UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Essays in Macroeconomics: Business Cycles and Earnings Inequality

Permalink
https://escholarship.org/uc/item/99q480ds

Author
Lee, Byoungchan

Publication Date
2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Essays in Macroeconomics: Business Cycles and Earnings Inequality

By

Byoungchan Lee

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Yuriy Gorodnichenko, Chair
Professor Emmanuel Saez
Professor Amir Kermani

Spring 2019

Abstract

Essays in Macroeconomics: Business Cycles and Earnings Inequality

by

Byoungchan Lee

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Yuriy Gorodnichenko, Chair

Does inequality react to stabilization policies and macroeconomic shocks at business cycle frequencies? Does an unanticipated innovation in inequality impact aggregate demand and drive cyclical fluctuations? Does the level of inequality influence the propagation of stabilization policies? To answer these questions, I first investigate which factors of earnings distributions are represented by a measure of inequality in Chapter 1. Chapter 2 develops an econometric tool to evaluate the contribution of macroeconomic shocks to the dynamics of an endogenous variable of interest. Finally, Chapter 3 deals with the questions above.

In Chapter 1, I derive principal components of log earnings distributions in the U.S. and propose a simple three-factor model to rationalize my empirical results. Using data on earnings distribution in the U.S. from 1978 to 2013, I find that more than 90 percent of the total variation in the distribution can be summarized by two underlying factors, which are related to the location and dispersion of the distribution, where the dispersion factor is tightly associated with the log P90/P10 index. Moreover, most of the remaining portion is due to another factor characterizing asymmetric components. To rationalize these findings, I suggest asymmetric Laplace distributions as a model of log earnings distributions. In this model, the right-tail of earnings always follows a Pareto distribution, unlike log normal distributions. Furthermore, it is a tractable distributional family featuring three parameters representing the location, dispersion, and degree of asymmetry, respectively. I describe the dynamics of those parameters in the U.S. both in trends and concerning business cycles. Finally, I illustrate how a conventional Gaussian AR(1) model for individual log earnings can be easily modified to admit asymmetric Laplace distributions.

Chapter 2 is based on joint work with Yuriy Gorodnichenko, which is forthcoming in *Journal of Business and Statistics* under the same title. We propose and study properties of

an estimator of the forecast error variance decomposition in the local projections framework. We find for empirically relevant sample sizes that, after being bias-corrected with bootstrap, our estimator performs well in simulations. We also illustrate the workings of our estimator empirically for monetary policy and productivity shocks.

Chapter 3 deals with questions on the relationship between business cycles and earnings inequality. For an empirical investigation, I construct a novel, high-quality, quarterly measure of earnings inequality and document the following facts. First, an expansionary productivity shock and a contractionary government expenditure shock reduce earnings inequality significantly at the medium-run, while monetary policy shocks have little effects. Second, an unanticipated positive innovation in earnings inequality, which summarizes redistribution from the poor to the rich, lowers aggregate demand substantially in a U-shaped manner. Lastly, the power of stabilization policies increases with the level of inequality. To rationalize these results, I develop a tractable, theoretical framework. I analytically illustrate that inequality in a simple two-agent model is related to demand shocks in a representative agent framework. To match the shape and magnitude of the empirical impulse responses, I further introduce new features including countercyclical earnings risk, an endogenous extensive margin of being credit constrained, and decreasing relative risk aversion preferences.

당연히 너에게

Contents

List of Figures

List of Tables

Acknowledgments

I am especially grateful to my advisor Yuriy Gorodnichenko for his continual support and invaluable guidance. He helped me in all stages of research with encouragement, patience, insights, and knowledge. He definitely made my life as a graduate student a pleasant and enjoyable one, and I feel tremendously indebted to him.

I would like to express my sincere appreciation to Berkeley professors and fellow graduate students for inspiring discussions, simulating comments, and priceless advice. My special gratitude goes out to Emmanuel Saez for important advice on including, but not limited to, the construction of inequality measures. I also thank ChaeWon Baek, Pierre-Olivier Gourinchas, Matthias Hoelzlein, Rupal Kamdar, Hyo Kang, Amir Kermani, Woojin Kim, Dmitri Koustas, Seongju Min, Emi Nakamura, Walker Ray, David Romer, Nick Sander, Benjamin Schoefer, Wendy Shin, Jón Steinsson, and Mauricio Ulate.

I was lucky to have opportunities to listen to people from other institutions which sharpened my thoughts, exposition, and arguments expressed in this dissertation. I am grateful to them for their time, discussions, and interests in my work. My special thanks go to Saroj Bhattarai, Olivier Blanchard, Nicholas Bloom, Nicolas Caramp, Todd Clark, Chris Edmond, Oscar Jordà, Young-hwan Lee, Mikkel Plagborg-Møller, Tommaso Porzio, David Price, Dmitriy Sergeyev, Ludwig Straub, Michael Weber, Christian Wolf, and seminar participants at Berkeley, City University of Hong Kong, Hong Kong University of Science and Technology, University of Hong Kong, and the EGSC at Washington University in St. Louis.

Korea Foundation for Advanced Studies kindly bestowed a fellowship for my Ph.D. study. I deeply appreciate their generous financial support.

Last, but certainly not the least, I would like to extend my deepest appreciation to my family. Their love always supports me and makes me want to be a better person.

Chapter 1

A three-factor Model of Earnings Distribution

1.1 Introduction

Understanding inequality has been one of the main tasks in our society as the concentration of earnings among the top workers has intensified over the past several decades. In a more short-run perspective, the Great Recession has revealed that distributional factors and individual heterogeneity may have substantial effects on macroeconomic fluctuations. While the problems of distribution in the long- and short-run have become important, there are many fundamental but unanswered questions. In this paper, I focus on one specific question about the essential dimension of a cross-section of resources. That is, "How many factors do underlie a distributional phenomenon?" My answer is three for earnings.

Distributions of economic resources such as earnings, income, and wealth are complicated, high-dimensional objects. For a full representation, in principle, we need a vector whose dimension equals the number of agents in an economy. However, that number is enormous in data, and even uncountably infinite in theoretical models assuming continuous distributions. Therefore, carrying the whole distribution as it stands is usually not feasible or meaningful. Finding an informative set of statistics (*i.e.,* a reduction of the dimension) is necessary for a better understanding of distributional issues.

I show that a substantial dimension reduction is possible using data on earnings distributions in the U.S. I find that more than 90 percent of the total variation in the distribution can be summarized by two underlying factors, measuring the location and dispersion. Furthermore, most of the remaining portion is due to another factor characterizing asymmetric components. Thus, the earnings distribution in the U.S. admits a three-dimensional representation.

To rationalize these findings, I propose asymmetric Laplace distributions as a model of log earnings distributions. This distributional family features a simple, functional form parameterized by three factors for the location, dispersion, and degree of asymmetry, respectively. I document that the estimated distribution fits the data remarkably well. Furthermore, I illustrate how a conventional Gaussian AR(1) model for individual log earnings can be adjusted for asymmetric Laplace distributions while maintaining the tractability of linear Gaussian models.

Throughout the empirical analysis in this paper, I use annual labor earnings data in the U.S. from 1978 to 2013. Song et al. (2018) extract every percentile of real earnings in each year from a confidential database constructed by the Social Security Administration. This database contains information on the universe of workers and compensations to their labor in the U.S. Therefore, it can shed light on the overall shape of the earnings distribution and its historical evolution in the past several decades.

A distribution is usually characterized by probability densities or cumulative probabilities. In this paper, I deal with an inverse cumulative distribution function (CDF) y_t such that $y_t(p)$ is the $100p$-th percentile of the log earnings distribution in the year t. Then the above data are values of $y_t(p)$ for p being 0.01 to 0.99. Hilbert space theories (Rudin, 1987) and orthogonal polynomials (Koornwinder et al., 2010) imply that the inverse CDF can be arbitrarily closely approximated by (finite-order) polynomials. This provides a mathematical background for reduced-dimension representations of distributions by proving that a few numbers may suffice to describe a function y_t. Indeed, I find that a 9th order polynomial fits the data almost perfectly. By relying on theoretical properties of orthogonal polynomials, I further document that there exist some asymmetric components in the U.S. log earnings distribution and their significance has been declining through the sample period. In other words, the log earnings distribution has become more symmetric, which is a novel finding up to my knowledge.

While polynomials are intuitive and informative, its coefficients may be correlated to each other. By extracting principal components out of them, the number of factors can be further reduced to three (see Stock and Watson, 2011, 2016, for an overview of the method). The first two leading principal components reflect the location and dispersion of the log earnings distribution up to a rotation, and they explain more than 90 percent of the dynamics of the distribution. The third factor, which captures most of the remaining, loads mainly on both tails and moves them into the same direction. Therefore, it measures the degree of asymmetry or skewness.

Asymmetric Laplace distributions are a simple, tractable distributional family characterized by the same three factors derived empirically (location, dispersion, and skewness). I introduce the definition and basic properties of this distribution later, while more comprehensive analysis is relegated to Kotz, Kozubowski and Podgórski (2001). This distribution features several attractive characteristics. First, it is directly parametrized by the three factors, and each of them has clear meanings. Second, the fit of the estimated log earnings distribution to the data is of impressively high-quality. Furthermore, when the cross-section of the log earnings is modeled by an asymmetric Laplace distribution, the right-tail of earnings follows a Pareto distribution. This is consistent with the empirical evidence in Atkinson, Piketty and Saez (2011) and Piketty and Saez (2003). Finally, making use of its tractability, one can easily embed asymmetric Laplace distributions into a conventional AR(1) model of idiosyncratic earnings dynamics.

Based on the estimated parameters, I find novel results that the log earnings distribution in the U.S. is *left*-skewed. Furthermore, there is a rising trend in the skewness as well as the location and dispersion. Therefore, the distribution has become less left-skewed, or more close to a symmetric one. In this sense, a rising earnings inequality in the U.S. has two components: (1) a secular trend in the dispersion factor which redistributes resources from the bottom to the top and (2) a similar trend in the skewness factor which drives the concentration of earnings in the right-tail while helping the left-tail too.

An increase in the dispersion and skewness of the *log* earnings distribution may contribute to a growth in *average* earnings per worker. Indeed, about 56 percent of the growth in average earnings per worker during the last several decades is due to the rise in the skewness factor. Accumulation in the location factor, which is distribution-neutral in a sense that it parallelly shifts everyone's log earnings, contributes to only 28 percent of the total growth. The dispersion factor explains the remaining 16 percent. As a result, most of the growth has been non-neutral to the allocation of resources across workers, and different income groups had disparate experiences on the economic growth (Goldin and Katz, 2009; Krusell et al., 2000; Piketty, Saez and Zucman, 2018).

Regarding business cycles, I study cyclical properties of the three factors of the earnings distribution. Among them, the skewness factor is the most sensitive to aggregate fluctuations in a procyclical manner. I further investigate the first four cross-sectional moments and find that the mean and skewness coefficients are procyclical while the standard deviation and kurtosis coefficient is countercyclical. This is because the log earnings distribution is left-skewed. As it becomes less left-skewed in expansions, the probability density on extremely large negative deviations decreases, and this lowers the standard deviation and kurtosis coefficient.

The high-dimensionality of distributions has been major obstacles to investigations of relationships between inequality and other components of an economy. In a heterogeneous agent model of Krusell and Smith (1998), a crucial step in the solution method is to summarize the distribution of cash-on-hand to a scalar (mean), which essentially serves a dimension reduction. Ahn et al. (2018) formally discuss dimension reduction problems in heterogeneous agent models, where reducing the dimension of cross-sectional distribution and detecting the most relevant factors are of the utmost importance for numerical tractability.

For an empirical study, the number of distributional factors included in the information set matters. As an example, suppose that one wants to know relationships between monetary policy and earnings inequality, and therefore constructs a small vector autoregression model with real GDP, inflation, federal funds rate, and a measure of earnings inequality, *e.g.,* log P90/P10 index. Here, one may wonder whether including a single inequality measure is sufficient to describe the dynamics of the whole distribution. Later in this paper, I will show that this information set can span most of the variation in the log earnings distribution. To achieve even higher coverage, one may add a variable tightly connected to the skewness factor. The main takeaway is that the factor analysis can reduce the dimension of distributional objects substantially, and this allows for a tractable investigation of distributional issues using more conventional tools developed for a system of aggregate variables. In this context, this paper is complementary to Chang, Chen and Schorfheide (2018), who reduce the dimension of distributions with sieve approximations.

Other applications of asymmetric Laplace distributions in economics include the maximum likelihood estimation of quantile regressions (Koenker and Machado, 1999), matching financial data (Linden, 2001), and the risk management (Taylor, 2019). However, this distributional family has not been widely studied concerning cross-sectional distributions of earnings, income, and wealth. Instead, other parametric families such as log normal, Pareto, and generalized beta are more frequently used in this context (see Cowell and Flachaire, 2015, for a review). Asymmetric Laplace distributions can be a useful model complementary to these alternatives. Another merit of the asymmetric Laplacian family is that it has a natural and tractable extension to a dynamic setup. This model complements existing dynamic models with mixture normal idiosyncratic innovations (Blundell and Preston, 1998; Kaplan, Moll and Violante, 2018; Pistaferri, 2001).

The rest of the paper is structured as follows. Section 1.2 introduces the log earnings data in the U.S. Section 1.3 covers the factor analysis. I first lay out the basics of dimension reduction in the space of inverse cumulative distribution functions based on orthogonal polynomials. Then I derive and study properties of principal components of the log earnings distribution. Section 1.4 presents asymmetric Laplace distributions and applies it to the U.S. data in a static and dynamic setup. Section 1.5 investigates dynamics of the log earnings distribution at business cycle frequencies. Section 1.6 concludes.

1.2 Data

Let $y_t(p)$ be the $100p$-th percentile of the log earnings distribution in the U.S. in the year t. I analyze $\{y_t(p)\}$ reported by Song et al. (2018) for $p = 0.01, \ldots, 0.99$ from 1978 to 2013. Thus, the earnings distribution in the U.S. is represented by a set of 99 percentiles in each year.

Below I briefly describe the data used to derive $y_t(p)$. Song et al. (2018) investigate a confidential database on earnings maintained by the U.S. Social Security Administration (SSA). It contains annual labor earnings records in Form W-2 of all individuals with a Social Security number. The earnings data are uncapped and include wages and salaries, bonuses, exercised stock options, and other fringe benefits.

In the sample of Song et al. (2018), individuals are aged between 20 and 60, and earn more than a minimum threshold. The threshold is set at one-fourth of a full-time Federal minimum wage, *i.e.*, 1/4 of 52 weeks for 40 hours at the minimum wage rate. All earnings are in 2013 real values based on the personal consumption expenditure (PCE) deflator. Other details about the database and sample can be found in Song et al. (2018) and references therein.

1.3 An empirical factor analysis

This section empirically studies the log earnings distribution in the U.S., dented by $\{y_t(p)\}$. To do so, I take a two-step approach. First, I change the basis and rewrite $y_t(p)$ in terms of orthogonal polynomials. This will show that we can substantially reduce the dimension and represent the whole distribution with a few numbers. Furthermore, it will clearly illustrate that there exists an asymmetric component in the log earnings distribution. Second, I derive principal components among the coefficients on orthogonal polynomials and find that more than 90 percent of the total variation in the distribution can be summarized by two underlying factors, which are related to the location and the dispersion of the distribution. Moreover, another factor characterizing the degree of asymmetry accounts most of the remaining variation.

1.3.1 A short introduction to orthogonal polynomials

Here I provide a short introduction to orthogonal polynomials. The discussions are minimal, where more details are relegated to Judd (1998).

For this subsection, let y_t be a function from $[0,1]$ to $\mathbb{R}$ mapping $p \in [0,1]$ to $100p$-th percentile of log earnings distribution in the U.S. in the year t. In other words, y_t is an inverse cumulative distribution function of log earnings. The goal of this subsection is to show that an infinite dimensional object y_t can be arbitrarily closely approximated by polynomials, which are characterized by only a finite number of coefficients.

An inner product between two square-integrable functions f and g is denoted by $\langle f, g\rangle = \int_0^1 fg$, where the integration is with respect to the Lebesgue measure. I define orthogonal polynomials as follows.[1]

Definition 1.1 (Orthogonal polynomial). *A function $P_n : [0,1] \to \mathbb{R}$ is the orthogonal polynomial of order n if*

(i) P_n *is a polynomial of order* n*,*

(ii) $\langle P_n, P_m\rangle = 1$ *when* $n \neq m$*,*

(iii) $||P_n|| = \sqrt{\langle P_n, P_n\rangle} = 1$,

[1]Given that the domain is a finite interval and that the weights are uniform in p, the orthogonal polynomials in this paper is a variant of standard Legendre polynomials (see Judd, 1998).

for all $n, m = 0, 1, \dots$.

Note that $\{P_n : n = 0, 1, \dots\}$ forms an orthogonal set due to (ii) and (iii) implies that P_n has a unit norm.

It is clear that $P_0 = 1$. For any $n > 0$, we can recursively find P_n by projecting p^n on $\{P_m : m < n\}$ and scaling the residual to make it a unit vector. For example, P_1 is proportional to $p - \langle p, P_0 \rangle P_0 = p - 0.5$. It is easy to check that $P_1(p) = \sqrt{12}(p - 0.5)$.

Figure 1.1 depicts the orthogonal polynomials of order 0 to 4. For an odd number n, $P_n(p)$ is symmetric to $(0.5, 0)$, while $P_n(p)$ is symmetric around $p = 0.5$ for an even number n. Now I state the main result in this subsection.

Theorem 1.1. $\{P_n : n = 0, 1, \dots\}$ *is an orthonormal basis of* $\mathbb{L}^2[0,1]$*, which is a set of square integrable functions from* $[0,1]$ *to* $\mathbb{R}$*. Furthermore, the following holds for any* $f \in \mathbb{L}^2[0,1]$*:*

$$f = \langle f, P_0 \rangle P_0 + \langle f, P_1 \rangle P_1 + \cdots = \sum_{n=0}^{\infty} \langle f, P_n \rangle P_n. \tag{1.1}$$

Finally, $||f||^2 = \sum_{n=0}^{\infty} |\langle f, P_n \rangle|^2$.

Proof. The first statement holds because of Weierstrass approximation theorem and Theorem 3.14 in Rudin (1987). For the other statements, see Rudin (1987, Ch. 4). □

Theorem 1.1 illustrates that any square integrable function f, which is an infinite dimensional object, can be represented by countably many coefficients $\{\langle f, P_n \rangle\}$. Furthermore, $\langle f, P_n \rangle$ is small for large n's because $||f||^2 = \sum_{n=0}^{\infty} |\langle f, P_n \rangle|^2 < \infty$. This implies that f can be closely approximated by an N-th order polynomial $\sum_{n=0}^{N} \langle f, P_n \rangle P_n$ for some N. In other words, we can characterize f with only a finite number of coefficients (with small errors).

For example, suppose that a random variable Z follows a Pareto distribution with a cumulative distribution function (CDF) $F_Z(z) = 1 - \left(\frac{z}{\underline{z}}\right)^{-\zeta}$ for $z \geq \underline{z} > 0$ and $\zeta > 0$. By writing $p = F_Z(z)$, an inverse CDF of $\log(Z)$ is given by $\log(z) = \log(\underline{z}) - \frac{1}{\zeta}\log(1 - p) = \log(\underline{z}) + \frac{1}{\zeta}\left(p + \frac{1}{2}p^2 + \frac{1}{3}p^3 + \dots\right)$, where the last expression is from the Taylor series of $\log(1-p)$. Therefore, a Pareto distribution, in a logarithm, can be easily approximated by a polynomial.

For another example, consider a random variable W whose probability density function (PDF) $f_W(w)$ is symmetric around $\bar{w}$. Then its CDF $F_W(w)$ becomes symmetric to $(\bar{w}, 0.5)$, and therefore its inverse CDF is symmetric to $(0.5, \bar{w})$. Note that P_n for an odd number n is also symmetric to $(0.5, 0)$ as shown in Figure 1.1.[2] Thus, the inverse CDF of W is spanned by $\{P_0, P_1, P_3, P_5, \dots\}$, where P_0 is for an upward parallel shift by $\bar{w}$. In other words, $\{P_0, P_1, P_3, P_5, \dots\}$ captures symmetric components in probability density functions, where $\{P_2, P_4, P_6, \dots\}$ reflects asymmetric parts.

1.3.2 First-stage factors and loadings

The main takeaway from the discussion above is straightforward: although earnings distributions are complicated high-dimensional objects, substantial dimension reduction is possible with the aid of orthogonal polynomials under a mild condition. I now apply this method to the U.S. earnings distribution.

From now on, y_t is a 99-dimensional column vector $(y_t(0.01), \dots, y_t(0.99))'$ representing the log earnings distribution in the U.S. in the year t. Similarly, $P_n = (P_n(0.01), \dots, P_n(0.99))'$ for all n. I consider the following model for y_t:

$$\begin{aligned} y_t &= \alpha_{t,0}P_0 + \alpha_{t,1}P_1 + \cdots + \alpha_{t,N}P_N + u_{t,N} \\ &= P^{(N)}\alpha_t^{(N)} + u_{t,N}, \end{aligned} \tag{1.2}$$

where $P^{(N)} = (P_0, \dots, P_N)$ and $\alpha_t^{(N)} = (\alpha_{t,0}, \dots, \alpha_{t,N})'$. The model (1.2) can be understood as a factor model with time-varying factors $\alpha_{t,n}$'s and corresponding loadings P_n's. I estimate $\alpha_{t,n}$ using an ordinary least squares (OLS) regression for each t, while the results are similar when $y_t'P_n/99$ is used instead in light of Equation (1.1). Note that $y_t'P_n/99$ does not depend on N. Similarly, the OLS estimates of the factors are not sensitive to N because P_n's are mutually orthogonal in the population.

Figure 1.2 shows y_t and its fitted values for $t = 1978$. Blue triangles depict $y_{1978}(p)$ for $p = 0.01, \dots, 0.99$, where these earnings are in 2013 real values. Black lines represent the fitted values based on Equation (1.2), where the order of polynomial N is 0, 1, 3, and 9, respectively. Note that even relatively low-order polynomials, *e.g.,* a cubic polynomial, provide a reasonable approximation. When $N = 9$, the fit becomes almost perfect in the tails as well as in the middle.

[2]Formally, P_n is symmetric to $(0.5, 0)$ when n is odd, while it is symmetric around $p = 0.5$ for an even n. One can prove it by induction using a recursion formula (6.3.1) in Judd (1998). In doing so, one should bear in mind that P_n there is defined on $[-1, 1]$ and not normalized to have a unit norm.

The results for other years are similar. The coefficient of determination R^2 of Equation (1.2) is greater than 95(99) percent in all years for $N = 2(3)$ in Table 1.1. Furthermore, the second-order orthogonal polynomial P_2 contributes to the fit by about 1.5 percent in 1980 and 0.5 percent in 2010. This implies that the log earnings distributions in the U.S. have some asymmetric components, and the degrees of asymmetry have been declining. I will further discuss the history of the degree and direction of asymmetry in Section 1.4.2. Finally, the fit is almost perfect for $N \geq 9$. Therefore, I fix N at 9 below.

The estimated first-stage factors $\alpha_{t,n}$ from 1978 to 2013 for n being 0 to 5 are illustrated in Figure 1.3. The zeroth order factor shifts every percentile parallelly, and therefore represents the location of the log earnings distribution in the U.S. A fast rise in $\alpha_{t,0}$ during the late 1990s and the early 2000s is consistent with the findings of Byrne, Fernald and Reinsdorf (2016) and Fernald (2015) that labor productivity growth was exceptionally high at that time when compared to the preceding and following periods. A secular increase in earnings inequality is reflected in upward trends in the first- and third-order factors, where the third-order factor focuses more on the tails than the first-order factor as is shown in Figure 1.1. While the two factors exhibit broadly similar patterns before the Great Recession, $\alpha_{t,3}$ decreased substantially after the Great Recession and did not bounce back to its pre-recession level, whereas $\alpha_{t,1}$ is about 0.88 in both 2008 and 2013. Finally, the magnitude of the second-order factor implies that there exist some asymmetric components in the log earnings distribution. However, the absolute value of the factor, and therefore the degree of asymmetry, has become smaller.

In sum, a small number of factors suffice to describe the log earnings distribution in the U.S. via the first-stage factor model (1.2). Furthermore, there exist some asymmetric components, whose contribution to the shape of the distribution has been decreasing. However, the correlation between $\alpha_{t,n}$ across n is non-zero, and therefore we can further reduce the number of factors by exploiting the comovements of those first-stage factors.

1.3.3 Deep factors and loadings

This subsection introduces a second-stage factor model for $\alpha_{t,n}$'s, where I employ a principal component analysis (PCA) for estimation. I investigate the loadings of these principal components on the log earnings distribution in the U.S. and conclude that three factors are enough to describe the variation of the whole distribution. Furthermore, I document that these factors are tightly related to the location, dispersion, and asymmetry (skewness) of the distribution up to factor rotations.

The second-stage factor model describes the first-stage factors $\alpha_t^{(N)}$ in terms of "deep" factors $f_{t,m}$:

$$\begin{aligned}\alpha_t^{(N)} &= a + bt + \lambda_1 f_{t,1} + \cdots + \lambda_M f_{t,M} + v_{t,M} \\ &= a + bt + \Lambda^{(M)} f_t^{(M)} + v_{t,M},\end{aligned} \tag{1.3}$$

where the second-stage loading λ_m is a N+1-dimensional vector for all m, $\Lambda^{(M)} = (\lambda_1, \ldots, \lambda_M)$, and $f_t^{(M)} = (f_{t,1}, \ldots, f_{t,M})'$. a and b are $N+1$-dimensional vectors capturing for trends in $\alpha_{t,n}$'s. For example, when only the first element of b is non-zero, a linearly detrended zeroth order factor $\alpha_{t,0}$ is examined with the demeaned $\alpha_{t,n}$'s for $n \neq 0$ in the principal component analysis. This corresponds to assuming a common linear trend to $y_t(p)$ for all p, where the trend is based on the pure location factor $\alpha_{t,0}$ affecting all the percentiles parallelly. On the other hand, detrending $\alpha_{t,n}$ for other n's, and thus having b whose elements other than the first one are non-zero, imposes different slopes to each percentile, which are given by $P^{(N)}b$. Finally, the elements of $f_t^{(M)}$ are normalized to have unit variances, where superscripts (N) and (M) are not specified hereafter.

By combining Equations (1.2) and (1.3), we have a representation of the log earnings distribution y_t based on the deep factors $(f_{t,m})$ and loadings (Ψ_m):

$$y_t = c_t + \Psi f_t + \eta_t, \tag{1.4}$$

where $c_t = P(a + bt)$, $\Psi = (\Psi_1, \ldots, \Psi_M) = P\Lambda$, and $\eta_t = Pv_t + u_t$. Also,

$$f_t = \Phi(L) f_{t-1} + \epsilon_t, \qquad \epsilon_t \sim iid(0, \Omega), \tag{1.5}$$

where L is the lag operator, $\Phi(L)$ is a lag polynomial matrix, and Ω is a diagonal matrix.

The deep factors $(f_{t,m})$ and second-stage loadings (λ_m) are estimated from Model (1.3) by deriving principal components of $\alpha_{t,n}$'s under specific assumptions on trends. Then the deep factor loadings (Ψ_m) follow directly from the fact that $\Psi_m = P\lambda_m$ for all m.

I first determine the number of meaningful factors in the log earnings distribution in the U.S. For the purpose, Table 1.2 shows eigenvalues corresponding to the leading principal components, normalized by the sum of all eigenvalues under various assumptions on trends.

This scree "table" illustrates the contribution of each factor $f_{t,m}$ to the variation in α_t in Equation (1.3). For example, when the first-stage factor $\alpha_{t,n}$ is directly investigated without detrending, the deep factors $f_{t,1}$, $f_{t,2}$, and $f_{t,3}$ explain 91, 6, and 2 percent of the total variance, respectively, where all the other principal components' contribution is less than 1 percent in total. For the second case, I use the linearly detrended zeroth order factor $\alpha_{t,0}$, and therefore b is proportional to a vector whose first element is one, and the other elements are zero (e_1). The third case is conditional on that $f_{1,t}$ equals the detrended $\alpha_{t,0}$ and therefore $\lambda_1 = e_1$. Because the principal components are derived from $(\alpha_{t,1}, \ldots, \alpha_{t,N})'$ without $\alpha_{t,0}$, the corresponding row in Table 1.2 begins with $n = 2$. The remaining two cases are for the detrended $\alpha_{t,m}$ for all m and the differenced first-stage factors $\Delta\alpha_{t,m}$, respectively.

Note that the first two factors $f_{t,1}$ and $f_{t,2}$ capture most of the variation in α_t and therefore the log earnings distribution y_t. For example, the contribution of $f_{t,1}$ and $f_{t,2}$ is greater than 90 percent in the second case. Furthermore, it seems to be sufficient to include just one more factor to take account of most of the remaining variances. Deriving principal components of y_t directly from Equation (1.4) yields almost identical results (*e.g.*, the results for Case 6 based on Δy_t is similar to that of Case 5).

Next, I turn to what these leading factors represent. I examine their factor loadings $\Psi_m = P\lambda_m$ for Case 2 and 3 in Table 1.2, while the results are similar for the other cases. The top three panels in Figure 1.4 are based on Case 2 (the linearly detrended zeroth order factor $\alpha_{t,0}$ and the demeaned $\alpha_{t,n}$'s for $n \neq 0$). The solid lines depict $P\lambda_m$, which is the effect of a unit increase in $f_{t,m}$ on $y_t(p)$ for each $p = 0.01, \ldots, 0.99$. Here λ_m is identified as the m-th eigenvector in the PCA times $\sqrt{Var(f_{t,m})}$, because $f_{t,m}$'s are normalized to have unit variance. In Figure 1.4, the first two deep factors seem to span the location and dispersion effects in combination, while the other one is related to the skewness.

However, f_t in Model (1.4) is identified only up to rotation, because $(\Psi R')(Rf_t) = \Psi f_t$ for any orthogonal matrix R. The dotted lines show an example of those loadings $\Psi R'$ on a rotated factors Rf_t. The dotted lines in the first two panels are obtained by rotating $(f_{t,1}, f_{t,2})'$ and the third one is based on a rotation of $(f_{t,3}, f_{t,4})'$.[3] It is more clear with these

[3] Specifically, I find the rotation matrices in the following way. Let $R_2(\zeta)$ be a two-dimensional rotation matrix with an angle ζ, *i.e.,* $R_2(\zeta) = \begin{pmatrix} \cos(\zeta) & -\sin(\zeta) \\ \sin(\zeta) & \cos(\zeta) \end{pmatrix}$. For $(f_{t,1}, f_{t,2})'$, the rotated loadings become $(\Psi_1, \Psi_2) R_2(\zeta_1)'$. I obtain ζ_1 by minimizing the dispersion in the effects of this rotated factor on different percentiles, *i.e.,* the variance of $\cos(\zeta_1)\Psi_1 - \sin(\zeta_1)\Psi_2$. I calibrate ζ_2 for $R_2(\zeta_2)(f_{t,3}, f_{t,4})'$ by making $\text{argmin}_p \cos(\zeta_2)\Psi_3 - \sin(\zeta_2)\Psi_4$ be around $p = 0.69$. This specific choice is related to the discussion in Section 1.4.2.

rotations that the first two factors jointly span the location and dispersion effects, while the remaining variation is mostly about skewness.

