, for a review). This chapter re-examines the presented evidence, under the same class of VAR models employed, adopting recent insights as to the specification of these models (Kilian and Zhou, 2023; Baumeister, 2025), with naturally more years of data, and exploits new econometric frameworks that provide new insights on the role of the central bank for the transmission of these shocks.

Four questions structure the chapter. First, what share of the macroeconomic response to an identified oil supply shock is driven by the systematic component of monetary policy rather than by the shock's direct transmission? Second, what does the central bank reaction function mostly load on along the oil-shock horizon—the real price of oil itself, the inflation pass-through it generates, real activity, or some combination? Third, following the suggestion of Kilian and Lewis (2011), how does the Fed reaction function look like had it not responded to the oil shock specifically, while still responding normally to inflation and real activity? Fourth, against three counterfactual rules—a nominal interest-rate peg, strict inflation targeting, and

an optimal dual-mandate rule that minimizes a quadratic loss in inflation and the unemployment gap—could the Fed have improved aggregate or distributional outcomes, and through which channel?

To answer the first three questions, I apply the generalized impulse response (GIR) framework of Dufour and Renault (1998) and the causal mediation interpretation of Dufour and Wang (2024). Together, these papers provide a structured econometric framework to quantify how different channels (mediators) contribute to the dynamic causal effects of an exogenous shock, offering a more nuanced view of the transmission mechanism. In my setting, the framework allows me to quantify the contribution of the ongoing stance of monetary policy to both the aggregate and distributional effects of oil supply shocks by decomposing impulse response functions (IRFs) at each horizon into the channels represented in the VAR.

My focus is the contribution of the Federal Funds Rate (FFR) to the level of the IRFs and whether it amplifies or dampens them. The FFR IRF itself captures the central bank's reaction to the shock. The same framework also allows me to decompose the FFR IRF and identify the variables driving the policy response, thereby recovering the central bank's reaction function. Under policy invariance in the Lucas critique sense, this further allows me to infer the central bank response absent the oil shock, as emphasized by Kilian and Lewis (2011).

The understanding I can obtain thus far is that of the typical Fed response and this is important because it answers the question of the average historical role played by the central bank in response to the history of oil supply shocks thus far. If the characterization of such oil supply shocks is unchanged, as well as the set of structural shocks and monetary regime in the economy, I can expect precisely this typical response; however, the central bank may at times want to deviate, either temporarily or permanently, from this typical response and it is important to understand whether such deviations lead to better economic stability or further contraction. For this, I use the sufficient-statistics counterfactual framework of Caravello, McKay, and Wolf (2024) (CMW) to ask how the aggregate and distributional consequences of an oil supply shock would have changed had the central bank operated under a different monetary policy rule.

Regarding the technicalities of the procedure, the construction of the counterfactuals requires at least as many auxiliary monetary policy innovations as there are independent constraints T in the rule. A nominal interest-rate peg, for example, imposes one constraint per horizon; the constraint being holding the deviation of the federal funds rate to zero from its baseline path. Strict inflation targeting analogously imposes T inflation constraints. The optimal dual-mandate rule is softer (a quadratic loss rather than a hard zero) but still pins down a T-dimensional optimum. In each case the rule violation is a T-vector that the counterfactual must project onto the span of the available monetary shocks. With a single monetary shock the span is one-dimensional and the projection collapses to least-squares onto a single instrument.

Increasing the number of shocks K expands the span, reduces the projection residual, and—provided the shocks are not collinear—makes the projection well-conditioned in the sense of Wang (2026, Sect. 3). The empirical payoff is that one does not have to litigate which monetary shock is "best": by including a battery of identified series, the framework leverages their joint information and is robust to mis-specification of any single proxy. The trade-off is that adding shocks too aggressively introduces multicollinearity and thus singularity issues. You will see in our setup, the shock space will be free of these concerns.

The number of constraints T is a choice and the convention of choosing T equal to H, the IRF horizon, is perhaps too stringent and unrealistic. Policymakers think about the implementation of policy in the near future; therefore, this chapter adopts the view of Wang (2026) and enforces rules over shorter horizons ($T \in \{4, 6, 8\}$ quarters) that are closer to following paths than rules. With the headline $K = 6$ specification the projection is admissible for $T \geq K = 6$; for the shorter $T = 4$ horizon I fall back to a $K = 4$ baseline that retains the standard MP-shock set (Luetticke, 2024; Bauer and Swanson, 2023; Hack, Istrefi, and Meier, 2024) and is exactly identified at $T = K = 4$. The dual-mandate rule is over-determined at every T in the sweep but nonetheless delivers a better fit than $T = 20$. Overdetermined systems are not automatically invalid; the diagnostics in Appendix C indicate the degree to which the Lucas critique bites. After T periods the rule no longer binds and the policy rate adjusts endogenously through the VAR coefficients, using the same systematic reaction function that produced the original IRF.

To enforce these rules in the short run I borrow *six* distinct monetary policy shocks from the literature for the headline $K = 6$ specification: **(i)** the Luetticke (2024) unified narrative series, which chains the conventional and unconventional halves of the post-1969 shock history into a single continuous measure; **(ii)–(v)** the four orthogonal high-frequency factors of Jarociński (2024), identified by sign and zero restrictions on FOMC-window surprises in FFR futures, forward-guidance futures, long-rate futures, and the S&P 500—the short-rate, forward guidance, large-scale-asset-purchase, and residual factors, respectively; and **(vi)** the Hack, Istrefi, and Meier (2024)-orthogonalized Romer and Romer (2004) narrative shock, which retains pre-2008 identification power that the Jarociński HF factors lack. A $K = 4$ baseline consisting of the Lütke unified series, Bauer–Swanson high-frequency surprises, and the Hack, Istrefi, and Meier (2024)-orthogonalized Romer and Romer (2004) and Aruoba and Drechsel (2024) narratives is retained for the $T_c = 4$ constraint horizon, where $K = 6$ would render the peg and strict-π projections under-determined ($T_c < K$). Each shock captures a different slice of the monetary-policy partition, and the joint span at $K = 6$ delivers the lowest pairwise correlations achievable from the union of identification schemes available in the literature (see Section 3.1).

Both methodologies will rely on a VAR, as did the conclusions of Bernanke, Gertler, and

Watson (1997) and Kilian and Lewis (2011), incorporating on top new insights as to the specification of these oil VARs (Baumeister and Hamilton, 2019; Kilian and Zhou, 2023) and more generally, structural VARs (Baumeister, 2025). The usual suspects remain, namely, the modeling of global demand and supply of oil, the relevant oil price, and the representation of domestic activity in the U.S.. What will be different here is (1) the choice of inventories variable, which will be more robust to measurement error concerns (Baumeister and Hamilton, 2019) (2) long-lags (De Graeve and Westermark, 2025) to (i) achieve a better representation of the covariances along the oil cycle (ii) adopt the mispecification robutness of LPs (Ludwig, 2024) (iii) without increased uncertainty around the estimates (De Graeve and Westermark, 2025) (3) a distributional block (Bayer, Calderon, and Kuhn, 2025) to further isolate the supply component of the shock, but of course to understand the broad distributional implications and (4) all under a Bayesian approach (Chan, 2022) to model richer environments without violating the regularity conditions of frequentist VARs. In Chapter 2, I realize and employ these insights, and additionally show they are important for the implications of oil price shocks.

Because this paper is extensively on counterfactuals, it is important to discuss the extent to which the Lucas critique (Lucas, 1976) applies. If agents form expectations using the prevailing policy rule, changing the *rule* should change the reduced-form dynamics of the economy, and IRFs estimated under one rule cannot be used to evaluate another. The recent sufficient-statistics literature, however, has converged on a common defense. Wang (2026) states it most explicitly. The construction holds the private-sector decision rules and the baseline policy rule fixed and alters only the sequence of policy shocks within a finite window; in his words, "the policy peg is not a new rule, it is a finite string of temporary deviations from the baseline rule" (Wang, 2026, p. 7). The counterfactual is credible when the implied shock sequence is short, moderate in magnitude, and balanced in sign; it is strained otherwise because it would be informative about a regime shift and may trigger belief revision in the sense of Lucas (1976).

Wang (2026) calls these conditions of short, moderate, and balanced deviations as the *policy invariance* restriction, and argues that his approach is appropriate. For our baseline counterfactuals, I will follow Caravello, McKay, and Wolf (2024), who also make the same restriction operational, but at the population level: the alternative rule is admissible if it lies in the linear span of the auxiliary monetary innovations identified in the VAR (Caravello, McKay, and Wolf, 2024). McKay and Wolf (2023b) show that, under linear-Gaussian dynamics, the resulting counterfactual is identified up to a remainder term that vanishes in the limit of small rule perturbations. In my setup, the rate peg, the strict-IT rule, and the dual-mandate rule are enforced through linear combinations of K monetary innovations whose pairwise correlations are low (Table 1); the implied shock sequences are unprecedented in neither magnitude nor sign-persistence, and they pass diagnostic tests proposed by Wang (2026). For now, I imple-

ment the counterfactual procedure of Caravello, McKay, and Wolf (2024). Down the line, I would like to implement Wang (2026) for comparison.

Findings. My application embeds this construction in a Bayesian VAR that jointly identifies a Känzig (2021) oil supply news shock and $K = 6$ monetary innovations (the headline specification; the $K = 4$ baseline is retained for the $T_c = 4$ robustness, see Section 3.1). The endogenous block contains eight macroeconomic series (the real oil price, world oil production, OECD oil inventories, world industrial production, real GDP, unemployment, CPI inflation, and the federal funds rate) and eight orthogonal factors that summarise the joint income–wealth–consumption distribution. I consider two policy rules:

(i) a *nominal rate peg* that pins the federal funds rate deviation to zero at every horizon, and

(ii) *strict inflation targeting* that pins CPI inflation deviation to zero at every horizon.

Both are intentionally extreme; they bound the space of admissible rules from two directions.[1]

The macro responses to the oil supply shock under the three counterfactual rules differ only modestly from the realized path. The federal funds rate, inflation, GDP, and unemployment under the rate peg, strict inflation targeting, and the dual-mandate rule all sit broadly within the 68% credible band of the realized response, differing mostly in second-order features of the medium-run path. The dual-mandate rule generates the largest macro deviation overall, trading off inflation and the unemployment gap; the rate peg becomes substantive only at the longer constraint horizons; strict-π tracks the realized FFR path closely because the realized response was already most of the way to fully offsetting oil-shock pass-through.

The distributional responses are more discriminating. The dual-mandate rule delivers Gini paths nearly indistinguishable from the realized response on all three margins. The rate peg and strict inflation targeting, in contrast, both *compress* inequality further than the realized path: strict-π amplifies the realized medium-run Gini compression on every margin (by roughly twenty per cent for income, two-and-a-half-fold for consumption, and threefold for wealth), and the rate peg compresses the income and consumption Ginis but tracks the realized response on wealth. The mechanism under strict-π is a large redistribution toward the bottom of the distribution rather than top-tail destruction; the top-decile/bottom-half decomposition (Section 5.3.2) makes the mechanism explicit.

Contribution. The chapter makes three contributions. First, it provides the first model-free counterfactual answer to the question Kilian and Lewis (2011) pose at the end of their paper. Where they criticize the BGW counterfactual and call for "models that allow policy responses

[1] The CMW machinery accommodates more flexible rules (Taylor relations, optimal targeting under loss functions); I leave those for future work.

to depend on the underlying causes of oil price shocks," the subsequent literature (Leduc and Sill, 2004; Carlstrom and Fuerst, 2006; Nakov and Pescatori, 2010; Bodenstein, Guerrieri, and Kilian, 2012; Plante, 2014; Gagliardone and Gertler, 2025) has answered the call only by committing to a structural DSGE model. My exercise computes oil-shock counterfactuals using only sufficient statistics for the monetary policy span, in the spirit of Caravello, McKay, and Wolf (2024), and so trades the structural interpretability of the DSGE exercises for a much weaker reliance on modeling assumptions. To my knowledge, while Caravello, McKay, and Wolf (2024) apply the sufficient-statistics framework empirically to monetary policy shocks, and Kilian and Lewis (2011) apply a coefficient-restriction counterfactual empirically to oil shocks, the chapter is the first to combine them—model-free CMW counterfactuals on identified oil shocks under multiple alternative rules—and to do so jointly with a household distributional dimension. Second, the multi-shock identification gives the alternative rules meaningful enforcement power, in contrast to the single-MP-shock VARs of the older counterfactual literature (Bernanke, Gertler, and Watson, 1997; Kilian and Lewis, 2011). Third, I show that the CMW linear-superposition formula extends, draw-by-draw, to the underlying distributional factor IRFs and hence to any non-linear function of the joint distribution; the substantive payoff is a quantitative estimate of the distributional cost of the two rules, which I find to be an order of magnitude larger than the macro cost suggests. The third contribution generalizes beyond the specific application: any household-level outcome that admits a finite-dimensional functional representation can be projected through.

Outline. The remainder of the chapter is organized as follows. Section 2 situates the chapter against the long oil–monetary-policy debate initiated by Bernanke, Gertler, and Watson (1997) and Kilian and Lewis (2011), the recent sufficient-statistics counterfactual literature (McKay and Wolf, 2023a; Caravello, McKay, and Wolf, 2024; Wang, 2026), and the literature on monetary policy and inequality. Section 3 describes the data: the K monetary policy shocks (Section 3.1) and the macroeconomic controls (Section 3.2); the oil supply shock and distributional block are set out in Chapter 2. Section 4 sets out the empirical framework in three parts: the multi-shock BVAR with the asymmetric conjugate prior of Chan (2022) (Section 4.1); the CMW counterfactual algebra (Section 4.2); and the three counterfactual rules in closed form (Section 4.3). Section 5 presents the empirical findings in three subsections: the macro counterfactuals (Section 5.1), the channel decomposition of the realized IRFs and the per-rule decomposition under each counterfactual (Section 5.2), and the distributional counterfactuals (Section 5.3). Section 6 concludes.

2 Related literature

The chapter overall is relatively underdeveloped with considerable literature to cover. With that, the chapter devotes a large part to the literature review to make clear the contributions one could make i.e., what is still missing in my understanding of the oil-monetary policy debate.

Kilian–Lewis (2011). An important empirical precedent is Kilian and Lewis (2011), henceforth KL, who revisit the influential claim of Bernanke, Gertler, and Watson (1997) (BGW) that the Federal Reserve's systematic tightening in response to oil-price shocks was responsible for *all* the negative impact of the oil shock. The claim is based on a proposed interest-rate peg counterfactual in which the central bank does not change the policy rate in response to the oil price shock. KL, together with Hamilton and Herrera (2004), show that the result is fragile across lag lengths, sample windows, and identification of the underlying oil shock. Hamilton and Herrera (2004) additionally show that the BGW peg required policy deviations of unprecedented magnitude, raising issues with the Lucas-critique problem that Bernanke, Gertler, and Watson (2004) also acknowledge. Overall, KL finds (i) no empirical evidence that systematic monetary policy is an important source of fluctuations in the US economy (ii) that in general, oil price shocks have negligible effects on the macroeconomy and (iii) argues that it would be a mistake to respond to oil price shocks rather than its underlying determinants e.g., the oil price responding to supply or demand factors.

DSGE literature. The KL prescription is to design models that take into account the endogeneity of the real price of oil and that allow policy responses to depend on these underlying causes of oil price shocks. This has been almost exclusively answered with DSGE models. Leduc and Sill (2004) attribute roughly 40% of the post-shock output drop to the policy response. Carlstrom and Fuerst (2006) attribute essentially none. Nakov and Pescatori (2010) and Plante (2014) try different rules and rank them by welfare; the latter finds that strict CPI-inflation targeting performs poorly relative to a core-inflation or output-gap rule once oil-shocks are decomposed by source. The most direct DSGE response to KL is Bodenstein, Guerrieri, and Kilian (2012), who derive shock-specific optimal responses and show that the Fed should ease against oil-driven aggregate-demand shocks but tighten only modestly against oil-supply shocks.

Most recently, Gagliardone and Gertler (2025) estimate a New Keynesian DSGE model with oil as a complementary good for households and a complementary input for firms, search-and-matching unemployment, and real wage rigidity, and match impulse responses to an SVAR identified with the Känzig (2021) oil supply news shock and the Bauer and Swanson (2023) monetary policy shock. Their historical decomposition over 2010–2023 attributes the post-2021

inflation surge mainly to oil shocks combined with accommodative monetary policy. Bernanke and Blanchard (2023)'s reduced-form decomposition of pandemic-era inflation finds that energy and supply-side factors account for the bulk of the headline CPI movement once expectations are accounted for. Their structural counterfactual under a strict inflation-stabilization rule produces sharply larger output and unemployment losses than the estimated baseline, while a near-accommodating rule moderates the real losses at the cost of higher inflation.

Caravello, McKay, and Wolf (2024). The methodological foundation for the model-free exercise performed herein is Caravello, McKay, and Wolf (2024) (CMW), with theoretical underpinnings in McKay and Wolf (2023b) and Barnichon and Mesters (2023). The CMW insight is that the counterfactual response of any endogenous variable to a structural shock under a new policy rule is a linear combination of its responses to that shock and to a set of auxiliary monetary policy innovations identified in the same VAR. As long as the rule is a linear constraint on outcomes and the auxiliary block spans a sufficiently rich policy-shock space, the counterfactual is identified without re-estimating the model and without the Lucas-critique problem that sank the BGW exercise: the rule is enforced through historically observed policy variation rather than through unprecedented shifts.

In practice, what this will mean is agents in period zero, in addition to the oil supply news shock, are hit by a single bundle of K auxiliary monetary policy shocks calibrated to enforce the rule; the economy then evolves under its historical reduced-form dynamics, with the systematic Fed reaction function continuing to operate. The appealing feature of this design is that it does not require the economy to be continuously surprised: agents experience one set of policy innovations at the moment of impact, not a fresh sequence of unanticipated deviations every quarter. The latter is the approach formalized by Sims and Zha (2006), which was actually applied informally in Bernanke and Mishkin (1997).

Under linear-Gaussian dynamics, McKay and Wolf (2023b) prove that the one-shot counterfactual (only date 0 shocks) is identified up to a Lucas-critique residual—the adjustment in agents' decision rules induced by the policy change—that is second-order in the size of the policy perturbation and therefore vanishes for small auxiliary impulses. Repeatedly surprising agents, as in the Sims–Zha sequential design, inflates the cumulative implied shocks and pushes this residual outside the small-perturbation regime where it remains bounded.

Dufour and Wang (2024). The aggregate part of my analysis—decomposing the realized IRFs of an oil supply shock into a direct effect and an indirect effect through the federal funds rate—has a precedent in the same KL paper. KL argue that "the relevant counterfactual" for the BGW question "is not one in which the Federal Reserve holds the interest rate constant in response to an oil price shock [...] but a counterfactual in which the Federal Reserve reacts to fluctua-

tions in other macroeconomic state variables (such as inflation and real output) as it normally would with only the direct response to the real price of oil being shut down." The construction is informal: KL reorder the VAR so the FFR depends on the oil price last, set the contemporaneous oil-price coefficient in the FFR equation to zero, and recompute IRFs. Dufour and Wang (2024) formalize the same idea using the short-run/long-run causality machinery of Dufour and Renault (1998): any structural shock admits a decomposition of its IRF into a direct effect on outcome Y that bypasses any specified set of mediator variables $\mathcal{M}$, and an indirect effect that travels through $\mathcal{M}$. KL's exercise is the special case in which $\mathcal{M}$ is the federal funds rate and the shock is oil-supply; my generalization adds two ingredients absent from KL: $\mathcal{M}$ can be any subset of the VAR's mediators, and the framework extends to the multi-shock VAR in which K monetary innovations enter alongside the oil disturbance. I adopt this formalism in Section 5.2, where I also discuss the conditions under which the mediation decomposition admits a counterfactual reading.

Most related literature. Presented here are papers that are empirical in nature and are the most recent empirical antecedents to my paper. Wang (2026) develops a local-projection (LP) counterpart to Sims and Zha (2006) *but also* Caravello, McKay, and Wolf (2024) and applies it, in his first empirical exercise, under counterfactual short-rate pegs of length three, twelve, and twenty-four months, using the Känzig (2021) supply news surprise and the Bauer and Swanson (2023) high-frequency monetary instrument. Broer, Kramer, and Mitman (2025) go further on the distributional side: using German administrative microdata, they ask how the systematic response of monetary policy to the oil supply news shock of Känzig (2021) shapes the earnings and employment distribution, conducting a policy rate non-response counterfactual under both Caravello, McKay, and Wolf (2024) and Sims and Zha (2006).

Closest in methodological spirit to my exercise is Ider et al. (2025), who apply the McKay and Wolf (2023b) sufficient-statistics counterfactual to the European Central Bank and energy prices. They document that ECB monetary policy decisions significantly influence global energy prices and show, under the same Lucas-critique-robust framework I adopt, that this ability roughly halves the tightening required to stabilize euro-area inflation after an energy supply shock. Their optimal-policy analysis under a stylised dual-mandate loss rationalises the ECB's observed "look-through" response to energy shocks as quasi-optimal precisely because the ECB *can* influence energy prices. Their setting is the ECB and they do not consider distributional outcomes; the $K = 6$ counterfactual rules and joint income–consumption–wealth distribution reported in Section 5 are complementary to their aggregate ECB-side evidence.

I differ from Wang (2026) and Broer, Kramer, and Mitman (2025) on three margins. First, the policy span. My auxiliary monetary block stacks K shocks—the Luetticke (2024) uni-

fied CMP/UMP narrative, Bauer and Swanson (2023) high-frequency surprises, and the Hack, Istrefi, and Meier (2024)-orthogonalized Aruoba–Drechsel and Romer–Romer narratives—giving the CMW projection more degrees of freedom to enforce the rule. Several papers, including Broer, Kramer, and Mitman (2025), build their span out of one or two monetary instruments, which leads to trouble in enforcing their supposed counterfactuals—a limitation Wang (2026) himself flags when discussing the cost of one-shot designs.

Second, the distributional object: Wang (2026) does not discuss distributional consequences and Broer, Kramer, and Mitman (2025) work with moments from marginal earnings/employment distribution. I use the joint income–consumption–wealth *functional* representation of Bayer, Calderon, and Kuhn (2025), which lets me evaluate counterfactuals for any moment of the joint distribution (Gini coefficients, top 10% shares, bottom 50% shares). Third, the rule set: I evaluate two boundary rules (a nominal rate peg and a strict CPI inflation target) and an optimal interior rule (a dual-mandate loss-minimizing response), rather than just a non-response benchmark.

Functional data in macroeconomics. The distributional data come from Bayer, Calderon, and Kuhn (2025), which fuses microdata at the household level with macroeconomic indicators of higher frequency in a Bayesian state-space framework. Their construction follows Sklar (1959): any joint distribution over $\mathbb{R}^d$ decomposes into its marginal CDFs and a copula encoding the dependence structure, and both objects can be projected onto an orthonormal Legendre basis to obtain a finite-dimensional state vector. The eight leading principal components underlying the functional data captures all the business-cycle variation in the joint distribution and enter my VAR as the distributional block. This functional-VAR strategy follows Chang, Chen, and Schorfheide (2024) and Chang and Schorfheide (2024), Lenza and Savoia (2024), and Bjørnland, Chang, and Cross (2023), but to my knowledge this chapter is the first to combine it with a CMW counterfactual exercise in oil context.

Monetary policy and inequality. An established literature documents that monetary contractions raise income and consumption inequality (Kaplan, Moll, and Violante, 2018; Coibion et al., 2017; Auclert, 2019; Bayer, Born, and Luetticke, 2020; Andersen et al., 2021); the mirror-image finding for monetary expansions is mixed (Lenza and Slacalek, 2018). This chapter speaks to the same question but with two differences. First, the impulse is an identified oil supply shock rather than a monetary shock, so the relevant object is whether the distributional incidence of the oil shock depends on the *systematic* component of the Fed's response. Second, the rule is varied counterfactually via the sufficient-statistics machinery of Section 4.3, holding the shock fixed. I find that the realized oil shock compresses the income, wealth, and consumption Ginis in the medium run; the dual-mandate rule delivers Gini paths nearly indistinguish-

able from the realized response; and the rate peg and strict inflation targeting both compress inequality further than the realized path, with strict-π the larger amplifier on all three margins. The top-decile/bottom-half decomposition (Section 5.3.2) attributes the strict-π compression to redistribution toward the bottom of the distribution rather than top-tail destruction.

3 Data

This section presents data unique to this chapter and refers the reader to Chapter 2 on the description of the distributional data I use for the estimation and the oil supply shocks as well. This section describes the monetary policy shocks used to discipline the counterfactuals using the methodology of Caravello, McKay, and Wolf (2024) and describe the macroeconomic aggregates, which have slightly changed from Chapter 2 to accommodate the rules specified in my counterfactuals i.e., any variable in some proposed rule must be within the variable space of the VAR.

3.1 Monetary policy shocks

The auxiliary monetary block enters the VAR (Section 4.1) as a set of externally identified shocks. I report two specifications: a $K = 6$ headline (used throughout Section 5 and the appendix robustness exercises) and a $K = 4$ baseline retained for when the number of constraints, T_c, is equal to 4.[2]

Headline ($K = 6$). The $K = 6$ specification stacks the following six monetary-policy shocks, chosen to span the 1969Q1–2023Q4 sample with as little overlap in identifying variation as possible:

(i) **Luetticke (2024) unified series**: a unified narrative monetary policy shock, a single continuous series spanning both the conventional and unconventional (ZLB) regimes. Constructed from using the Romer-Romer econometric setup on shadow rates from Wu and Xia (2015). Available 1969Q1–2015Q4. I refer to them later as Lütke for short.

(ii)–(v) **Jarociński (2024) orthogonal factors**: the four factors u_1 (short rate), u_2 (forward guidance), u_3 (large-scale asset purchases), and u_4 (residual long-rate factor) from Jarociński (2024). The factors are identified by sign and zero restrictions on 30-minute FOMC-window surprises in FFR futures, forward-guidance futures, long-rate futures, and the S&P 500. They are orthogonal to each other by construction; the maximum pairwise correlation among the four is 0.12 (Table 1). Available 1990Q1–2024Q3.

[2] $K = 6$ would render the peg and strict-π projections under-determined ($T_c < K$).

(vi) **Romer–Romer (HIM-cleaned)**: the Romer and Romer (2004) narrative shock with the systematic component of monetary policy removed following Hack, Istrefi, and Meier (2024). Retains pre-2008 identification power for the 1970s and Volcker-era oil-shock episodes that the HF Jarociński factors do not. Available 1969Q1–2007Q4.

As mentioned earlier, for the $T_c = 4$ specification, we can only use systems with $K \leq 4$—otherwise, they are under-determined ($T_c = 4 < K = 6$). The $K = 4$ specification replaces the four Jarociński factors with the more standard pair (Bauer and Swanson, 2023; Hack, Istrefi, and Meier, 2024; Aruoba and Drechsel, 2024):

(ii) **Bauer–Swanson high-frequency surprises**: Bauer and Swanson (2023) FOMC announcement-window surprises in the four-quarter-ahead Eurodollar future, orthogonalized against the Greenbook information set. Available 1988Q1–2023Q4.

(iii) **Aruoba–Drechsel (HIM-cleaned)**: the Aruoba and Drechsel (2024) narrative shock with the systematic component of monetary policy removed following Hack, Istrefi, and Meier (2024). Available 1982Q4–2007Q4.

together with the same Lütke unified series (i) and HIM-cleaned Romer–Romer series listed above.[3]

Why these shocks. The choice of monetary proxies is the central identification decision of the chapter, and Table 1 makes the reasoning concrete. The table reports pairwise Pearson correlations across ten candidate shocks; the diagonal is shaded grey (each shock's correlation with itself is unity by construction) and off-diagonal cells with $|\rho| > 0.5$ are highlighted yellow, flagging pairs that move toward collinearity.

The $K = 6$ headline uses the four orthogonal HF factors u_1, u_2, u_3, u_4 of Jarociński (2024) (rows five through eight of the table). The four factors are orthogonal to each other by construction; the 4×4 block among $u_1, \ldots, u_4$ is essentially diagonal (max $|\rho| = 0.12$). They are moderately correlated with Bauer–Swanson (e.g. BS–$u_1 = 0.51$, BS–$u_2 = 0.57$) which is expected—both factor models extract structured directions from the same FOMC-window HF surprise space—but Bauer–Swanson is no longer in the $K = 6$ set, so these cross-correlations do not appear among the included shocks.

Picking out the $K = 6$ entries from Table 1, the maximum pairwise correlation among the six is 0.63, occurring on the Lütke–HIM-RR pair; the other nine pairs in the $K = 6$ block are at $|\rho| \leq 0.36$, none of them yellow. In the end, the multiple shocks setup give the rule meaningful

[3] Because the monetary proxies have non-overlapping or only partially overlapping coverage windows (e.g. HIM-AD ends 2007Q4; HIM-RR ends 2007Q4; Jarociński factors begin 1990Q1), I follow Caravello, McKay, and Wolf (2024) and impute missing values with zero outside each shock's native coverage window before estimation. The same zero-padding convention is used for the K=4 baseline, K=6 headline, and all robustness specifications.

enforcement power: with $K = 1$ the rule is satisfied only in least-squares sense across $T = 20$ horizons, and the resulting counterfactual may barely deviate from the baseline at any horizon other than the impact response.