The bottom three panels illustrate the results when I consider principal components of $(\alpha_{t,1}, \ldots, \alpha_{t,M})'$ conditioning on $f_{t,1}$ being equal to the detrended $\alpha_{t,0}$. When the first deep factor is pinned down to a pure location effect that shifts all the log percentiles equally, the leading component in the remaining variation becomes the dispersion, followed by the skewness. Furthermore, the deep factor loadings, with rotations, are robust to specification details on trends. This is evident when the top and bottom panels in Figure 1.4 are compared.

The results above have a powerful implication on how empirical studies with distributional considerations can be conducted. Because the location factor is highly correlated with aggregate variables like real GDP per capita in a standard information set, adding a *single* variable capturing the dispersion or inequality to the information set is enough to span most of the variation in the earnings distribution. For example, one may include a log P90/P10 index of earnings $y_t(0.9) - y_t(0.1)$ to the information set to span the second deep factor and take account of more than 90 percent of the variation in the distribution.[4] When more precision is required, one may consider adding a measure of skewness. Most importantly, carrying over the entire distribution is unnecessary.

This subsection investigated the second-stage factor model and the implied representation of y_t in terms of the deep factors f_t and loadings Ψ. It was shown that a substantial dimension reduction relative to the first-stage model is possible. The two leading principal components can explain more than 90 percent of the dynamics in the whole log earnings distribution in the U.S., where they reflect the location and dispersion components. Furthermore, the remaining part is tightly related to a force affecting the skewness by impacting both tails in the same direction. Therefore, it would be useful to have a distributional family featuring these three factors for a systematic understanding of distributional issues. It would be even better if the model is tractable like a log normal distribution, and a Pareto right-tail of earnings is incorporated in the model.

[4] Indeed, the log P90/P10 index is highly correlated with the second deep factor $f_{t,2}$. To see this, let ι be a 99-dimensional vector such that $\iota' y_t$ is the log P90/P10 index. From Equation (1.4), I obtain $Var(\iota' y_t - \iota' c_t) = \sum_{m=1}^{M} \beta_m^2 + Var(\iota' \eta_t)$, where $\beta = (\beta_1, \ldots, \beta_M)' = \Psi' \iota$. Thus, the contribution of the m-th deep factor on the log P90/P10 index can be measured by $s_m = \beta_m^2 / Var(\iota' y_t - \iota' c_t)$. In Case 3, $\iota' c_t$ is a constant, $s_1 = 0$ by construction, and $s_2 = 0.98$. Thus, the log P90/P10 index reflects the second deep factor quite precisely.

1.4 A simple model of earnings distribution

I propose asymmetric Laplace distributions as a model of the log earnings distributions in the U.S. in this section. An asymmetric Laplace distribution has a simple probability density function (PDF) characterized by three parameters. A notable property of this distributional family is that those parameters correspond to the three deep factors discussed in the previous section. Furthermore, I show that the fit of the estimated distribution from the asymmetric Laplacian family to the data (y_t) is impressive. Finally, I illustrate how a conventional dynamic model for individual log earnings based on a Gaussian AR(1) process can be adjusted to yield an asymmetric Laplace distribution.

1.4.1 Elementary properties of asymmetric Laplace distributions

I now define an asymmetric Laplace distribution and study basic properties of it. The materials here are mostly borrowed from Chapter 3 and 6 in Kotz, Kozubowski and Podgórski (2001). Most of the proofs are omitted below, and an interested reader may refer to Kotz, Kozubowski and Podgórski (2001).

Definition 1.2 (Asymmetric Laplace distribution)**.** *A random variable X has a univariate asymmetric Laplace distribution if there exist parameters $\theta \in \mathbb{R}$, $\kappa > 0$, and $\sigma > 0$ such that the probability density function of X has the following form*

$$g(x) = \frac{\sqrt{2}}{\sigma}\frac{\kappa}{1+\kappa^2}\begin{cases}\exp\left(-\frac{\sqrt{2}}{\sigma\kappa}|x-\theta|\right), & \text{if } x \geq \theta \\ \exp\left(-\frac{\sqrt{2}\kappa}{\sigma}|x-\theta|\right), & \text{if } x < \theta.\end{cases} \tag{1.6}$$

The distribution of X is denoted by $\mathcal{AL}(\theta, \sigma^2, \kappa)$.[5]

Among the three parameters, θ shifts the location of the distribution without affecting the shape. The second parameter σ captures the dispersion. A larger σ implies a more dispersed distribution because the inputs to the exponential function decrease more slowly as $|x-\theta|$ increases for a larger σ. Finally, κ reflects the degree of asymmetry, or skewness. For example, when $\kappa = 1$, I have a symmetric Laplace distribution with mean θ and variance σ^2. When κ is less than 1, the distribution becomes skewed to the left. Note that if $\kappa < 1$,

[5] κ in Equation (1.6) is $\frac{1}{\kappa}$ in the notation of Kotz, Kozubowski and Podgórski (2001). I change the definition of this parameter to make an increase in κ corresponds to an increase in a skewness coefficient.

$-\frac{\sqrt{2}\kappa}{\sigma} > -\frac{\sqrt{2}}{\sigma\kappa}$. Therefore, for the same deviation of x from θ, $g(x)$ is larger in the left-"wing" ($x < \theta$) than in the right-wing ($x > \theta$), which implies a negative skewness.

Figure 1.5 depicts PDFs of three different asymmetric Laplace distributions and a standard normal distribution. A dotted line is for a standard normal distribution, where a dash-dot line represents a symmetric Laplace distribution with the same mean (0) and variance (1). The Laplace distribution has more probability masses around zero, which is the value of its location parameter, and in both tails than the normal distribution. A left-skewed density function, denoted by a solid line, is based on $\theta = -2$, $\sigma = 1$, and $\kappa = 0.7 < 1$. Finally, the last one is illustrated by a dashed line, where its $\theta = 2$, $\sigma = 2$, and $\kappa = 1$. Because its σ is greater than that of the other distributions, its PDF is more spread than the others.

Since the PDF $g(\cdot)$ in (1.6) is simple and tractable, so is its cumulative distribution function (CDF), denoted by $G(\cdot)$.

Proposition 1.1. *Let $X \sim \mathcal{AL}(\theta, \sigma^2, \kappa)$. Then*

$$G(x) = Pr(X \leq x) = \begin{cases} 1 - \frac{\kappa^2}{1+\kappa^2} \exp\left(-\frac{\sqrt{2}}{\sigma\kappa}|x-\theta|\right) & \text{if } x \geq \theta, \\ \frac{1}{1+\kappa^2} \exp\left(-\frac{\sqrt{2}\kappa}{\sigma}|x-\theta|\right) & \text{if } x < \theta. \end{cases} \tag{1.7}$$

Specifically, the probability mass in the left-wing is given by

$$G(\theta) = Pr(X \leq \theta) = \frac{1}{1+\kappa^2}. \tag{1.8}$$

And the inverse CDF also admits an analytic expression.

$$G^{-1}(p) = \begin{cases} -\frac{\sigma\kappa}{\sqrt{2}} \log\left(\frac{1-p}{1-G(\theta)}\right) + \theta & \text{if } G(\theta) \leq p \leq 1, \\ \frac{\sigma}{\sqrt{2}\kappa} \log\left(\frac{p}{G(\theta)}\right) + \theta & \text{if } 0 \leq p \leq G(\theta). \end{cases} \tag{1.9}$$

Another important property of asymmetric Laplace distributions is its relationship to a Pareto distribution.

Proposition 1.2. *Let $X \sim \mathcal{AL}(\theta, \sigma^2, \kappa)$. Conditional on X being on the right-wing,* i.e., *$X \geq \theta$, $U = \exp(X)$ has a Pareto distribution.*

Proof. For $x \geq \theta$, $Pr(X \leq x | X \geq \theta) = \frac{Pr(\theta \leq X \leq x)}{Pr(X \geq \theta} = \frac{G(x) - G(\theta)}{1 - G(\theta)} = 1 - exp(-\xi(x-\theta))$, where $\xi = \frac{\sqrt{2}}{\sigma\kappa}$. Therefore, for $u \geq \exp(\theta)$, $Pr(U \leq u | X \geq \theta) = Pr(X \leq \log(u) | X \geq \theta) = 1 - \left(\frac{exp(\theta)}{u}\right)^{\xi}$, which is a CDF of a Pareto distribution. □

Proposition 1.2 implies that when the cross-section of log earnings has an asymmetric Laplace distribution, earnings at the top have a Pareto tail. This is consistent with empirical evidence in the U.S. and other countries (Atkinson, Piketty and Saez, 2011; Piketty and Saez, 2003).

Next, I turn to moments and the characteristic function. While $(\theta, \sigma^2, \kappa)$ parametrization is intuitive, these results can be stated more succinctly with a different parametrization.

Proposition 1.3. *Let* $X \sim \mathcal{AL}(\theta, \sigma^2, \kappa)$ *and* $\mu = \frac{\sigma}{\sqrt{2}}\left(\kappa - \frac{1}{\kappa}\right)$. *Then*

(i) $E(X) = \theta + \mu$.

(ii) $Var(X) = \sigma^2 + \mu^2$.

(iii) (Coefficient of skewness) $\frac{E(X-E(X))^3}{Var(X)^{3/2}} = 2\frac{\kappa^3 - 1/\kappa^3}{(\kappa^2+1/\kappa^2)^{3/2}}$. *This increases monotonically from* -2 *to* 2 *in* κ.

(iv) (Coefficient of kurtosis) $\frac{E(X-E(X))^4}{Var(X)^2} - 3 = 6 - \frac{12}{(\kappa^2+1/\kappa^2)^2}$. *It varies from 3 (when* $\kappa = 1$*) to 6 (when* κ *is either 0 or* ∞*).*

(v) (Characteristic function) $\psi(\tau) = E(\exp(i\tau X)) = \frac{\exp(i\theta\tau)}{1+\frac{1}{2}\sigma^2\tau^2 - i\mu\tau}$ *for* $\tau \in \mathbb{R}$.

For symmetric Laplace distributions, $\mu = 0$ because $\kappa = 1$. Then the mean and variance are θ and σ^2. When the distribution is skewed, both moments are adjusted accordingly via μ. Proposition 1.3 (iii) demonstrates how κ reflects the skewness of the distribution.

Another useful characteristic of asymmetric Laplace distributions is that it has a natural extension to multivariate cases based on the characteristic function.

Definition 1.3 (Multivariate asymmetric Laplace distribution)**.** *A d-dimensional random vector X has a multivariate asymmetric Laplace distribution (MAL) with parameters* $\theta \in \mathbb{R}^d$, *a positive semi-definite matrix* $\Sigma \in \mathbb{R}^{d\times d}$, *and* $\mu \in \mathbb{R}^d$ *if its characteristic function has the form*

$$\psi(t) = E(\exp(i\tau' X)) = \frac{\exp(i\theta'\tau)}{1 + \frac{1}{2}\tau'\Sigma\tau - i\mu'\tau} \qquad \text{for } \tau \in \mathbb{R}^d. \tag{1.10}$$

I denote the distribution by $\mathcal{MAL}(\theta, \Sigma, \mu)$.

Note that univariate cases in Definition 1.2 are embedded in Definition 1.3. Similar to Proposition 1.3, it can be shown that $E(X) = \theta + \mu$ and $Var(X) = \Sigma + \mu\mu'$. Furthermore, an affine transformation of X also has a MAL distribution, similar to the multivariate normal family being closed under affine transformations. This property substantially enhances tractability of linear dynamic models in Section 1.4.3, and allows for simple representations of transitions between asymmetric Laplace distributions of log earnings in the U.S. in different years.

Proposition 1.4. *Let $X \sim \mathcal{MAL}(\theta, \Sigma, \mu)$. For a matrix Γ and a vector δ with compatible sizes,*

$$\Gamma X + \delta \sim \mathcal{MAL}\left(\Gamma\theta + \delta, \Gamma\Sigma\Gamma', \Gamma\mu\right). \tag{1.11}$$

Proof. $\psi_{\Gamma X+\delta}(t) = E\left[\exp\left\{it'(\Gamma X + \delta)\right\}\right] = \exp\left(it'\delta\right) E\left[\exp\left\{i(\Gamma' t)' X\right\}\right] = \exp\left(it'\delta\right) \psi_X(\Gamma' t)$. Now replace $\psi_X(\Gamma' t)$ with Equation (1.10) and compare the result with the characteristic function of $\mathcal{MAL}\left(\Gamma\theta + \delta, \Gamma\Sigma\Gamma', \Gamma\mu\right)$. □

1.4.2 Earnings distribution in the U.S.

The main finding in Section 1.3 was that there exist three factors in the log earnings distribution in the U.S. And I introduced a specific parametric family above whose three parameters characterize the location, dispersion, and skewness, respectively. The purpose of this subsection is to show that an asymmetric Laplace distribution does fit the data remarkably well. Furthermore, I document a secular trend in the skewness as well as the dispersion of the cross-section of the log earnings, which is a novel finding up to my knowledge. I also decompose growth rates of average earnings into parts due to each parameter.

I find asymmetric Laplace distributions $\mathcal{AL}(\theta_t, \sigma_t^2, \kappa_t)$ that best describe the U.S. earnings data in years 1978 to 2013. Because Song et al. (2018) set a minimum threshold (ν_t) and drop observations below it, I use a logarithm of translated earnings $\tilde{y}_t(p) = \log\left\{\exp(y_t(p)) - \nu_t\right\}$ for all t and p.[6] I estimate parameters for each t by solving the following non-linear least squares problem:

$$\min \sum_p \left\{\tilde{y}_t(p) - G^{-1}(p; \theta_t, \sigma_t, \kappa_t)\right\}^2, \tag{1.12}$$

where the inverse CDF G^{-1} is from Equation (1.9).

[6] Setting ν_t at 0 deteriorates the fit only at the bottom (*e.g.,* $p \leq 0.04$) without changing the other results substantially.

To clearly illustrate the fit of asymmetric Laplace distributions to the log earnings, I focus on the real earnings data in 1978 in the U.S. In the top panel of Figure 1.6, each diamond represent $\left(\log\left(\frac{p}{G(\theta_1)}\right) \text{ or } \log\left(\frac{1-G(\theta_1)}{1-p}\right), \tilde{y}_t(p)\right)$ for $p = 0.01, \ldots, 0.99$. I use $\log\left(\frac{p}{G(\theta_1)}\right)$ if $p < G(\theta_1)$, and therefore the log translated earnings $\tilde{y}_t(p)$ lies in the left-wing. Otherwise, $\tilde{y}_t(p)$ is plotted against $\log\left(\frac{1-G(\theta_1)}{1-p}\right)$ for the right-wing. A solid line is based on the fitted distribution with a similar transformation. I obtain a piecewise linear graph in this manner that characterizes an asymmetric Laplace distribution (see Equation (1.9)). It is impressive to note the resemblance between the data and the model-induced distribution.

The bottom panel covers empirical and model-based CDFs of real earnings in 1978. Diamonds depict several selected percentiles of earnings $\exp(y_t(p))$ in the data. A solid line is for the CDF implied by the estimated asymmetric Laplace distribution of the log earnings, whereas a dotted line presents the results for a normal distribution estimated similarly to Equation (1.12). While the log normal distribution provides a reasonable approximation to the data, it overestimates the CDF in the range between P20 and P75 and underestimates in both tails. The log asymmetric Laplace distribution provides a better fit.

A major implication of the results above is that it suffices to keep track of three variables θ_t, σ_t, and κ_t when handling the entire U.S. log earnings distribution. Each panel in Figure 1.7 illustrates these three parameters of the estimated asymmetric Laplace distributions (Equation (1.12)) and the three factors that are inferred from the PCA (Equation (1.3)). For the PCA, I calculate the principal components conditional on $f_{t,1} = \alpha_{t,0}$ as is the case in the second row in Figure 1.4. $f_{t,3}$ is subject to a rotation explained in footnote 3. The ticks on the left (right) vertical axis are for θ_t, σ_t, and κ_t ($f_{t,1}$, $f_{t,2}$, and $f_{t,3}$), which are denoted by solid (dotted) lines.

A component of economic growth that is "neutral" to the dispersion or inequality is captured by the location parameter θ_t.[7] The dispersion parameter σ_t has a rising trend, while it is rather muted during the late 1990s. This reflects a secular trend in earnings inequality, which is documented by Autor, Katz and Kearney (2008), Piketty and Saez (2003), and many others. While σ_t does not react sensitively to the Great Recession, the asymmetry (skewness) parameter κ_t does decline after 2007. Until 2012, it does not return to its pre-recession level, meaning that the Great Recession hurt the top and the bottom more than the others, conditional on being employed. Furthermore, κ_t is less than one throughout

[7] A proportional change in everyone's earnings does not affect measures of inequality such as Gini coefficient, top 10% share, P90/P10 index, coefficient of variation, etc.

the sample periods. This implies that the U.S. log earnings distribution has more probability mass in the *left*-wing and is *left*-skewed. However, the degree of asymmetry has been declining as κ_t increases to one. The log earnings distribution in the U.S. is becoming more symmetric, consistent with the finding in Section 1.3.2. The share of workers in the left-wing declined from 71 percent in 1980 to 67 percent in 2013 when calculated using Equation (1.8), and the average throughout the sample periods is about 69 percent.

Therefore, there is an important trend in the skewness as well as the dispersion of the cross-section of log earnings. However, this trend does not appear in $f_{t,3}$, which is based on the PCA in Section 1.3.3. This is because the PCA allocate the correlated components in both trends to $f_{t,2}$ when orthogonalizing the factors. Therefore, $f_{t,3}$ looks almost like a detrended κ_t up to a scale. Apart from that, $(\theta_t, \sigma_t, \kappa_t)$ shares a similar dynamic to $(f_{t,1}, f_{t,2}, f_{t,3})$. Thus, the parameters of asymmetric Laplace distributions in this section match the empirical factors I find in Section 1.3.3 quite successfully.

The location parameter θ_t has a homogeneous effect on everyone in the economy. However, such a distribution-neutral component has not been a dominant driver of the aggregate earnings growth in the U.S. For example, the slope of a fitted linear trend of θ_t is only 43 basis points per year, which is too low relative to the growth rate of real GDP per capita. Importantly, the other factors can also impact on the growth of average earnings. It is clear that given the same mode (θ), a less left-skewed distribution with a larger κ has greater average earnings. Furthermore, increasing the dispersion (σ) in *log* earnings can raise average earnings.

I provide a decomposition of the growth rate of average earnings into components based on each parameter. Suppose Y_t represent a cross-section of earnings in the year t: $Y_t = \exp(y_t) = \exp(\tilde{y}_t) + \nu_t$. Similar to Proposition 1.3 (v), one can show that $E(Y_t) = \frac{\exp(\theta_t)}{1-\frac{1}{2}\sigma_t^2-\mu_t} + \nu_t$, where $\mu_t = \frac{\sigma_t}{\sqrt{2}}\left(\kappa_t - \frac{1}{\kappa_t}\right)$.[8] This average earnings per worker grew at a 0.93 percent annual rate on average from 1981 to 2013 (Table 1.3). $E(Y_t|\sigma_1, \kappa_1)$ calculates this quantity under the assumption that σ and κ do not change from their 1981 values. Therefore, by comparing $E(Y_t|\sigma_1, \kappa_1)$ in 1981 and 2013, we can detect the contribution of θ_t on the growth of average earnings per worker. The growth rate of this quantity is only 26 basis points, taking up only 28 percent of the total average growth rate of $E(Y_t)$. $E(Y_t|\kappa_1)$ is similar, but it fixes only κ_t enabling us to assess the effects of σ_t. A rise in σ_t explains an additional 16 basis points in

[8] $E(Y_t^\tau)$ exists when $-\frac{\sqrt{2}\kappa_t}{\sigma_t} < \tau < \frac{\sqrt{2}}{\sigma_t \kappa_t}$, which is satisfied for all t in the sample when $\tau = 1$.

the growth rate (0.42% - 0.26%), and θ_t and σ_t together explain 44 percent of the growth rate of $E(Y_t)$. Therefore, more than half of the growth in average earnings is based on a rise in the skewness of the log earnings distribution (κ_t). In other words, a thickening right-tail of earnings distribution was a major driver of a rise in average earnings. When calculated similarly, average earnings of the bottom 50% grew at a 0.28 percent annual rate, which is much lower than 0.93 percent.[9] In short, different income groups in the U.S. had starkly diverging experiences on economic growth (Piketty, Saez and Zucman, 2018).

On the other hand, the growth rate of real GDP per capita is 1.71 percent, which is almost a double of that of $E(Y_t)$, 0.93 percent. This huge difference is originated mostly from two sources: a faster rise in the number of workers in the SSA's data than the population and a declining labor share in a total income (Elsby, Hobijn and Şahin, 2013; Karabarbounis and Neiman, 2013; Koh, Santaeulalia-Llopis and Zheng, 2018). The number of workers in the sample of Song et al. (2018) grew faster than population on average by 0.34 basis points per year. Furthermore, labor's shares in gross domestic income decrease on average by 0.41 percent per year. These two changes can lower the growth rate of average earnings relative to that of real GDP per capita by 75 basis points per year, which account for almost all differences. Therefore, the growth rate of real GDP per capita may overestimate a rise in average earnings and experiences of workers belonging to the left-wing of the log earnings distribution.

So far, we saw that asymmetric Laplace distributions successfully match the log earnings data in the U.S. I also documented a trend in the log earnings distribution becoming less left-skewed as well as trends in the dispersion and location. Furthermore, a simple functional form of the PDF of the asymmetric Laplacian family enabled an easy decomposition of the growth of average earnings into different factors and different income groups. While the analysis in this subsection is mostly based on a repeated application of static models to the earnings data in each year, and therefore highlights tractability of asymmetric Laplacian family in a static setup, this distribution family can provide a useful framework in a dynamic setup too.

[9] By combining Equations (1.6), (1.8), and (1.9), for $p \leq G(\theta)$, one can show that $E(Y|\text{bottom } 100p\%) = \frac{1}{1+\kappa^2}\frac{\sqrt{2}\kappa}{\sigma+\sqrt{2}\kappa}\left\{\exp(\theta)\left[p(1+\kappa^2)\right]^{1+\frac{\sigma}{\sqrt{2}\kappa}} - \exp\left(-\frac{\sqrt{2}\kappa}{\sigma}\theta\right)\right\} + \nu$.

1.4.3 A dynamic model

This subsection provides a simple, linear dynamic model of individual log earnings whose cross-sectional distribution always has an asymmetric Laplace density. I start from a conventional Gaussian AR(1) model and make a slight adjustment to achieve asymmetric Laplace distributions. In doing so, I rely on Proposition 1.4, which allows an easy analysis with substantial tractability.

I consider the following standard AR(1) model of individual log earnings:

$$y_{i,t+1} = (1-\rho)E(y_{i,t}) + \rho y_{i,t} + \epsilon_{i,t+1}, \tag{1.13}$$

where $\epsilon_{i,t+1}$ is an idiosyncratic innovation for all individuals i. When $y_{i,t}$ and $\epsilon_{i,t+1}$ are jointly normally distributed (*e.g.*, when each of them has a normal marginal density, and they are independent, they become jointly normally distributed), $y_{i,t+1}$ also has a normal distribution. A similar property holds for asymmetric Laplacian random variables.

Proposition 1.5. *Suppose that*

$$\begin{pmatrix} y_{i,t} \\ \epsilon_{i,t+1} \end{pmatrix} \sim \mathcal{MAL}\left(\begin{pmatrix} \theta_t \\ \bar{\theta}_{t+1} \end{pmatrix}, \begin{pmatrix} \sigma_t^2 & \bar{\rho}_{t+1}\sigma_t\bar{\sigma}_{t+1} \\ \bar{\rho}_{t+1}\sigma_t\bar{\sigma}_{t+1} & \bar{\sigma}_{t+1}^2 \end{pmatrix}, \begin{pmatrix} \mu_t \\ \bar{\mu}_{t+1} \end{pmatrix} \right), \tag{1.14}$$

for all i. Then

$$\begin{pmatrix} y_{i,t} \\ y_{i,t+1} \end{pmatrix} \sim \mathcal{MAL}\left(\begin{pmatrix} \theta_t \\ \theta_{t+1} \end{pmatrix}, \begin{pmatrix} \sigma_t^2 & \sigma_{t+1,12} \\ \sigma_{t+1,12} & \sigma_{t+1}^2 \end{pmatrix}, \begin{pmatrix} \mu_t \\ \mu_{t+1} \end{pmatrix} \right), \tag{1.15}$$

where

$$\theta_{t+1} = \rho\theta_t + \bar{\theta}_{t+1} + (1-\rho)(\theta_t + \mu_t), \tag{1.16a}$$

$$\sigma_{t+1,12} = \rho\sigma_t^2 + \bar{\rho}_{t+1}\sigma_t\bar{\sigma}_{t+1}, \tag{1.16b}$$

$$\sigma_{t+1}^2 = \bar{\sigma}_{t+1}^2 + 2\rho\bar{\rho}_{t+1}\sigma_t\bar{\sigma}_{t+1} + \rho^2\sigma_t^2, \tag{1.16c}$$

$$\mu_{t+1} = \rho\mu_t + \bar{\mu}_{t+1}. \tag{1.16d}$$

Specifically, $y_{i,t+1} \sim \mathcal{AL}(\theta_{t+1}, \sigma_{t+1}^2, \kappa_{t+1})$ *where* $\kappa_{t+1} = \frac{\sqrt{2}\sigma_{t+1}}{\sqrt{2\sigma_{t+1}^2 + \mu_{t+1}^2} - \mu_{t+1}}$.

Proof. Apply Proposition 1.4 to Equation (1.14) with $\Gamma = \begin{pmatrix} 1 & 0 \\ \rho & 1 \end{pmatrix}$ and $\delta = \begin{pmatrix} 0 \\ (1-\rho)(\theta_t + \mu_t) \end{pmatrix}$. □

Therefore, we can make the cross-section of log earnings always follow an asymmetric Laplace distribution in this manner given the AR(1) model (1.13). This further implies that the right-tail of the earnings distribution follows a Pareto distribution in every period (Proposition 1.2). Gabaix et al. (2016) show that with normal innovations, the canonical model (1.13) fails to achieve such a property. A simple refinement in Proposition 1.5 is to consider innovations having distributions with fatter tails than Gaussian tails without changing the basic structure of Model (1.13). This is also consistent with empirical findings in Guvenen et al. (2015) that earnings shocks are fat-tailed (and left-skewed). Instead, one may develop a model of fat-tailed innovations by mixing two Gaussian processes, usually understood as transitory and permanent earnings (Blundell and Preston, 1998; Kaplan, Moll and Violante, 2018; Pistaferri, 2001). Asymmetric Laplacian innovations in Model (1.14) provide a different framework in which either large (seemingly permanent) shocks or small (seemingly transitory) shocks are drawn from a single distribution. Note that Laplace distributions have more probability densities both around the mode (θ) and in tails than normal distributions (see Figure 1.5). Therefore, a mixture of two normal distributions with a small and large standard deviation can be closely approximated by a single Laplace distribution.

Next, I turn to the estimation of newly introduced parameters $\bar{\theta}_t$, $\bar{\rho}_t$, $\bar{\sigma}_t$, $\bar{\mu}_t$, and $\bar{\kappa}_t$, where $\epsilon_{i,t} \sim \mathcal{AL}(\bar{\theta}_t, \bar{\sigma}_t^2, \bar{\kappa}_t)$. In doing so, I fix ρ in Equation (1.13) at 0.9136, which is an estimate in Floden and Lindé (2001) based on U.S. panel data. To recover the parameters with a 'bar' from the estimates of θ_t, σ_t, κ_t, and $\mu_t = \frac{\sigma_t}{\sqrt{2}}\left(\kappa_t - \frac{1}{\kappa_t}\right)$ obtained in the previous subsection, I assume that $y_{i,t}$ and $\epsilon_{i,t+1}$ are not correlated.

Proposition 1.6. *Suppose that* $Cov(y_{i,t}, \epsilon_{i,t+1}) = 0$. *Then*

$$\bar{\theta}_{t+1} = \theta_{t+1} - \theta_t - (1-\rho)\mu_t, \tag{1.17a}$$

$$\bar{\sigma}_{t+1}^2 = \sigma_{t+1}^2 + 2\rho\mu_t(\mu_{t+1} - \rho\mu_t) - \rho^2\sigma_t^2, \tag{1.17b}$$

$$\bar{\mu}_{t+1} = \mu_{t+1} - \rho\mu_t, \tag{1.17c}$$

$$\bar{\kappa}_{t+1} = \frac{\sqrt{2}\bar{\sigma}_{t+1}}{\sqrt{2\bar{\sigma}_{t+1}^2 + \bar{\mu}_{t+1}^2} - \bar{\mu}_{t+1}}, \tag{1.17d}$$

$$\bar{\rho}_{t+1} = -\frac{\mu_t(\mu_{t+1} - \rho_t \mu_t)}{\sigma_t \bar{\sigma}_{t+1}}. \tag{1.17e}$$

Proof. $\bar{\theta}_{t+1}$ and $\bar{\mu}_{t+1}$ follow directly from Equations (1.16a) and (1.16d), respectively. Because $Cov(y_{i,t}, \epsilon_{i,t+1}) = \bar{\rho}_{t+1}\sigma_t\bar{\sigma}_{t+1} + \mu_t\bar{\mu}_{t+1} = 0$, $\bar{\rho}_{t+1}\bar{\sigma}_{t+1} = -\frac{\mu_t\bar{\mu}_{t+1}}{\sigma_t} = -\frac{\mu_t}{\sigma_t}(\mu_{t+1} - \rho_t\mu_t)$. By plugging the last expression to Equation (1.16c), I derive $\bar{\sigma}_{t+1}$ and $\bar{\rho}_{t+1}$. $\bar{\kappa}_{t+1}$ is obtained by inverting the definition of μ in terms of κ given σ. □

The above parameters describing idiosyncratic risks are plotted in Figure 1.8 with the implied mean and standard deviation of $\epsilon_{i,t}$ being equal to $\bar{\theta}_t + \bar{\mu}_t$ and $\sqrt{\bar{\sigma}_t^2 + \bar{\mu}_t^2}$. Correlation between $y_{i,t-1}$ and $y_{i,t}$ is based on Equation (1.15). Because $\bar{\kappa}_t$ is less than 1 (a solid line in the top-right panel), distributions of idiosyncratic innovations are left-skewed. However, it is more close to a symmetric distribution than $y_{i,t}$, because $\kappa_t < 0.8 < \bar{\kappa}_t < 1$ for all t. Given less-skewed innovations in every year, the log earnings distribution is becoming more symmetric as is discussed in Section 1.4.2. $E(\epsilon_{i,t})$, depicted in the bottom-left panel by a sold line, is procyclical, conforming to intuitions. $\bar{\sigma}_t$ shows an upward trend (a dotted line in the top-left panel), which leads to a positive trend in the standard deviation of $\epsilon_{i,t}$ (a dotted line in the bottom-left panel) and therefore the dispersion of $y_{i,t}$. Idiosyncratic risks should increase to match rising inequality in the data using Model (1.13). As a result, the correlation between $y_{i,t}$ and $y_{i,t+1}$ reduces.

This subsection serves an illustrative purpose of explaining how to handle linear dynamic models with asymmetric Laplace distributions. The main takeaway from Proposition 1.5 and 1.6 is that one can work with the same logic and idea that are valid for normal distributions while maintaining asymmetric Laplace distributions for the cross-section of log earnings.

Applications in the last two subsections deal with the U.S. log earnings distribution mainly with a focus on the fit and tractability of asymmetric Laplace distributions. Furthermore, I document trends in the skewness factor as well as the location and dispersion factors of the log earnings distribution. However, certain aspects of distributional issues may be more relevant in the short-run as they interact with the aggregate economy at the business cycle frequencies (Auclert, 2017; Kaplan, Moll and Violante, 2018; Hagedorn, Manovskii and Mitman, 2019).

1.5 Cyclical properties of the earnings distribution

This section covers basic statistics characterizing the short-run dynamics of the earnings distribution in the U.S. in relations to business cycles. I investigate cyclical properties of the three underlying factors of the log earnings distribution in the U.S. and their implications on the first four moments of the cross-section of log earnings.

To isolate the cyclical components, I employ Hodrick-Prescott (HP) filters (Hodrick and Prescott, 1997). I calibrate the smoothness parameter of the filter at 6.25 following the suggestion of Ravn and Uhlig (2002) for annual data. The results do not change qualitatively when different methods for detrending are used, or the smoothness parameter for the HP filter is 100.

Ten variables are analyzed in Table 1.4: log real GDP per capita, θ_t, σ_t, κ_t, moments of $\mathcal{AL}(\theta_t, \sigma_t^2, \kappa_t)$ including mean, standard deviation, skewness coefficient, and kurtosis coefficient, correlation between $y_{i,t-1}$ and $y_{i,t}$ implied by Equation (1.15), and log P90/P10 index in the data ($y_t(0.9) - y_t(0.1)$). The second column displays standard deviations of the HP-filtered variables. For example, the standard deviation of θ_t is $0.80\% = 0.008$. The next three columns are for correlations between log real GDP per capita at $t + h$ and the corresponding variable at t. When interpreting the results, one should bear in mind that these moments are *unconditional* quantities aggregating the effects of various structural shocks.

Among the three underlying factors, only the asymmetry (skewness) factor κ_t is procyclical. Its correlation with real GDP per capita is 33%, where θ_t and σ_t are almost orthogonal to log real GDP per capita. When I use the smoothness parameter of 100, the correlation between θ_t and log real GDP per capita at t becomes 39%, implying procyclicality of the location factor. Even in this case, σ_t is very weakly correlated with output dynamics. Therefore, degrees of asymmetry, or skewness is the major driving force of cyclical fluctuations of the earnings distribution. This is in line with Guvenen, Ozkan and Song (2014) emphasizing the importance of time-varying skewness of idiosyncratic risks with the business cycles.

The procyclicality of κ_t (and arguably θ_t) leads to procyclical movements in the mean of the estimated log earnings distribution $\mathcal{AL}(\theta_t, \sigma_t^2, \kappa_t)$. This is because as κ_t increases, more workers lie in the right-wing of the distribution (see Equation (1.8) and Proposition 1.3 (i)). This further yields a positive correlation between the skewness coefficient (Proposition 1.3 (iii)) and log real GDP per capita. On the other hand, the standard deviation and kurtosis coefficient tend to decrease in recessions (Proposition 1.3 (ii) and (iv)). Because $\kappa_t < 1$, the log earnings distribution is left-skewed. As κ_t increases in expansions, the distribution becomes less left-skewed, or closer to a symmetric distribution. As probability densities move from a thicker left-wing to a thinner right-wing, the number of workers with extremely negative log earnings reduces while that of workers with (relatively) moderately positive log earnings rises. Therefore, the standard deviation and kurtosis coefficient decrease. As an extremely large downward movement becomes less likely in a similar vein ($\bar{\kappa}_t$ in Figure 1.8 is procyclical), the correlation between $y_{i,t-1}$ and $y_{i,t}$ implied by the model in Section 1.4.3 rises in expansions. This implies that a recession is when $y_{i,t-1}$ and $y_{i,t}$ become less correlated, consistent with the idea of counter-cyclical earnings risks (Guvenen, Ozkan and Song, 2014; Storesletten, Telmer and Yaron, 2004).

In short, most of the cyclical variations in the moments of log earnings distributions are driven by fluctuations in κ_t. By tilting both tails and changing probability densities at large deviations from the center of the distribution, κ_t impacts on the cross-sectional moments substantially. However, it is less clear whether the middle part of the distribution is also similarly affected by those forces inducing such dynamics in the tails. Indeed, the log P90/P10 index ($y_t(0.9) - y_t(0.1)$) is only weakly correlated with log real GDP per capita unlike the other cross-sectional moments in Table 1.4. This is consistent with the findings in footnote 4 that the log P90/P10 index is tightly related to the dispersion factor σ_t, which is almost acyclical. In other words, the cyclical properties of the log earnings distributions in Table 1.4 are due to the tail dynamics.

1.6 Conclusion

Understanding distributional issues is important because individual heterogeneity matters for not only individual decisions but also the aggregate economy. While cross-sectional distributions are complicated, high-dimensional objectives, I show that it is possible to represent the whole distribution with a small number of the underlying factors. The log earnings distribution in the U.S. have only three major factors, each of them measures the location, dispersion, and skewness, respectively. I propose asymmetric Laplace distributions as a simple model of the log earnings distribution. It is parametrized by these three factors and has a tractable functional form in both static and dynamic environments. Furthermore, asymmetric Laplace distributions achieve a better fit to the data than conventional Gaussian models.

Given the impressive quality of the three-factor model of the log earnings distribution, shedding lights on the underlying economic mechanisms which enable such a representation is an important direction for future research. Developing a method for finite state Markov-chain approximations of dynamic models such as Rouwenhorst (1995), Tauchen (1986) and Tauchen and Hussey (1991), but are tailored to asymmetric Laplace distributions is another topic that would be useful for applied macroeconomic research. From a theoretical angle, while being speculative, the characterization of a probability measure by a sequence of real numbers in Section 1.3.1 might provide a convenient technique when dealing with infinite-dimensional state-variables in dynamic systems (*e.g.*, a cross-section of wealth in incomplete market models à la Krusell and Smith (1998)).

Table 1.1: R^2 of the first stage factor models with a different N.

Year / Order	1980	1990	2000	2010
$N = 1$	0.9457	0.9421	0.9444	0.9616
$N = 2$	0.9616	0.9609	0.9528	0.9667
$N = 3$	0.9943	0.9954	0.9946	0.9966
$N = 4$	0.9945	0.9957	0.9947	0.9966
$N = 5$	0.9986	0.9991	0.9990	0.9991
$N = 9$	0.9999	0.9999	0.9999	0.9999

Notes: Equation (1.2) includes an intercept on the right-hand side because $P_0 = 1$, and therefore the coefficient of determination R^2 is defined in a usual manner. When $N = 0$, R^2 is zero by definition. With a relatively small number of orthogonal polynomials, we can approximate the log earnings distribution quite closely. For example, the R^2 of Equation (1.2) in 1980 when a third order polynomial ($N = 3$) is employed is greater than 99 percent. Furthermore, the second-order orthogonal polynomial P_2 makes some contributions to the fit, which implies that log earnings distributions in the U.S. have some asymmetric components. Finally, the fit is almost perfect for $N \geq 9$.

Table 1.2: Scree table.

	$m=1$	$m=2$	$m=3$	$m=4+$
Case 1: no detrending ($b=0$)	91	6	2	1
Case 2: detrended $\alpha_{t,0}$ (b is proportional to e_1)	47	45	7	2
Case 3: $f_{t,1}$ = detrended $\alpha_{t,0}$ ($\lambda_1 = e_1$)	-	82	12	6
Case 4: detrended $\alpha_{t,m}$ for all m	78	15	5	2
Case 5: differenced $\alpha_{t,m}$ ($\Delta\alpha_t = \Lambda f_t + v_t$)	58	28	10	4
Case 6: differenced y_t ($\Delta y_t = \Psi f_t + \eta_t$)	60	27	9	4

Notes: This table is about the contribution of each factor $f_{t,m}$ to the variation in α_t in Equation (1.3) based on principal component analyses. For example, when the first-stage factor $\alpha_{t,n}$ is directly investigated without detrending, the deep factors $f_{t,1}$, $f_{t,2}$, and $f_{t,3}$ explain 91, 6, and 2 percent of the total variance, respectively, where all the other principal components' contribution is less than 1 percent in total. For the second case, I use the linearly detrended zeroth order factor $\alpha_{t,0}$ by assuming that b is a vector whose first element is one and the other elements are zero (e_1). The third case is conditional on $f_{1,t}$ being equal to the detrended $\alpha_{t,0}$, and therefore $\lambda_1 = e_1$. Because the principal components are derived from $(\alpha_{t,1}, \ldots, \alpha_{t,N})'$ without $\alpha_{t,0}$, the corresponding row in Table 1.2 begins with $n = 2$. The remaining two cases are for the detrended $\alpha_{t,m}$ for all m and the differenced first-stage factors $\Delta\alpha_{t,m}$, respectively. Note that the first two factors $f_{t,1}$ and $f_{t,2}$ capture most of the variation in α_t and therefore the log earnings distribution y_t. For example, the contribution of $f_{t,1}$ and $f_{t,2}$ is greater than 90 percent in the second case. Furthermore, it seems to be sufficient to include just one more factor to take account of most of the remaining variances. Deriving principal components of y_t directly from Equation (1.4) yields almost identical results (*e.g.*, the results for case 6 based on Δy_t is similar to that of case 5).

Table 1.3: Decomposition of the growth from 1981 to 2013.

Variable	$E(Y_t)$	$E(Y_t\|\sigma_1, \kappa_1)$	$E(Y_t\|\kappa_1)$	RGDP per capita
Units	2013 USD	2013 USD	2013 USD	2013 USD
1981	39,388	39,388	39,388	30,661
2013	53,121	42,800	44,994	53,045
Avg. growth rate (%)	0.93	0.26	0.42	1.71
Decomposition (%)	100	28	44	183
Variable	Population	SSA Sample	Population / SSA Sample	Labor's Share
Units	Millions	Millions	%	%
1981	230.0	55.5	414	47.7
2013	316.4	85.2	371	41.9
Avg. growth rate (%)	1.00	1.34	-0.34	-0.41
Decomposition (%)	107	143	-37	-43

Notes: Y_t represents a cross-section of earnings in the year t: $Y_t = \exp(y_t) = \exp(\tilde{y}_t) + \nu_t$. Similar to the characteristic function in Proposition 1.3 (v), one can show that $E(Y_t) = \frac{\exp(\theta_t)}{1-\frac{1}{2}\sigma_t^2-\mu_t} + \nu_t$, where $\mu_t = \frac{\sigma_t}{\sqrt{2}}\left(\kappa_t - \frac{1}{\kappa_t}\right)$. This average earnings per worker grew at a 0.93 percent annual rate on average from 1981 to 2013. $E(Y_t|\sigma_1, \kappa_1)$ calculates this quantity under the assumption that σ and κ do not change from their 1981 values. Therefore, by comparing $E(Y_t|\sigma_1, \kappa_1)$ in 1981 and 2013, we can detect the contribution of θ_t on the growth of average earnings per worker. The growth rate of this quantity is only 26 basis points, taking up only 28 percent of the total average growth rate of $E(Y_t)$. $E(Y_t|\kappa_1)$ is similar, but it fixes only κ_t enabling us to assess the effects of σ_t. A rise in σ_t explains an additional 16 basis points in the growth rate (0.42% - 0.26%), and θ_t and σ_t together explain 44 percent of the growth rate of $E(Y_t)$. Therefore, more than half of the growth in the average earnings is based on a rise in the skewness of the log earnings distribution (κ_t). On the other hand, the growth rate of real GDP per capita is 1.71 percent, which is almost a double of that of $E(Y_t)$. This huge difference is originated mostly from two sources: a faster rise in the number of workers in the SSA's data than the population and a declining labor's share in a total income. The number of workers in the sample of Song et al. (2018) grew faster than population in the U.S. by 0.34 basis points per year. Furthermore, labor's shares in gross domestic income decrease by 0.41 percent on average per year. These two change can lower the growth rate of average earnings relative to that of real GDP per capita by 75 basis points per year, which account for almost all differences.

Table 1.4: Cyclical properties of the log earnings distribution in the U.S.

Unit: %	$std(\cdot)$	$corr$(log RGDP per capita at $t+h,\ \cdot$)		
		$h=-1$	$h=0$	$h=1$
log RGDP per capita at t	1.23	31	100	31
Factors				
Location (θ_t)	0.80	-3	4	-25
Dispersion (σ_t)	0.97	-22	-4	12
Asymmetry (κ_t)	0.65	23	33	19
Moments of $\mathcal{AL}(\theta_t, \sigma_t^2, \kappa_t)$				
Mean	1.04	42	51	0
Standard deviation	0.71	-58	-38	0
Skewness coefficient	2.15	24	34	19
Kurtosis coefficient	4.31	-23	-33	-19
$corr(y_{i,t-1}, y_{i,t})$	0.69	7	41	41
Log P90/P10 index	1.35	0	15	-5

Notes: Ten variables are analyzed in Table 1.4: log real GDP per capita, θ_t, σ_t, κ_t, moments of $\mathcal{AL}(\theta_t, \sigma_t^2, \kappa_t)$ including mean, standard deviation, skewness coefficient, and kurtosis coefficient, correlation between $y_{i,t-1}$ and $y_{i,t}$ implied by Equation (1.15), and log P90/P10 index in the data ($y_t(0.9) - y_t(0.1)$). All the variables are HP-filtered where the smoothness parameter is 6.25. The sample period is from 1978 to 2013. The second column displays standard deviations of the HP filtered variables. For example, the standard deviation of θ_t is 0.80% = 0.008. The next three columns are for correlations between log real GDP per capita at $t+h$ and the corresponding variable at t.

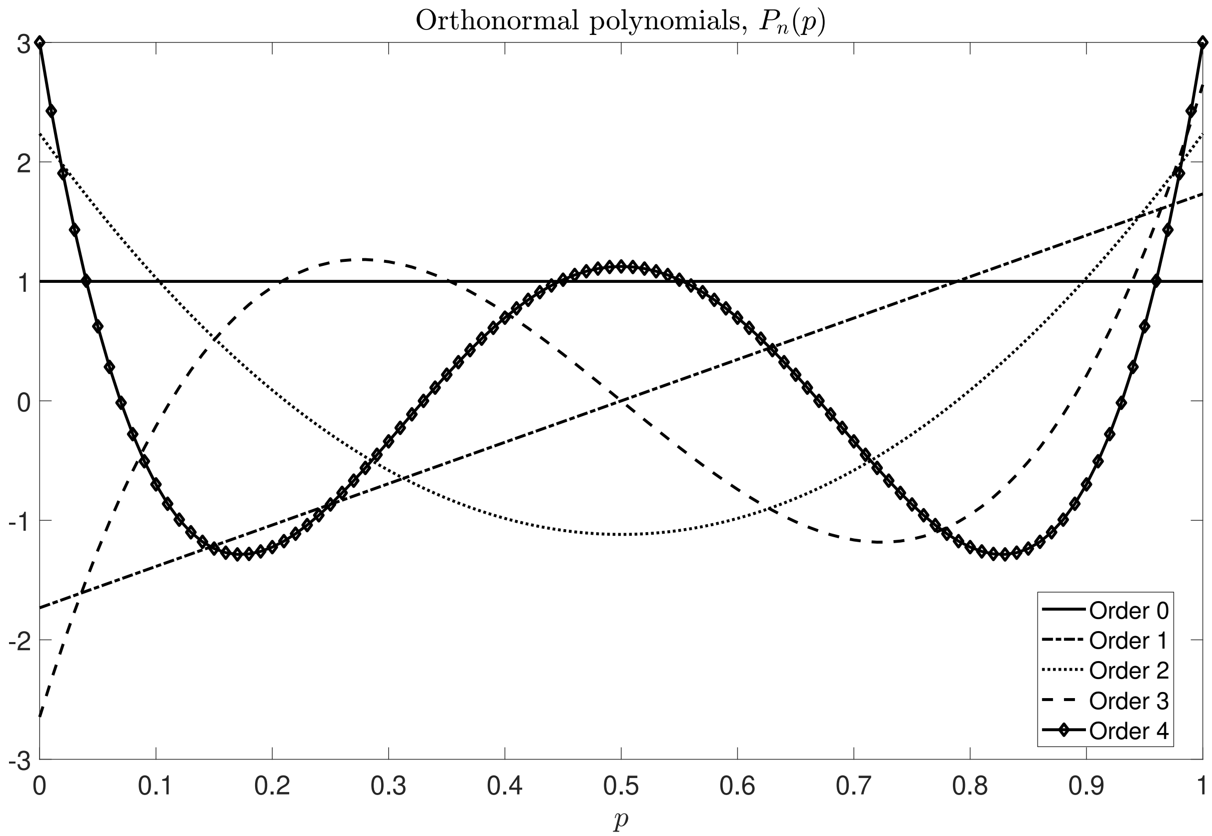

Figure 1.1: Orthogonal polynomials, $P_n(p)$.

Notes: $P_0 = 1$ and $P_1(p) = \sqrt{12}(p-0.5)$ as discussed in Section 1.3.1. P_2, P_3, and P_4 are represented by a dotted line, a dashed line, and a solid line with diamonds, respectively.

Figure 1.2: Goodness-of-fit of the first-stage factor model, 1978.

Notes: Blue triangles depict $y_{1978}(p)$, log percentiles of earnings in 1978 for $p = 0.01, \ldots, 0.99$. These earnings are in 2013 real values. Black lines represent the fitted values based on Equation (1.2), where the order of polynomial N is 0, 1, 3, and 9, respectively. Note that even relatively low-order polynomials, *e.g.,* a cubic polynomial, provide a reasonable approximation. When $N = 9$, the fit becomes almost perfect in the tails as well as in the middle.

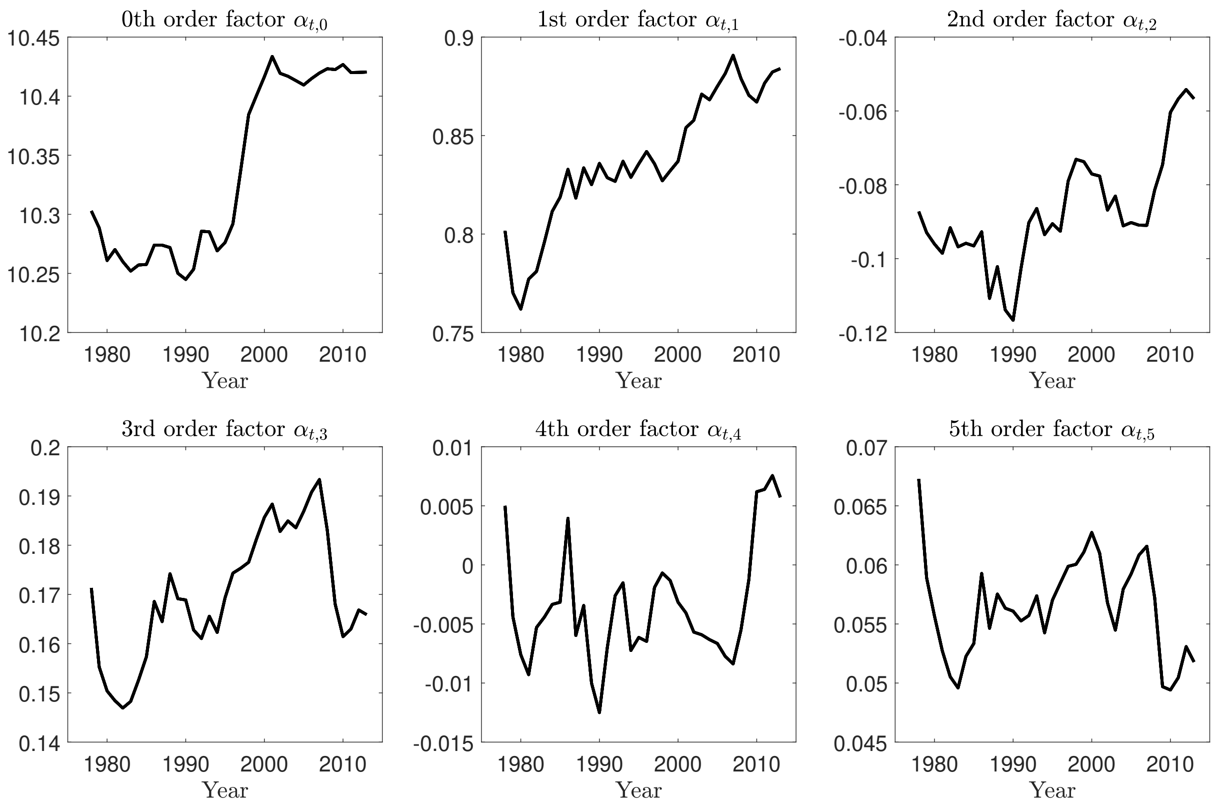

Figure 1.3: First-stage factors $\alpha_{t,n}$ for $n = 0, \ldots, 5$ when $N = 9$.

Notes: Each panel illustrates the first-stage factors $\alpha_{t,n}$ in Equation (1.2) for each n from 0 to 5. The order of polynomial on the right-hand side N is 9. The zeroth order factor represents the location of the log earnings distribution in the U.S., where the rising trend captures economic growth. The secular rise in earnings inequality is reflected in upward trends in the first- and third-order factors, where the third-order factor focuses more on the tails than the first-order factor as is shown in Figure 1.1. The magnitude of the second-order factor implies that there exist some asymmetric components in the log earnings distribution. However, the absolute value of the factor has become smaller, and therefore the degree of asymmetry has been declining.

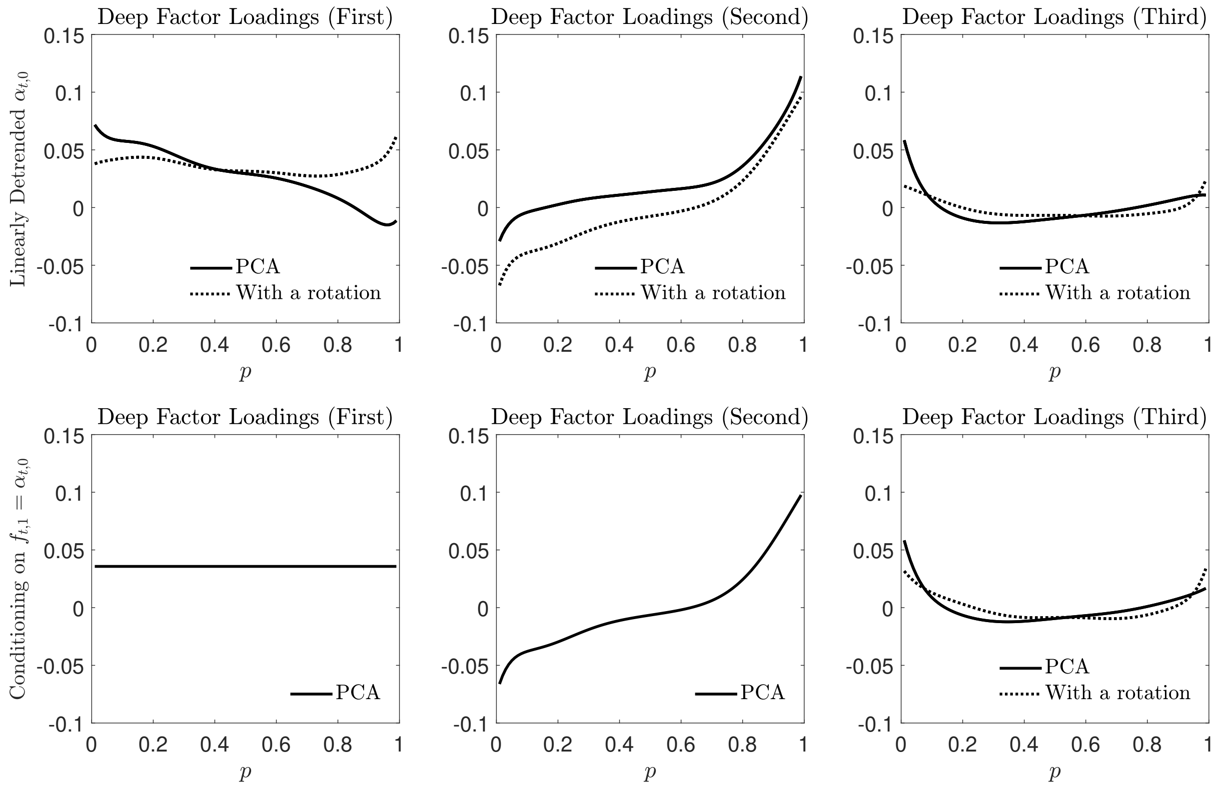

Figure 1.4: Deep factor loadings $P\lambda_m$ for $m = 1$, 2, and 3 with and without factor rotations.

Notes: The top three panels are based on Case 2 in Table 1.2, where the linearly detrended zeroth order factor $\alpha_{t,0}$ are investigated in Model (1.3) and the other $\alpha_{t,n}$'s are demeaned. The solid lines depict $P\lambda_m$, which is the effect of a unit increase in $f_{t,m}$ on $y_t(p)$ for each $p = 0.01, \ldots, 0.99$. Here λ_m is identified as the m-th eigenvector in the PCA times $\sqrt{Var(f_{t,m})}$, because $f_{t,m}$'s are normalized to have a unit variance. However, in Model (1.4), f_t is identified only up to a rotation, because $(\Psi R')(Rf_t) = \Psi f_t$ for any orthogonal matrix R. The dotted lines show an example of those loadings $\Psi R'$ on a rotated factors Rf_t. The dotted lines in the first two panels are obtained by rotating $(f_{t,1}, f_{t,2})'$ and the third one is based on a rotation of $(f_{t,3}, f_{t,4})'$. The bottom three panels are similar, but illustrate the results for Case 3 in Table 1.2. That is, I derive principal components of $(\alpha_{t,1}, \ldots, \alpha_{t,M})'$ conditioning on $f_{t,1}$ being equal to the detrended $\alpha_{t,0}$.

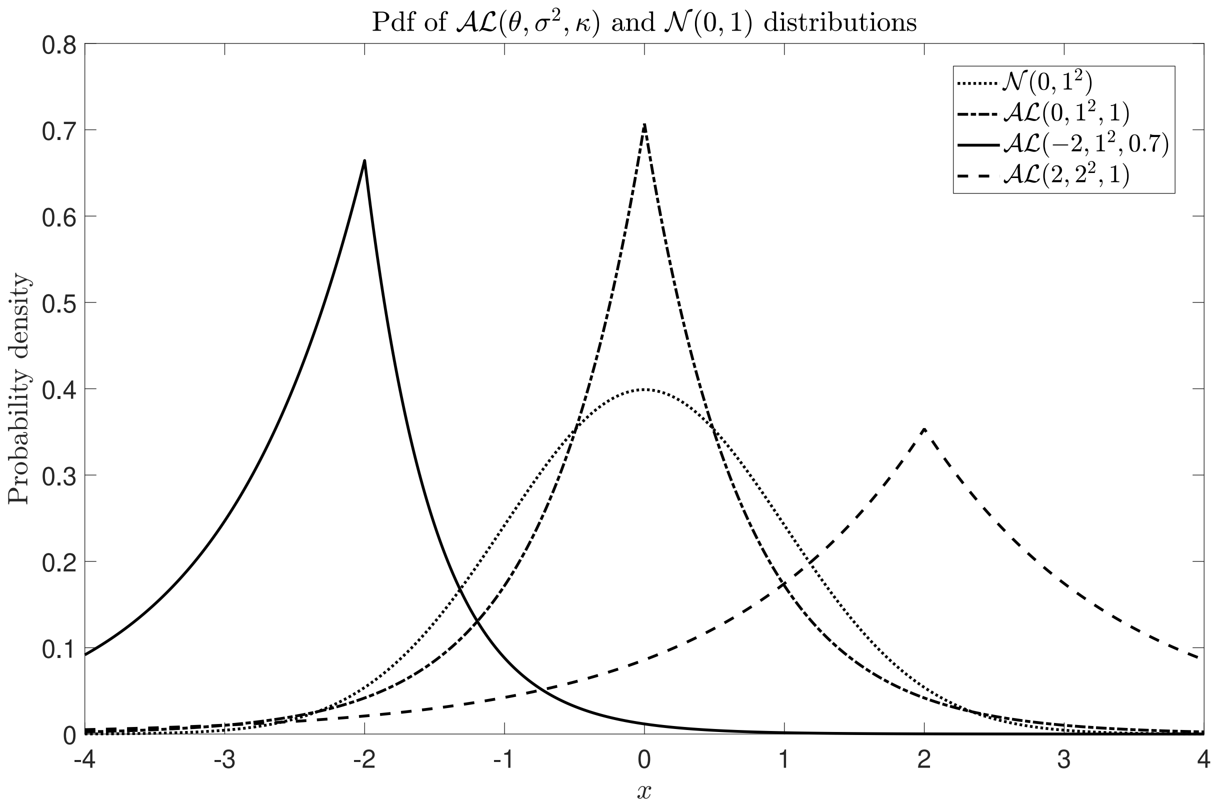

Figure 1.5: Probability densities of $\mathcal{AL}(\theta, \sigma^2, \kappa)$ and a standard normal distribution.

Notes: This figure depicts four different PDFs. A dotted line is for a standard normal distribution, where a dash-dot line represents a symmetric Laplace distribution with the same mean (0) and variance (1). The Laplace distribution has more probability masses around zero, which is the value of its location parameter, and in both tails than the normal distribution. A left-skewed density function, denoted by a solid line, is based on $\theta = -2$, $\sigma = 1$, and $\kappa = 0.7 < 1$. Finally, the last one is illustrated by a dashed line, where its $\theta = 2$, $\sigma = 2$, and $\kappa = 1$.