Table 1: Pairwise correlations of monetary policy shocks (overlapping observations only).

	BS	Lütke	HIM-R	HIM-A	u_1	u_2	u_3	u_4	JK-MP	JK-CBI
Bauer–Swanson	1.00	0.16	0.06	0.12	0.51	0.57	0.07	−0.10	0.69	0.30
Lütke unified	0.16	1.00	0.63	0.54	0.27	0.22	−0.11	−0.08	0.27	0.14
HIM-Romer	0.06	0.63	1.00	0.59	0.15	0.11	−0.36	0.17	0.05	0.27
HIM-Aruoba	0.12	0.54	0.59	1.00	0.19	0.01	−0.15	−0.11	0.14	0.05
u_1 short rate	0.51	0.27	0.15	0.19	1.00	−0.03	0.12	−0.11	0.72	0.53
u_2 forward guidance	0.57	0.22	0.11	0.01	−0.03	1.00	−0.02	−0.08	0.51	0.01
u_3 LSAP	0.07	−0.11	−0.36	−0.15	0.12	−0.02	1.00	0.07	0.31	−0.05
u_4 residual factor	−0.10	−0.08	0.17	−0.11	−0.11	−0.08	0.07	1.00	−0.41	0.64
JK pure MP	0.69	0.27	0.05	0.14	0.72	0.51	0.31	−0.41	1.00	0.23
JK CB-information	0.30	0.14	0.27	0.05	0.53	0.01	−0.05	0.64	0.23	1.00

Notes. Pearson correlations over the pairwise overlap of observation windows. JK-MP and JK-CBI are not mentioned in the text—they are median monetary-policy and central-bank-information shocks of Jarociński and Karadi (2020), identified by sign-restrictions on the co-movement of rates and equity prices around FOMC announcements (1990–2024, n=145).

3.2 Macro variables

The macro block follows a standard oil-VAR specification (Kilian, 2009; Baumeister and Hamilton, 2019; Känzig, 2021) augmented with the U.S. real side (GDP, unemployment), inflation, and the policy rate. The eight series enter the VAR in the following order: the real WTI crude oil price (deflated by CPI-U), world crude oil production, the OECD-proxy crude oil inventories series of Baumeister–Hamilton, the OECD-plus-six global industrial production index, U.S. real GDP, the U.S. civilian unemployment rate, CPI inflation (quarterly log-difference of CPI-U), and the effective federal funds rate. The five log-level series (real oil price, oil production, oil inventories, world industrial production, U.S. real GDP) enter in log levels and are scaled by 100 for interpretability so that percent-deviation IRFs read off the y-axis directly; unemployment, inflation, and the federal funds rate enter in levels (percent).

Normalisation. Before turning to the counterfactual, *all* impulse responses are rescaled per posterior draw so that the impact response of the real oil price to the oil shock is $+10\%$. The CMW projection in Eq. (7) below then operates on these scaled IRFs, so the auxiliary m_{aux} is interpreted as the bundle of monetary innovations the Fed would have to deliver to enforce the rule *given* a 10pp oil-price impact.

4 Methodology

4.1 The multi-shock BVAR

Let $y_t \in \mathbb{R}^{n_y}$ stack the variables in the following order: the identified oil supply shock first, the K monetary policy shocks second, the eight macro variables third, and the eight distributional factors last. Placing the oil shock at position one of y_t is the natural recursive ordering: the oil shock is the impulse whose transmission the chapter studies, and within-quarter macro and distributional outcomes are assumed not to feed back into the externally measured shock series. The monetary policy shocks occupy positions $2, \ldots, K+1$ so that the CMW auxiliary policy span (Section 4.2) reads off the first K columns of the structural impulse-response matrix Θ_h immediately to the right of the oil-shock column. I estimate the reduced-form BVAR with the same machinery as in Chapter 2

$$y_t = c + B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t, \qquad u_t \sim \mathcal{N}(0, \Sigma), \tag{1}$$

with $p = 16$ lags, an asymmetric Minnesota-style prior with per-equation hyperparameters (Chan, 2022), priors for the long-run (Giannone, Lenza, and Primiceri, 2019), and $D = 5{,}000$ posterior samples that satisfy stability requirements of the VAR after burn-in. Again, the asymmetric structure matters because the five-shock block has very different prior information than the macro and distributional blocks: shock equations are tightly centered on white noise (diagonal $\delta_i = 0$), while macro and distributional equations may adhere to the standard Minnesota persistence prior.

The five shocks are already structurally identified upstream of the VAR: the oil supply shock by Känzig (2021) (with Baumeister and Hamilton (2019)'s sudden oil-production shortfalls as a robustness identification), and the K monetary policy shocks by their respective external identification strategies. Each enters y_t as an externally measured proxy in the first $K+1$ positions, and the recursive structure of B_0 in those rows is a *timing assumption*, not an identifying restriction: within-quarter macro and distributional outcomes are assumed not to feed back contemporaneously into the externally measured shock series (Stock and Watson, 2018; Plagborg-Møller and Wolf, 2021). Stacking the impact matrix B_0 such that the first $K+1$ rows are diagonal yields the structural representation

$$y_t = \sum_{h=0}^{\infty} \Theta_h \, \varepsilon_{t-h}, \tag{2}$$

where $\Theta_h \in \mathbb{R}^{n_y \times n_y}$ is the h-step structural impulse-response coefficient matrix: its (i, j) entry is the response of variable i at horizon h to a unit shock to ε^j at t, for j some structural shock.

For each posterior draw I extract two slices of Θ_h that the counterfactual algebra needs:

$$\Psi_i^{\text{oil}}(h) \equiv \Theta_{h,i,1}, \tag{3}$$

$$m_{i,k}^{\text{base}}(h) \equiv \Theta_{h,i,k+1}, \quad k = 1, \dots, K, \tag{4}$$

i.e., the response of the i-th endogenous variable to the oil shock (Ψ_i^{oil}) and to the k-th monetary policy shock ($m_{i,k}^{\text{base}}$) at horizon h.

4.2 CMW counterfactual algebra

Fix a posterior draw and a length of enforcement T_c. For some variable i, write the T-vector of impulse responses to the oil shock as $\Psi_i \in \mathbb{R}^T$ and the $T \times K$ matrix of responses to the K monetary shocks as M_i. A linear policy rule that the central bank follows under the counterfactual is encoded by a collection of $T \times T$ constraint matrices $\{A_i\}_{i=1}^{n_y}$ such that the target is to satisfy

$$\sum_{i=1}^{n_y} A_i y_i^{\text{cnf}} = \mathbf{0}. \tag{5}$$

Counterfactual responses are obtained by adding to the original oil IRF a linear combination of the monetary IRFs:

$$y_i^{\text{cnf}} = \Psi_i + M_i m_{\text{aux}}, \qquad m_{\text{aux}} \in \mathbb{R}^K. \tag{6}$$

Substituting Eq. (6) into Eq. (5) and stacking yields the linear system

$$\underbrace{\Big[\sum_i A_i M_i\Big]}_{\equiv M_1\ (T\times K)} m_{\text{aux}} = \underbrace{-\sum_i A_i \Psi_i}_{\equiv M_2\ (T\times 1)},$$

whose least-squares solution is

$$\boxed{m_{\text{aux}} = -(M_1' M_1)^{-1} M_1' M_2.} \tag{7}$$

This is the closed-form CMW formula: the auxiliary monetary policy shock vector is a linear projection of the rule violation under the oil shock onto the space of monetary IRFs. Once m_{aux} is known, Eq. (6) delivers the counterfactual IRF of every variable in the VAR, including the eight distributional factors.

4.3 Counterfactual policy rules

I evaluate three counterfactual rules. A rate-peg and a strict-inflation-target rule, as well as a closed-form optimal dual-mandate rule that nests both extremes in a single quadratic loss. The first two rules impose hard linear constraints on a *single* target row: the $T \times T$ constraint blocks A_i are zero except for the target variable. With I_T the T-dimensional identity:

- **Zero-rate peg.** $A_{\text{ffr}} = I_T$, all other $A_i = 0$. The rule pins the federal funds rate deviation to zero at every horizon $h \in \{0, 1, \dots, T-1\}$. The Fed uses the linear combination m_{aux} of monetary innovations to fully neutralise the oil-shock-induced FFR path.
- **Strict inflation targeting.** $A_{\text{infl}} = I_T$, all other $A_i = 0$. The rule pins CPI inflation deviation to zero at every horizon. The Fed uses m_{aux} to fully offset oil-shock pass-through to inflation.

For these two rules the projection is admissible iff $M_1' M_1$ has full column rank, i.e. the IRFs of FFR (resp. inflation) to the four monetary shocks span $\mathbb{R}^K$ in their first T horizons. This is the rank condition emphasized by Wang (2026, Sect. 3) as the identification requirement for any local-projection or VAR-based counterfactual: without enough *independent* monetary instruments, the projection of the rule violation onto the policy span is not unique.

For the third rule is an *optimal* interior rule that does not impose a hard constraint on any row but instead minimizes a standard central-bank quadratic loss in inflation and the unemployment gap:

- **Optimal dual-mandate.** The Fed minimizes $\mathcal{L}(m_{\text{aux}}) = \|\Psi_{\pi}^{\text{cnf}}\|^2 + \lambda \|\Psi_{u_{\text{gap}}}^{\text{cnf}}\|^2$ over m_{aux}, where Ψ_{π}^{cnf} and $\Psi_{u_{\text{gap}}}^{\text{cnf}}$ are the counterfactual inflation and unemployment-gap IRFs. Substituting Eq. (6) into the loss yields a quadratic in m_{aux} whose first-order condition is the stacked least-squares problem

$$m_{\text{aux}}^{\text{dm}} = -\left(M_{1,\pi}' M_{1,\pi} + \lambda\, M_{1,u}' M_{1,u}\right)^{-1} \left(M_{1,\pi}' M_{2,\pi} + \lambda\, M_{1,u}' M_{2,u}\right), \tag{8}$$

  where $M_{1,\pi}, M_{2,\pi}$ (resp. $M_{1,u}, M_{2,u}$) are the M_1 and M_2 objects of Eq. (7) built from the inflation (resp. unemployment) row of the IRF cube. I set $\lambda = 1$ (equal weights, the textbook benchmark); $\lambda \to 0$ collapses to strict inflation targeting.

The first two bound the space of monetary stances from two sides: a peg is the most accommodative rule that an inflation-targeting central bank could plausibly entertain (it accepts unbounded inflation pass-through to keep the rate fixed), and strict inflation targeting is the most aggressive (it accepts unbounded rate movements to keep inflation at the target). The third sits between them and maps directly to the textbook central-bank loss function: it lets the Fed trade off inflation stabilization against employment stabilization according to λ rather

than committing to one pole. Differences between the realized response and the three rules tell me how much of the realized macro and distributional response was driven by the systematic component of monetary policy, and whether a welfare-maximizing central bank with a standard loss function could have done substantially better.

The FFR IRF decomposition is a finer-grained counterfactual. The rate-peg rule is the *full-non-response* limit: the Fed does not move the funds rate in response to anything — oil shock, oil price, inflation, real activity, or anything else — for the full T quarters (the FFR is kept at its steady state). Kilian and Lewis (2011) insist that the substantively interesting counterfactual is finer than a full peg:

> "a counterfactual in which the Federal Reserve reacts to fluctuations in other macroeconomic state variables (such as inflation and real output) as it normally would with only the direct response to the real price of oil being shut down."

The Fed equation in the VAR has loadings on lagged macro state variables, including the real oil price itself. KL's counterfactual zeroes out the FFR equation's loading on the oil price specifically, leaving every other loading (inflation, real activity, distribution, lagged FFR, etc.) intact. The Fed continues to react to inflation and output as it normally does—including the inflation and output movements that the oil shock has produced—but its direct rule-of-thumb response to the real oil price is removed.

The Dufour and Wang (2024) channel decomposition of the FFR IRF (Section 5.2) operationalizes exactly this question. The oil-price block of the decomposition isolates the portion of the FFR response that flows through the FFR equation's loading on the oil price. Removing only that block and propagating the remaining (inflation-, activity-, distribution-, own-FFR-driven) path through to any other variable Y gives Kilian and Lewis (2011)'s preferred counterfactual.

The Dufour–Wang shutdown, however, imposes the restriction via some ex-post algebra of the realized IRF; it is in some sense BGW's coefficient-zeroing exercise. The CMW peg implements non-response through a sequence of auxiliary monetary innovations that satisfy the policy-invariance restriction up to a small remainder (McKay and Wolf, 2023b; Wang, 2026). The two methods answer *different* substantive questions (the full peg vs. the shock-specific shutdown). The channel decomposition has no built-in defense against the Lucas critique, so results there are to be interpreted with care. One could do this differently and model a different authority that, as a rule, stabilizes the oil price response, as done in Ider et al. (2025).

5 Results

This section reports the empirical findings in three subsections. Section 5.1 reports the baseline IRFs along with the macro counterfactual IRFs under the three rules across constraint horizons $T_c \in \{4, 8\}$. Appendix 7 and 8 show results for $T_c \in \{6, 20\}$. Section 5.2 decomposes the realized IRFs into mediator channels: first the FFR and oil-price baseline decompositions that recover the Fed's reaction function and the oil-shock pass-through, then the full per-rule decomposition of the four macro outcomes. Section 5.3 reports the distributional counterfactual paths.

Summary of main findings. Four results stand out and motivate the subsections that follow.

(i) On Macro Sensitivity. At first glance, under the Känzig (2021) oil supply news shock and the $K = 6$ baseline, all three counterfactual rules deliver macro paths that are quantitatively close to the realized response on every variable other than the rule's own target (Figure 1). This continues the Kilian and Lewis (2011) verdict *against* Bernanke, Gertler, and Watson (1997)'s "systematic monetary tightening explains most of the macro impact" claim, now with a model-free (Caravello, McKay, and Wolf, 2024) sufficient-statistics counterfactual rather than a coefficient-zeroing exercise. I double-down on Kilian and Lewis (2011) by showing the FFR has little to no role in the propagation of aggregate responses.

(ii) Policy Duration. Policy duration will be an important nuance. At $T_c = 4$, the strict inflation targeting and the dual-mandate generate the largest deviations. Both policies will exchange slightly higher initial rates for greater output and unemployment recovery (medium-run) and less inflation overall. At $T_c = 8$, for dual-mandate, it becomes more salient. The rate-peg would do nothing if held for $T_c = 4$ periods, but convincingly achieves greater output, inflation, and unemployment stability if held for the entire horizon, $T_c = H = 20$. I show that a bank operating under a dual-mandate for 20 periods will converge to something close to a rate-peg.