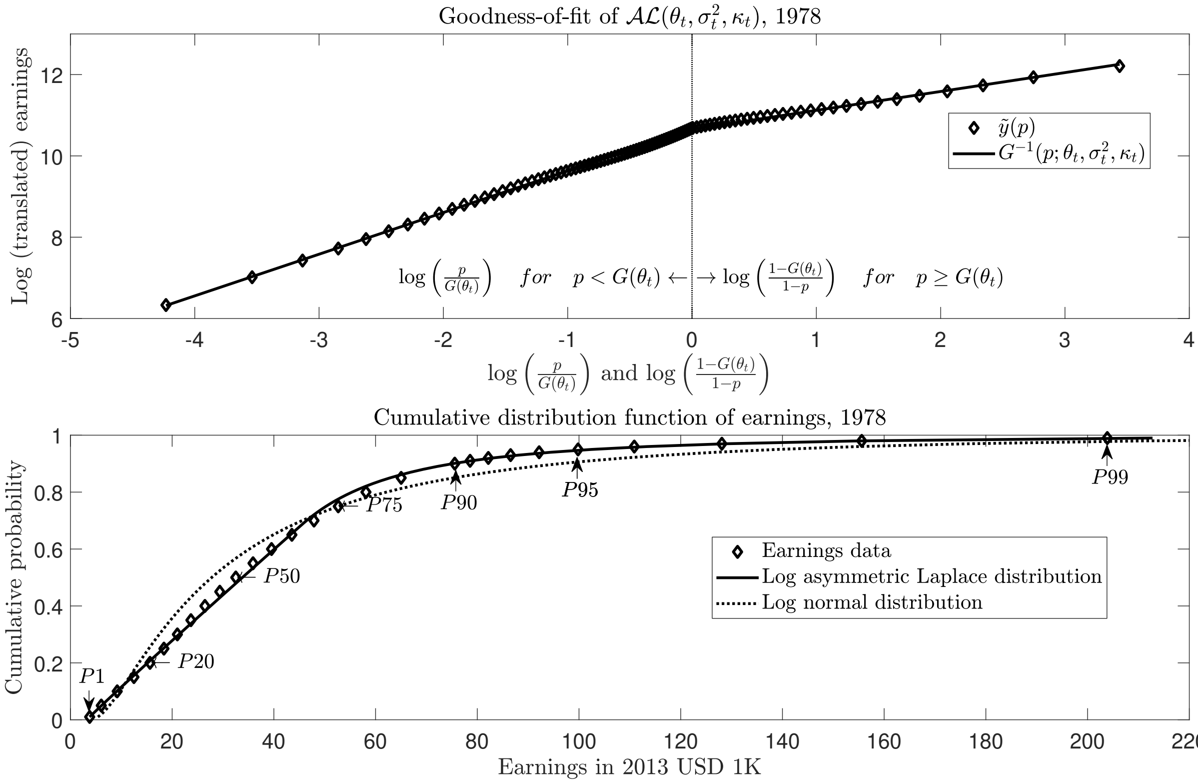

Figure 1.6: Goodness-of-fit of asymmetric Laplace distributions to the U.S. earnings data.

Notes: In the top panel, each diamond represent $\left(\log\left(\frac{p}{G(\theta_1)}\right) \text{ or } \log\left(\frac{1-G(\theta_1)}{1-p}\right), \tilde{y}_t(p)\right)$ for $p = 0.01, \ldots, 0.99$ and $t = 1978$. I use $\log\left(\frac{p}{G(\theta_1)}\right)$ if $p < G(\theta_1)$, and therefore the log translated earnings $\tilde{y}_t(p)$ lies in the left-wing. Otherwise, $\tilde{y}_t(p)$ is plotted against $\log\left(\frac{1-G(\theta_1)}{1-p}\right)$ for the right-wing. A solid line is based on the fitted distribution with a similar transformation. In this manner, I obtain a piecewise linear graph that characterizes an asymmetric Laplace distribution (see Equation (1.9)). The bottom panel covers a CDF of real earnings in 1978 in 2013 real values. Diamonds depict several percentiles of earnings $\exp(y_t(p))$ in the data. A solid line is for the CDF implied by the estimated asymmetric Laplace distribution of the log earnings, whereas a dotted line presents the results for a normal distribution estimated similarly to Equation (1.12). While the log normal distribution provides a reasonable approximation to the data, it overestimates the CDF in the range between $P20$ and $P75$ and therefore underestimates in both tails. The log asymmetric Laplace distribution provides a better fit.

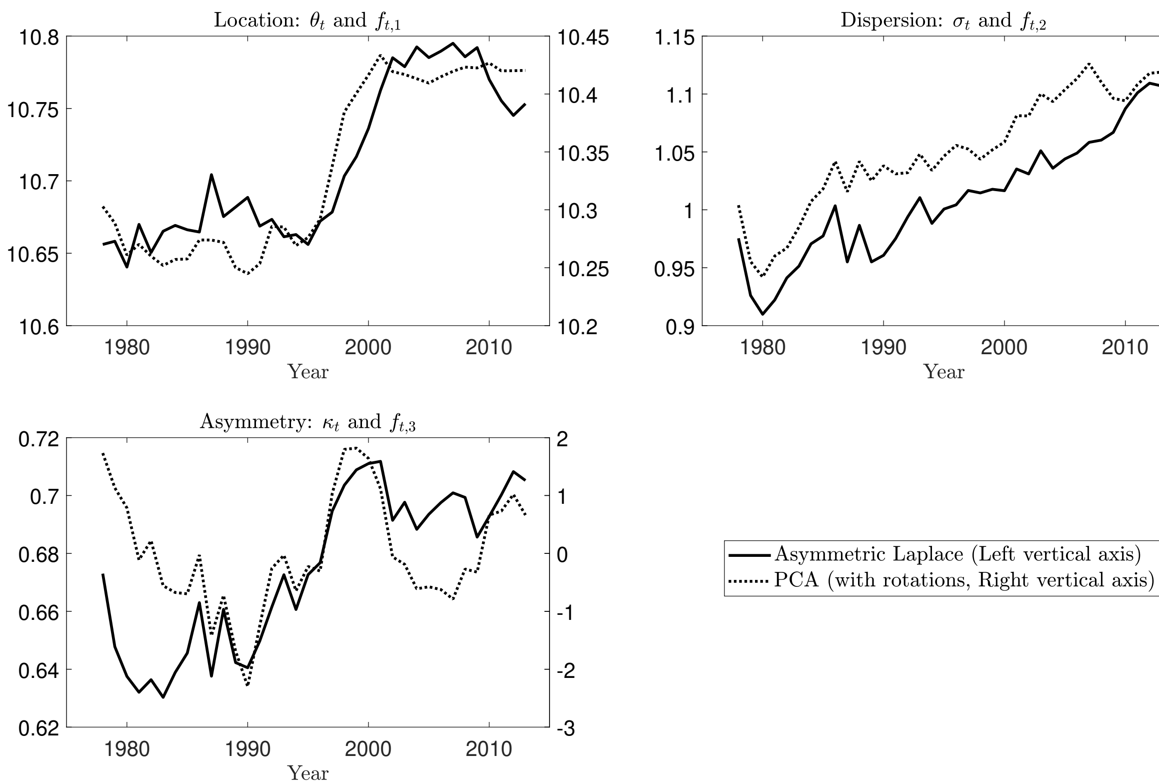

Figure 1.7: Parameters of asymmetric Laplace distributions and factors based on the PCA.

Notes: Each panel illustrates the three parameters of the estimated asymmetric Laplace distributions (Equation (1.12)) and the three factors that the PCA (Equation (1.3)) induces. For the PCA, I calculate the principal components conditional on $f_{t,1} = \alpha_{t,0}$ as is the case in the second row in Figure 1.4. $f_{t,3}$ is subject to a rotation explained in footnote 3. The ticks on the left (right) vertical axis are for θ_t, σ_t, and κ_t ($f_{t,1}$, $f_{t,2}$, and $f_{t,3}$), which are denoted by solid (dotted) lines.

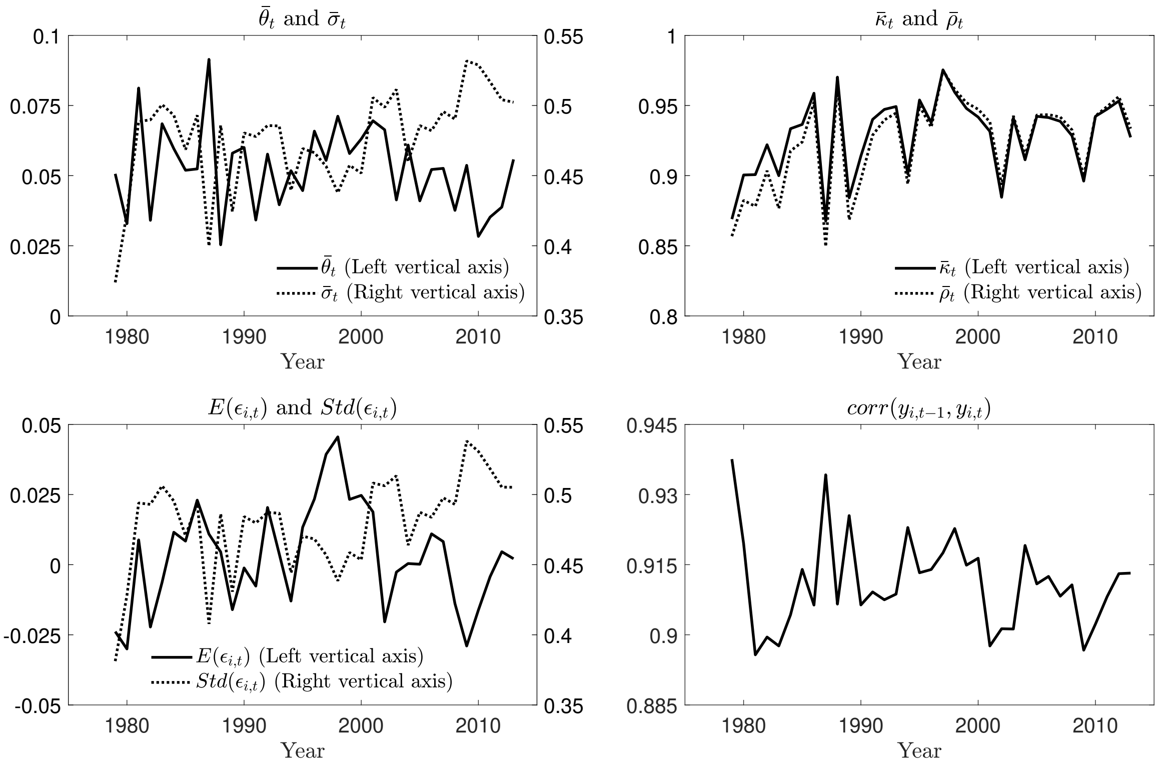

Figure 1.8: Parameters for idiosyncratic innovations and implied moments.

Notes: $\bar{\theta}_t$, $\bar{\sigma}_t$, $\bar{\kappa}_t$, and $\bar{\rho}_t$ are constructed as explained in Proposition 1.6. Mean and standard deviation of $\epsilon_{i,t}$ equal $\bar{\theta}_t + \bar{\mu}_t$ and $\sqrt{\bar{\sigma}_t^2 + \bar{\mu}_t^2}$. Correlation between $y_{i,t-1}$ and $y_{i,t}$ is based on Equation (1.15). Because $\bar{\kappa}_t$ is less than 1 (a solid line in the top-right panel), distributions of idiosyncratic innovations are left-skewed. However, it is more close to a symmetric distribution than $y_{i,t}$, because $\kappa_t < 0.8 < \bar{\kappa}_t < 1$ for all t. Given less-skewed innovations in every year, the log earnings distribution is becoming more symmetric as is discussed in Section 1.4.2. $E(\epsilon_{i,t})$, depicted in the bottom-left panel by a sold line, is procyclical, conforming to intuitions. $\bar{\sigma}_t$ shows an upward trend (a dotted line in the top-left panel), which leads to a positive trend in the standard deviation of $\epsilon_{i,t}$ (a dotted line in the bottom-left panel) and therefore the dispersion of $y_{i,t}$. Idiosyncratic risks should increase to match rising inequality in the data using Model (1.13). As a result, the correlation between $y_{i,t}$ and $y_{i,t+1}$ reduces.

Chapter 2

Forecast Error Variance Decompositions with Local Projections

2.1 Introduction

Macroeconomists have been long interested in estimating dynamic responses of output, inflation, and other aggregates to structural shocks.[1] While many analyses use vector autoregressions (VARs) or dynamic stochastic general equilibrium (DSGE) models to construct estimated responses, an increasing number of researchers focus on a single structural shock and employ single-equation methods to study the dynamic responses. This approach allows concentrating on well-identified shocks and leaving other sources of variation unspecified. In addition, these approaches often impose no restrictions on the shape of the impulse response function. As a result, the local projections (LP) method (Jordà, 2005; Stock and Watson, 2007) has gained prominence in applied macroeconomic research.

The properties of impulse responses estimated with these methods are well studied (see, e.g., Coibion, 2012; Kilian and Kim, 2011), but little is known about how one can reliably estimate the quantitative significance of shocks in the single-equation framework. While some methods for constructing the forecast error variance decompositions (FEVDs) have been suggested, it usually has been done without investigation of their econometric properties,

[1] This chapter is based on joint work with Yuriy Gorodnichenko.

especially for empirically relevant sample sizes.[2] As a result, the vast majority of studies using single-equation approaches do not report the FEVD for the variable of interest, and hence one does not know if a given shock accounts for a large share of variation for the variable.[3] This practice contrasts sharply with the nearly universal convention to report FEVDs in VARs and DSGE models. In this paper, we propose and study finite-sample and asymptotic properties of a method to construct forecast error variance decompositions in the local projections framework.

We show that local projections lead to a simple and intuitive way to assess the contribution of identified shocks to the variation of forecast errors at different horizons. While there are several options to implement this insight, we mostly focus on an estimator based on the coefficient of determination, or R^2. To illustrate the properties of this method, we use several data generating processes (DPGs), including the Smets and Wouters (2007) model. These DGPs cover main profiles of FEVDs documented in previous works. We show that estimated contributions to the variation of forecast errors may be biased in small samples and one should use bootstrap to correct for possible biases in the FEVDs estimated by local projections. We also show that, in simulations, our estimator performs better than alternative approaches based on sums of squared estimates of impulse responses. We further illustrate the performance of our method with actual data and commonly used identified shocks. In short, our contribution is to develop a new estimator of FEVDs and to assess finite-sample properties of our estimator and alternative estimators.

We assume in this paper that the researcher has a series of identified shocks. However, these shocks may be measured with error in practice because, e.g., they are estimated rather than directly observed. We show that our estimator of FEVDs is downward biased when the shocks are imperfectly observed. Thus, our point estimates are conservative and likely provide a lower bound. In a concurrent and complementary work, Plagborg-Møller and Wolf (2018*a*) provide set-identified FEVDs given measurement errors in the local projections framework. Their partially identified untestable bounds could be useful tools for the

[2]For example, Jordà (2005) suggests an estimator close in spirit to LP-A and LP-B estimators that we cover in Section 2.3.1 and Appendix B. Our baseline estimator of FEVDs performs better than these estimators for empirically relevant sample sizes. Another method is to compute FEVDs by using VARs that directly include a structural shock (Plagborg-Møller and Wolf, 2018*b*). While this method identifies the same population FEVDs, it requires a large number of lags (Baek and Lee, 2019), a feature that may be too costly in practice given the curse of dimensionality in VARs and the noise generated by many estimated parameters.

[3]Coibion et al. (2017) is among the very few papers reporting FEVDs based on the local projection method.

researcher who is interested in upper bounds of the FEVDs.

The rest of the paper is structured as follows. Section 2.2 lays out a basic setting to derive the estimator. Section 2.3 introduces our estimator and illustrates its econometric properties. Section 2.4 presents simulation results for bivariate and multivariate settings. Section 2.5 applies our method to measuring the contribution of monetary policy and productivity shocks to the forecast error variance of output and inflation in the local projections framework. Section 2.6 concludes.

2.2 Basics of the forecast error variance decomposition

Consider a generic setup encountered in studies using local projections. Let y_t be an endogenous variable of interest. An identified white-noise shocks series $\{z_t\}$ has mean zero and variance σ_z^2. We assume that variation in y due to z is represented by $\psi_z(L)z_t = \sum_{i=0}^{\infty} \psi_{z,i} z_{t-i}$, where coefficients $\{\psi_{z,i}\}$ provide us with the impulse response function of y to z.

The forecast error for the h-period ahead value of the endogenous variable is given by

$$f_{t+h|t-1} \equiv (y_{t+h} - y_{t-1}) - P[y_{t+h} - y_{t-1}|\Omega_{t-1}], \tag{2.1}$$

where $P[y_{t+h} - y_{t-1}|\Omega_{t-1}]$ is the projection of $y_{t+h} - y_{t-1}$ on the information set $\Omega_{t-1} \equiv \{\Delta y_{t-1}, z_{t-1}, \Delta y_{t-2}, z_{t-2}\}$. To keep the exposition as simple as possible, we focus only on a single shock and a single endogenous variable for now, but in Section 2.3.6 we consider the case where the information set includes other ("control") variables. We can decompose the forecast errors due to innovations in z and other sources of variation as follows:

$$f_{t+h|t-1} = \psi_{z,0} z_{t+h} + \cdots + \psi_{z,h} z_t + v_{t+h|t-1}, \tag{2.2}$$

where $v_{t+h|t-1}$ is the error term due to innovations orthogonal to $\{z_t, z_{t+1}, \ldots, z_{t+h}\}$ and Ω_{t-1}.

Following Sims (1980), we can define the population share of the variances explained by the contemporaneous and future innovations in z_t to the total variations in $f_{t+h|t-1}$:

$$s_h = \frac{Var(\psi_{z,0} z_{t+h} + \cdots + \psi_{z,h} z_t)}{Var(f_{t+h|t-1})}. \tag{2.3}$$

In what follows, we propose and evaluate a method to estimate s_h based on equation (2.3).

Note that, if we use definitions of Plagborg-Møller and Wolf (2018*a*), the object of our analysis is the forecast variance ratio. Although this definition of s_h seems natural, one should bear in mind several caveats. First, s_h depends on Ω_t: adding more control variables changes the population parameter s_h (see Section 2.3.6). Second, the forecast error variance decomposition for a structural VAR model or a DSGE model is usually defined given an information set which includes all structural shocks, while s_h above is purely based on the observables. These two definitions might not coincide if two information sets differ. For example, if a data generating process is not invertible for structural shocks (the shocks are not recoverable from the history of observable variables), forecast variance ratio is different from variance decomposition (see Plagborg-Møller and Wolf, 2018*a*, for details on this point).

2.3 Estimator

In this section, we introduce our estimator of FEVDs using the coefficient of determination, or R^2 of local projections. We discuss asymptotic properties of our estimator and address issues that may be encountered in practice. Those issues include measurement errors in z_t, small-sample refinements with a focus on biases, and other control variables in the information set.

2.3.1 R^2 method

Let $Z_t^h = (z_{t+h}, \dots, z_t)'$. It can be shown with some algebra that equation (2.3) can be written as

$$s_h = \frac{\mathbf{Cov}(f_{t+h|t-1}, Z_t^h) \left[\mathbf{Var}(Z_t^h)\right]^{-1} \mathbf{Cov}(Z_t^h, f_{t+h|t-1})}{Var(f_{t+h|t-1})}. \tag{2.4}$$

In the numerator, the first **Cov** term is a row vector, the **Var** in the middle is a matrix, and the last **Cov** is a column vector. This quantity can be understood as an R^2 of the population projection of $f_{t+h|t-1}$ on Z_t^h, or the probability limit of sample R^2's. This observation suggests a natural estimator of s_h. First, the forecast errors for each horizon h are estimated using local projections. Second, the estimated forecast errors for the horizon h at time t are regressed on shocks that happen between t and $t+h$. The R^2 in this regression is an estimate of s_h.

More precisely, the estimated forecast error $f_{t+h|t-1}$ is the residual of the following regression:

$$y_{t+h} - y_{t-1} = c_h + \sum_{i=1}^{L_y} \gamma_i^h \Delta y_{t-i} + \sum_{i=1}^{L_z} \beta_i^h z_{t-i} + f_{t+h|t-1}, \tag{2.5}$$

which is an approximation to $y_{t+h} - y_{t-1} = c_h + \sum_{i=1}^{\infty} \gamma_i^h \Delta y_{t-i} + \sum_{i=1}^{\infty} \beta_i^h z_{t-i} + f_{t+h|t-1}$ in population. Then we run the following regression and calculate its R^2:

$$\hat{f}_{t+h|t-1} = \alpha_{z,0} z_{t+h} + \cdots + \alpha_{z,h} z_t + \tilde{v}_{t+h|t-1}. \tag{2.6}$$

Thus, our estimator $\hat{s}_h^{R2}$ is R^2 of equation (2.6) which, by construction, is between 0 and 1. Note that $\alpha_{z,i}$ in equation (2.6) corresponds to the impulse response coefficient $\psi_{z,i}$. Because $\hat{f}_{t+h|t-1}$ in equation (2.6) is a residual of an OLS regression with an intercept in equation (2.5) and the mean of z_t is zero, an intercept term in equation (2.6) is not required. Moreover, the population mean of both $f_{t+h|t-1}$ and Z_t^h are zeros, and so both centered and non-centered R^2's are the same in the population. We report results for the non-centered R^2 below, but properties are similar when we use the centered R^2.

Note that one may implement this estimator by augmenting equation (2.5) with shocks $z_t, \ldots, z_{t+h}$ and calculating the partial R^2. This modification ensures that any predictable variation in $z_t, \ldots, z_{t+h}$ is removed. In practice, this step likely makes little difference since z_t is typically constructed in a way such that z_t is not predictable by lags of macroeconomic variables.

LP-A and LP-B estimators of s_h. While we concentrate on the R^2 estimator, there are other options for estimating s_h. For example, note that s_h admits the following representations:

$$s_h = \frac{\left(\sum_{i=0}^{h} \psi_{z,i}^2\right) \sigma_z^2}{Var\left(f_{t+h|t-1}\right)} \tag{2.7}$$

$$= \frac{\left(\sum_{i=0}^{h} \psi_{z,i}^2\right) \sigma_z^2}{\left(\sum_{i=0}^{h} \psi_{z,i}^2\right) \sigma_z^2 + Var\left(v_{t+h|t-1}\right)}. \tag{2.7'}$$

Thus, one may estimate s_h by plugging estimates of $\psi_{z,i}$'s, σ_z^2, $Var(f_{t+h|t-1})$, or $Var(v_{t+h|t-1})$

into either (2.7) or (2.7').

We estimate $\psi_{z,h}$ with local projections by running the following regression:

$$y_{t+h} - y_{t-1} = c_h^{LP} + \sum_{i=1}^{L_y} \gamma_i^{h,LP} \Delta y_{t-i} + \sum_{i=0}^{L_z} \beta_i^{h,LP} z_{t-i} + r_{t+h|t-1}, \tag{2.8}$$

where $\hat{\beta}_0^{h,LP}$ is an estimator of $\psi_{z,h}$ (Jordà, 2005). Note that, in contrast to equation (2.5), equation (2.8) includes the current value of z_t. Since we can estimate σ_z^2 directly from the time series of z, we can estimate $\left(\sum_{i=0}^{h} \psi_{z,i}^2\right) \sigma_z^2$ in equation (2.7) or (2.7'). For the denominator in equation (2.7), we note that the residual in equation (2.8) can be related to the forecast error $f_{t+h|t-1}$ in equation (2.5). By comparing equations (2.5) and (2.8), it becomes clear that $f_{t+h|t-1} = \beta_0^{h,LP} z_t + r_{t+h|t-1}$ for each h. Therefore, we can construct estimates of the forecast errors, denoted by $\hat{f}_{t+h|t-1}^{LP}$, by adding $\hat{\beta}_0^{h,LP} z_t$ to $\hat{r}_{t+h|t-1}$. Then we can compute $\widehat{Var}(f_{t+h|t-1}^{LP})$ that is an estimate of the denominator in equation (2.7), where $\widehat{Var}(\cdot)$ denotes a sample variance. We now define a local projection estimator of FEVDs, which we call "LP-A" estimators, as

$$\hat{s}_h^{LPA} = \frac{\left(\sum_{i=0}^{h} \left\{\hat{\beta}_0^{i,LP}\right\}^2\right) \hat{\sigma}_z^2}{\widehat{Var}\left(\hat{\beta}_0^{h,LP} z_t + \hat{r}_{t+h|t-1}\right)}, \tag{2.9}$$

where $\hat{\sigma}_z^2 \equiv \widehat{Var}(z_t)$.

Although simple, the LP-A estimator does not guarantee that the estimated s_h is between 0 and 1. A simple solution to this issue is to split the denominator into variation due to z and due to v so that $\left(\sum_{i=0}^{h} \psi_{z,i}^2\right) \sigma_z^2$ appears in both the numerator and the denominator as in equation (2.7'). Note that

$$\begin{aligned} \hat{v}_{t+h|t-1} &= \hat{f}_{t+h|t-1}^{LP} - \hat{\beta}_0^{h,LP} z_t - \hat{\beta}^{h-1,LP} z_{t+1} - \cdots - \hat{\beta}_0^{0,LP} z_{t+h} \\ &= \hat{r}_{t+h|t-1} - \hat{\beta}_0^{h-1,LP} z_{t+1} - \cdots - \hat{\beta}_0^{0,LP} z_{t+h}, \end{aligned}$$

and that $\widehat{VAR}\left(\hat{r}_{t+h|t-1} - \hat{\beta}_0^{h-1,LP} z_{t+1} - \cdots - \hat{\beta}_0^{0,LP} z_{t+h}\right)$ is an estimate of $Var(v_{t+h|t-1})$. We use this quantity to define another local projection estimator of FEVDs, or “LP-B”:

$$\hat{s}_h^{LPB} = \frac{\left(\sum_{i=0}^{h}\left\{\hat{\beta}_0^{i,LP}\right\}^2\right)\hat{\sigma}_z^2}{\left(\sum_{i=0}^{h}\left\{\hat{\beta}_0^{i,LP}\right\}^2\right)\hat{\sigma}_z^2 + \widehat{VAR}\left(\hat{r}_{t+h|t-1} - \hat{\beta}_0^{h-1,LP} z_{t+1} - \cdots - \hat{\beta}_0^{0,LP} z_{t+h}\right)}. \tag{2.10}$$

The LP-A and LP-B estimators are based on a single regression (2.8) for each horizon, while the R2 estimator requires two regressions (2.5) and (2.6). While the LP-A and the LP-B are in some sense simpler (they estimate only one equation and they correspond more closely to the conventional way to compute FEVD, that is, use squares of estimated impulse responses to compute variance contributions), we find that the R2 estimator has weakly better finite-sample performances. To preserve space, we focus on the R2 estimator in the rest of the paper and relegate the details for the LP-A and LP-B estimators to Appendix B.

2.3.2 Asymptotics

To derive the asymptotic properties of our R2 estimator, we begin with the case where the forecast errors are observable, not generated. Then we show that using the estimated forecast errors does not alter the asymptotic distribution. Readers more interested in the implementation of the estimator may want to skip to the next subsection.

For now, we suppose that $f_h = (f_{T|T-h-1}, f_{T-1|T-h-2}, \ldots, f_{L_{max}+h+1|L_{max}})'$ is observable for any $h \geq 0$, where $L_{max} = max\{L_z, L_y\}$. We write $Z_t^h = (z_{t+h}, \ldots, z_t)'$ for all t and h and define a matrix $\mathbf{Z}_h = (Z_{T-h}^h, Z_{T-1}^h, \ldots, Z_{L_{max}+1}^h)'$. The (non-centered) R^2 of the regression of $f_{t+h|t-1}$ on Z_t^h is given by $(f_h' P_{\mathbf{Z}_h} f_h)/(f_h' f_h)$, where $P_{\mathbf{Z}_h} = \mathbf{Z}_h(\mathbf{Z}_h'\mathbf{Z}_h)^{-1}\mathbf{Z}_h'$.

Let $\theta_0 = \left(\theta_{1,0}', \theta_{2,0}', \theta_{3,0}'\right)'$, where $\theta_{1,0} = \left(E\left[Z_t^h Z_t^{h'}\right]\right)^{-1}\left(E\left[Z_t^h f_{t+h|t-1}\right]\right) = (\psi_{z,0}, \psi_{z,1}, \ldots, \psi_{z,h})'$, $\theta_{2,0} = E\left[Z_t^h f_{t+h|t-1}\right] = \theta_{1,0}\sigma_z^2$, and $\theta_{3,0} = E\left[f_{t+h|t-1}^2\right] \equiv \sigma_{f,h}^2$. A method of moments estimator $\hat{\theta} = \left(\hat{\theta}_1', \hat{\theta}_2', \hat{\theta}_3'\right)'$ is as follows: $\hat{\theta}_1 = (\mathbf{Z}_h'\mathbf{Z}_h)^{-1}(\mathbf{Z}_h' f_h)$, $\hat{\theta}_2 = \frac{\mathbf{Z}_h' f_h}{T_h}$, $\hat{\theta}_3 = \frac{f_h' f_h}{T_h}$, where $T_h = T - (L_{max} + h)$. For $\xi(\theta) = \xi(\theta_1, \theta_2, \theta_3) = \frac{\theta_2'\theta_1}{\theta_3}$, we have $s_h = \xi(\theta_0)$ and $\frac{f_h' P_{\mathbf{Z}_h} f_h}{f_h' f_h} = \xi\left(\hat{\theta}\right)$. Therefore, we first derive the asymptotic distribution of $\sqrt{T}\left(\hat{\theta} - \theta_0\right)$ and then apply the delta method to obtain the asymptotic distribution of $\sqrt{T}\left(\xi\left(\hat{\theta}\right) - \xi(\theta_0)\right) = \sqrt{T}\left(\frac{f_h' P_{\mathbf{Z}_h} f_h}{f_h' f_h} - s_h\right)$.

The moment conditions above can be summarized as $E\left[g_{t+h}(\theta)\right]=0$, where

$$g_{t+h}(\theta)\equiv g\left(f_{t+h|t-1},Z_t^h,\theta\right)=\begin{pmatrix} Z_t^h\left(f_{t+h|t-1}-(Z_t^h)'\theta_1\right)\\ Z_t^h f_{t+h|t-1}-\theta_2\\ f_{t+h|t-1}^2-\theta_3\end{pmatrix}. \tag{2.11}$$

It is clear that the conditions are satisfied only when $\theta=\theta_0$ and the system is just-identified. Therefore, $\sqrt{T}\left(\hat{\theta}-\theta_0\right)\xrightarrow{d}\mathcal{N}\left(0,(G_{h,R^2})^{-1}\Omega_{h,R^2}\left(G'_{h,R^2}\right)^{-1}\right)$, where '$\xrightarrow{d}$' denotes convergence in distribution, $G_{h,R^2}=E\left[\nabla_\theta g_{t+h}(\theta_0)\right]$, $\Omega_{h,R^2}=\sum_{l=-\infty}^{\infty}\Gamma(l)$, and $\Gamma(l)$ is the autocovariance of $g_{t+h}(\theta_0)$ at lag l (Hansen, 1982). With some algebra, we can further show that $ghr=-diag\left(\sigma_z^2 I_{h+1},I_{h+2}\right)$ where $diag(A,B)$ is the block diagonal matrix whose diagonal components are A and B in order, and I_h is the h-dimensional identity matrix .

Now we define Δ_{h,R^2} as $\frac{\partial\xi(\theta_0)}{\partial\theta'}=\frac{1}{\theta_{3,0}}\left(\theta'_{2,0},\theta'_{1,0},-s_h\right)$. By combining the above derivations, we derive the asymptotic distribution of $\frac{f_h' P_{\mathbf{Z}_h} f_h}{f_h' f_h}$.

Proposition 2.1. *Let* $f_h=\left(f_{T|T-h-1},f_{T-1|T-h-2},\ldots,f_{L_{max}+h+1|L_{max}}\right)'$ *and* $\mathbf{Z}_h=(Z_{T-h}^h,Z_{T-1}^h,\ldots,Z_{L_{max}+1}^h)'$ *for all* $h\geq 0$*, where* $L_{max}=max\{L_z,L_y\}$*. The* R^2 *of the regression of* $f_{t+h|t-1}$ *on* Z_t^h *is given by* $(f_h' P_{\mathbf{Z}_h} f_h)/(f_h' f_h)$*, where* $P_{\mathbf{Z}_h}=\mathbf{Z}_h(\mathbf{Z}_h'\mathbf{Z}_h)^{-1}\mathbf{Z}_h'$*. Furthermore, the following holds:*

$$\sqrt{T}\left(\frac{f_h' P_{\mathbf{Z}_h} f_h}{f_h' f_h}-s_h\right)\xrightarrow{d}\mathcal{N}\left(0,\Delta_{h,R^2}\left(G_{h,R^2}\right)^{-1}\Omega_{h,R^2}\left(G'_{h,R^2}\right)^{-1}\Delta'_{h,R^2}\right), \tag{2.12}$$

where $\Delta_{h,R^2}=\frac{1}{\sigma_{f,h}^2}\left(\psi_{z,0}\sigma_z^2,\ldots,\psi_{z,h}\sigma_z^2,\psi_{z,0},\ldots,\psi_{z,h},-s_h\right)$*,* $G_{h,R^2}=-diag(\sigma_z^2 I_{h+1},I_{h+2})$*, and* Ω_{h,R^2} *is the long-run variance of* $g_{t+h}(\theta_0)$ *in equation* (2.11)*. We denote the variance in equation* (2.12) *by* $V_{h,R^2}=\Delta_{h,R^2}\left(G_{h,R^2}\right)^{-1}\Omega_{h,R^2}\left(G'_{h,R^2}\right)^{-1}\Delta'_{h,R^2}$*.*

However, f_h is not directly observable in practice. We use its estimate $\hat{f}_h$ instead, which is based on equation (2.5). Next, we show that the feasible estimator $\frac{\hat{f}_h' P_{\mathbf{Z}_h}\hat{f}_h}{\hat{f}_h'\hat{f}_h}$ has the same asymptotic variance V_{h,R^2} in Proposition 1.

To separate issues from truncation and estimation of the forecast errors, we now assume that L_y and L_z are large enough, and the population residual of equation (2.5) is the true forecast error. In other words, we assume that $(z_t,\Delta y_t)'$ follows a finite-order Markov process and focus on the variability in $\hat{f}_h$ due to the estimation of the forecast errors.

For a simple notation, we rewrite equation (2.5) as $y_{t+h}-y_{t-1}=W'_{t-1}\phi+f_{t+h|t-1}$, where $W_{t-1}\equiv\left(1,\Delta y_{t-1},\ldots,\Delta y_{t-L_y},z_{t-1},\ldots,z_{t-L_z}\right)'$ and $\phi\equiv\left(c_h,\gamma_1^h,\ldots,\gamma_{L_y}^h,\beta_1^h,\ldots,\beta_{L_z}^h\right)'$. For $\hat{\phi}$

being the OLS estimator of ϕ, we have $\hat{f}_{t+h|t-1} = f_{t+h|t-1} - W'_{t-1}\left(\hat{\phi} - \phi\right)$. By stacking up and defining $\mathbf{W}$ matrix accordingly, we obtain $\hat{f}_h = f_h - \mathbf{W}\left(\hat{\phi} - \phi\right)$.

A feasible estimator $\tilde{\theta}$ of θ_0 based on $\hat{f}_h$ is given by $\tilde{\theta}_1 = \left(\mathbf{Z}'_h\mathbf{Z}_h\right)^{-1}\left(\mathbf{Z}'_h\hat{f}_h\right)$, $\tilde{\theta}_2 = \frac{\mathbf{Z}'_h\hat{f}_h}{T_h}$, and $\tilde{\theta}_3 = \frac{\hat{f}'_h\hat{f}_h}{T_h}$. We will show that $\tilde{\theta} = \hat{\theta} + O_p\left(\frac{1}{T_h}\right)$, and therefore the feasible estimator $\tilde{\theta}$ converges to the infeasible estimator $\hat{\theta}$ fast enough not to change the asymptotic distribution of $\sqrt{T}\left(\tilde{\theta} - \theta\right)$, and more specifically, the asymptotic variance. Note that $\tilde{\theta}_1 = \left(\frac{\mathbf{Z}'_h\mathbf{Z}_h}{T}\right)^{-1}\left(\frac{\mathbf{Z}'_h\hat{f}_h}{T}\right) = \hat{\theta}_1 - \left(\frac{\mathbf{Z}'_h\mathbf{Z}_h}{T}\right)^{-1}\left(\frac{\mathbf{Z}'_h\mathbf{W}}{T}\right)\left(\hat{\phi} - \phi\right) = \hat{\theta}_1 - O_p(1)\left(E\left[Z^h_t W'_{t-1}\right] + O_p\left(\frac{1}{\sqrt{T}}\right)\right) O_p\left(\frac{1}{\sqrt{T}}\right)$, which follows from the law of large numbers, the central limit theorem, and standard asymptotics of OLS estimators. Because $W_{t-1} \in \Omega_{t-1}$, $Z^h_t = (z_{t+h}, \ldots, z_t)'$ is orthogonal to W_{t-1}. In other words, $E\left[Z^h_t W'_{t-1}\right] = 0$. Thus, $\tilde{\theta}_1 = \hat{\theta}_1 - O_p(1)O_p\left(\frac{1}{\sqrt{T}}\right)O_p\left(\frac{1}{\sqrt{T}}\right) = \hat{\theta}_1 + O_p\left(\frac{1}{T}\right)$. One can similarly show that $\tilde{\theta}_2 = \hat{\theta}_2 + O_p\left(\frac{1}{T}\right)$ and $\tilde{\theta}_3 = \hat{\theta}_3 + O_p\left(\frac{1}{T}\right)$ using $E\left[Z^h_t W'_{t-1}\right] = 0$ and $E\left[f_{t+h|t-1}W'_{t-1}\right] = 0$. We summarize these results in the following proposition.

Proposition 2.2. *Suppose that $(z_t, \Delta y_t)'$ follows a finite-order Markov process, and therefore the true residual in equation 2.5 coincides with the population forecast error for large enough L_y and L_z. In this case, the feasible R^2 estimator has the same asymptotic distribution as the infeasible estimator in Proposition 1. That is,*

$$\sqrt{T}\left(\frac{\hat{f}'_h P_{\mathbf{Z}_h}\hat{f}_h}{\hat{f}'_h\hat{f}_h} - s_h\right) = \sqrt{T}\left(\frac{f'_h P_{\mathbf{Z}_h} f_h}{f'_h f_h} - s_h\right) + o_p(1) \xrightarrow{d} \mathcal{N}(0, V_{h,R^2}), \tag{2.13}$$

where $V_{h,R^2} = \Delta_{h,R^2}\left(G_{h,R^2}\right)^{-1}\Omega_{h,R^2}\left(G'_{h,R^2}\right)^{-1}\Delta'_{h,R^2}$ is the asymptotic variance in Proposition 1.

2.3.3 Measurement errors

Empirically identified shocks z_t could be measured with errors since e.g., these shocks are often estimates rather than direct observations. One may handle this issue by considering noisy measures of underlying structural shocks as external instruments as is the case in Plagborg-Møller and Wolf (2018*a*) who derive partial-identification results and set-identified s_h.

Our approach is different. Given measurement errors, we show in Appendix D that asymptotic biases of our estimators are negative. Therefore, our methods underestimate the true s_h without further refinements to tackle measurement errors. Furthermore, although shocks are often estimated and thus are generated regressors, the researcher is often interested in testing the null of no responses (i.e., $s_h = 0$), and there is no need to adjust inference for this exercise (Pagan, 1984).

Specifically, we can decompose the true shock into two parts as $z_t = z_t^o + z_t^u$, where superscripts o and u denote observable and unobservable components, respectively. We assume that $\{(z_t^o, z_t^u)'\}$ is a white noise process with $\sigma_o^2 = Var(z_t^o)$, $\sigma_u^2 = Var(z_t^u)$, and $\rho_{o,u} = corr(z_t^o, z_t^u)$. For example, a measurement error m_t can be modelled as $z_t^o = z_t + m_t$ and $z_t^u = -m_t$, and so $\rho_{o,u} < 0$. Denote the full information set with $\Omega_{t-1} = \{z_{t-1}^o, z_{t-1}^u, \Delta y_{t-1}, \dots\}$ for now and the econometrician's information set with $\Omega_{t-1}^e \equiv \{z_{t-1}^o, \Delta y_{t-1}, \dots\}$. The econometrician's forecast error $f_{t+h|t-1}^e$ is given by $f_{t+h|t-1}^e = y_{t+h} - y_{t-1} - P\left[y_{t+h} - y_{t-1} | \Omega_{t-1}^e\right]$. Note that we project $y_{t+h} - y_{t-1}$ on Ω_{t-1}^e, while the full-information forecast error $f_{t+h|t-1}$ is based on Ω_{t-1}. Finally, the econometrician's regressor is denoted by $Z_t^{h,e} = \left(z_{t+h}^o, \dots, z_t^o\right)'$.

Proposition 2.3. *Given the assumptions above, the followings hold for any* $|\rho_{o,u}| \leq 1$.

(a) $Var\left(f_{t+h|t-1}^e\right) \geq Var\left(f_{t+h|t-1}\right)$.

(b) $Var\left(\psi_{z,0} z_{t+h} + \dots + \psi_{z,h} z_t\right)$
$= \mathbf{Cov}\left(f_{t+h|t-1}^e, Z_t^{h,e}\right)\left[\mathbf{Var}\left(Z_t^{h,e}\right)\right]^{-1}\mathbf{Cov}\left(Z_t^{h,e}, f_{t+h|t-1}^e\right) + \sum_{i=0}^h \psi_{z,i}^2 \left(1 - \rho_{o,u}^2\right)\sigma_u^2$.

(c) $s_h = \frac{Var\left(\psi_{z,0} z_{t+h} + \dots + \psi_{z,h} z_t\right)}{Var\left(f_{t+h|t-1}\right)} \geq \frac{\mathbf{Cov}\left(f_{t+h|t-1}^e, Z_t^{h,e}\right)\left[\mathbf{Var}\left(Z_t^{h,e}\right)\right]^{-1}\mathbf{Cov}\left(Z_t^{h,e}, f_{t+h|t-1}^e\right)}{Var\left(f_{t+h|t-1}^e\right)}$.

For a formal proof, please see Appendix D. Proposition 3(a) covers the forecast error variance, which is the denominator of s_h in equation (4). The result implies that the econometrician's forecast error variance is greater than that based on the full information set. Furthermore, one can show that the equality holds only for (uninteresting) special cases such as $\psi_z(L) = 0$, $\rho_{o,u} = \pm 1$, and $\sigma_u^2 = 0$. We discuss the numerator of s_h in Proposition 3(b). When estimated without taking z_t^u into consideration, the econometrician's numerator in equation (2.4) is less than that under the full information set by $\sum_{i=0}^h \psi_{z,i}^2 \left(1 - \rho_{o,u}^2\right)\sigma_u^2$. Similarly, the difference reduces when $\psi_{z,i}'$s are close to 0, when the observable component and the unobservable component are highly correlated, and when the variance of the unobservable component σ_u^2 is small. Because the econometrician's denominator is greater and the numerator is less than those based on the full information set, the econometrician's FEVDs are downward biased to zero as illustrated in Proposition 3(c). In other words, our point estimate is conservative in favor of the hypothesis $s_h = 0$.

2.3.4 Small-sample refinements

While $\hat{s}_h^{R2}$ is asymptotically unbiased as illustrated in Proposition 1 and 2, there may exist substantial finite-sample biases. Note that the OLS estimator in equation (2.6) is obtained by maximizing the sum of explained variation, or R^2, which may lead to an upward bias in $\hat{s}_h^{R2}$ (Cramer, 1987).

To correct for potential small-sample biases in the estimates of s_h and to enhance coverage rates for confidence bands, we employ a VAR-based bootstrap, where the VAR includes two variables $(z_t, \Delta y_t)'$. We use a VAR-based bootstrap to address challenges associated with bootstrapping highly persistent data but researches may utilize alternative approaches.[4] [5]

We now discuss the details of the bootstrap procedure. First, we need to choose the order of the VAR model L_{VAR}. In simulations below, we rely on the Hannan-Quinn information criterion (HQIC) for the purpose. We simulate the estimated $VAR\,(L_{VAR})$ model $Y_t = \hat{\mu} + \hat{\phi}_1 Y_{t-1} + \cdots + \hat{\phi}_{L_{VAR}} Y_{t-L_{VAR}} + \epsilon_t$ to generate artificial time series B times, where $Y_t \equiv (z_t, \Delta y_t)'$. And we use this model to compute s_h^*, the true contribution of z to the forecast error variance of y at the horizon h for this data generating process. For each $b \leq B$, we randomly choose t between $1 + L_{VAR}$ and T to initiate the simulation. Then $(z_t, \Delta y_t)', \ldots, (z_{t-L_{VAR}}, \Delta y_{t-L_{VAR}})'$ are used as $Y_0^{(b)}, \ldots, Y_{-L_{VAR}}^{(b)}$. Given the initial condition, we randomly draw $\left\{\epsilon_t^{(b)}\right\}$ from the estimated reduced form residuals $\{\hat{\epsilon}_t\}$ with replacement. Using the estimated model with the above initial conditions and the shuffled residuals, we obtain the simulated series $\left\{\left(z_t^{(b)}, \Delta y_t^{(b)}\right)'\right\}$, where the first T_{BurnIn} number of observations are discarded as burn-in. We apply our estimator to $\left\{\left(z_t^{(b)}, \Delta y_t^{(b)}\right)'\right\}$ and obtain the bootstrap estimate $\hat{s}_h^{R2,(b)}$ for each b. Then we estimate the bias in $\hat{s}_h^{R2}$ with $bias_h \equiv \frac{1}{B}\sum_{b=1}^{B} \hat{s}_h^{R2,(b)} - s_h^*$ and compute bias-corrected estimates $\hat{s}_h^{R2,BC} \equiv \hat{s}_h^{R2} - bias_h$. The procedure is similar for VARs.

[4] One may use alternative implementations of bootstrap to refine asymptotic inference. We tried the block bootstrap for local projections following Kilian and Kim (2011). However, this block bootstrap method performs worse than the VAR-based bootstrap in simulations. Results are in Appendix E1.

[5] Our bootstrap procedure implicitly assumes homoscedasticity of shocks. If a researcher suspects important heteroskedasticity in shocks, one should use alternative bootstrap methods (e.g., Gonçalves and Kilian, 2004). An extensive discussion of practical considerations for various bootstrap methods is in Kilian and Lütkepohl (2017, Ch. 12).

2.3.5 Standard errors and confidence intervals

We have several options to construct standard errors and confidence intervals. For example, one may directly estimate V_{h,R^2} in equations (2.12) and (2.13) and derive a symmetric confidence interval based on the estimated vhr (see Appendix A for details including implementation of pre-whitening following Andrews and Monahan, 1992). While this works asymptotically, its finite-sample performance is not better than bootstrap confidence intervals as discussed in Appendix E3. Furthermore, the estimated standard errors are often spiky across h's, which induce non-smooth and erratic confidence bands.

Therefore, we employ a different approach for the simulations and the application in this paper. To study finite-sample properties of our estimator, we rely on the distribution of the bootstrap estimates $\hat{s}_h^{R2,(b)}$. The standard error can be easily obtained from a standard deviation of $\hat{s}_h^{R2,(b)}$ across B replications. Constructing a symmetric confidence interval is also straightforward. On the other hand, one may want to take the shape of the bootstrap distribution into consideration when constructing confidence intervals. Let $\hat{q}_{h,\alpha/2}^{R2}$ and $\hat{q}_{h,1-\alpha/2}^{R2}$ refer to the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles of the distribution of $\hat{s}_h^{R2,(b)} - \frac{1}{B}\sum_{b=1}^{B}\hat{s}_h^{R2,(b)}$. Then the $100(1-\alpha)\%$ confidence interval is given by $\left[\hat{q}_{h,\alpha/2}^{R2} + \hat{s}_h^{R2,BC}, \hat{q}_{h,1-\alpha/2}^{R2} + \hat{s}_h^{R2,BC}\right]$. Note that we consider the distribution of $\hat{s}_h^{R2,(b)} - \frac{1}{B}\sum_{b=1}^{B}\hat{s}_h^{R2,(b)}$ to make the confidence interval centered around the estimated FEVD with bias-correction.

2.3.6 Extension

While our analysis has focused on the bivariate case, this framework can be readily generalized to include more controls in equation (2.5):

$$y_{t+h} - y_{t-1} = \sum_{i=1}^{L_z}\beta_i^h z_{t-i} + \sum_{i=1}^{L_C} C_{t-i}'\Gamma_i^h + f_{t+h|t-1}, \tag{2.14}$$

where C_t is the vector of control variables which may include lags of additional variables and structural shocks other than z_t. In the base case, C_t consists only of Δy_t. Note that for VAR-based bootstraps, one has to include z_t and all variables in C_t to simulate data.[6]

[6] As the number of variables in C_t increases, the number of parameters in the VAR increases rapidly. When C_t is a large vector, or when a VAR is not a good representation of the DGP for control variables, VAR-based bootstrap might not be an appealing option. In this case, one may consider other forms of bootstrap (e.g., block bootstrap). Alternatively, one may correct for biases by simulating asymptotic distributions of primitive quantities in equations (2.3), (2.7), and (2.7') such as $\hat{\psi}_{z,i}$, $\hat{\sigma}_z^2$, and $\widehat{Var}\left(\hat{v}_{t+h|t-1}\right)$. By considering

One should bear in mind that, although including or excluding C_t or changing the composition of variables in C_t should make little difference of impulse responses estimated with local projections (provided z_t is uncorrelated with other shocks), what goes in C_t is potentially important for FEVDs. Intuitively, by including more controls in C_t (that is, information set Ω_t expands), we (weakly) reduce the size of the forecast error, and hence the amount of variation to be explained shrinks. In other words, the regressand in equation (2.6) and therefore s_h change with the list of variables in C_t. Thus, one should not be surprised to observe that the share of variation explained by shocks $\{z_t, \ldots, z_{t+h}\}$ may be sensitive to C_t.

Similar to the simple case considered in Section 2.3.1, for equation (2.6), one may want to use residuals from projecting z_t on lags of z_t and C_t rather than the "raw" shock z_t. For example, when the Cholesky orderings are an identifying assumption, such a procedure is essential to guarantee that forecastable movements in $z_t, \ldots, z_{t+h}$ are not used to account for variation in $\hat{f}_{t+h|t-1}$. In practice, however, shocks z_t are constructed in ways to ensure that z_t is not predictable by current values and lags of macroeconomic variables. As a result, we find in our simulations and applications that purifying structural shocks in this manner makes little difference.

2.3.7 Taking stock: A cookbook for FEVDs

To summarize our discussion so far, we suggest that the researcher should take the following steps to estimate FEVDs:

Step 1 Estimate the forecast errors for the horizon h from local projections (2.5) or (2.14) depending on the information set.

Step 2 Regress the estimated forecast errors on the shocks from t to $t+h$ as in equation (2.6). The R^2 of this regression measures the share of the forecast error variance explained by the shock at the horizon h.

Step 3 To improve the small-sample performance of the estimator, a bias-correction step is recommended for empirically relevant sample sizes. One may rely on a VAR-based bootstrap to do so, where the lag order can be selected via an information criterion.

s_h as a non-linear function of those parameters, such simulations would detect biases due to the non-linearity. See Appendices A and B for implementation and F and G for the results.

Step 4 For inference, we can calculate the standard error from either the analytical expression for the asymptotic variance in equation (2.12) or the distribution of the bootstrap estimates in Step 3. Similarly, we may construct the confidence interval by using either the standard error or the quantiles of the bootstrap distribution. We recommend using bootstrap to construct confidence bands, but one may choose a different approach depending on the data generating process and the sample size.

2.4 Simulations

This section presents two sets of simulations. The first set shows results for the baseline bivariate case and studies the performances of R2 methods and VARs for various profiles of the contribution of z to the forecast error variance of y at different horizons. The second set uses the estimated Smets and Wouters (2007) model to investigate the performance in a setting with many control variables.

For each data generating process (DGP), we simulate data 2,000 times. When we employ bootstrap to correct for biases, the number of bootstrap replications is set to $B = 2,000$ and $T_{BurnIn} = 100$. As a benchmark, we also report results based on a corresponding VAR. This benchmark corresponds to the practice of including shocks into VARs directly (e.g., Barakchian and Crowe, 2013; Basu, Fernald and Kimball, 2006; Ramey, 2011; Romer and Romer, 2004, 2010). For the simulations below, we order z_t as the first variable in VARs as is the case in Section 2.3.4. We choose the Hannan-Quinn information criterion (HQIC) as our benchmark criterion to determine the number of lags in VAR. To make VAR and LP models comparable, we use HQIC number of lags in the VAR for L_z and L_y (Plagborg-Møller and Wolf, 2018*b*). Results are similar when we use higher-order VARs, where the lag order is selected by Akaike information criterion instead of HQIC (Appendix E2).

The sample size for simulated data is $T = 160$, which is common in applied macroeconomic analyses. Results for other sample sizes are reported in Appendices F and G. The coverage rates are calculated as $Pr\left(\hat{q}^{R2}_{h,\alpha/2} + \hat{s}^{R2,BC}_h \leq s_h \leq \hat{q}^{R2}_{h,1-\alpha/2} + \hat{s}^{R2,BC}_h\right)$ where $\alpha = 0.1$, and therefore the nominal coverage rate is 90%.[7]

[7]We also considered percentile-t bootstrap and found similar results.

2.4.1 Bivariate Data Generating Processes

We study three data generating processes (DGPs) to cover different shapes of s_h. The basic structure is as follows:

$$
\begin{aligned}
y_t &= \psi_z(L)z_t + u_t, & z_t &\sim iid\ \mathcal{N}(0,\sigma_z^2),\\
u_t &= p_t + a_t, & &\\
(\Delta p_t - g_y) &= \rho_p(\Delta p_{t-1} - g_y) + e_t^p, & e_t^p &\sim iid\ \mathcal{N}(0,\sigma_p^2),\\
a_t &= \rho_a a_{t-1} + e_t^a, & e_t^a &\sim iid\ \mathcal{N}(0,\sigma_a^2),
\end{aligned}
$$

where z, e^p and e^a are mutually independent. p and a are permanent and transitory components of u. To find the value of s_h based on $\Omega_{t-1} = \{\Delta y_{t-1}, z_{t-1}, \Delta y_{t-2}, z_{t-2}, \dots\}$, we need to find the population $MA(\infty)$ representation of $\Delta u_t = g_y + \sum_{i=0}^{\infty} \psi_{e,i} e_{t-i}$, where $\{e_t\}$ is a zero mean white-noise series with variance σ_e^2, $\sum_{i=0}^{\infty} \psi_{e,i}^2 < \infty$, and $e_t \in \Omega_t$. We assume that $\psi_{e,0} = 1$ without loss of generality, and the Wold Decomposition implies that such representation exists uniquely. Because z_t and e_t are uncorrelated at all leads and lags, we can write s_h in equation (2.3) in terms of $\{\psi_{z,i}\}$, $\{\psi_{e,i}\}$, σ_z^2, and σ_e^2. Appendix C discusses how one can use a Kalman filter to derive $\{\psi_{e,i}\}$ and σ_e^2 from ρ_p, ρ_a, σ_p^2, and σ_a^2.

DGP1 is characterized by hump-shaped ψ_z and s_h. We assume that $\{\psi_z(L)z_t\}$ follows an $MA(100)$ process with the maximum response of 3 after 8 periods.[8] The resulting profile of s_h is consistent with e.g., predictions about how monetary shocks contribute to variation in output: there is little to no response of output in the short-run due to various rigidities, then the response is strong in the medium-run, and the long-run response is zero due to nominal neutrality (e.g., Christiano, Eichenbaum and Evans, 2005). DGP2 has a strong response of y to z only in the short-run, and thus the shape of s_h is downward-sloping. This profile is consistent with e.g., how temporary fiscal shocks influence output: the effect of a government spending increase or a tax cut is large on impact but then the effect gradually wears out (e.g., Smets and Wouters, 2007). Finally, DGP3 assumes $\lim_{h\to\infty} \psi_{z,h} > 0$, so that z has persistent effects on y and the shape of s_h is upward-sloping. This profile is consistent with e.g., models emphasizing that technology shocks are a key (or even exclusive) source of variation in output at long horizons (e.g., Blanchard and Quah, 1989). Table 2.1 reports

[8]This value and pattern are motivated by a 3 percent response of real GDP to a 100bp monetary policy shock estimated in Coibion (2012).

parameter values for each DGP. Figure 2.1 plots true impulse responses of y to z (Panel A) and the contribution of z to forecast error variances of y at different horizons (Panel B).

For DGP1, we find (Table 2.2 that local projections capture the hump-shaped impulse response correctly but $\hat{s}_h^{R2}$ without bias-correction fails to match the hump-share dynamics of s_h: $\hat{s}_h^{R2}$ tends to monotonically increase with the horizon. When we use a VAR to estimate impulse responses and FEVDs, the VAR misses the hump both in the impulse response and FEVDs as HQIC selects too few lags (on average the number of lags is 1.24). Confidence bands yield poor coverage rates. This performance reflects the fact that, by construction, z contributes little to the forecast error variation in y for this DGP at short horizons with $h \leq 4$. Since s_h is between zero and one, we effectively have estimates close to the boundary, and therefore standard methods are likely to fail. While bootstrap appears to provide some improvement (e.g., the bias at long horizons as $h \geq 12$ when z accounts for a larger share of the forecast error variance in y is corrected),[9] it does not perform consistently better in terms of the coverage rates because the parameter is at the boundary. When we allow z to explain 5 percent or more of the forecast error variance in y at short horizons, bootstrap brings coverage rates close to nominal (results are available upon request). Note that, although the VAR estimators $\left(\hat{s}_h^{VAR}\right)$ are strongly biased, they tend to have smaller variances so that the root mean squared error (RMSE) is similar in magnitude to that of the $\hat{s}^{R2}$ estimator. The large RMSEs underscore difficulties in estimating R^2 (Cramer, 1987) and hence s_h.

Because DGP2 permits an exact, finite-order VAR representation,[10] $\hat{s}_h^{VAR}$ has good properties in terms of bias, RMSE, and coverage rates (Table 2.3). The local projections recover the impulse responses properly, but the estimates of FEVDs again overstate the contribution of z to the unforecasted variation in y at long horizons as $h \geq 12$. Note that bootstrap can correct for this bias. Given that the VAR nests the DGP and that the VAR is more parsimonious than local projections, the VAR has a better performance than the $\hat{s}^{R2}$ estimators.

In the case of DGP3, z has long-lasting effects on y and the VAR underestimates the responses at long horizons as $h \geq 16$ in small samples. Impulse responses estimated with local projections perform better but also exhibit a downward bias at long horizons. In a similar spirit, $\hat{s}_h^{VAR}$ shows a strong downward bias and $\hat{s}_h^{R2}$ is downward biased by a smaller, but still considerable amount (this is the case even after we use bootstrap to correct for possible

[9]The bias can be further reduced by using higher values of L_z and L_y by reducing errors in $\hat{f}_{t+h|t-1}$ due to the truncation.

[10]Given the parameter values in Table 2.1, $\Delta y_t = g_y + (1-L)(1-0.9L)^{-1} z_t + (1-0.9L)^{-1} e_t^p$. By pre-multiplying $(1-0.9L)$, we have $\Delta y_t = 0.1 g_y + 0.9 \Delta y_{t-1} - z_{t-1} + z_t + e_t^p$.

biases). This performance reflects the fact that HQIC chooses a low number of lags (1.29 lags on average across simulations). As a result, VARs used to simulate bootstrap samples fail to capture the degree of persistence in the data. To demonstrate the importance of the lag order, we report results (Table 2.4) when we use VAR(5) and VAR(10) for bootstrap. As the number of lags increases, we observe some improvement (e.g., the remaining bias in the bias-corrected $\hat{s}_h^{R2}$ is smaller for VAR(10) than VAR(5)), but these enhancements are achieved at the price of higher variances in the estimates (e.g., the RMSEs of the bias-corrected $\hat{s}_h^{R2}$ are similar for both VARs used for bootstrap). These results suggest that one may want to overfit VAR for persistent processes at the bootstrap stage.

In summary, we find for small samples that estimating s_h precisely is not easy. Nonetheless, we also note that the $\hat{s}_h^{R2}$ estimator performs reasonably well across the DGPs and that bootstrap helps to improve the estimator's properties. In contrast, VARs that include structural shocks z_t tend to perform poorly when a DGP is not nested in a small-order VAR.

2.4.2 Smets-Wouters model

While the bivariate DGPs provide important insights on how the R^2 estimator performs, researchers face potentially more complex DGPs and often have more information in practice. In this section, we use the Smets and Wouters (2007) model to study the performance of our estimator in an environment with multiple shocks and many control variables.

As discussed above, different information sets determine different population s_h. In the simulations, we assume that the researcher is interested in explaining variation in output and that the researcher observes output growth rate, inflation, federal funds rate, and monetary policy shocks.[11] This choice of variables is motivated by the popularity of small VARs which include output, inflation, and a policy rate to study the effects of monetary policies on the economy. In this exercise, the shock is ordered first because the Smets-Wouters model allows contemporaneous responses of macroeconomic variables to policy shocks. When estimating impulse responses using local projections, we augment equation (2.14) with $\psi_{z,h} z_t$ on the

[11] For this information set, we construct the true FEVD using a stationary Kalman filter similar to the method in Appendix C. We also tried various combinations of shocks and endogenous variables in the information set and found similar results. Figures for inflation and results with large samples are in Appendix G. Note that monetary policy shocks are nearly invertible in the Smets-Wouters model (see Wolf, 2017, for mode details). While this may be a problem if we use shocks identified and recovered from a DSGE model, the spirit of our exercise is to assume that we have access to other information (as in e.g., Romer and Romer, 2004) so that we can observe monetary policy shocks directly.

right-hand side.

We find (Figure 2.2) that local projections correctly recover the responses of output to monetary policy shocks, while a low order VAR (lag length is chosen with HQIC) fails to capture the transitory effect of monetary shocks on output. Consistent with our bivariate analysis, $\hat{s}_h^{R2}$ increase with the horizon while the true s_h exhibits hump-shaped dynamics. s_h estimated with a VAR also fails to capture the true dynamics as $\hat{s}_h^{R2}$ flattens out after about $h = 5$. Similar to our results in the previous section, we find that bias correction helps $\hat{s}_h^{R2}$ to recover the true hump-shaped profile of s_h. Coverage rates are close to nominal at all horizons after bias-correction. Again, although the VAR estimator of s_h is strongly biased, the variance of the estimator is low so that RMSEs are broadly similar across methods. We conclude that our proposed methods to estimate FEVDs work reasonably well in more complex settings.