(iii) Reaction-function. The Dufour and Wang (2024) mediator decomposition of the FFR IRF identifies what the Fed is reacting to: a considerable share of the FFR response is mediated by inflation and real activity, but also the oil price.

(iv) Distributional sensitivity. The headline macro aggregates—FFR, inflation, GDP, unemployment—have similar dynamics across the four rules under $T_c \in \{4, 6, 8\}$, differing mostly in the medium-run path. The three Ginis and distributional shares tell a different story. The dual-mandate rule delivers Gini paths nearly indistinguishable from the realized response. The rate peg and strict inflation targeting both compress inequality further than the realized path, with strict-π the larger amplifier on every margin (by roughly twenty per cent for income, two-and-a-half-fold for consumption, and threefold for wealth); the mechanism is a redistribution toward the

bottom of the distribution rather than top-tail destruction. A central bank evaluated on macro aggregates alone would not separate the four rules cleanly; one evaluated on the cross-section can.

5.1 Macro counterfactuals

Figures 1 and 2 report the counterfactual macro responses to the Känzig (2021) oil supply news shock under two constraint horizons: $T_c = 4$ and $T_c = 8$. For the $T_c = 4$ case, I use the $K = 4$ unified specification referenced above, which provides an exact-fit for the rate-peg and inflation-peg and leaves it free thereafter. For the $T_c = 8$ case, I use the $K = 6$ baseline, where the rule may not be enforced perfectly.[4] Each figure reports the four core macro variables (FFR, CPI inflation, real GDP, unemployment) under the realized policy (solid red) and the three counterfactual rules: rate peg (orange dash), strict π-target (green dot), dual mandate (blue dash-dot). The 68% credible band is shaded only for the realized response. The oil-price column and the additional constraint horizons $T_c \in \{6, 20\}$ are reported in Appendix A (Figures 7, 8, 9).

What the figures show. Three observations stand out across the two figures. *First, the rate peg neutralises the FFR over the constraint window.* By construction the peg rule pins the FFR-deviation to zero over the first T_c horizons. At $T_c = 4$ (Figure 1) the peg path is exactly flat for the first four quarters—the $T_c = K = 4$ system is square and the projection is exact—and drifts back toward the realized response as the systematic Fed reaction function carries on. At $T_c = 8$ (Figure 2) the peg flattens FFR over a wider window, but overall looks no different than the baseline. The rate-peg in our setting is active once it deviates from the baseline—pegging for more than a two years. With $T_c = 20$, the interest-rate is not perfectly pegged, but stays close to zero. In this setting, *it does* achieve greater output, inflation, and unemployment stability and only deviates (lowers) to fight medium-run recessionary movements.

Second, the strict-π counterfactual sits close to the realized path on every variable other than its target. Under strict-π targeting, when enforced perfectly ($T_c \in \{4, 6\}$), we can achieve better medium-run output, but no difference elsewhere. In general, these counterfactuals follow similar FFR paths, most different in the short-run for the $T_c = 4$ case, where the rules are enforced *perfectly*, exchanging higher initial rates for greater output recovery (medium-run) and less inflation overall though not by much.

Third, the dual-mandate rule delivers the empirically largest macro deviation. It minimizes a

[4] The $K = 6$ baseline cannot be evaluated at $T_c = 4$ for the hard-constraint rules because $T_c \geq K$ is required for M_1 to have full column rank; at $T_c = 4$ the $K = 4$ unified specification is exactly identified ($T_c = K = 4$, square system). The dual-mandate rule does compute at $T_c = 4$ under $K = 6$ (the stacked $2T_c \times K$ system is over-determined) and is reported in the appendix.

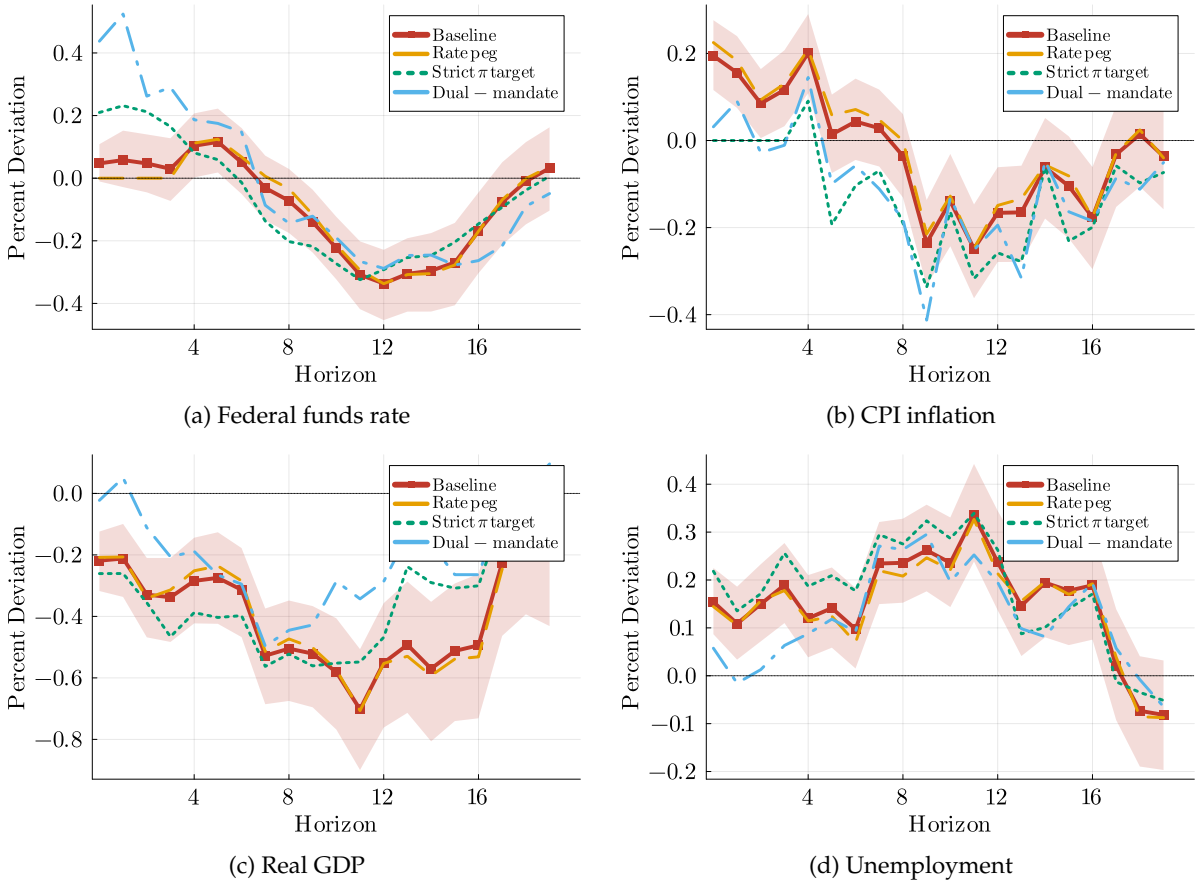
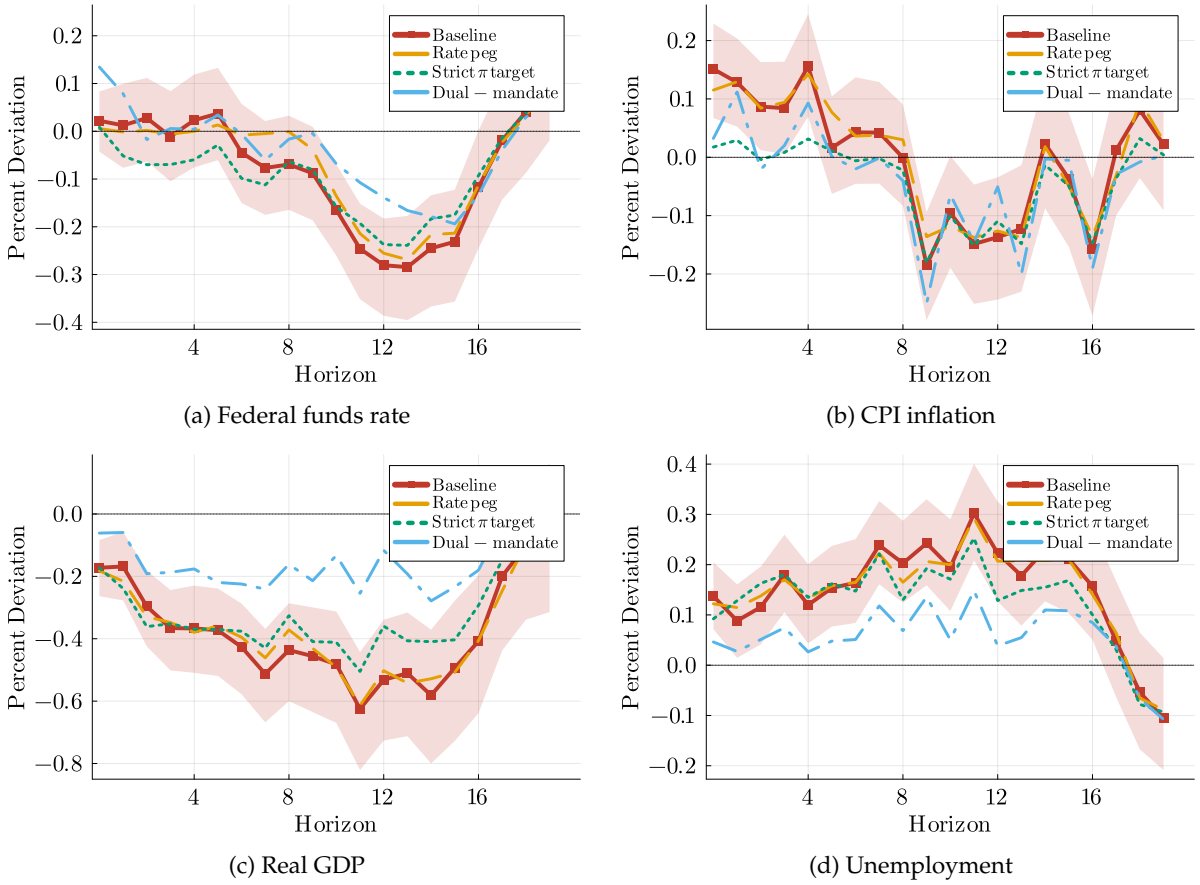


Figure 1: Macro counterfactual responses at $T_c = 4$ ($K = 4$ exact-fit case for peg and strict-π). Solid red: baseline response; orange dash: rate peg; green dot: strict π-target; blue dash-dot: dual-mandate. Shaded band: 68% credible interval on the realized response. The peg rule pins FFR to zero exactly over the first four quarters; the strict π rule pins inflation analogously. $H = 20$.

quadratic loss in both inflation and the unemployment gap, and its m_{aux} path splits the difference between the peg and strict-π extremes. At $T_c = 4$, it exchanges much higher rates near impact to soften the inflation hike, but without the typically associated output drop. The output response is milder relative to baseline and the unemployment response, in the short-run, is improved. At $T_c = 6$, the same pattern emerges. At $T_c = 8$ the dual-mandate path softens the GDP trough even further and the unemployment hump as well, *but* does not hike rates as much as the previous two cases, accepting a bit more inflation. At $T_c = 20$, it even wants to *lower* rates relative to baseline in the short-run, but in exchange for smaller (than baseline) cuts in the medium-run and in some sense, resembles the rate-peg at $T_c = 20$, which achieves the best outcome at the macro level.



(a) Federal funds rate

(b) CPI inflation

(c) Real GDP

(d) Unemployment

Figure 2: Macro counterfactual responses at $T_c = 8$ ($K = 6$ baseline, over-determined system). Lines and bands as in Figure 1.

5.2 Channel decomposition

5.2.1 What is the Fed responding to?

Figure 3 reports the channel decomposition of the realized FFR IRF to the Känzig (2021) oil supply news shock under the in-sample policy. This is the empirical answer to the Kilian and Lewis (2011) question: *what is the Fed actually reacting to when an oil supply shock hits?* Each bar at horizon h stacks the contribution of each mediator block to the FFR response; the sum equals the posterior median IRF (the white markers).

First note the baseline FFR IRF in Figure 3. It hovers around zero in the short run, then decreases strongly after two years. This near-zero short-run response is the net effect of several mediator channels pulling in opposite directions, which is unobserved otherwise with just an IRF. In Chapter 2 this is covered, but as a reminder, for some horizon h, a single bar in Figure 3 aggregates the contribution of one mediator (or channel) summed over several mediation-lag pathways ($n = 0, \ldots, 16$), following the truncated Dufour and Wang (2024) decomposition. As an example, the Oil Price bar at horizon $h = 8$ sums the eight pathways 'oil shock $\rightarrow$ oil price at $t = n \rightarrow$ FFR at $t = 8$' for $n \in 0, 1, 2, ..., 8$ — i.e., the oil shock activating the

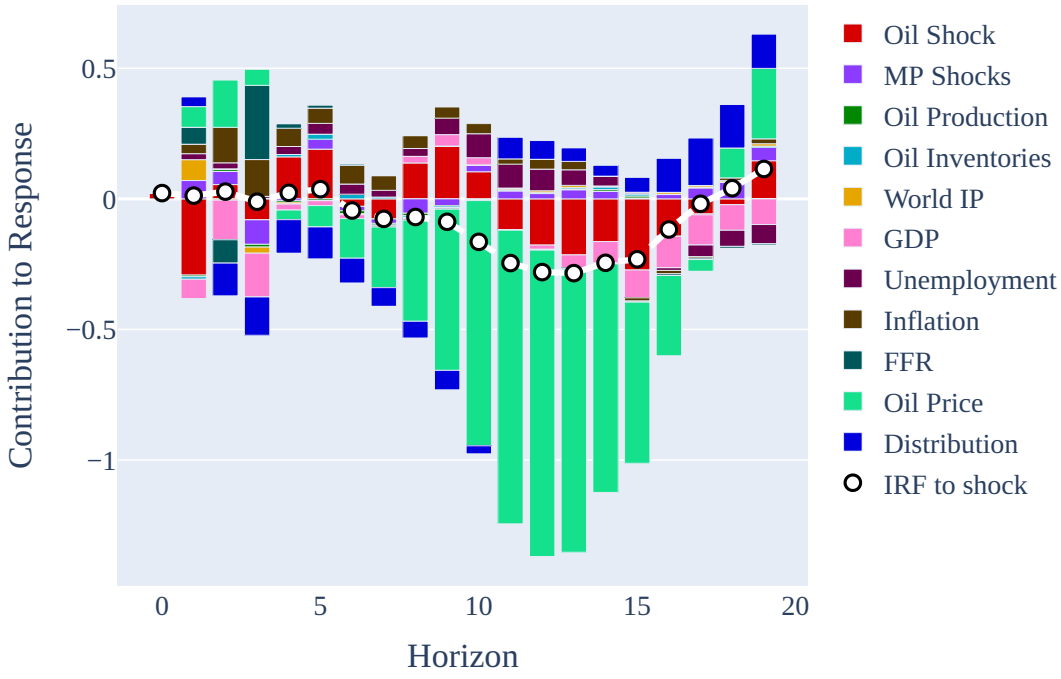


Figure 3: Channel decomposition of the realized FFR IRF to the Känzig (2021) oil supply news shock under the in-sample policy, $K = 6$ baseline. Each bar at horizon h stacks the contribution of each mediator block; the white-circle markers trace the posterior median IRF. The MP Shocks block is zero by construction in the baseline; in counterfactual panels (Figure 4) it carries the contribution of the auxiliary monetary innovations m_{aux}.

oil-price mediator at impact and at each of the next eight quarters, with the resulting price movement then propagating through the VAR's autoregressive structure from $t = n$ to the response horizon h.