2.5 Application

To illustrate the properties of our estimators, we use two structural shocks identified in the literature. The first shock is the monetary policy (MP) innovation identified as in Romer and Romer (2004) and extended in Coibion et al. (2017). The second shock is the total factor productivity (TFP) change identified as in Fernald (2014).[12] The sample autocorrelations and the sample partial autocorrelations at non-zero lags are close to zero for both shocks, that is, the shocks are white noises. The correlation between the shocks is -0.059. Our objective is to quantify the contribution of these shocks to the variation of output and inflation. The sample covers 1969Q1-2008Q4 which excludes the period of binding zero lower bound. The set of variables for local projections includes inflation (annualized growth rate of GDP deflator, i.e., $400\Delta \ln(P_t)$), annual GDP growth rate ($400\Delta \ln(Y_t)$), federal funds rate, and both identified shocks. We set $L_C = L_z = 4$ in equation (2.14) and add control variables similarly when estimating impulse responses. In the benchmark VAR, we have all five variables and allow four lags.[13]

Consistent with previous studies, we find (Figures 2.3 and 2.4) that a contractionary

[12] Appendix H presents results for military spending shocks constructed in Ramey and Zubairy (2018).

[13] The ordering of variables in the VAR is TFP measure (from Fernald, 2014), output growth rate, inflation, monetary policy innovations (from Coibion et al., 2017), and fed funds rate. For the VAR-based analysis, we follow the practice and compute FEVDs using shocks in these variables where shocks are identified recursively from reduced-form residuals.

monetary policy shock lowers output and prices, and that a positive TFP shock raises output and lowers prices. Impulse responses estimated with a VAR and local projections are similar at horizons $h \leq L_{VAR} = 4$. However, the estimated impulse responses differ at longer horizons, and therefore the peak effects and the overall shapes are different. The VAR estimates of the FEVDs suggest that TFP (MP) shocks account for approximately 10 (3.5) percent of the forecast error variances of output at horizons longer than 2 years. For inflation, MP shocks contribute up to 19 percent of the variation in the forecast error of inflation at the 5-year horizon and little variation at shorter horizons while the contribution of TFP shocks is generally small. Bias-correction makes no material difference for the forecast error variance decomposition estimates for all cases but one: the bias-corrected estimate of the contribution of MP shocks to the variation in the forecast error of inflation at the 5-year horizon increases to 32 percent.

The local projections estimates of the contribution of the two shocks to the forecast error variances of output are much larger than the VAR estimates. Moreover, bias-correction tends to generate lower contributions, consistent with simulations. For example, monetary policy shocks account for 18 percent of the forecast error variance of output according to the R2 estimate (28 percent without bias-correction) and only 3.5 percent according to the VAR estimate at the 5-year horizon. Similarly, the VAR estimate of the contribution of MP shocks to inflation at the 5-year horizon is less than 20 percent, which is a surprising result given Milton Friedman's "inflation is always and everywhere a monetary phenomenon." In contrast, the R2 estimate of the same FEVD with bias-correction amounts to 44 percent. Also, while the profile of $\hat{s}_h^{VAR}$ for output is generally flat after $h = 5$, $\hat{s}_h^{R2}$ has richer dynamics. This is consistent with what we find in our simulations for DGP1: when the true s_h is close to zero for small h's, $\hat{s}_h^{VAR}$ fails to match the shape, while $\hat{s}_h^{R2}$ is much more successful. The profiles of $\hat{s}_h^{R2}$ and $\hat{s}_h^{VAR}$ for output also differ remarkably for TFP shocks. While $\hat{s}_h^{R2}$ increases in h, $\hat{s}_h^{VAR}$ flattens around 10 percent after $h = 10$. At the 5-year horizon, TFP shocks contribute to 28 percent of the forecast error variance of output based on the R2 estimate after bias-correction, where the VAR estimate without bias-correction is only 11 percent.

2.6 Concluding remarks

Single-equation methods can offer flexibility and parsimony that many economists seek. The increasing popularity of these methods, specifically the local projections, calls for further development of these tools. An important limitation for practitioners using this framework has been a lack of simple tools with well-known econometric properties especially in small samples to assess quantitative significance of a given set of shocks, that is, the contribution of the shocks to the forecast error variance of the variable of interest. We propose a method to provide such a metric. In a series of simulation exercises, we document that our method has good small-sample properties. We also show that conventional approaches to assess the quantitative significance of two popular structural shocks (monetary policy shocks and total factor productivity shocks) could have understated the importance of these two shocks.

Table 2.1: Parameter values for data generating processes (DGPs) used in simulations.

	$\psi_z(L)$	σ_z	g_y	ρ_p	σ_p	ρ_a	σ_a
DGP1	Hump-shaped	1	0.5	0.9	0.5	0.9	3
DGP2	$(1-0.9L)^{-1}$	3	0.5	0.9	1.5	-	-
DGP3	$(1-L)^{-1}(1-0.9L)^{-1}$	1	0.5	0.5	2	0.9	3

Table 2.2: Simulation results for DGP 1.

	Horizon h					
	0	4	8	12	16	20
Impulse Response						
True	0.00	1.39	3.00	2.06	0.88	0.29
Local projections	0.00	1.39	3.00	2.05	0.87	0.29
VAR(HQIC)	0.00	0.18	0.24	0.25	0.25	0.25
Forecast Error Variance Decomposition						
True	0.00	0.04	0.19	0.21	0.18	0.14
Average estimate						
R2	0.01	0.06	0.20	0.25	0.26	0.27
VAR(HQIC)	0.01	0.02	0.02	0.02	0.03	0.03
Root mean squared error						
R2	0.01	0.05	0.11	0.15	0.19	0.22
VAR(HQIC)	0.01	0.03	0.17	0.20	0.16	0.13
Coverage (90 % level, asymptotic)						
R2	0.99	0.81	0.69	0.65	0.63	0.61
VAR(HQIC)	0.99	0.75	0.06	0.06	0.07	0.10
Forecast Error Variance Decomposition (bias-corrected, VAR(HQIC))						
True	0.00	0.04	0.19	0.21	0.18	0.14
Average estimate						
R2	0.00	0.02	0.13	0.16	0.13	0.11
VAR(HQIC)	0.00	0.00	0.01	0.01	0.01	0.01
Root mean squared error						
R2	0.01	0.05	0.12	0.16	0.17	0.18
VAR(HQIC)	0.01	0.04	0.19	0.21	0.17	0.14
Coverage (90 % level, asymptotic)						
R2	0.99	0.95	0.64	0.64	0.72	0.81
VAR(HQIC)	1.00	0.53	0.06	0.05	0.07	0.09

Notes: The table reports the performance of estimators introduced in Section 2.3 for DGP1. The sample size is $T = 160$, and the number of simulations is 2,000. R2 and VAR stand for $\hat{s}_h^{R2}$ and $\hat{s}_h^{VAR}$ estimators of forecast error variance decompositions. The lag order is selected by the Hannan-Quinn information criterion (HQIC). Confidence intervals for the bias-corrected R2 estimator are given by $\left[\hat{q}_{h,\alpha/2}^{R2} + \hat{s}_h^{R2,BC}, \hat{q}_{h,1-\alpha/2}^{R2} + \hat{s}_h^{R2,BC}\right]$ as discussed in Section 2.3.5, where $\alpha = 0.1$. Confidence intervals for the other estimators are constructed similarly.

Table 2.3: Simulation results for DGP 2.

	Horizon h					
	0	4	8	12	16	20
Impulse Response						
True	3.00	1.97	1.29	0.85	0.56	0.36
Local projections	2.99	1.83	1.07	0.57	0.22	0.06
VAR(HQIC)	2.96	1.93	1.33	0.95	0.71	0.56
Forecast Error Variance Decomposition						
True	0.80	0.25	0.10	0.05	0.03	0.02
Average estimate						
R2	0.79	0.26	0.15	0.14	0.15	0.19
VAR(HQIC)	0.80	0.27	0.12	0.08	0.06	0.05
Root mean squared error						
R2	0.03	0.11	0.12	0.14	0.17	0.21
VAR(HQIC)	0.03	0.08	0.06	0.06	0.05	0.05
Coverage (90 % level, asymptotic)						
R2	0.90	0.89	0.89	0.82	0.73	0.67
VAR(HQIC)	0.88	0.90	0.92	0.96	0.97	0.98
Forecast Error Variance Decomposition (bias-corrected, VAR(HQIC))						
True	0.80	0.25	0.10	0.05	0.03	0.02
Average estimate						
R2	0.81	0.24	0.09	0.03	0.01	0.00
VAR(HQIC)	0.80	0.25	0.10	0.05	0.03	0.02
Root mean squared error						
R2	0.03	0.10	0.09	0.09	0.10	0.12
VAR(HQIC)	0.03	0.07	0.06	0.05	0.04	0.04
Coverage (90 % level, asymptotic)						
R2	0.92	0.90	0.97	0.97	0.95	0.94
VAR(HQIC)	0.88	0.89	0.91	0.96	0.99	0.99

Notes: The table reports the performance of estimators introduced in Section 2.3 for DGP2. The sample size is $T = 160$, and the number of simulations is 2,000. R2 and VAR stand for $\hat{s}_h^{R2}$ and $\hat{s}_h^{VAR}$ estimators of forecast error variance decompositions. The lag order is selected by the Hannan-Quinn information criterion (HQIC). Confidence intervals for the bias-corrected R2 estimator are given by $\left[\hat{q}_{h,\alpha/2}^{R2} + \hat{s}_h^{R2,BC}, \hat{q}_{h,1-\alpha/2}^{R2} + \hat{s}_h^{R2,BC}\right]$ as discussed in Section 2.3.5, where $\alpha = 0.1$. Confidence intervals for the other estimators are constructed similarly.

Table 2.4: Simulation results for DGP 3 with alternative lag orders in VARs.

	Horizon h					
	0	4	8	12	16	20
Impulse Response						
True	1.00	4.10	6.13	7.46	8.33	8.91
Local projections	0.98	3.93	5.75	6.86	7.46	7.70
VAR(5)	0.93	3.71	4.70	4.94	5.04	5.08
VAR(10)	0.91	3.65	5.33	6.05	6.17	6.27
Forecast Error Variance Decomposition (bias-corrected, VAR(5))						
True	0.06	0.29	0.47	0.58	0.65	0.70
Average estimate						
R2	0.06	0.26	0.41	0.49	0.55	0.57
VAR(HQIC)	0.06	0.24	0.32	0.36	0.37	0.38
Root mean squared error						
R2	0.04	0.12	0.16	0.19	0.21	0.23
VAR(HQIC)	0.04	0.12	0.20	0.27	0.32	0.35
Coverage (90 % level, asymptotic)						
R2	0.81	0.82	0.82	0.84	0.83	0.83
VAR(HQIC)	0.87	0.81	0.66	0.50	0.38	0.32
Forecast Error Variance Decomposition (bias-corrected, VAR(10))						
True	0.06	0.29	0.47	0.58	0.65	0.70
Average estimate						
R2	0.07	0.29	0.46	0.56	0.62	0.65
VAR(HQIC)	0.06	0.27	0.41	0.49	0.53	0.55
Root mean squared error						
R2	0.05	0.12	0.16	0.19	0.21	0.23
VAR(HQIC)	0.04	0.11	0.16	0.19	0.22	0.24
Coverage (90 % level, asymptotic)						
R2	00.74	0.78	0.79	0.80	0.82	0.82
VAR(HQIC)	0.87	0.85	0.83	0.81	0.78	0.75

Notes: The table reports the performance of estimators introduced in Section 2.3 for DGP3. The sample size is $T = 160$, and the number of simulations is 2,000. R2 and VAR stand for $\hat{s}_h^{R2}$ and $\hat{s}_h^{VAR}$ estimators of forecast error variance decompositions. L_z and L_y are selected by the Hannan-Quinn information criterion (HQIC) and L_{VAR} is either 5 or 10. Confidence intervals for the bias-corrected R2 estimator are given by $\left[\hat{q}_{h,\alpha/2}^{R2} + \hat{s}_h^{R2,BC}, \hat{q}_{h,1-\alpha/2}^{R2} + \hat{s}_h^{R2,BC}\right]$ as discussed in Section 2.3.5, where $\alpha = 0.1$. Confidence intervals for the other estimators are constructed similarly.

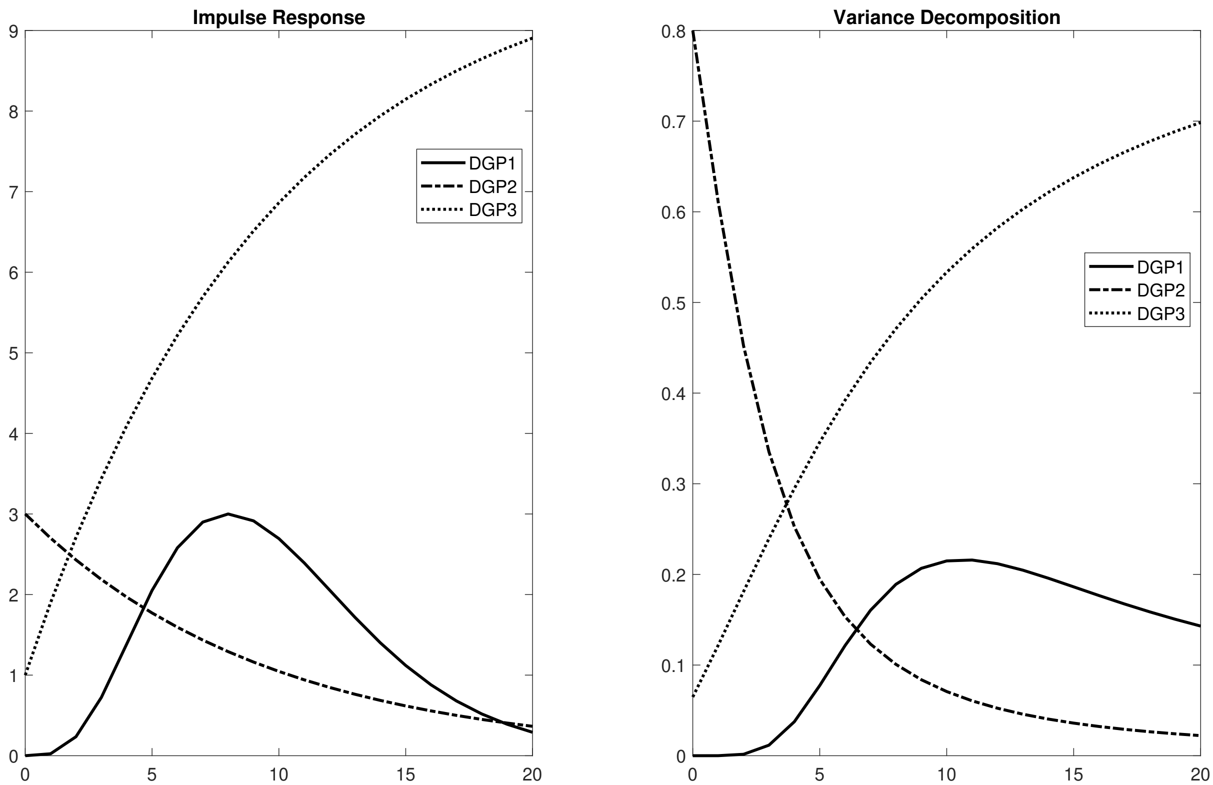

Figure 2.1: Population impulse responses and forecast error variance decompositions for each DGP.

Notes: The left panel shows the impulse response functions for three bivariate data generating processes (DGPs) in Section 2.4.1. The right panel shows the contribution of the structural shocks to the forecast error variances of an outcome variable for the DGPs.

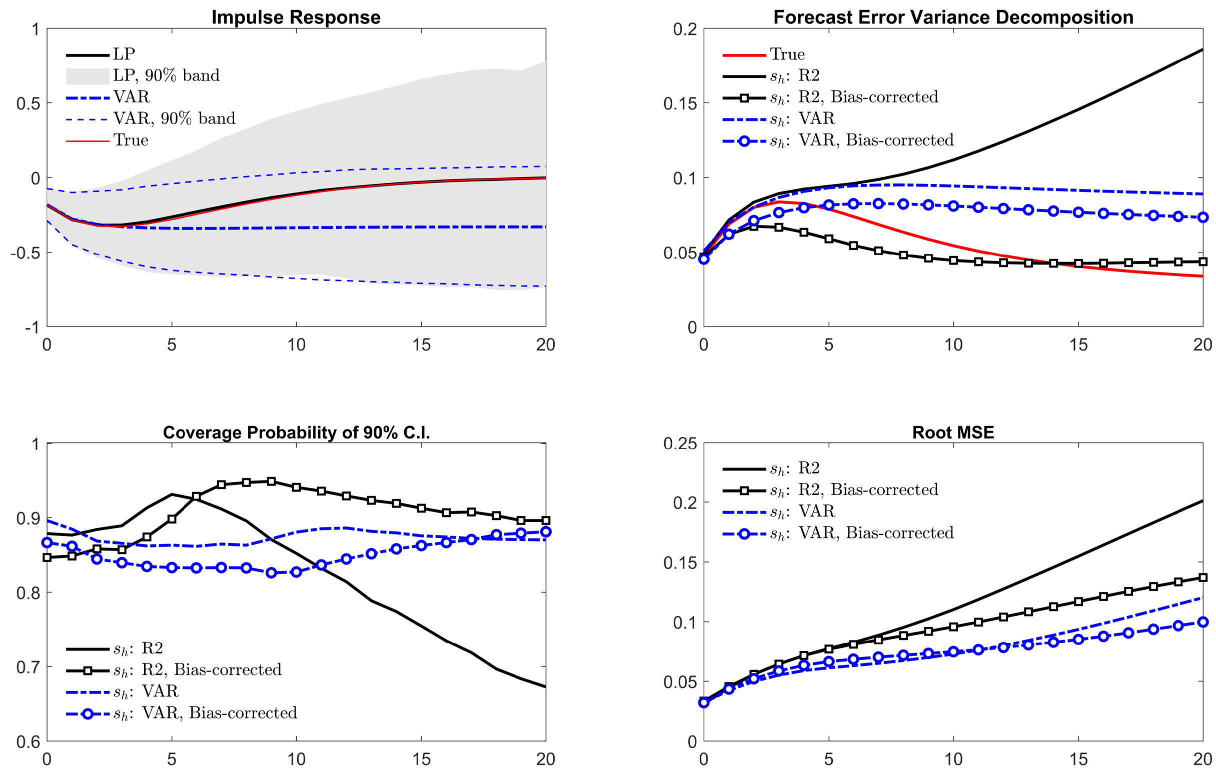

Figure 2.2: Smets and Wouters (2007) model, real GDP and monetary policy shock, $T = 160$.

Notes: We simulate the Smets and Wouters (2007) model to evaluate the performance of our estimators as discussed in Section 2.4.2. The top-left panel covers local projections (LP) and VAR estimators of impulse responses where lag lengths are determined with the Hannan-Quinn information criterion (HQIC). The shaded area and the dashed lines represent the 5th and 95th percentiles of the simulated LP and VAR estimates, respectively. For forecast error variance decompositions (FEVDs), $E\left[\hat{s}_h^{R2}\right]$ and $E\left[\hat{s}_h^{VAR}\right]$ with or without bias-correction can be found in the top-right panel. Coverage probabilities of 90% confidence intervals around $\hat{s}_h^{R2}$ and $\hat{s}_h^{VAR}$ with or without bias-correction are shown in the bottom-left panel. We construct the confidence intervals using $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ percentiles of bootstrapped estimates (see Section 2.3.5). The bottom-right panel illustrates root mean squared errors of the estimators of the FEVDs.

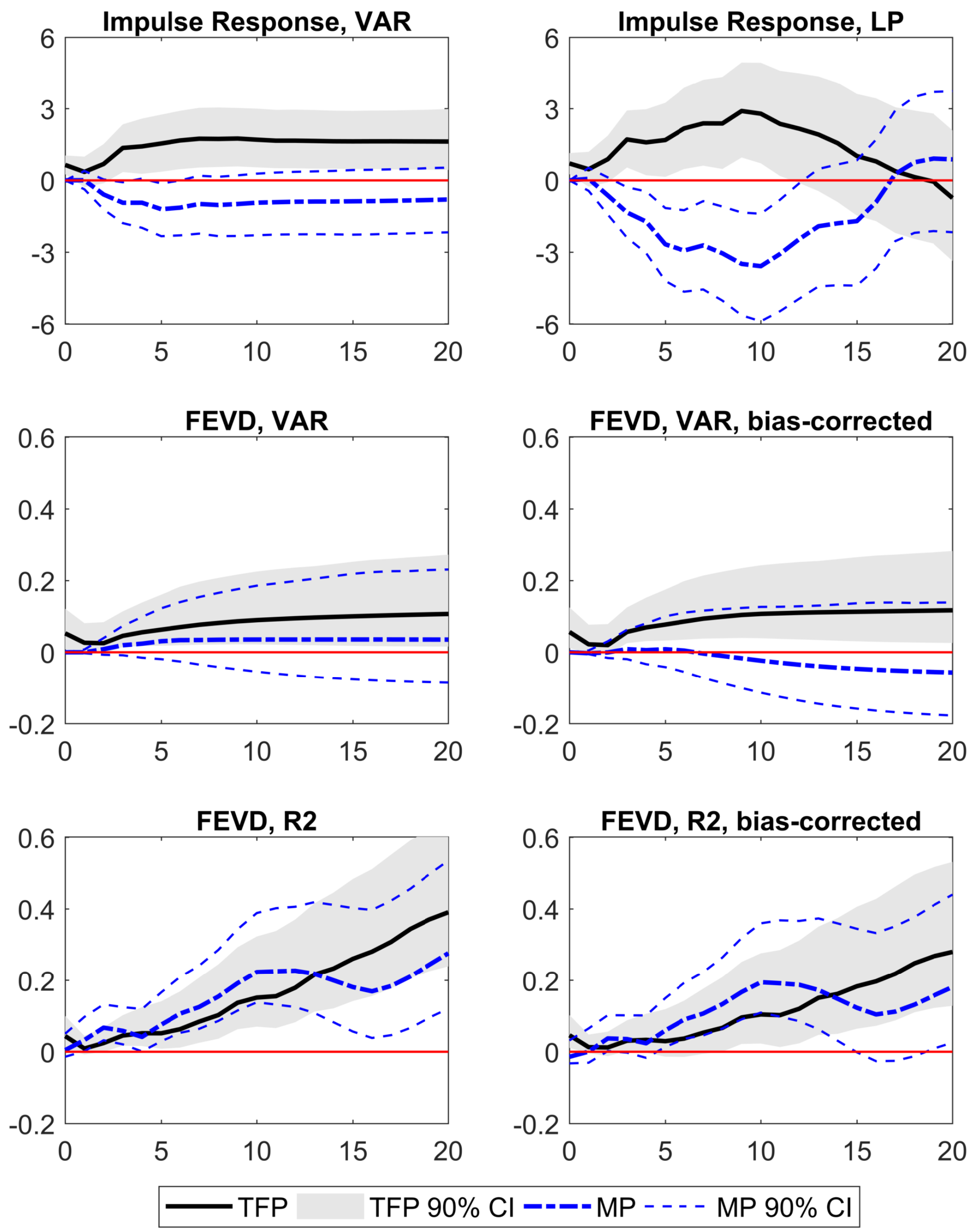

Figure 2.3: Real GDP. Sample: 1969:Q1-2008:Q4.

Notes: We estimate impulse responses and forecast error variance decompositions (FEVDs) of real GDP in Section 2.5. We focus on total factor productivity (TFP) shocks identified as in Fernald (2014) and monetary policy (MP) shocks of Romer and Romer (2004) extended by Coibion et al. (2017). The first row covers the estimated impulse responses and 90% bootstrap confidence intervals in response to a one standard deviation shock to TFP and MP. We depict the results for VARs (top-left panel) and local projections (LP, top-right panel). The unit of the y-axis is annualized percent. The second row shows $\hat{s}_h^{VAR}$ and 90% bootstrap confidence intervals with and without bias-correction. The last row is for $\hat{s}_h^{R2}$ and 90% bootstrap confidence intervals with and without bias-correction.

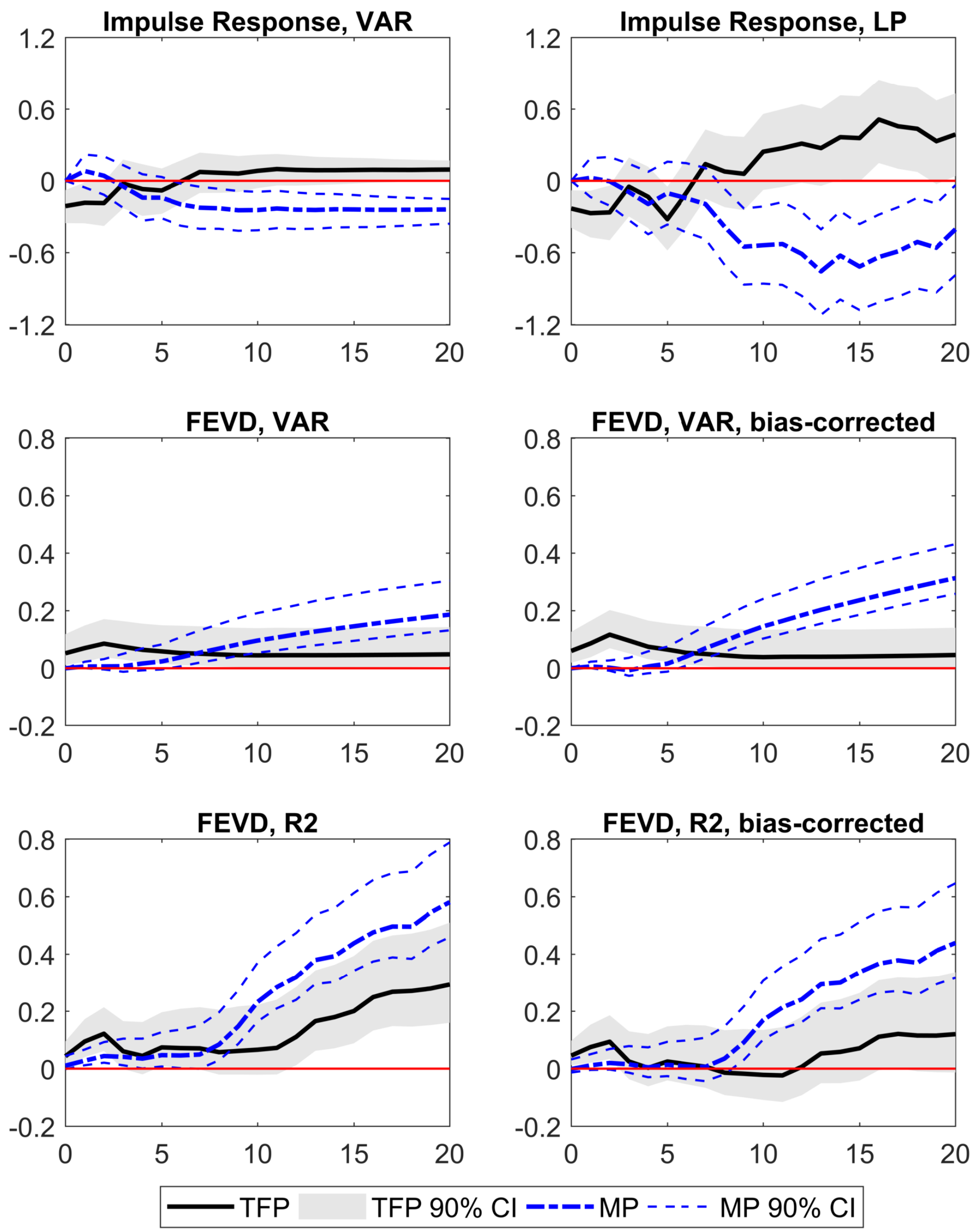

Figure 2.4: Inflation. Sample: 1969:Q1-2008:Q4.

Notes: We estimate impulse responses and forecast error variance decompositions (FEVDs) of inflation in Section 2.5. We focus on total factor productivity (TFP) shocks identified as in Fernald (2014) and monetary policy (MP) shocks of Romer and Romer (2004) extended by Coibion et al. (2017). The first row covers the estimated impulse responses and 90% bootstrap confidence intervals in response to a one standard deviation shock to TFP and MP. We depict the results for VARs (top-left panel) and local projections (LP, top-right panel). The unit of the y-axis is annualized percent. The second row shows $\hat{s}_h^{VAR}$ and 90% bootstrap confidence intervals with and without bias-correction. The last row is for $\hat{s}_h^{R2}$ and 90% bootstrap confidence intervals with and without bias-correction.

Chapter 3

Business Cycles and Earnings Inequality

> *[T]he various linkages between heterogeneity and aggregate demand are not yet well understood, either empirically or theoretically. – Yellen (2016)*

3.1 Introduction

The Great Recession was a pivotal moment for modern business cycle research. One of the key elements revealed by the recession was that distributional factors could have significant effects on macroeconomic fluctuations. Indeed, a major objective of policymakers has since become understanding the interplay between inequality and business cycles and analyzing what distributional effects various macroeconomic stabilization policies have while achieving their intended aggregate goals. But this is not the only question that the Great Recession poses. Another important issue is whether inequality and redistribution contribute to variation in aggregate demand. If distributional forces can initiate demand-driven business cycles, appropriate policies should be taken to stabilize the economy. In this regard, it is central to understand how the power of stabilization policies varies with the level of inequality. Although the Great Recession spurred interest in these questions, research in this area is still in its infancy. We have limited understanding of the relationship between business cycles, inequality, and stabilization policies, either empirically or theoretically as underscored by former Fed chair Janet Yellen (2016).

This paper develops a new framework for studying the linkages between inequality and aggregate dynamics. The model highlights key mechanisms for the interaction between a cross-sectional distribution and aggregate demand while maintaining tractability. The theory is further connected to the novel empirical findings based on a high-quality quarterly measure of inequality that I construct. The structural interpretation of my results provides new insights on the interplay of inequality, business cycles, and stabilization policies.

I investigate the U.S. data where the recent rise in inequality has been most prominent. I empirically study how drivers of business cycles including shocks to total factor productivity, monetary policy, and fiscal policy cause variation in inequality at cyclical frequencies. I also explore the other direction from inequality to macroeconomic fluctuations. I document that changes in the shape of cross-sectional distributions in combination with heterogeneous marginal propensities to consume (MPC) across agents may influence aggregate demand. These results illustrate why inequality matters for both business cycles and policymakers, and vice versa.

To shed light on the mechanisms through which inequality impacts aggregate demand, I develop a new theoretical framework. My model captures how inequality, MPCs, and aggregate demand interact in a parsimonious manner. The empirical results are also successfully rationalized by the model with novel insights on how aggregate consumption demand is determined. An intriguing policy implication of the model is that the power of monetary and fiscal policies increases with the level of inequality, where the interaction between inequality and MPC plays a key role for the result. This provides yet another reason why inequality is relevant for stabilization policies and why policymakers should be aware of the distributional outcomes of their policies, even if their objectives are based only on aggregate economic conditions.

For the empirical analysis, the biggest hurdle is to find a high-frequency measure of inequality.[1] I resolve the problem by constructing a new quarterly inequality index based on the Quarterly Census of Employment and Wages (QCEW), a quarterly, publicly available, administrative database featuring wide coverage. The QCEW publishes counts of employment and total pre-tax earnings at the U.S. county level by detailed industry classification codes. The earnings include bonuses, stock options, profit distributions, and some fringe

[1] Most of the existing measures of inequality are annual such as the top income share of Piketty and Saez (2003), the top wealth share of Saez and Zucman (2016), the log P90/P10 wage ratio of Autor, Katz and Kearney (2008), and the Gini coefficient prepared by the U.S. Census Bureau. However, annual data are not fitting for the time series analysis in this paper due to small sample sizes and difficulties in identifying high-frequency variations.

benefits such as cash value of meals and lodging.