With that, what does the decomposition show? The decomposition shows the bank indeed faces a trade-off between prices and output in the presence of an oil supply shock. Inflation (brown) and the oil price (light teal) carry positive contributions, meaning it induces the bank to want to increase rates, while the demand side (GDP-pink, the distribution-dark blue) induce the bank to decrease rates. Both sets of channels cancel each other out, with slight dominance for price stability. The inflation channel itself would remain positive throughout (the Fed always wants to lean against inflation pass-through), but its magnitude is small relative to the oil-price channel; over the medium horizon the central bank "looks through" headline inflation and responds primarily to the oil-price-induced output contraction (as explicitly shown in the decompositions later). This is the textbook cost-channel pattern (Bernanke and Blanchard, 2023; Gagliardone and Gertler, 2025): short-run hawkish, medium-run dovish, long-run neutral.

Upon an oil supply shock, output and unemployment as mediators would play a secondary role in the Fed's reaction function relative to the oil-price channel. The realized reaction

function is also not purely a textbook two-target Taylor rule but also responds to distributional shifts induced by the oil shock. The distributional contribution turns mildly negative through the medium horizon (along with the oil-price channel) and recovers at long horizons. The oil-block channels — oil production, oil inventories, and world industrial production — carry virtually zero contribution throughout the horizon. The interpretation is straightforward: the Fed responds to the oil shock primarily through the relative-price and real-activity transmission, not through the global supply-side dynamics of the oil market itself. The Fed's reaction function is essentially silent on the upstream oil-market block conditional on the downstream macroeconomic state.

Mediation and counterfactual: two views on the same object. The Dufour and Wang (2024) decomposition is identified from the reduced-form VAR alone and is purely descriptive of the realized IRF under the in-sample policy. Reading the indirect effect as a counterfactual—"what would the IRF look like if the channel $\mathcal{M}$ were shut off"—requires the policy-invariance assumption underlying the CMW machinery (McKay and Wolf, 2023b; Caravello, McKay, and Wolf, 2024; Wang, 2026). Under that assumption, when the FFR-channel is shut off through the indirect-effect computation, the resulting paths coincide—in the linear-Gaussian limit, and up to the McKay and Wolf (2023b) small-perturbation residual—with the IRFs that would obtain under a nominal interest-rate peg implemented through the auxiliary policy shocks. The mediation decomposition and the CMW counterfactual are two views on the same object: the former through a channel-attribution lens that does not need to solve for the implied policy shocks, the latter through an alternative-rule lens that does. Their agreement is itself a robustness check on policy invariance, and their *disagreement* is a measure of the Lucas-critique residual. The honest statement of what policy invariance buys is the one Kilian and Lewis (2011) make in their paper: it is not a *defense* of the Lucas critique, it is a *waiver*, and the waiver is credible only insofar as the contemplated rules can be implemented through monetary innovations that look like ones the data have already shown me.

The FFR decomposition can also be used (with caution) to implement an "implicit peg via mediation" counterfactual described in Section 4.3: removing the *oil-price block* of the FFR response (light teal) and propagating the remaining FFR path (inflation-, activity-, distribution-, own-FFR-driven) through to any other variable's IRF. This delivers Kilian and Lewis (2011, Section 3.2.2)'s preferred counterfactual: "the Federal Reserve reacts to fluctuations in other macroeconomic state variables as it normally would with only the direct response to the real price of oil being shut down." In this world, the FFR would definitely increase, leaning and eventually falling into the wind.

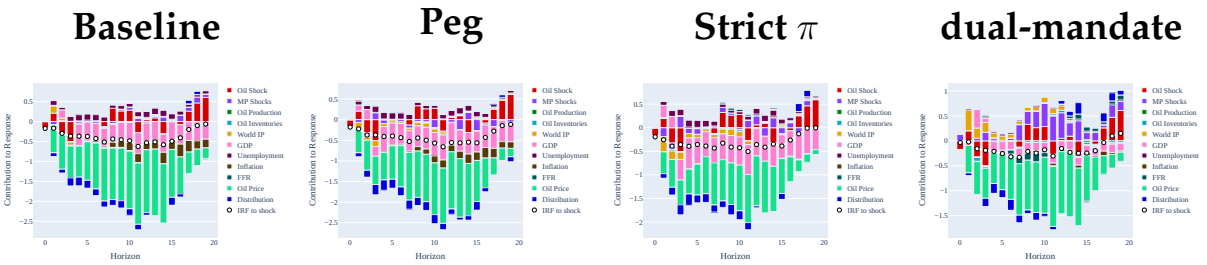
5.2.2 Channel decomposition under each rule

Figure 4 reports the full channel decomposition of the four key macro outcomes (GDP, inflation, unemployment, FFR) under the realized policy and the three counterfactual rules, all at the $K = 6$ baseline with $T_c = 6$. A lot can be learned from these decompositions. I will make two points on the machinery, particularly how the implementation of each rule is exemplified through the lens of the decomposition. I close with some discussion on the most prominent channels driving the observed effects.

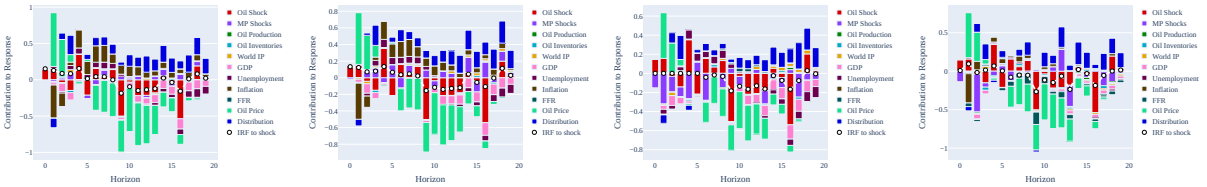
Rules do what they were set out to do. This speaks to the functionality of the two machinery (Dufour and Wang, 2024; Caravello, McKay, and Wolf, 2024). Under Panel (a) Real GDP, the inflation channel that *is* observed under *Baseline* and *Peg* is *completely gone* under strict inflation targeting and the dual-mandate—precisely what the rules were set out to do. This is also observed when looking at CPI inflation for Panel (b) CPI inflation. For Panel (c) Unemployment, the unemployment and inflation channel are indeed completely suppressed for the dual-mandate and for Panel (d) Federal funds rate, the central bank has no feedback contribution through its own lag dynamics.

MP shocks shows the degree of intervention. The size and sign of MP shocks channel (purple) under each counterfactual give the contribution of the auxiliary monetary innovations m_{aux} to the outcome's response at each horizon. A larger bar means the Fed is intervening more aggressively to enforce the rule; the MP Shocks bar is zero by construction in the baseline column. From all the rules, the dual-mandate relies on more intervention.

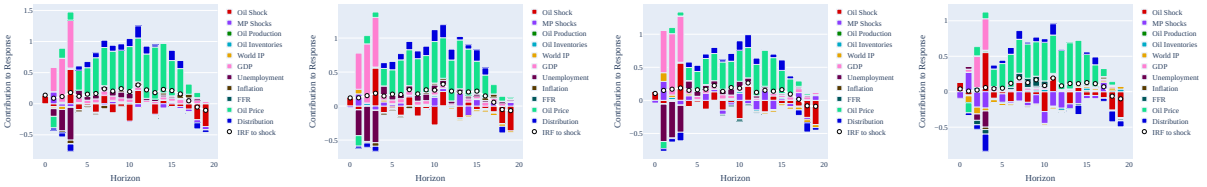
The oil price, the distribution, and the FFR. The oil price has considerable impact across all aggregates shown. Without even zooming in on each plot, you can observe that the oil price channel dominates and drives all responses, especially in the medium-run. The effect is strongest for Real GDP and weakest, ironically, for CPI inflation. The distribution seems to amplify inflation, output, and unemployment, with the largest contribution to inflation dynamics. The FFR plays little to no role in the propagation of aggregate responses, confirming and strengthening results by Kilian and Lewis (2011).



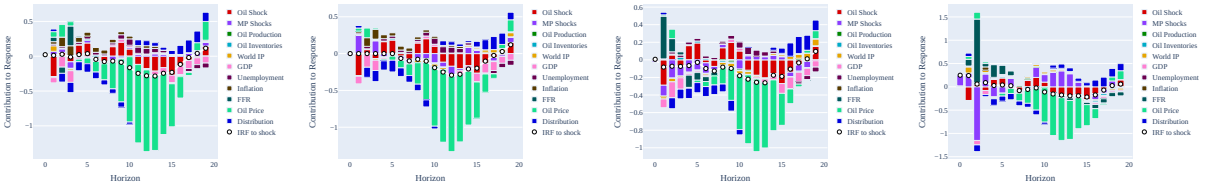
(a) Real GDP



(b) CPI inflation



(c) Unemployment



(d) Federal funds rate

Figure 4: Channel decomposition of the four macro outcomes under each policy rule, $K = 6$ specification at $T_c = 6$. Columns correspond to the realized baseline and the three counterfactual rules. The MP Shocks block (zero in the baseline column by construction) gives the share of each outcome's response mediated by the auxiliary monetary innovations m_{aux} under the rule.

5.3 Distributional counterfactuals

This section turns to the distributional responses. I report Gini coefficients, top 10% shares, and bottom 50% shares for income, wealth, and consumption. The main-text figures fix the constraint horizon at $T_c = 6$; the same panels at $T_c = 20$ are reported in Appendix B.

5.3.1 Gini coefficients

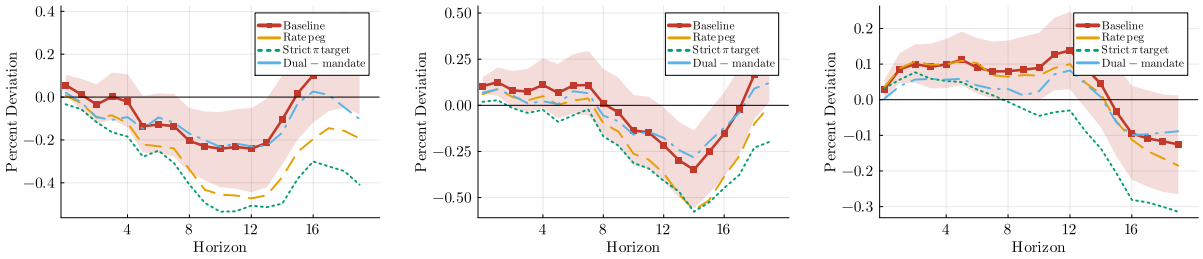
Figure 5 reports the three Ginis under the four monetary rules at $T_c = 6$ (the constraint binds for six quarters; the rules then lift and the system evolves under the realized policy). Two features structure the figure: an early-horizon bunching of all four lines while the constraint binds, and a medium-run fan-out in which strict inflation targeting stands apart from the other three rules.

The realized path is itself compressionary. A point worth surfacing before discussing the rules: under realized policy the oil shock generates a small positive Gini bump in the first year on all three dimensions, then turns negative through the medium horizon, troughing in the second-to-fourth year and recovering toward zero. The oil shock therefore *compresses* all three Ginis in the medium run, on average.

Income Gini. All four lines bunch tightly through the constraint window ($h \leq 6$): a small positive deviation of $+0.1$ to $+0.25$ percentage points in the first year, then crossing zero. The rules fan out from $h \approx 8$ onward. The rate peg tracks the realized response in the short-run, but deviates in the medium-run. The dual-mandate rule sits slightly above the realized response in the medium horizon (i.e. marginally less compression). The strict-π target is the standout: it deepens the realized trough by roughly twenty per cent, bottoming at about -0.50 at $h = 14$ against the realized trough of about -0.35, and recovers more slowly. The strict-π rule therefore *amplifies the compressionary effect* of the oil shock on the income Gini.

Wealth Gini. The realized wealth Gini path is muted – a small positive bump ($+0.1$) over the first year, dipping to about -0.1 at $h \approx 10$, recovering. The rate peg and dual-mandate rules sit close to the realized path throughout. Strict inflation targeting, again, is the standout: it pulls the wealth Gini further down to a trough of about -0.30 at $h \approx 14$, a roughly three-fold amplification of the realized compression.

Consumption Gini. The realized consumption Gini follows the income pattern at smaller amplitude: a brief positive bump, then a negative dip troughing around -0.20 at $h \approx 10$, recovery toward zero by $h = 18$. The dual-mandate again tracks the realized path. Strict



(a) Consumption Gini (b) Income Gini (c) Wealth Gini

Figure 5: Counterfactual Gini responses to the Känzig (2021) oil supply news shock, $K = 6$ baseline, $T_c = 6$. Lines: realized (solid red), rate peg (dashed orange), strict π-target (dotted teal), dual-mandate (dash-dot light blue). Bands shown only for the realized response. Vertical axis: percent deviation of the Gini coefficient from steady state.

inflation targeting and the rate-peg amplifies the trough to about -0.5 at $h \approx 11$, recovering thereafter. The amplification factor under strict-π is the largest of the three Ginis: roughly two-and-a-half-fold relative to the realized trough.

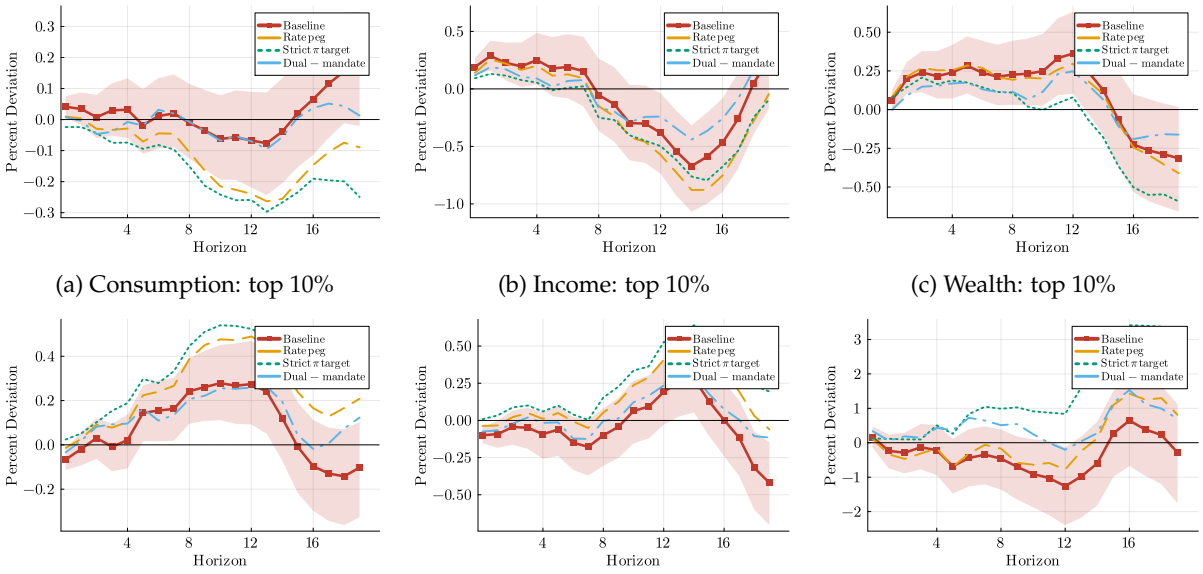
Cross-rule comparison. The summary statement is sharper than at the macro level. One of the three rules – dual-mandate – delivers near-indistinguishable Gini paths on all three margins. The strict-π rule amplifies the medium-run compression on all three margins, with the largest amplification for consumption and the smallest for income. The rate-peg equally compresses, but follows the realized path for wealth.

The trade-off the strict-π rule purchases at the macro level shows up in the distribution as a deeper and more persistent Gini compression. Whether this is welfare-improving or welfare-reducing depends on the social welfare function: a planner that values inequality compression mechanically prefers strict π, but the compression may be from declines at the top, not by gains for the bottom – a point I visit in the next section.

5.3.2 Top decile and bottom half shares

To understand the mechanism behind the Gini results, I decompose each Gini into its top 10% and bottom 50% components. Figure 6 reports the top 10% (row 1) and bottom 50% (row 2) shares for income, consumption, and wealth. The patterns clarify why all three Ginis amplify under strict inflation targeting.

Decomposition of the Gini movement. The three Ginis may differ in which margin does the heavier work. First, I discuss the strict inflation targeting rule (dotted-green line). On consumption, we observe a negative change in the top-10% share and a larger positive change for the bottom-50%. On income, the top-10% movement is slightly more pronounced than the bottom-50% movement, which explains the small difference observed in the ginis. On wealth,



(a) Consumption: top 10% (b) Income: top 10% (c) Wealth: top 10%

(d) Consumption: bottom 50% (e) Income: bottom 50% (f) Wealth: bottom 50%

Figure 6: Counterfactual top 10% (row 1) and bottom 50% (row 2) shares for income, consumption, and wealth, $K = 6$ baseline, Känzig (2021) oil supply news shock, $T_c = 6$. Lines as in Figure 5.

the bottom-50% movement dominates by a large margin across the IRF horizon, which is why the wealth Gini under strict-π shows the largest amplification of the three. Overall, the rate-peg stays close to baseline, except for consumption, while the strict-inflation target compresses inequality through large redistribution to the bottom. Dual-mandate follows the realized path.

6 Conclusion

I have shown how the CMW sufficient-statistics framework can be married to a multi-shock distributional VAR to recover the counterfactual income, wealth, and consumption distributions that would have prevailed under three alternative monetary rules—a nominal rate peg, strict inflation targeting, and an optimal dual-mandate rule—in response to an identified Känzig (2021) oil supply news shock. The key methodological contribution is the recognition that the linear superposition that delivers the CMW macro counterfactual extends, draw-by-draw, to the underlying factor IRFs and hence to any non-linear function of the joint distribution. The key empirical contribution is a quantitative estimate of the distributional cost of the three counterfactual rules at a six-quarter constraint window. The dual-mandate rule delivers Gini paths nearly indistinguishable from the realized response. The rate peg and strict inflation targeting both compress inequality further than the realized path, with strict-π the larger amplifier on every margin (by roughly twenty per cent for income, two-and-a-half-fold for consumption, and threefold for wealth). The top-decile/bottom-half decomposition attributes the strict-π compression to redistribution toward the bottom of the distribution rather than

top-tail destruction.

The substantive lesson is that the cross-section discriminates between rules that the headline macro aggregates cannot. The three counterfactual rules deliver broadly similar inflation, GDP, and unemployment paths; their distributional footprints, by contrast, separate strict-π and the rate peg from the dual-mandate in a quantitatively non-negligible way. A central bank that incorporates distributional moments into its loss function would rank the rules differently than one that does not. I see this as a strong case for incorporating distributional moments directly into central-bank loss functions, along the lines proposed by Bhandari et al. (2021) and others; the CMW framework offers a tractable empirical complement to that theoretical literature.

Future work. Four extensions are immediate. First, discriminating between *types* of supply shocks matters: an oil-supply news shock, an oil-demand shock, and a broader non-oil supply disturbance can trigger very different Fed reaction functions and distributional responses, and the chapter's machinery is well-suited to that shock-by-shock comparison. Second, the Wang (2026) local-projection counterpart to the CMW machinery offers a complementary identification of the auxiliary monetary block, with different small-sample and misspecification properties; running the same rule set under his counterfactual machinery would provide an informative robustness check on the BVAR-based results reported here. Third, the chapter has stipulated the central bank as the sole policy authority responding to the oil shock; an oil-side authority—OPEC quota adjustments, strategic-reserve releases, or energy taxes—is a complementary lever whose counterfactual could be enforced through the same sufficient-statistics machinery, identifying the rule through historically observed oil-supply or oil-inventory innovations rather than through monetary innovations. The cross-authority comparison would clarify whether inequality stabilisation is best pursued on the monetary or the oil-policy margin. Fourth, the same exercise can be run with monetary or fiscal impulses in place of the oil shock; the distributional counterfactuals under alternative fiscal rules would be a useful complement to the recent literature on fiscal multipliers and inequality.

References

[1] Asger Lau Andersen et al. "Monetary policy and inequality". In: (2021).

[2] S Borağan Aruoba and Thomas Drechsel. *Identifying monetary policy shocks: A natural language approach*. Tech. rep. National Bureau of Economic Research, 2024.

[3] Adrien Auclert. "Monetary policy and the redistribution channel". In: *American Economic Review* 109.6 (2019), pp. 2333–2367.

[4] Regis Barnichon and Geert Mesters. "A sufficient statistics approach for macro policy". In: *American Economic Review* 113.11 (2023), pp. 2809–2845.

[5] Michael D Bauer and Eric T Swanson. "A reassessment of monetary policy surprises and high-frequency identification". In: *NBER Macroeconomics Annual* 37.1 (2023), pp. 87–155.

[6] Christiane Baumeister. *Discussion of" Local Projections or VARs? A Primer for Macroeconomists" by José Luis Montiel Olea, Mikkel Plagborg-Møller, Eric Qian, and Christian K.* Tech. rep. Wolf, Working paper, Prepared for the NBER Macroeconomics Annual, 2025.

[7] Christiane Baumeister and James D. Hamilton. "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks". In: *American Economic Review* 109.5 (2019), pp. 1873–1910.

[8] Christian Bayer, Benjamin Born, and Ralph Luetticke. "Shocks, frictions, and inequality in US business cycles". In: (2020).

[9] Christian Bayer, Luis Calderon, and Moritz Kuhn. *Distributional Dynamics*. Tech. rep. 2025.

[10] Ben S. Bernanke and Olivier J. Blanchard. *What Caused the U.S. Pandemic-Era Inflation?* NBER Working Paper 31417. National Bureau of Economic Research, 2023.

[11] Ben S Bernanke, Mark Gertler, and Mark Watson. "Systematic monetary policy and the effects of oil price shocks". In: *Brookings papers on economic activity* 1997.1 (1997), pp. 91–157.

[12] Ben S Bernanke, Mark Gertler, and Mark W Watson. "Oil shocks and aggregate macroeconomic behavior: the role of monetary policy: a reply". In: *Journal of Money, Credit, and Banking* 36.2 (2004), pp. 287–291.

[13] Ben S Bernanke and Frederic S Mishkin. "Inflation targeting: a new framework for monetary policy?" In: *Journal of Economic Perspectives* 11.2 (1997), pp. 97–116.

[14] Anmol Bhandari et al. "Inequality, Business Cycles, and Monetary-Fiscal Policy". In: *Econometrica* 89.6 (2021), pp. 2559–2599.

[15] Hilde C Bjørnland, Yoosoon Chang, and Jamie Cross. "Oil and the stock market revisited: A mixed functional var approach". In: *Available at SSRN* (2023).

[16] Martin Bodenstein, Luca Guerrieri, and Lutz Kilian. "Monetary policy responses to oil price fluctuations". In: *IMF Economic Review* 60.4 (2012), pp. 470–504.

[17] Tobias Broer, John Kramer, and Kurt Mitman. "The Distributional Effects of Oil Shocks". In: *IMF Economic Review* 73.3 (2025), pp. 851–889.

[18] Tomás E Caravello, Alisdair McKay, and Christian K Wolf. *Evaluating policy counterfactuals: A var-plus approach*. Tech. rep. National Bureau of Economic Research, 2024.

[19] Charles T Carlstrom and Timothy S Fuerst. "Oil prices, monetary policy, and counterfactual experiments". In: *Journal of Money, Credit and Banking* 38.7 (2006), pp. 1945–1958.

[20] Joshua C C Chan. "Asymmetric conjugate priors for large Bayesian VARs". In: *Quantitative Economics* 13.3 (2022), pp. 1145–1169.

[21] Minsu Chang, Xiaohong Chen, and Frank Schorfheide. "Heterogeneity and Aggregate Fluctuations". In: *Journal of Political Economy* 132.12 (2024), pp. 4021–4067.

[22] Minsu Chang and Frank Schorfheide. *On the Effects of Monetary Policy Shocks on Income and Consumption Heterogeneity*. Working Paper 32166. National Bureau of Economic Research, 2024.

[23] Olivier Coibion et al. "Innocent Bystanders? Monetary policy and inequality". In: *Journal of Monetary Economics* 88 (2017), pp. 70–89.

[24] Ferre De Graeve and Andreas Westermark. *How long is long enough? Long-lag VARs*. Tech. rep. Sveriges Riksbank Working Paper Series, 2025.

[25] Jean-Marie Dufour and Eric Renault. "Short run and long run causality in time series: theory". In: *Econometrica* (1998), pp. 1099–1125.

[26] Jean-Marie Dufour and Endong Wang. *Causal mechanism and mediation analysis for macroeconomics dynamics: a bridge of Granger and Sims causality*. Tech. rep. 2024.

[27] Luca Gagliardone and Mark Gertler. "Oil Prices, Monetary Policy and Inflation Surges". In: *American Economic Journal: Macroeconomics* (2025). Forthcoming. Previous version: NBER Working Paper 31263, 2023.

[28] Domenico Giannone, Michele Lenza, and Giorgio E Primiceri. "Priors for the long run". In: *Journal of the American Statistical Association* 114.526 (2019), pp. 565–580.

[29] Lukas Hack, Klodiana Istrefi, and Matthias Meier. *The Systematic Origins of Monetary Policy Shocks*. Working paper. https://cepr.org/publications/dp19063. CEPR Discussion Paper No. 19063 / CRC TR 224 Discussion Paper No. 557, University of Bonn, 2024.

[30] James D Hamilton and Ana Maria Herrera. "Comment: oil shocks and aggregate macroeconomic behavior: the role of monetary policy". In: *Journal of Money, credit and Banking* (2004), pp. 265–286.

[31] Gökhan Ider et al. *Friend, Not Foe? Monetary Policy and Energy Prices*. DIW Discussion Paper. May 2025. DIW Berlin, 2025.

[32] Marek Jarociński. "Estimating the Fed's unconventional policy shocks". In: *Journal of Monetary Economics* (2024).

[33] Marek Jarociński and Peter Karadi. "Deconstructing monetary policy surprises—the role of information shocks". In: *American Economic Journal: Macroeconomics* 12.2 (2020), pp. 1–43.

[34] Diego R Känzig. "The macroeconomic effects of oil supply news: Evidence from OPEC announcements". In: *American Economic Review* 111.4 (2021), pp. 1092–1125.

[35] Greg Kaplan, Benjamin Moll, and Giovanni L Violante. "Monetary policy according to HANK". In: *American Economic Review* 108.3 (2018), pp. 697–743.

[36] Lutz Kilian. "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market". In: *American economic review* 99.3 (2009), pp. 1053–1069.

[37] Lutz Kilian and Logan T Lewis. "Does the Fed respond to oil price shocks?" In: *The Economic Journal* 121.555 (2011), pp. 1047–1072.

[38] Lutz Kilian and Xiaoqing Zhou. "The econometrics of oil market VAR models". In: (2023).

[39] Sylvain Leduc and Keith Sill. "A quantitative analysis of oil-price shocks, systematic monetary policy, and economic downturns". In: *Journal of Monetary Economics* 51.4 (2004), pp. 781–808.

[40] Michele Lenza and Ettore Savoia. *Do we need firm data to understand macroeconomic dynamics?* Working Paper Series 438. Sveriges Riksbank (Central Bank of Sweden), 2024.

[41] Michele Lenza and Jiri Slacalek. *How does monetary policy affect income and wealth inequality? Evidence from quantitative easing in the euro area*. Working paper 2190. Published as Lenza and Slacalek (2024), *Journal of Applied Econometrics*. European Central Bank, 2018.

[42] Robert E. Lucas. "Econometric policy evaluation: a critique". In: *The Phillips Curve and Labor Markets*. Ed. by Karl Brunner and Allan H. Meltzer. Vol. 1. Carnegie–Rochester Conference Series on Public Policy. North-Holland, 1976, pp. 19–46.

[43] Julian F. Ludwig. *Local Projections Are VAR Predictions of Increasing Order*. Tech. rep. SSRN Working Paper 4882149, 2024.