I extract an earnings distribution in each quarter from the microdata. Although the QCEW is not at the individual level, it is disaggregated enough to capture major dynamics of earnings inequality. Indeed, the number of observations is enormous given the administrative nature of the data source. Also, inequality series based on the QCEW show similar historical trends to existing ones based on individual but annual data. Lastly, it is important that my benchmark measure is the log P90/P10 index, which does not require measuring earnings within the tails. Instead, my focus is on the "middle class," who contribute to aggregate variables significantly.

I study how driving forces of business cycles influence earnings inequality using this new, high-quality, quarterly time series. I report impulse responses and forecast error variance decompositions to illustrate the relationships between earnings inequality and shocks to total factor productivity, monetary policy, and fiscal policy. I employ local projections of Jordà (2005) to estimate the impulse response functions and find that an unanticipated expansion in government spending raises earnings inequality, while a positive productivity shock lowers it. However, the responses are small and statistically insignificant for the first two years for both shocks. On the other hand, shocks to monetary policy have little effects on earnings inequality. For the forecast error variance decompositions, I apply a new and flexible method with local projections developed by Gorodnichenko and Lee (2017). The results are consistent with the impulse responses in the sense that only technology and fiscal policy shocks contribute to earnings inequality significantly in the medium-run. Also, most of the short-run fluctuations in the new inequality measure are not explained by those shocks. These facts may provide useful empirical inputs to theoretical heterogeneous agent models (for example, Gornemann, Kuester and Nakajima, 2016; Guerrieri and Lorenzoni, 2017; Kaplan, Moll and Violante, 2018; McKay, Nakamura and Steinsson, 2016; McKay and Reis, 2016).

The next part of the paper investigates the opposite direction, from inequality to business cycles. Researchers have spent an enormous amount of time and effort to detect and evaluate sources of business cycles. While this literature typically focuses on level shocks on aggregates, I propose to use innovations in inequality as a measure of "redistribution" shocks. Rising inequality or redistribution from the poor to the rich may reduce aggregate demand and impact on aggregate variables because marginal propensities to consume (MPC) decrease in income or wealth (see Dynan, Skinner and Zeldes, 2004; Johnson, Parker and

Souleles, 2006; Parker et al., 2013; Zidar, 2018).

Specifically, I rely on unanticipated innovations in the time series of earnings inequality, which are orthogonal to aggregate shocks and macroeconomic variables. These innovations summarize redistributive forces shifting earnings from the bottom to the top while maintaining aggregate earnings contemporaneously. I show that such redistribution that increases earnings inequality lowers aggregate demand substantially. Major macroeconomic variables such as real GDP, consumption, investment, price levels, and the federal funds rate decline in a U-shaped manner in response to the positive unanticipated innovations. Furthermore, the responses are large. For example, 35 percent of the forecast error variance of real GDP per capita at a four-year horizon is due to these innovations. In short, redistribution shocks seem to be an important driving force of regular business cycle dynamics similar to standard level shocks to aggregates.

To illustrate the mechanisms through which shocks to inequality affect an economy, I develop New Keynesian dynamic stochastic general equilibrium (DSGE) models. I study two models, a simple one for the intuition based on analytical results and a medium-sized one for the quantitative analysis rationalizing the large, negative, U-shaped responses. The models feature two agents who are either hand-to-mouth or intertemporal in line with Campbell and Mankiw (1989) and Galí, López-Salido and Vallés (2007). Unlike usual two-agent models, I assume that the labor productivity of both agents differs, where the hand-to-mouth agent is less productive. This setup is in accordance with the data in the sense that MPCs decrease in income or wealth, that the probability of being credit constrained decreases in income (Crook, 2001, 2006), and that there is limited participation in financial markets among households below median wealth (Guiso and Sodini, 2013).

I consider a shock increasing the dispersion of the idiosyncratic labor productivity, which makes the rich richer and the poor poorer. The main analytical result based on the simple two-agent New Keynesian (TANK) model is that this earnings inequality shock is isomorphic to a discount rate shock in a textbook representative agent New Keynesian (RANK) model. This new finding illustrates in a simple way why individual heterogeneity associated with earnings inequality can be a micro-foundation for an aggregate demand shock in a RANK framework.

However, this simple TANK model cannot rationalize the empirical impulse responses, especially the U-shaped patterns. Thus, I extend the simple model and develop a two-agent medium-sized DSGE model with three novel features affecting aggregate demand: an endogenous extensive margin between two agents, decreasing relative risk aversion (DRRA) consumption utility, and a small amount of financial income for the credit constrained agents. In a recession, there may be more consumers subject to credit constraints due to unemployment risk (Ravn and Sterk, 2017) or countercyclical idiosyncratic earnings risk conditional on being employed (Guvenen, Ozkan and Song, 2014; Storesletten, Telmer and Yaron, 2004). This leads to more consumers having higher MPCs during deeper recessions. Indeed, Mian, Rao and Sufi (2013) and Mian and Sufi (2015) document that limited access to credit and related MPC heterogeneity played a central role in the development of the Great Recession. I provide a parsimonious characterization of this channel in the model as well as a microfoundation for it. Another new feature is that the degree of relative risk aversion (RRA) of both agents may differ. When the model is estimated, the coefficient of RRA of the hand-to-mouth agents is higher than that of the intertemporal agents. As agents move from the credit constrained state to the intertemporal state, their consumption increases and their coefficients of RRA decrease. This is exactly DRRA preferences because the same agent alternates between the two states in my model. Note that DRRA preferences also conform to empirical findings of Calvet, Campbell and Sodini (2009, Section IV.C). Finally, I assume that the credit-constrained agents also receive non-zero (albeit small) financial income for three reasons. First, the wealthy hand-to-mouth agents of Kaplan and Violante (2014) and Kaplan, Violante and Weidner (2014) would hold a significant amount of assets and receive dividends while they are credit constrained. Second, some wealth poor agents engage in financial investment (Guiso and Sodini, 2013). Lastly, this may represents government transfers and pensions in a parsimonious way. My model further incorporates characteristics of medium-sized RANK models such as investment or capital utilization adjustment costs, sticky prices and wages, and habit preferences (Christiano, Eichenbaum and Evans, 2005; Smets and Wouters, 2007). Because a consumer is temporarily hand-to-mouth or temporarily intertemporal, I dub the model "the Temporarily Hand-to-mouth and Intertemporal agent New Keynesian model," or in short, "the THINK."

I estimate the model using a Bayesian impulse response matching method of Christiano, Trabandt and Walentin (2010). This approach enables me to focus on structural shocks with clearly identified empirical counterparts. The estimated model generates large and U-shaped impulse responses to the earnings inequality shocks, comparable to the empirical responses to the unanticipated innovations in inequality. In doing so, the model relies on the interplay of the three new features discussed above, which induces intriguing dynamics in discount factors and aggregate demand. For example, the number of credit constrained agents is time-varying because of the endogenous extensive margin of being credit constrained and countercyclical earnings risk. This adds a new component in aggregate consumption demand in the following way. When aggregate demand decreases in response to an inequality shock, the economy goes into a recession. Then some agents are hit by large negative idiosyncratic shocks due to countercyclical earnings risk and unemployment risk. As they become credit constrained and reduce consumption, aggregate consumption demand decreases and the economy falls into a deeper recession. Then more agents become credit constrained and so on. Such distributional effects are crucial for rationalizing the shape and magnitude of the empirical responses of aggregate consumption to the unanticipated innovations in inequality.

Another important prediction of my theory is that inequality affects the power of stabilization policies. Intuitively, in an economy with high inequality, there may be more people at the bottom of either income or wealth distribution. Also, these people have higher MPCs because they do not have enough buffers to absorb shocks. An interaction effect between more people and higher MPCs makes aggregate consumption demand more sensitive to economic conditions including monetary and fiscal policies. This channel is relevant to the U.S. economy because the share of households with negative net wealth has been increasing since 1969 (Wolff, 2017). Consistent with these insights, the non-linear dynamics of the THINK model predicts that the economy responds more strongly to a monetary or government spending shock when there are more credit constrained agents. If the interaction effect also applies to other structural shocks, the aggregate economy may fluctuate more, and so cyclical volatility in general may be elevated. But on the bright side, stabilization policies become more powerful too.

Given the policy implication based on the model, I empirically test whether aggregate variables react differently to policy shocks conditional on the level of inequality. Using a variety of datasets (state-level, aggregate, various identified shock series, and sample periods), I find that the U.S. economy responds more strongly to either a monetary or fiscal policy shock of the same magnitude when income is distributed more unequally.

There are several empirical studies on cyclical variations of inequality. Some focus on the effects of inflation on poverty or redistribution of nominal wealth (Blank and Blinder, 1986; Doepke and Schneider, 2006; Romer and Romer, 1999). Others look at differential exposure of individual consumption, earnings, and income to aggregate fluctuations (Parker and Vissing-Jorgensen, 2009; Guvenen, Ozkan and Song, 2014; Guvenen et al., 2017). Coibion et al. (2017) deals with the dynamics of inequality conditional on monetary policy shocks, which is the most related previous work to this paper. However, this paper differs from Coibion et al. in several respects. First, I study how inequality impacts business cycles as well as whether major structural shocks affect inequality, whereas Coibion et al. concentrates solely on the effects of monetary policy shocks on inequality. Second, I articulate mechanisms at play using structural models, while the analysis of Coibion et al. is purely empirical. Lastly, data sources are different. Coibion et al. constructs measures of inequality based on the Consumer Expenditure Survey, while I use the QCEW.

My model features both hand-to-mouth and intertemporal agents. Such models have been used to study monetary policy rules (Bilbiie, 2008; Galí, López-Salido and Vallés, 2004) and effects of government spending shocks (Bilbiie, Meier and Müller, 2008; Galí, López-Salido and Vallés, 2007) when there are hand-to-mouth consumers following Campbell and Mankiw (1989). While these models usually assume equally productive agents and ignore distributional factors, I introduce earnings inequality with heterogeneous labor productivity. This provides a new, simple theoretical framework for studying inequality and macroeconomic fluctuations. Furthermore, this parsimonious framework is consistent with the empirical evidence that less productive workers have higher MPCs and are more likely to be credit constrained (see Crook, 2001, 2006; Dynan, Skinner and Zeldes, 2004; Johnson, Parker and Souleles, 2006; Parker et al., 2013; Zidar, 2018).

There have been papers summarizing individual heterogeneity by a wedge to a discount factor in an aggregate consumption Euler equation (Braun and Nakajima, 2012; Constantinides and Duffie, 1996; Werning, 2015). These papers show exact aggregation is possible under some restrictive assumptions. I derive a similar result for a canonical TANK model in a first-order approximation and connect earnings inequality to aggregate demand and discount rate shocks more explicitly.

Another framework for studying economic fluctuations with distributional issues is based on heterogeneous agent New Keynesian (HANK) models. These quantitative models generate a realistic description of cross-sectional distributions of households as an equilibrium outcome (see Kaplan and Violante, 2018, for a review). Others propose models with two (or a finite number of) agents as a middle ground between tractable RANK and rich HANK models and provide analytical expressions highlighting HANK mechanisms (Acharya and Dogra, 2018; Bilbiie, 2017; Debortoli and Galí, 2017; Ragot, 2018; Ravn and Sterk, 2018). I take a similar approach to emphasize insights based on analytical results while utilizing efficient tools developed for solving and estimating medium-sized models.

The THINK model features an extensive margin between two agents. Bilbiie (2017) considers an analogous channel with fixed transition probabilities in his analytically tractable HANK model and illustrates how it relates to the discounted Euler equation of McKay, Nakamura and Steinsson (2017). Because the transition probabilities are fixed, the number of credit constrained agents in his model is constant. My paper goes one step further and makes transition probabilities vary endogenously with aggregate fluctuations. Because earnings risk is countercyclical as reported by Guvenen, Ozkan and Song (2014), Ravn and Sterk (2017), and Storesletten, Telmer and Yaron (2004), it is harder for credit constrained agents to escape from their constraints during a recessions. This would increase the number of credit constrained agents during economic downturns, constituting a new channel for aggregate consumption dynamics.

Auclert and Rognlie (2018) also investigate the effects of redistribution shocks on economic output in their HANK model. However, the THINK model differs from the model of Auclert and Rognlie in several respects. First, Auclert and Rognlie work with a continuum of heterogeneous agents, whereas the THINK model is based on two agents. Second, Auclert and Rognlie assume a constant RRA (CRRA) utility function, while I assume a DRRA preference with habit formation. For the supply block, Auclert and Rognlie let downward nominal wage rigidities induce room for monetary policies, whereas I introduce both price and wage stickiness a la Rotemberg (1982). Lastly, my model includes an autoregressive term in monetary policy rule which does not exist in the model of Auclert and Rognlie. When all the differences are combined, the models generate divergent predictions on the effects of earnings inequality shocks. Auclert and Rognlie find that such shocks have little aggregate effects in their model, which is contrary to the predictions of the THINK model and my empirical results.

While previous research (*e.g.*, Alesina and Perotti, 1996; Bordo and Meissner, 2012; Cairó and Sim, 2017; Kumhof, Rancière and Winant, 2015) covers why a financial or political crisis may be related to inequality, little work has been done about the power of stabilization policies and volatility of regular business cycles given various degrees of inequality. Debortoli and Galí (2017) is a notable exception. Debortoli and Galí compare TANK models with the fixed but different steady state shares of the hand-to-mouth agents and find that the effects of monetary policy shocks are significantly larger around the steady state with more hand-to-mouth agents. In the THINK model, I focus on a non-linear interaction effect between more people and higher MPCs around the same steady state. I study fiscal policy shocks as well as monetary policy shocks, and I also find empirical results consistent with the theoretical predictions of my THINK model.

The remainder of this paper is organized as follows. Section 3.2 covers the construction of the new, high-quality, high-frequency measure of earnings inequality. Section 3.3 deals with the responses of earnings inequality to shocks to stabilization policies and total factor productivity. In Section 3.4, I study the direction from earnings inequality to key aggregate variables and illustrate that an unanticipated positive innovation in inequality decreases aggregate demand substantially in a U-shaped manner. Section 3.5 analyzes the mechanisms through which an inequality shock reduces aggregate demand and generates large, negative, U-shaped responses in DSGE models. Section 3.6 discusses the relationship between the power of stabilization policies and the level of inequality. Section 3.7 concludes.

3.2 A New Quarterly Measure of Inequality

3.2.1 Data

The Quarterly Census of Employment and Wages (QCEW) is a quarterly, publicly available, administrative database. The Bureau of Labor Statistics and the State Employment Security Agencies prepare the data based on reports filed by employers, collected for the unemployment insurance programs.

The employment series includes all forms of jobs: full-time, part-time, temporary, and permanent. The wages in the data are pre-tax earnings including bonuses, stock options, profit distributions, and some fringe benefits such as cash value of meals and lodging.

The main advantages of the QCEW are frequency, coverage, and accuracy. First, the QCEW is quarterly, whereas most of the other data previously used for studying inequality are annual.[2] Moreover, the QCEW covers all counties and industries. Finally, the data are administrative and therefore observed with little measurement error.

[2]For example, Guvenen et al. (2015), Guvenen, Ozkan and Song (2014), and Song et al. (2018) use the Master Earnings File of the U.S. Social Security Administration. Piketty and Saez (2003) rely on tax returns statistics of the Internal Revenue Service. The Current Population Survey (CPS) is analyzed by Autor, Katz and Kearney (2008). The CPS has two types of earnings data. The first one is collected annually in the March annual demographic survey. The other is based on merged outgoing rotation groups (MORG) available monthly. However, the MORG data are about *usual* weekly earnings, and therefore it is not suitable for identifying high-frequency variation in inequality.

However, the data are not perfect. First, the data are not at the individual-level. The most granular information available is average earnings and the number of workers in a cell, where a cell is an industry/county/ownership-type combination.[3] Thus, measures in this paper represent between-cell, not within-cell inequality. Moreover, self-employed workers are not included, and some observations are suppressed due to confidentiality. Finally, the data cover only earnings.[4]

Despite these limitations, I will show in Section 3.2.2 that the log P90/P10 index based on the QCEW is consistent with the same measure based on the March annual demographic survey in the Current Population Survey (CPS), which is annual and individual-level (Autor, Katz and Kearney, 2008). In other words, the QCEW is sufficiently disaggregated, and so the measurement errors due to the unobservable within-cell inequality seem to be small.

I use several filters to attenuate the potential adverse effects of extreme observations and seasonality. First, observations with unreasonably small earnings are dropped. Following Guvenen et al. (2015), the threshold is what can be earned by working one-quarter of full-time at half of the legal minimum wage rate. Second, I seasonally adjust the percentiles of the log earnings distribution. These are available at three different levels of aggregation depending on period: SIC 2-digit for 1975:Q1-2000:Q4, SIC 4-digit for 1984:Q1-2000:Q4, and NAICS 6-digit for 1990:Q1-2014:Q4, where SIC and NAICS stand for the Standard Industrial Classification codes and the North American Industry Classification System, respectively. I splice these three series and deflate the combined one using the GDP implicit deflator (see Appendix A.1 for details).[5]

Table 3.1 shows summary statistics for selected quarters. The top half of Table 3.1 displays the number of observations and coverage. The number of cells is greater than two hundred thousand after a few early quarters, which far exceeds the number of respondents in a typical survey. The bottom half of the table shows the sizes of the cells. For example, there are around 66 workers in a median-sized cell, and this corresponds to only 0.00007% of the total number of workers in the first quarter of 2014. In other words, the sizes of most of the cells, in which I assume workers earn uniformly divided compensations, are small when

[3]The ownership code differentiates establishments owned privately, by a local government, by a state government, by the federal government, and by an international government.

[4]However, taking capital income into accounts might not affect the log P90/P10 index (the benchmark measure in this paper) significantly, because capital income is extremely concentrated above the top 10th percentile.

[5]All standard macroeconomic variables are obtained from the FRED run by the Federal Reserve Bank of St. Louis.

we consider the cross-section of earnings.

3.2.2 Percentiles and Inequality Index

In the right panel of Figure 3.1, I plot the log of selected percentiles of the real earnings distributions. Median real earnings have not grown as fast as the upper half of the distribution for the last few decades, and therefore the U.S. real earnings distribution has widened. Similarly, the gap between the median and the bottom 10th percentile increased throughout most of the periods (Figure A2 in Appendix A). The late 1990s was an exception during which the gap was stable. Furthermore, the imprints of historical events such as the dot-com bubble around 2000 and the sub-prime crisis around 2008 are evident among the top percentiles.

The log P90/P10 index is on the left panel. When it is compared with an existing, annual measure reported by Autor, Katz and Kearney (2008), not only the historical pattern but also the values are similar.[6] Because Autor, Katz and Kearney use the CPS which is individual-level data, this similarity indicates that my quarterly log P90/P10 index is of high-quality.

The new quarterly log P90/P10 index, which is my benchmark inequality measure, has desirable properties for the following reasons. First, the QCEW is a large administrative dataset. Second, although within-cell inequality is not observable, the size of most cells is small. Furthermore, the P90/P10 index is rather robust to changes in within-cell inequality because the index utilizes only two points in the entire distribution.[7] Finally, considering the log P90/P10 index allows us to circumvent measuring inequality within the extreme tails and to focus on inequality in the "middle class," who affect aggregate variables significantly.

3.3 From Aggregate Shocks to Earnings Inequality

This section investigates how earnings inequality reacts to major drivers of business cycles, using the new, high-quality, quarterly measure of earnings inequality that I construct. The

[6] I construct three other measures: (i) cross-sectional standard deviation of the log real earnings, (ii) Gini coefficients of the real earnings, and (iii) top 10% earnings share. Although these series replicate historical patterns successfully, the levels of them are lower than the corresponding measures based on individual-level data (Figure A3 and A4 in Appendix A).

[7] Relatedly, Song et al. (2018) argue that changes in earnings inequality in the U.S. have been primarily a between-firm phenomenon, not within-firm. This might explain why ignoring within-cell inequality leads to little distortions in time series variation.

estimated impulse response functions and the forecast error variance decompositions constitute novel empirical facts regarding dynamics of earnings inequality.

3.3.1 Shocks and Sample Period

I analyze three structural shocks in relation to the inequality index. The identified shock series I employ are shocks to total factor productivity (TFP), monetary policy (MP), and fiscal policy (FP). Fernald (2014) provides a quarterly, utilization-adjusted series of the TFP. He deals with both capital and labor utilization, where the adjustment process is similar to how Basu, Fernald and Kimball (2006) purify annual measures. For a monetary policy shock, Romer and Romer (2004) identify the MP shock as an orthogonal component in the federal funds rate to the Federal Reserves' information set around the Federal Open Market Committee meetings. I use an updated version of the shocks from Coibion et al. (2017), who extend the series to 2008. Finally, I rely on the FP shock series in Auerbach and Gorodnichenko (2012), which is constructed from comparison of the realized and forecasted growth rates of government spending. They use the forecasts from the Greenbook and the Survey of Professional Forecasts.

I select the first quarter of 1978 as the first period in the benchmark sample, when there was a significant change to coverage of the QCEW.[8] The last period of the sample is the fourth quarter of 2008, when the updated MP shock series ends.

3.3.2 Impulse Responses

Let y_t, $x_{t,1}$, $x_{t,2}$, and $x_{t,3}$ be the inequality index, TFP, MP, and FP shocks in period t, respectively. The response of y_{t+h} to a unit impulse in $x_{t,j}$ is denoted by $\psi_{h,j}$:

$$\psi_{h,j} = \frac{\partial y_{t+h}}{\partial x_{t,j}} \qquad \text{for all } h \text{ and } j. \tag{3.1}$$

The impulse response coefficients $\{\psi_{h,j}\}$ are estimated by local projections of Jordà (2005):

$$y_{t+h} - y_{t-1} = c_h + \sum_{i=1}^{L_y} \rho_i^{(h)} \Delta y_{t-i} + \sum_{i=0}^{L_x} \sum_{j=1}^{3} \beta_{i,j}^{(h)} x_{t-i,j} + u_{t,h}^{(y)}, \tag{3.2}$$

[8]Specifically, the Federal Unemployment Compensation Amendments of 1976 became effective on January 1, 1978. This incorporates major changes to state unemployment insurance program on which raw data of the QCEW are based on. See https://www.bls.gov/cew/cewbultncur.htm#Coverage.

where $\beta_{0,j}^{(h)}$ captures $\psi_{h,j}$ for each h and j. In other words, $\{\beta_{0,j}^{(h)} : h = 0, 1, \ldots\}$ represents how the inequality index responds to $x_{t,j}$.

In Equation (2), lags of Δy_t are included on the right-hand side to absorb the predictable variation. I set L_y and L_x at six, but the results are robust to the lag length and various other specifications. Lastly, the identified shocks in Equation (2) are orthogonal. For every pair of the three shocks, the null of zero correlation is not rejected at the 5% level. Details of these statistical tests and sensitiveness analysis are in Appendix B.1-B.4.

I similarly estimate how the aggregate earnings in the QCEW react to the shocks and depict the results with that of the inequality index in Figure 3.2. Given a one standard deviation positive TFP shock (3 percent, annualized), the aggregate earnings increase with a peak of around 3 percent (annualized) after 10 quarters, but the inequality index decreases by around 2.5 log points (annualized) after 3 to 4 years. Thus, the earnings distribution shifts to the right, while the dispersion among the middle workers shrinks.

This finding may sound contradictory to a view that rising earnings inequality in recent decades is due to skill-biased technological progress (Goldin and Katz, 2009; Krusell et al., 2000). However, the results in Figure 3.2 are about cyclical relationships between productivity and inequality around trends, not about the trends themselves. Furthermore, a dynamic stochastic general equilibrium model of Gornemann, Kuester and Nakajima (2016) similarly predicts that earnings inequality decreases when a positive productivity shock hits an economy.

Reduction in the inequality index is mostly because of compression among the upper half of the distribution, not the lower half. The log P90/P50 index decreases statistically significantly at the 10% level while the P50/P10 index does not as illustrated in Figure 3.3. However, the right-tail above P90 reacts differently. Indeed, the P99/P50 index increases by 5 log points (annualized) at the peak in response to a one standard deviation positive TFP shock.[9] The top 10% share also rises, contrary to the log P90/P10 index (see Figure B7).

In short, the earnings distribution becomes more right-skewed in response to a positive TFP shock. While the middle 80% shrinks, (especially the upper part), the right-tail diverges.

A contractionary MP shock decreases the aggregate earnings while it has little effects on the earnings dispersion among the employed. Coibion et al. (2017) also reports similar

[9] See figures in Appendix B.5 for how various percentiles respond to the shocks. This specific observation regarding P99 is in Figure B13.

results based on a different dataset, especially in their Figure 3.3. Thus, any redistribution channel of monetary policy should be from either unemployment risk or financial income, not from labor earnings conditional on being employed (see Auclert, 2017; Kaplan, Moll and Violante, 2018, for the redistribution channel). In theory, earnings inequality may respond in either direction to a monetary policy shock. For example, earnings inequality increases given a contractionary monetary policy shock in the model of Gornemann, Kuester and Nakajima (2016). On the other hand, Dolado, Motyovszki and Pappa (2018) illustrate how earnings inequality between high and low-skilled workers could decline in response to a contractionary monetary policy shock in a NK model with search and matching frictions and capital-skill complementarity. However, neither of those theoretical predictions is consistent with my empirical result that the MP shock has little effects on the earnings inequality index.

The earnings distribution widens when government expenditures expand. The responses in Figure 3.2 are delayed and persistent like those for the TFP shocks. The peak effects are 3.8 log points (annualized) after 15 quarters given a one standard deviation shock (4.2 percent, annualized). Similarly, rising dispersion among the upper half is a key to the reaction because the P90/P50 index increases statistically significantly, while the P50/P10 index does not in Figure 3.3. Qualitatively, this result is consistent with the prediction of the model in Heer and Scharrer (2016). Heer and Scharrer find that an expansionary government spending shock raises income inequality in an overlapping generations model with both hand-to-mouth and intertemporal agents.

3.3.3 Forecast Error Variance Decompositions

Next, I evaluate the economic significance of each shock as a driver of earnings inequality at business cycle frequencies. I decompose the forecast error variance of the inequality index in relation to each shock. The parameters of interests are

$$s_{h,j} = \frac{Var\left(\sum_{i=0}^{h} \psi_{i,j} x_{t+h-i,j}\right)}{Var\left(y_{t+h} - y_{t-1} - P_{t-1}\left(y_{t+h} - y_{t-1}\right)\right)}, \tag{3.3}$$

where the subscript j indexes the type of shocks (TFP, MP, or FP), and P_t means a projection on a period t information set. The forecast error $y_{t+h} - y_{t-1} - P_{t-1}(y_{t+h} - y_{t-1})$ consists of the effects of $\{x_{t,j}\}$ and an unrelated component: $y_{t+h} - y_{t-1} - P_{t-1}(y_{t+h} - y_{t-1}) = \psi_{0,j} x_{t+h,j} + \cdots + \psi_{h,j} x_{t,j} + u_{t,h,j}^{(FE)}$. Then the contribution of the shock j's to the total variance of the forecast error is captured by $s_{h,j}$. In other words, it measures the importance of the shock j in explaining the dynamics of y_t at a horizon h.

I employ a bias-corrected R^2 estimator of Gorodnichenko and Lee (2017) who develop flexible methods for estimating the forecast error variance decompositions (FEVDs) with local projections. For the projection $P_{t-1}(\cdot)$ in Equation (3), I use the three shocks and Δy_t at lags 1 to 4.

Unlike other empirical results in this paper, the FEVDs for the FP shock are sensitive to the periods when the Fed targeted the quantity of non-borrowed reserves between 1979 and 1982.[10] Therefore, I plot the results in Figure 3.4 based on the sample both with and without the early Volcker period, where the latter sample spans from 1983:Q1 to 2008:Q4. Except for the sensitivity to the early Volcker period, the results are robust to other modifications in specification (Appendix B.6).

The TFP shock is a major determinant of earnings inequality at a 4-year horizon, explaining about 20-30 percent of the forecast error variances of the inequality index. The FP shock is another important factor. About 20 percent of the forecast error variances of the inequality index at the three to four-year horizons is due to the FP shock after the early Volcker period. Note that the results for the TFP and FP shocks are consistent with the delayed and persistent impulse response functions in Figure 3.2. For the MP shock, the estimated FEVDs are statistically insignificant, similar to the impulse responses in Figure 3.2.

In sum, expansionary fiscal policy shocks raise earnings inequality substantially at the medium-run. On the other hand, earnings inequality does not react to monetary policy shocks, which is contrary to the predictions of some theoretical heterogeneous agent models. This further implies that monetary actions are more suitable when policymaker's objective is to design earnings distribution-neutral stabilization policies. Finally, total factor productivity shocks also have the statistically and economically significant medium-run effects on earnings inequality.

Although macroeconomic factors contribute to cyclical variations in inequality significantly, they have little effects on the short-run dynamics. I will show in Section 3.4 that a considerable fraction of the short-run movements is similarly unpredictable when the information set is substantially extended. The next section investigates a role of this unanticipated

[10]Relatedly, Coibion (2012) and Romer and Romer (2004) find that the estimated effects of the MP shock on output is sensitive to several observations in this period.

variation in the inequality measure as a potential source of business cycles.

3.4 From Earnings Inequality to Business Cycles

The previous section highlights that drivers of business cycles, (especially shocks to TFP and fiscal policy), affect earnings inequality. Now I focus on the other direction, from earnings inequality to business cycles. I show that inequality itself impacts aggregate demand substantially by redistributing economic resources across agents with different MPCs, and so policies are called for to stabilize business cycles.

This section begins with heuristics of how shocks to earnings inequality can be related to aggregate demand shocks. For empirical analyses, I rely on unanticipated innovations in the inequality index, which summarize shocks to individual heterogeneity and redistributive factors in the economy in a parsimonious manner. In response to an unanticipated innovation in inequality that represents redistribution of earnings from the bottom to the top, aggregate variables such as real GDP, price level, and interest rates decline substantially in a U-shaped manner. The signs of the estimated impulse responses imply that redistribution shocks reduce aggregate demand. The forecast error variance decompositions further highlight that these redistributive forces may be an important source of macroeconomic fluctuations.

3.4.1 Inequality, Redistribution, and Aggregate Demand

How can redistribution shocks generate aggregate fluctuations? Rothschild and Stiglitz (1970, 1971) show that a mean-preserving spread can reduce aggregate consumption demand given a concave consumption function despite aggregate earnings remaining the same. Empirical evidence strongly supports the concavity of a consumption function (see Dynan, Skinner and Zeldes, 2004; Johnson, Parker and Souleles, 2006; Parker et al., 2013; Zidar, 2018). Therefore, an inequality shock constitutes a negative demand shock in a system of aggregate variables. Note that two factors are essential for this heuristics. First, the inequality shock reflects redistribution from the bottom to the top. Second, marginal propensity to consume decreases in income.