[44] Ralph Luetticke. *A Unified Narrative of Conventional and Unconventional Monetary Policy Shocks*. Working paper. Replication archive: https://github.com/ralphluet/unified_narrative_mp_shocks. University College London, 2024.

[45] Alisdair McKay and Christian K Wolf. "Monetary policy and inequality". In: *Journal of Economic Perspectives* 37.1 (2023), pp. 121–144.

[46] Alisdair McKay and Christian K Wolf. "What Can Time-Series Regressions Tell Us About Policy Counterfactuals?" In: *Econometrica* 91.5 (2023), pp. 1695–1725.

[47] Anton Nakov and Andrea Pescatori. "Oil and the Great Moderation". In: *The Economic Journal* 120.543 (2010), pp. 131–156.

[48] Mikkel Plagborg-Møller and Christian K Wolf. "Local projections and VARs estimate the same impulse responses". In: *Econometrica* 89.2 (2021), pp. 955–980.

[49] Michael Plante. "How should monetary policy respond to changes in the relative price of oil? Considering supply and demand shocks". In: *Journal of Economic Dynamics and Control* 44 (2014), pp. 1–19.

[50] Christina D Romer and David H Romer. "A new measure of monetary shocks: Derivation and implications". In: *American economic review* 94.4 (2004), pp. 1055–1084.

[51] Christopher A Sims and Tao Zha. "Does monetary policy generate recessions?" In: *Macroeconomic Dynamics* 10.2 (2006), pp. 231–272.

[52] Abe Sklar. "Vol. 8 of Fonctions de Répartition à n Dimensions et Leurs Marges, 229–231". In: *Paris: Publications de l'Institut de statistique de l'Université de Paris* (1959).

[53] James H Stock and Mark W Watson. "Identification and estimation of dynamic causal effects in macroeconomics using external instruments". In: *The Economic Journal* 128.610 (2018), pp. 917–948.

[54] Endong Wang. *Local projections identify the same policy counterfactuals as empirical and structural models*. Working paper. https://arxiv.org/abs/2409.09577. University of Mannheim, 2026.

A Additional macro counterfactual robustness

This appendix reports the macro counterfactual responses at the two remaining constraint horizons in the robustness sweep ($T_c = 6$ in Figure 7, $T_c = 20$ in Figure 8) and the real oil-price column for all four constraint horizons (Figure 9). The oil-price column is collected separately because monetary policy has at most a small effect on the global oil market at quarterly frequency (Kilian, 2009; Baumeister and Hamilton, 2019), so the counterfactual deviations on the oil-price are mechanically small and do not discriminate between rules.

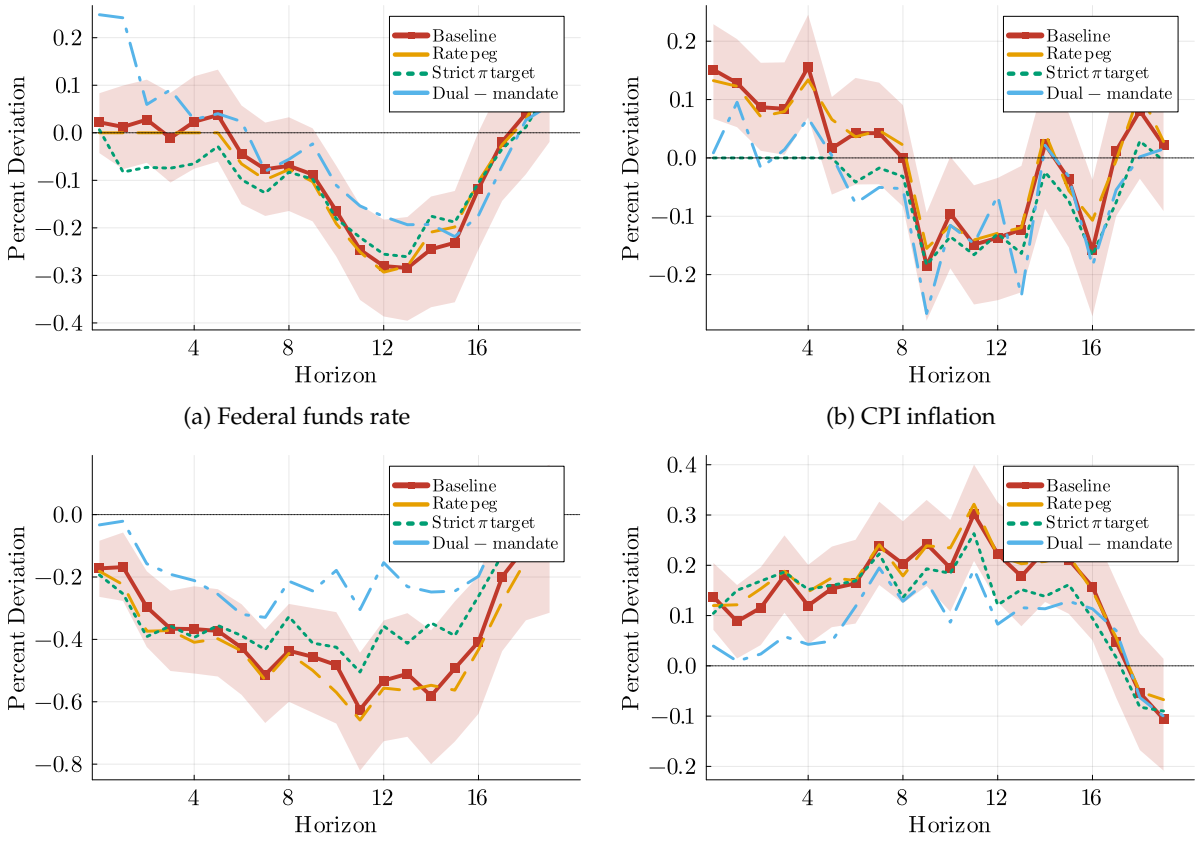


Figure 7: Macro counterfactual responses at $T_c = 6$ ($K = 6$ baseline, exact-fit case where $T_c = K$). Lines and bands as in Figure 1. The square-system case for the $K = 6$ specification: peg and strict-π projections are exact over the first six quarters.

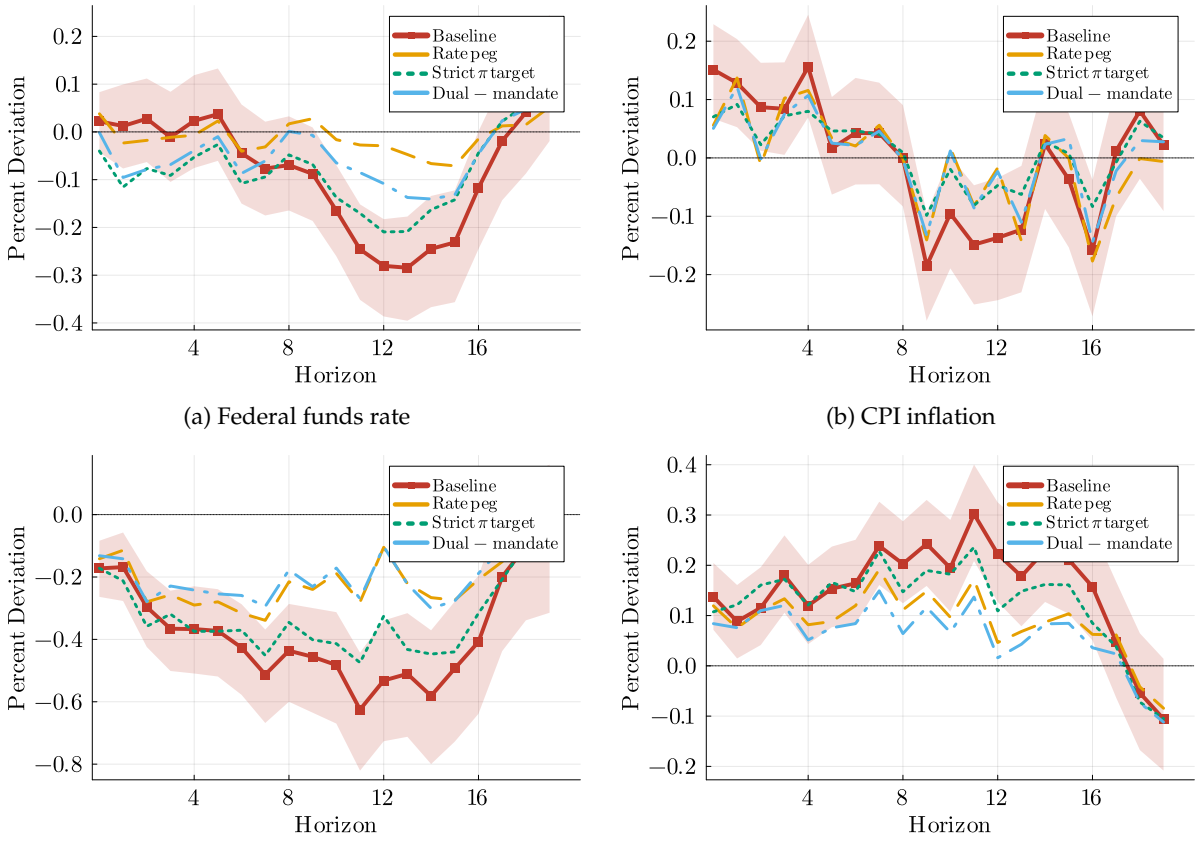
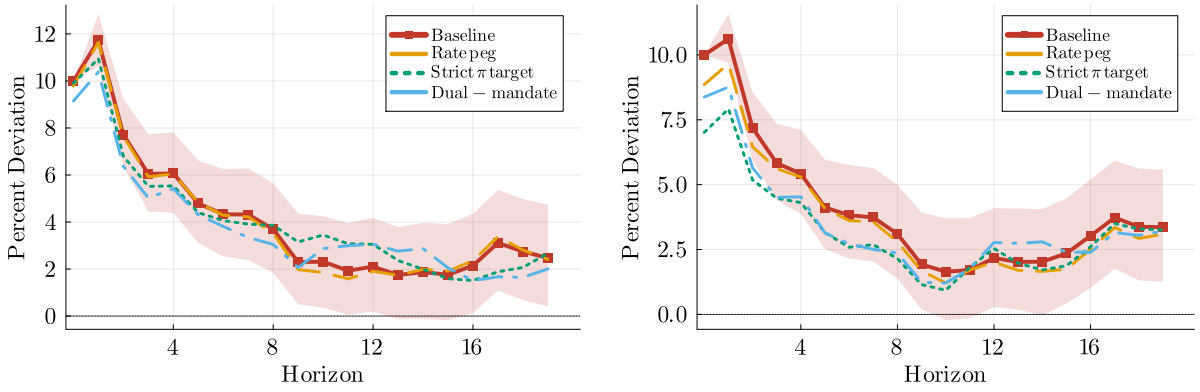
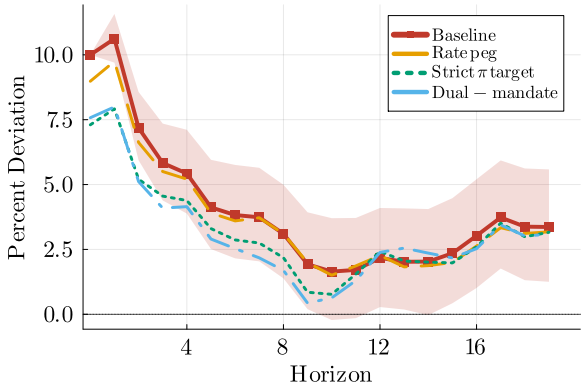


Figure 8: Macro counterfactual responses at $T_c = 20$ ($K = 6$ baseline, full-horizon enforcement). Lines and bands as in Figure 1. This is the largest-T_c case in the robustness sweep; the rule is enforced over the entire plotted horizon. The conclusions are qualitatively identical to Figure 2.

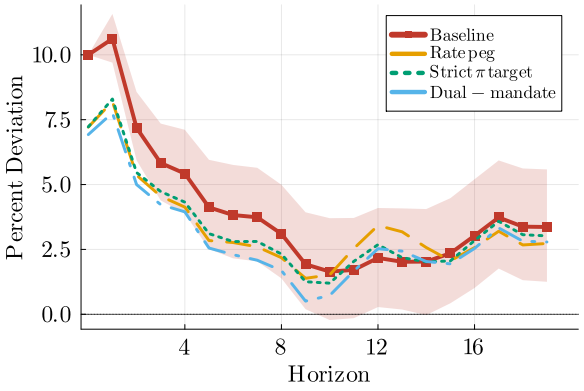


(a) $T_c = 4$ ($K = 6$ unified)

(b) $T_c = 6$ ($K = 6$ baseline)



(c) $T_c = 8$ ($K = 6$ baseline)

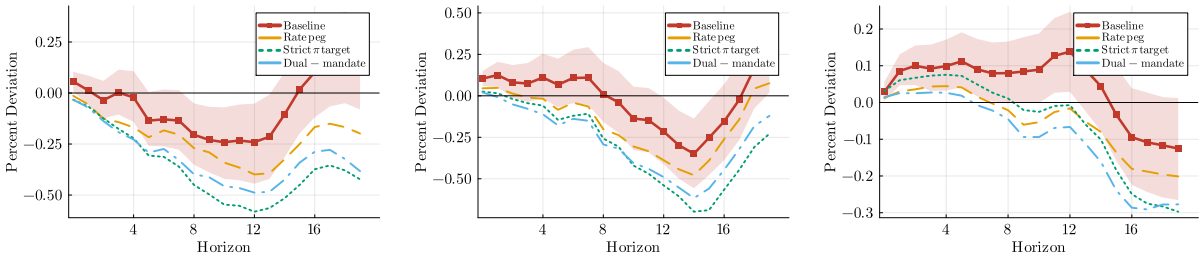


(d) $T_c = 20$ ($K = 6$ baseline)

Figure 9: Real-oil-price counterfactual responses by constraint horizon T_c. Lines and bands as in Figure 1. Counterfactual deviations on the real oil price are small under all rules at every T_c, consistent with the standard finding that domestic monetary policy has at most a modest effect on the global oil market at quarterly frequency (Kilian, 2009; Baumeister and Hamilton, 2019).

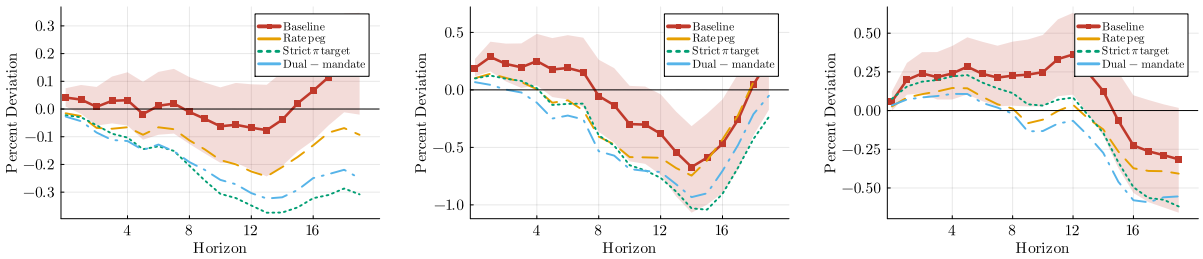
B Distributional counterfactuals at $T_c = 20$

This appendix replicates the distributional figures of Section 5.3 at the long constraint horizon $T_c = 20$, in which the rule binds for the entire response window. The main-text figures report $T_c = 6$ as the headline; the $T_c = 20$ panels here serve as a long-horizon robustness check and amplify the same qualitative ordering across rules.

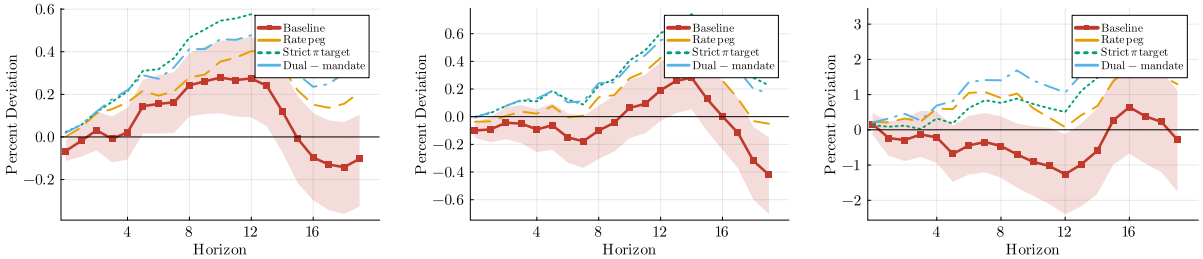


(a) Consumption Gini (b) Income Gini (c) Wealth Gini

Figure 10: Counterfactual Gini responses at $T_c = 20$. Känzig (2021) oil supply news shock, $K = 6$ baseline. Lines as in Figure 5.



(a) Consumption: top 10% (b) Income: top 10% (c) Wealth: top 10%



(d) Consumption: bottom 50% (e) Income: bottom 50% (f) Wealth: bottom 50%

Figure 11: Top 10% (row 1) and bottom 50% (row 2) shares at $T_c = 20$, $K = 6$ baseline, Känzig (2021) oil supply news shock. Lines as in Figure 5.

C T_c-robustness diagnostics and the Lucas-critique residual

Table 2 reports the CMW diagnostics for every (specification, T_c, rule) combination in the robustness sweep, separately for the BH and Kanzig oil shocks. Each row reports three quantities at the posterior median (with the 16th–84th percentile band in brackets): the L2 norm of the auxiliary monetary impulse vector m_{aux}, the condition number of $M_1'M_1$ (the policy-span Gram matrix), and the rule-fit ratio $\|M_1 m_{\text{aux}} + M_2\|/\|M_2\|$ (zero = rule exactly enforced, one = no enforcement). All entries use the same 5% per-rule trim on ill-conditioned draws described in Section 4.3; the $K = 6$ entries at $T_c = 4$ for the hard-constraint rules show `NaN` because the projection is under-determined ($T_c < K$) and all draws are masked.

Reading the table. The diagnostics behave as the algebra predicts. As T_c falls toward K, the rule-fit ratio drops monotonically; at $T_c = K$ (unified $T_c = 4$, K6 $T_c = 6$) the projection is exact for the hard-constraint rules and the ratio is essentially zero. For $T_c > K$ the projection is

over-determined and the ratio rises with $T_c - K$ until it stabilizes around the values reported at $T_c = 8$. The condition number $\text{cond}(M_1' M_1)$ is largest at $T_c = K$ (where the square system has condition-squared sensitivity to the underlying M_1) and falls as additional rows over-determine the projection.

The Lucas-critique residual. The economically interesting quantity is $\|m_{\text{aux}}\|_2$, because McKay and Wolf (2023b) prove that the Lucas-critique residual—the part of agents' decision rules that would actually adjust under a policy change but that the sufficient-statistics counterfactual cannot capture—is $O(\|m_{\text{aux}}\|^2)$. Smaller $\|m_{\text{aux}}\|$ therefore means a smaller Lucas residual and a more credible policy-invariance restriction.

Three patterns in Table 2 bear on how hard the Lucas critique bites in this exercise.

(i) Kanzig is comfortably inside the small-perturbation regime. Under the Kanzig shock, $\|m_{\text{aux}}\|_2$ medians range from 0.14 to 0.85 across all specifications and rules, with all upper-tail (84th-percentile) values below 3 except for the two $K = 6$ $T_c = 6$ square-system entries. Squaring the medians to bound the McKay–Wolf residual gives Lucas-residual upper bounds under 0.04 for the $K = 4$ unified spec and under 0.7 for the largest $K = 6$ cases. These are small relative to the size of the counterfactual deviations themselves (FFR deviations of roughly 0.4 percentage points at peak), so the Lucas residual is quantitatively negligible for the headline Kanzig results.

(ii) The BH shock sits on the edge of the regime. BH $\|m_{\text{aux}}\|_2$ medians are roughly four to five times larger than Kanzig's (e.g. 1.18 for the unified $T_c = 4$ peg, 2.16 for the $K = 6$ $T_c = 6$ peg) with upper-tail values reaching 7.5. Squaring these gives Lucas-residual upper bounds of 1.4 to 30, no longer negligible relative to the counterfactual deviations. The BH results reported alongside the headline should therefore be read as *quantitatively* more sensitive to the Lucas critique than the Kanzig headline; the qualitative ordering across rules is preserved, but the magnitudes carry a larger interpretation tax. This is the empirical reason for adopting the Kanzig HF identification as the headline shock and reporting BH as robustness rather than the other way around.

(iii) Smaller T_c does not always mean smaller $\|m_{aux}\|$. The natural intuition is that pinning the rule over a shorter window requires a smaller impulse. The table shows that this is true comparing $T_c = 20$ to $T_c = 8$ within an over-determined spec, but is reversed in the exact-fit case ($T_c = K$). At $T_c = K$ the system is square and the projection must absorb *all* of the rule violation M_2 through m_{aux}, with no over-determination to soften the load; the result is a larger $\|m_{\text{aux}}\|$ and a worse $\text{cond}(M_1' M_1)$ than the over-determined neighbours. Under the BH shock, $\|m_{\text{aux}}\|_2$ for the $K = 6$ $T_c = 6$ peg is 2.16 vs. 1.08 at $T_c = 8$ – a factor of two reduction by moving one quarter further over-determined. The empirically Lucas-safest specification under the BH

shock is therefore $T_c = 8$ with $K = 6$, not the exact-fit $T_c = 6$.

Implication for the headline results. The two main-text specifications in Section 5.1 ($T_c = 4$ $K = 4$ and $T_c = 8$ $K = 6$) sit at posterior-median $\|m_{\text{aux}}\|_2$ values of 0.19 and 0.14 under Kanzig, respectively, with corresponding Lucas-residual upper bounds of 0.04 and 0.02. These are well inside the small-perturbation regime in which McKay and Wolf (2023b)'s identification result is tight, and the policy-invariance restriction is empirically defensible in the sense of Wang (2026)'s $\Delta_{\mathcal{Q}}$ persistence test. The Lucas critique bites quantitatively when the exercise is pushed to large-T_c $K = 6$ specifications or to the BH shock, but neither configuration is what the headline reports.

Table 2: CMW diagnostics across T_c and rule, by oil shock. Median [16th, 84th percentile] across posterior draws (5% top-cond trim).

			$\|m_{\text{aux}}\|_2$		cond$(M_1'M_1)$		rule-fit ratio	
Variant	T_c	Rule	med	[16,84]	med	[16,84]	med	[16,84]
Panel A. BH supply shock								
unified	4	peg	1.18	[0.42, 4.07]	$3.3{\cdot}10^4$	$[3.4{\cdot}10^3, 9.3{\cdot}10^5]$	0.00	[0.00, 0.00]
unified	4	strict π	1.54	[0.59, 5.39]	2285	$[219, 7.2{\cdot}10^4]$	0.00	[0.00, 0.00]
unified	4	dual-mandate	0.59	[0.29, 1.25]	510	[83, 5938]	0.70	[0.52, 0.85]
unified	6	peg	0.48	[0.20, 1.18]	7193	$[1143, 9.3{\cdot}10^4]$	0.37	[0.18, 0.61]
unified	6	strict π	0.59	[0.29, 1.32]	510	[84, 6636]	0.53	[0.30, 0.75]
unified	6	dual-mandate	0.40	[0.19, 0.75]	341	[58, 3987]	0.80	[0.67, 0.91]
unified	8	peg	0.49	[0.20, 1.00]	3969	$[732, 5.2{\cdot}10^4]$	0.48	[0.29, 0.69]
unified	8	strict π	0.45	[0.21, 0.89]	358	[64, 4485]	0.67	[0.47, 0.83]
unified	8	dual-mandate	0.32	[0.15, 0.64]	278	[49, 3297]	0.86	[0.76, 0.94]
K6	4	peg		NaN		NaN		NaN
K6	4	strict π		NaN		NaN		NaN
K6	4	dual-mandate	1.22	[0.65, 2.42]	2207	$[318, 3.0{\cdot}10^4]$	0.36	[0.19, 0.55]
K6	6	peg	2.16	[0.94, 7.45]	$1.5{\cdot}10^5$	$[1.4{\cdot}10^4, 5.8{\cdot}10^6]$	0.00	[0.00, 0.00]
K6	6	strict π	1.91	[0.83, 6.73]	7833	$[699, 2.4{\cdot}10^5]$	0.00	[0.00, 0.00]
K6	6	dual-mandate	0.72	[0.40, 1.20]	726	$[129, 1.0{\cdot}10^4]$	0.60	[0.44, 0.74]
K6	8	peg	1.08	[0.54, 2.11]	$1.8{\cdot}10^4$	$[3048, 2.8{\cdot}10^5]$	0.27	[0.13, 0.45]
K6	8	strict π	1.05	[0.58, 1.99]	1298	$[222, 2.0{\cdot}10^4]$	0.44	[0.22, 0.66]
K6	8	dual-mandate	0.68	[0.38, 1.14]	534	[97, 7427]	0.71	[0.58, 0.83]
Panel B. Kanzig oil supply news shock								
unified	4	peg	0.19	[0.07, 0.67]	$1.3{\cdot}10^4$	$[1301, 3.7{\cdot}10^5]$	0.00	[0.00, 0.00]
unified	4	strict π	0.40	[0.17, 1.32]	1379	$[130, 4.5{\cdot}10^4]$	0.00	[0.00, 0.00]
unified	4	dual-mandate	0.23	[0.12, 0.42]	240	[44, 3230]	0.50	[0.33, 0.67]
unified	6	peg	0.15	[0.07, 0.36]	2641	$[416, 3.1{\cdot}10^4]$	0.30	[0.14, 0.53]
unified	6	strict π	0.24	[0.12, 0.49]	274	[48, 4075]	0.46	[0.23, 0.67]
unified	6	dual-mandate	0.20	[0.10, 0.37]	160	[29, 1957]	0.65	[0.50, 0.79]
unified	8	peg	0.14	[0.06, 0.30]	1633	$[276, 2.0{\cdot}10^4]$	0.46	[0.28, 0.68]
unified	8	strict π	0.18	[0.09, 0.34]	202	[36, 2267]	0.58	[0.39, 0.77]
unified	8	dual-mandate	0.19	[0.10, 0.35]	128	[24, 1587]	0.73	[0.59, 0.86]
K6	4	peg		NaN		NaN		NaN
K6	4	strict π		NaN		NaN		NaN
K6	4	dual-mandate	0.50	[0.27, 0.93]	1400	$[220, 2.2{\cdot}10^4]$	0.33	[0.17, 0.52]
K6	6	peg	0.68	[0.27, 2.30]	$1.2{\cdot}10^5$	$[1.2{\cdot}10^4, 3.2{\cdot}10^6]$	0.00	[0.00, 0.00]
K6	6	strict π	0.85	[0.36, 2.87]	5255	$[463, 1.6{\cdot}10^5]$	0.00	[0.00, 0.00]
K6	6	dual-mandate	0.33	[0.20, 0.56]	459	[92, 7540]	0.57	[0.41, 0.73]
K6	8	peg	0.39	[0.19, 0.77]	$1.4{\cdot}10^4$	$[2429, 2.0{\cdot}10^5]$	0.26	[0.12, 0.47]
K6	8	strict π	0.39	[0.21, 0.72]	946	$[155, 1.6{\cdot}10^4]$	0.41	[0.22, 0.62]
K6	8	dual-mandate	0.33	[0.19, 0.51]	359	[75, 6267]	0.64	[0.50, 0.79]

Notes. All entries pool 2000 posterior draws after the per-rule 5% top-cond trim. $K = 4$ unified spec uses Lütke + BS + HIM-RR + HIM-AD; $K = 6$ spec uses Lütke + 4 Jarociński factors + HIM-RR. `NaN` = under-determined system ($T_c < K$ for hard-constraint rules); all draws masked. The square-system case ($T_c = K$) has rule-fit ≈ 0 for peg and strict π by construction.

Lebenslauf

Luis Calderon

geboren in Miami, FL, USA

Akademischer Werdegang

seit 10/2020	Promotionsstudium (Ph.D. in Economics), Bonn Graduate School of Economics, Rheinische Friedrich-Wilhelms-Universität Bonn
10/2018 – 10/2020	M.Sc. in Economics, Rheinische Friedrich-Wilhelms-Universität Bonn
08/2013 – 12/2017	B.B.A. in Finance & B.A. in Economics (Minor in Mathematics), Florida International University, Miami, FL, USA

Wissenschaftliche Tätigkeit

seit 06/2020	Wissenschaftlicher Mitarbeiter, Institute of Macroeconomics and Econometrics, Universität Bonn
05/2019 – 06/2020	Research Assistant, briq Institute on Behavior & Inequality, Bonn

Lehre (Auswahl)

seit Wintersemester 2021	*Coding in Julia and Stata* (Ph.D.), Bonn Graduate School of Economics
Sommersemester 2024–2026	*Macroeconomics I* (B.Sc.), Universität Bonn
Wintersemester 2023/24	*Topics in Inequality* (B.Sc.), Universität Bonn

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertation mit dem Titel

Data, Inequality, and Oil

selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die den herangezogenen Werken wörtlich oder sinngemäß entnommenen Stellen sind als solche kenntlich gemacht.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Ort, Datum	Luis Calderon