3.4.2 Unanticipated Innovations in Inequality

To empirically evaluate the mechanism above, I begin with identifying redistribution shocks from time-series variation. Specifically, I use an unanticipated innovation $x_{t,ineq}$ in the inequality index y_t:

$$y_t - y_{t-1} = \Gamma_x' \mathbf{Z}_t^{(x)} + x_{t,ineq}. \tag{3.4}$$

The unanticipated innovation in earnings inequality, or in short, inequality shock, is a component of y_t orthogonal to the information set denoted by $\mathbf{Z}_t^{(x)}$, which includes key macroeconomic variables such as effective federal funds rate (EFFR), inflation, and growth rate of real GDP, consumption, and investment, and the structural shocks in Section 3.3: the TFP, MP, and FP shocks. Throughout this paper, real GDP, consumption, and investment are measured in per capita terms. $\mathbf{Z}_t^{(x)}$ also contains an intercept and 6 lags of Δy_t and the variables above. I include a sufficient number of lags to remove predictable variation as much as possible.

Note that $\mathbf{Z}_t^{(x)}$ has contemporaneous values of the variables except for Δy_t. Thus, the identification of $x_{t,ineq}$ is equivalent to that of a structural vector autoregression model with Cholesky ordering where Δy_t is the last variable. By purging all contemporaneous co-movements, I define $x_{t,ineq}$ in a conservative manner.

An omitted variable bias might be a potential threat to my identification. If there is a demand shock not originating from, but affecting, earnings inequality, this may distort my empirical results. In this regard, I consider three probable confounding factors: shocks to an excess bond premium (EBP), news, and consumer confidence. For the EBP, I add a series built by Gilchrist and Zakrajšek (2012) to $\mathbf{Z}_t^{(x)}$, which is an average corporate bond premiums unrelated to the systematic default risk of individual firms. Identification of a news shock is based on stock prices $\ln S_t$ and TFP_t, similar to Beaudry and Portier (2006). The idea is that a component of the stock price unrelated to current productivity reflects news about the future. Lastly, I employ a measure of Barsky and Sims (2012) on consumer confidence, E5Y. Barsky and Sims show that the E5Y contains information on animal spirits in the sense of Lorenzoni (2009).

Although an uncertainty shock may be another confounding factor, it is unlikely to quantitatively affect my estimates. The identified $x_{t,ineq}$ based on the $\mathbf{Z}_t^{(x)}$ above is orthogonal to the uncertainty shock of Jurado, Ludvigson and Ng (2015). Furthermore, the shocks do not Granger-cause each other (see Appendix C.1).

Figure 3.5 depicts the identified inequality shocks. It follows a white noise process in the sense that the autocorrelations and the partial-autocorrelations at every lag are statistically insignificant. The inequality shock does not Granger-cause the TFP, MP, FP, and uncertainty shocks, and vice versa. Lastly, either including a dummy variable for the early Volcker period in $\mathbf{Z}_t^{(x)}$ or using a sample from 1983:Q1 delivers effectively identical shock series (see Appendix C.1).

While it is not easy to rationalize the realized shocks, some of them have narratives related to distribution of tax changes. The identified series is consistent with leading tax reforms where the shades in Figure 3.5 denote when they are signed into law. For example, the Tax Reform Act of 1986, or Reagan II in Figure 3.5, reduced the top marginal income tax rates from 50% to 28%. Piketty and Saez (2003) note that the earnings distribution widened as a result at least temporarily. It was signed into law in the middle of the fourth quarter of 1986, and $x_{t,ineq}$ was positive in the following quarters. In a similar vein, the Economic Recovery Tax Act of 1981, or Reagan I, lowered the top tax rates from 70% to 50%, and the positive unanticipated innovations followed. Another example is the Omnibus Budget Reconciliation Act of 1993 during the Clinton administration. It raised the top income tax rates from 31% to 39.6% and the negative $x_{t,ineq}$'s in 1993:Q4 and the following quarter may be related to the reform. Lastly, the Jobs and Growth Tax Relief Reconciliation Act of 2003, or the Bush tax cut lowered the top rates from 38.6% to 35%. The positive unanticipated innovations in the third and fourth quarters of 2003 might reflect this change.

3.4.3 Impulse Responses

In the beginning of this section, I raised the hypothesis that more inequality may reduce aggregate demand by redistributing resources from the bottom to the top. Here I empirically evaluate the hypothesis by looking at how key macroeconomic variables respond to the unanticipated innovations in inequality. My results are consistent with the hypothesis in the sense that real GDP, price level, and interest rates decline at the same time in response to $x_{t,ineq}$.

I employ the following local projections to estimate the impulse response functions:

$$m_{t+h} - m_{t-1} = \psi_h^{(m)} x_{t,ineq} + \Gamma_m' \mathbf{Z}_t^{(m)} + u_{t,h}^{(m)}, \tag{3.5}$$

where $\psi_h^{(m)}$ is the parameter of interest, and $\{\psi_h^{(m)} : \ h = 0, 1, \ldots\}$ represents how m_{t+h}

responds to a unit shock in inequality. $\mathbf{Z}_t^{(m)}$ includes macroeconomic variables such as effective federal funds rate, inflation based on GDP deflator, and growth rates of real GDP, consumption, and investment, their lags, lags of $x_{t,ineq}$, and an intercept. Lag length is 6 and the results are robust to the lag specification. When estimating the responses of the inequality index y_t itself in response to the shock $x_{t,ineq}$, lags of Δy_t are further added to $\mathbf{Z}_t^{(m)}$.

The results in Figure 3.6 are consistent with the hypothesis that redistribution shocks reduce aggregate demand. A one standard deviation unanticipated innovation in earnings inequality lowers real GDP by 1.64 percent (annualized) after two years.[11] Similarly, real consumption, investment, and the EFFR decrease. While negative responses of the GDP deflator after 3 to 4 years are weak, this depends on the inclusion of the early Volcker period in the sample. When I exclude those periods from the sample, the estimated peak effect becomes -0.84 percent (annualized) and statistically significant (see Figure C5). The comovement that real GDP, consumption, investment, price level, and the policy rate decrease at the same time is in line with redistribution shocks being negative demand shocks. Note further that these variables react in a U-shaped manner, where the peak level of the responses is reached after about 2 years.

The responses are not only statistically significant, but also economically significant. The magnitudes of the responses are comparable to other prominent structural shocks. For example, a one standard deviation contractionary monetary policy shock of Romer and Romer (2004) reduces real GDP by about 2 percent (annualized) at the peak when estimated similarly (Figure D2). The TFP shock of Fernald (2014) also has a similar peak effect on real GDP (Figure D3). In short, inequality matters for aggregate fluctuations. More inequality increases the amount of slack in an economy by reducing aggregate demand substantially, and the results are robust to various modifications to the baseline specification including different lag length, exclusion of the early Volcker period, and using inequality measures other than the log P90/P10 index (see Appendix C.2).

Straub (2018) notes that aggregate implications of inequality may depend on whether it is based on permanent income or transitory income. Because consumption may be approximately linear in permanent income, rising permanent income inequality may have little imprints on aggregate demand. In this regard, it is intriguing that my unanticipated inno-

[11] Although $x_{t,ineq}$ is a generated regressor, we do not need to adjust the inference when the null hypothesis is of no effect. See Coibion and Gorodnichenko (2012, Appendix D) and Pagan (1984).

vations raise the inequality index only temporarily in Figure 3.6. A one standard deviation innovation in inequality increases the log P90/P10 index approximately by 2 log points (annualized) concurrently, and the responses gradually return to zero similar to an AR(1) process. Thus, my series presumably represents shocks to transitory earnings.

3.4.4 Forecast Error Variance Decompositions

This subsection examines the economic importance of the redistribution shocks as a source of the U.S. business cycles. Specifically, I estimate how much forecast error variances of aggregate variables are attributable to the unanticipated innovations in earnings inequality.

I use a bias-corrected R^2 estimator of Gorodnichenko and Lee (2017) as in Section 3.3.3. The estimates for real GDP at a four-year horizon is 35 percent with the lower bound of its 90 percent confidence band being around 20 percent as depicted in Figure 3.7. The results are similar for real consumption and investment, (25 percent and 20 percent at a four-year horizon, respectively), implying that redistributive forces may be an important driver of aggregate fluctuations. The unanticipated innovations explain large variation of the log P90/P10 index in the short-run, consistent with the impulse responses in Figure 3.6. This further resembles the result in Section 3.3.3 that a significant fraction of the short-run variation in the log P90/P10 index is not predictable by shocks to the TFP, MP, and FP. On the other hand, the EFFR and GDP deflator are mostly driven by other factors. The results are not sensitive to the specification details (see Appendix C.3).[12]

Given the results so far, the main conclusion in Section 3.4 is that the redistribution shocks can reduce aggregate demand substantially in a U-shaped manner. This novel empirical finding leads to natural follow-up questions on mechanisms. The next section develops DSGE models to investigate the amplification and propagation mechanisms of shocks to inequality and illustrate how the shape and magnitude of the empirical impulse responses can be rationalized.

[12] Because it is not easy to estimate FEVDs precisely based on a finite sample, caution needs to be exercised when interpreting the results. In particular, the inequality shock might encompass measurement errors because it is a generated variable. However, Gorodnichenko and Lee (2017) show that measurement errors incur only negative asymptotic biases, and therefore my estimates are conservative in favor of no effect.

3.5 Inequality Shocks in DSGE Models

This section introduces inequality shocks into DSGE models. I show that an inequality shock in a simple two-agent New Keynesian (TANK) model is isomorphic to a discount rate shock in a textbook representative agent New Keynesian (RANK) model. This implies that earnings inequality can be a primitive source of an aggregate demand shock in a representative agent framework. For the quantitative evaluation, I develop the temporarily hand-to-mouth and intertemporal agent New Keynesian (THINK) model. I demonstrate how the model can replicate the large, negative, U-shaped, empirical impulse responses in Section 3.4.

3.5.1 Inequality Shocks in a Simple Two-Agent New Keynesian Model

Suppose that there are two types of households. The first type is a hand-to-mouth agent while the other type can smooth their consumption intertemporally. Following Debortoli and Galí (2017), I call the hand-to-mouth agents Keynesians and the others Ricardians.

The Keynesians are credit constrained and cannot engage in intertemporal optimization. Thus, their consumption is determined by labor earnings:

$$P_t C_t^K = Z_t^K W_t N_t^K, \tag{3.6}$$

where $P_t = \left(\int_0^1 {P_{j,t}}^{1-\epsilon_P} dj\right)^{1/(1-\epsilon_P)}$ is an aggregate price level, W_t is a nominal wage rate, $C_t^K = \left(\int_0^1 \left(C_{j,t}^K\right)^{(\epsilon_P-1)/\epsilon_P} dj\right)^{\epsilon_P/(\epsilon_P-1)}$ is a composite consumption bundle, and Z_t^K denotes labor productivity of the Keynesians. They pick hours of work N_t^K to equate a real wage rate and a marginal rate of substitution:

$$Z_t^K \frac{W_t}{P_t} = \frac{v_N(N_t^K)}{u_C(C_t^K)}, \tag{3.7}$$

where a period utility function is $U(C^K, N^K) = u(C^K) - v(N^K)$, and subscripts C and N denote the first-derivative with respect to C and N, respectively.

On the other hand, the Ricardians maximize $E_t\left[\sum_{\tau=0}^{\infty} \beta^\tau U(C_{t+\tau}^R, N_{t+\tau}^R)\right]$ subject to flow budget constraints:

$$P_{t+\tau}C^R_{t+\tau} + \frac{B^R_{t+\tau}}{1+i_{t+\tau}} = B^R_{t+\tau-1} + Z^R_{t+\tau}W_{t+\tau}N^R_{t+\tau} + \theta^R_D D_{t+\tau} - T_{t+\tau}, \tag{3.8}$$

where C^R_t and N^R_t are consumption and labor supply of the Ricardian agent, B^R_t is an amount of risk-free nominal bonds, i_t is a nominal interest rate, and Z^R_t is productivity of the Ricardians. D_t denotes aggregate dividends, and I assume that each Ricardian agent owns θ^R_D share of all of the firms. $\bar{s}^K$ and $\bar{s}^R$ represent population shares of the Keynesians and Ricardians, and therefore $\theta^R_D = 1/\bar{s}^R$. T_t is lump-sum taxes. The Ricardian's problem leads to the following optimality conditions:

$$1 = E_t\left[\beta \frac{u_C(C^R_{t+1})}{u_C(C^R_t)} \frac{1+i_t}{1+\pi^P_{t+1}}\right], \tag{3.9}$$

$$Z^R_t \frac{W_t}{P_t} = \frac{v_N(N^R_t)}{u_C(C^R_t)}, \tag{3.10}$$

where π^P_t is price inflation.

Usually in TANK models, Z^K_t and Z^R_t are the same, and so earnings inequality is excluded from the analysis. I assume instead that $Z^K_t < Z^R_t$ to introduce distributional factors to the model. As Keynesians earn less and consume a larger fraction of marginal earnings increases than Ricardians, the MPCs decrease in earnings in my model, consistent with empirical evidence.

I assume that the coefficient of RRA of the consumption utility function $u(\cdot)$ at both $\bar{C}^K$ and $\bar{C}^R$ is the same and is denoted by $\gamma = -\frac{u_{CC}(\bar{C}^K)\bar{C}^K}{u_C(\bar{C}^K)} = -\frac{u_{CC}(\bar{C}^R)\bar{C}^R}{u_C(\bar{C}^R)}$, where a variable with a bar means its value at the steady state and double subscripts are for the second-derivative. Similarly, the inverse Frisch elasticity of labor supply at the steady state is denoted by $\varphi = \frac{v_{NN}(\bar{N}^K)\bar{N}^K}{v_N(\bar{N}^K)} = \frac{v_{NN}(\bar{N}^R)\bar{N}^R}{v_N(\bar{N}^R)}$.

Note that aggregate consumption and labor in efficiency unit can be written as follows:

$$C_t = \bar{s}^K C^K_t + \bar{s}^R C^R_t, \tag{3.11}$$

$$N_t = \bar{s}^K Z^K_t N^K_t + \bar{s}^R Z^R_t N^R_t. \tag{3.12}$$

I denote the consumption and labor shares of the Keynesians at the steady state by $\bar{s}^K_C = \frac{\bar{s}^K\bar{C}^K}{\bar{C}}$ and $\bar{s}^K_N = \frac{\bar{s}^K\bar{Z}^K\bar{N}^K}{\bar{N}}$, where $\bar{s}^R_C$ and $\bar{s}^R_N$ are defined accordingly.

Monopolistically competitive firms produce intermediate goods indexed by $j \in [0,1]$. Firm j takes a demand curve $Y_{j,t} = \left(\frac{P_{j,t}}{P_t}\right)^{-\epsilon_P} Y_t$ as given when choosing its price $P_{j,t}$. Profit $D_{j,t}$ is given by $P_{j,t}Y_{j,t} - W_t N_{j,t} - \frac{\psi_P}{2}\left(\frac{P_{j,t}}{P_{j,t-1}} - 1\right)^2 P_t Y_t$ where the last term represents quadratic price-adjustment costs a la Rotemberg (1982). Each firm maximizes $E_t\left[\sum_{\tau=0}^{\infty} Q^D_{t,t+\tau} D_{j,t+\tau}\right]$ subject to the demand curve and a production function $Y_{j,t+\tau} = A_{t+\tau}N_{j,t+\tau}$, where $Q^D_{t,t+\tau} = \beta^\tau \frac{u_C(C^R_{t+\tau})}{u_C(C^R_t)}$. The first-order condition at a symmetric equilibrium is as follows:

$$Y_t - \mathcal{M}_P w_t \frac{Y_t}{A_t} + \frac{\psi_P}{\epsilon_P - 1}\pi^P_t\left(1 + \pi^P_t\right)Y_t - E_t\left[\frac{\psi_P}{\epsilon_P - 1} Q^D_{t,t+1}\pi^P_{t+1}\left(1 + \pi^P_{t+1}\right)^2 Y_{t+1}\right] = 0, \quad (3.13)$$

where $\mathcal{M}_P = \frac{\epsilon_P}{\epsilon_P - 1}$ is the steady state markup and w_t is a real wage rate.

Finally, the model's aggregate resource constraint and policy rule for the central bank are standard:

$$Y_t = C_t + G_t + \frac{\psi_P}{2}\left(\pi^P_t\right)^2 Y_t, \quad (3.14)$$

$$i_t = (1-\rho_i)\bar{i} + \rho_i i_{t-1} + (1-\rho_i)\left(\zeta_\pi \pi^P_t + \zeta_x x_t\right) + \sigma_i u^i_t, \quad (3.15)$$

where G_t represents government expenditure, x_t is an output gap $\check{Y}_t - \check{Y}^n_t = \log(Y_t/\bar{Y}) - \log(Y^n_t/\bar{Y})$, and Y^n_t is the level of output when prices are fully flexible. Similarly, other variables with a check mean log-deviations from their steady state values. I further define ϕ_C as $\bar{C}/\bar{Y}$ and ϕ_G as $\bar{G}/\bar{Y}$.

An inequality shock is an exogenous force decreasing Z^K_t and increasing Z^R_t in such a way that $\bar{s}^K_N \check{Z}^K_t + \bar{s}^R_N \check{Z}^R_t = 0$ for all t. Thus, this shock that increases earnings inequality is a mean-preserving spread when working hours are equal to the steady state level. The first-order dynamics of the model can be described by three equations (15)-(17):

$$x_t = E_t\left[x_{t+1}\right] - \frac{1}{\tilde{\gamma}}\left(i_t - E_t\left[\pi^P_{t+1}\right] - r^n_t\right), \quad (3.16)$$

$$\pi^P_t = \beta E_t\left[\pi^P_{t+1}\right] + \tilde{\lambda} x_t, \quad (3.17)$$

where r^n_t is the real interest rate under flexible prices, $\tilde{\lambda} = \frac{\epsilon_P - 1}{\psi_P}\Delta$, $\tilde{\gamma} = \gamma\left(1 - \bar{s}^K_C \phi_C \frac{1+\varphi}{\gamma+\varphi}\Delta\right) / \left(\bar{s}^R_C \phi_C\right)$, and $\Delta = \left(\varphi + \frac{\bar{s}^R_N \gamma}{\bar{s}^R_C \phi_C}\right) / \left[1 - \left(\bar{s}^K_N - \frac{\bar{s}^R_N}{\bar{s}^R_C}\bar{s}^K_C\right)\frac{\gamma(1+\varphi)}{\gamma+\varphi}\right]$. The derivations of the equations are in Appendix D.1. Note that these equations are observationally equivalent to a standard three-equations NK model of Galí (2015) and Woodford (2003).

Inequality matters in this model in two ways. First, distributional parameters $\bar{s}_C^K$, $\bar{s}_N^K$, $\bar{s}_C^R$, and $\bar{s}_N^R$ affect how shocks are propagated by changing the slopes of the dynamic IS equation (16) and the Phillips curve (17). Second, the inequality shock has an effect on $\check{Y}_t^n$ and r_t^n.

I start with the slopes. While $\tilde{\lambda}$ and $\tilde{\gamma}$ are related to the distributional parameters in a complicated manner, there is an interesting special case when the consumption share and the earnings share of the Keynesians are the same, *i.e.*, $\bar{s}_C^K = \bar{s}_N^K$.[13] In this case, $\Delta = \varphi + \frac{\gamma}{\phi_C}$ and $\tilde{\lambda}$ is independent of the distributional parameters. However, the slope of the dynamic IS equation still depends on inequality. When $\bar{s}_C^K = 0$ (*i.e.*, there are no Keynesians), the aggregate elasticity of intertemporal substitution (EIS) $\frac{1}{\tilde{\gamma}}$ is $\frac{\phi_C}{\gamma}$, recovering the RANK model of Woodford (2003, p.80). As $\bar{s}_C^K$ increases, $\tilde{\gamma}$ decreases or the aggregate EIS increases, if the coefficient of RRA γ is greater than 1.[14] This implies that the presence of hand-to-mouth agents amplifies the effects of real interest rates on aggregate demand.[15]

To study the effects of the inequality shocks, suppose that $\check{Z}_t^K$ follows an AR(1) process, $\check{Z}_t^K = \rho_Z \check{Z}_{t-1}^K - \sigma_Z u_t^Z$, where $0 < \rho_Z < 1$. The mean-preserving spread assumption, $\bar{s}_N^K \check{Z}_t^K + \bar{s}_N^R \check{Z}_t^R = 0$, implies that that $\check{Z}_t^R = \rho_Z \check{Z}_{t-1}^R + \sigma_Z \frac{\bar{s}_N^K}{\bar{s}_N^R} u_t^Z$, and so u_t^Z is a redistribution shock increasing earnings inequality. This shock affects the economy via both $\check{Y}_t^n$ and r_t^n. However, when $\bar{s}_C^K = \bar{s}_N^K$, $\check{Y}_t^n$ becomes unrelated to the inequality shocks, and therefore u_t^Z propagates only through the natural rate of interest r_t^n.[16] One can show that

$$\frac{\partial E_t[r_{t+\tau}^n]}{\partial u_t^Z} = -\rho_Z^\tau \frac{\bar{s}_C^K}{\bar{s}_C^R} \frac{1+\varphi}{\gamma+\varphi} \gamma (1-\rho_Z)\sigma_Z \qquad < \qquad 0. \tag{3.18}$$

Note that this resembles how a contractionary discount rate shock in a RANK model works: when utility in the future is less discounted, r_t^n decreases and a representative agent consumes less as consumption in the future becomes more important. Therefore, the inequality shock in the simple TANK model is isomorphic to a demand shock in a RANK model. This further illustrates why individual heterogeneity can be a source of aggregate demand shocks in a representative agent framework.

Intuitively, $\check{C}_t^K$ is similar to $\check{Z}_t^K$, because the Keynesians are hand-to-mouth. On the

[13] A sufficient condition for $\bar{s}_C^K = \bar{s}_N^K$ is $\phi_G = \frac{\bar{G}}{\bar{Y}} = \frac{1}{\epsilon_P}$. When the steady state price markup is 20 percent (Rotemberg and Woodford, 1997), this corresponds to ϕ_G being equal to 17 percent.

[14] Precisely, the condition is that $\gamma + \phi_C(1+\varphi) > 1$.

[15] When $\bar{s}_C^K$ is very large, $\tilde{\gamma}$ becomes negative and an inverted aggregate demand logic of Bilbiie (2008) prevails.

[16] When $\bar{s}_C^K \neq \bar{s}_N^K$, the inequality shock has a supply-side effect of altering $\check{Y}_t^n$. However, this effect is quite small as long as $\bar{s}_C^K$ is close to $\bar{s}_N^K$. See Appendix D.1 for an analysis of this general case.

other hand, the Ricardians want to smooth their consumption intertemporarily, and so $\check{C}_t^R$ is less volatile than $\check{Z}_t^R$. Given a large decrease in $\check{C}_t^K$ and a small increase in $\check{C}_t^R$ in response to the inequality shock, there are some negative responses in aggregate consumption $\check{C}_t$. This illustrates why u_t^Z is a negative aggregate demand shock.

Although the simple TANK model above is useful to build intuition, it cannot quantitatively rationalize the U-shaped responses of aggregate consumption estimated in Section 3.3. In the simple TANK model, $\check{C}_t$ decreases contemporaneously and returns monotonically to zero. While introducing habit formation in preferences for consumption is useful to induce hump-shaped dynamics in response to monetary policy shocks (Woodford, 2003), this is not the case for the inequality shocks. Because the Keynesians consume all of their labor earnings every period, habit formation does not play a central role and $\check{C}_t^K$ closely follows $\check{Z}_t^K$. While the dynamics of $\check{C}_t^R$ are affected by the consumption habit, it responds positively given an inequality shock that increases $\check{Z}_t^R$. By combining negative AR(1)-like dynamics of $\check{C}_t^K$ and positive hump-shaped responses of $\check{C}_t^R$, it is less likely that $\check{C}_t$ would decline in a U-shaped manner. To rationalize the empirical impulse responses and understand the mechanisms through which inequality shocks propagate, further enhancement is required in a model.

3.5.2 The THINK Model and Its Quantitative Evaluation

The previous subsection analytically shows that the inequality shock reduces aggregate consumption demand. Here I examine the effects of the inequality shock quantitatively using a two-agent medium-sized DSGE model building on Christiano, Eichenbaum and Evans (2005) and Smets and Wouters (2007). The model combines temporarily hand-to-mouth and intertemporal (THI) agents and New Keynesian (NK) characteristics. When estimated by a Bayesian impulse response matching method of Christiano, Trabandt and Walentin (2010), the THINK model successfully generates large, U-shaped impulse responses comparable to the empirical ones.

Model

The THINK model extends the simple TANK model in several aspects. For individual heterogeneity, three new features are introduced: an endogenous extensive margin between the Keynesian and the Ricardian "families," a small amount of financial income for the

Keynesians, and a decreasing relative risk aversion (DRRA) consumption utility. Various characteristics of medium-sized RANK models are further incorporated such as investment and capital utilization adjustment costs, sticky wages, and habit formation in consumption preferences.

Keynesian and Ricardian Families I introduce an extensive margin of being a credit constrained or unconstrained agent in the model, which makes the population shares of both families determined endogenously. Suppose that s_t^K and s_t^R are the number of members in each family in period t. The transition probability of becoming a Keynesian in period t among who were a Ricardian in period $t-1$ is denoted by q_t^{RK} and the other transition probabilities are denoted accordingly. The Keynesian family consists of agents who were either a Keynesian or a Ricardian in the previous period:

$$s_t^K = s_{t-1}^K q_t^{KK} + s_{t-1}^R q_t^{RK}. \tag{3.19}$$

It is clear that $q_t^{KR} = 1 - q_t^{KK}$, $q_t^{RR} = 1 - q_t^{RK}$, and $s_t^R = 1 - s_t^K$.

I assume that the probability of staying in the Keynesian family for an agent who was a Keynesian in the previous period is as follows:

$$q_t^{KK} = \bar{q}^{KK} \left(\frac{Y_t}{\bar{Y}} \right)^{-\eta_Y} \left(\frac{s_{t-1}^K}{\bar{s}^K} \right)^{-\eta_s}, \qquad \eta_Y \geq 0 \quad \text{and} \quad \eta_s \in \mathbb{R}. \tag{3.20}$$

For special cases, the type of an agent is fixed when $\bar{q}^{KK} = 1$, $\eta_Y = 0$, $\eta_s = 0$, and $q_t^{RR} = 1$. If $\bar{q}^{KK} = \bar{s}^K$, $\eta_Y = 0$, $\eta_s = 0$, and $q_t^{RR} = \bar{s}^R$, agents are credit constrained in an identically and independently distributed manner. The parameter η_Y governs the cyclicality of q_t^{KK}. As documented by Guvenen, Ozkan and Song (2014), Ravn and Sterk (2017), and Storesletten, Telmer and Yaron (2004), unemployment risk and idiosyncratic earnings risk are countercyclical. This implies that more people receive large negative idiosyncratic shocks and become credit constrained during recessions. A positive η_Y captures such dynamics by increasing q_t^{KK} when output Y_t is low. That is, it is hard to escape from a credit constraint during economic downturns. On the other hand, η_s influences the persistence of the number of credit constrained agents. For example, when $s_{t-1}^K > \bar{s}^K$, a positive η_s lowers the probability of staying in the Keynesian family, which increases the degree of mean-reversion in the number of credit constrained agents s_t^K.

The parameter η_Y can be micro-founded as follows. Suppose that earnings of agents who were credit constrained in the previous period are represented by an inverse Pareto distribution $\upsilon_{i,t}^{-1} Y_t$, where $\upsilon_{i,t} \sim \text{Pareto}(\eta_Y)$ for $\upsilon_{i,t} \geq \upsilon_m$. I assume further that one needs to earn more than a threshold to circumvent the credit constraint, where the threshold is an aggregate variable. In this setup, q_t^{KK} becomes proportional to $Y_t^{-\eta_Y}$. Intuitively, an increase in aggregate income affects individual earnings positively and this leads to fewer credit constrained agents. That is, "a rising tide lifts all boats." Furthermore, η_s can be related to the (negative) elasticity of the threshold earnings to the number of credit constrained agents at the steady state. For example, consider a case where there are more credit constrained agents than in the steady state. The additional credit constrained agents would have enough resources not to be constrained at the steady state, and therefore they are likely to be wealthier on average than those who would be credit constrained at the steady state. Because those additional credit constrained agents can sell illiquid assets for cash or pledgeable collateral, less earnings may be enough for these agents to escape from credit constraints. While such actions are not explicitly modeled here, a positive η_s reflects this channel by lowering the threshold earnings and making more agents circumvent credit constraints. On the other hand, banks may become reluctant to issue additional loans to households when many households are already borrowing from banks. Some of the potential borrowers may have poor credit condition, and therefore banks may have to pay additional efforts in screening. This implies that more earnings are required for some consumers not to be credit constrained. If this channel is important, η_s may be negative. Because the sign of η_s is not clear a priori, I let the support of this parameter include both positive and negative values for now and let the estimation later pin down a value. See Appendix D.2.3 for details of the microfoundation.

While one can impose a similar structure on q_t^{RK}, I shut down this channel and let $q_t^{RK} = \bar{q}^{RK}$ and $q_t^{RR} = \bar{q}^{RR}$. This is to keep the model parsimonious and keep my analysis focused. Furthermore, several log points deviations of q_t^{RK} from $\bar{q}^{RK}$ have little effect on s_t^K because $\bar{q}^{RK}$ is small in the benchmark parameters.[17]

I assume that each Keynesian receives a positive share θ_D^K of dividends. This is because even the wealth-poor households have some financial investments (Guiso and Sodini, 2013).

[17] Specifically, the log-linearized Equation (19) is that $\check{s}_t^K = \bar{q}^{KK}(\check{s}_{t-1}^K + \check{q}_t^{KK}) + \bar{q}^{KR}(\check{s}_{t-1}^R + \check{q}_t^{RK})$. Because $\check{s}_{t-1}^R = -\frac{\bar{s}^K}{\bar{s}^R}\check{s}_{t-1}^K$, the contribution of the time-varying $\check{q}_t^{RK}$ to $\check{s}_t^K$ depends on $\bar{q}^{KR} = \bar{q}^{RK}\frac{\bar{s}^R}{\bar{s}^K}$. Both $\bar{q}^{RK}$ and $\bar{q}^{KR}$ are tiny in the benchmark calibration, and therefore $\bar{q}^{KR}\check{q}_t^{RK}$ is negligible.

One may also consider this income as including government transfers to the poor, or pensions. Finally, Kaplan, Violante and Weidner (2014) and Kaplan and Violante (2014) argue that there are agents who own a significant amount of assets but credit constrained because most of their wealth is illiquid. It is natural to suppose that such agents are credit constrained but receive some financial income.

Similarly, each Ricardian holds $\theta_{D,t}^R$ share of the stocks. Because the population shares of two families are time-varying, at least one of θ_D^K and $\theta_{D,t}^R$ should be also time-varying to satisfy $s_t^K \theta_D^K + s_t^R \theta_{D,t}^R = 1$. For simplicity, I fix θ_D^K and let $\theta_{D,t}^R$ be determined by s_t^K and s_t^R. Note that $\theta_{D,t}^R = \frac{1-\theta_D^K}{1-s_t^K} + \theta_D^K$, and therefore $\theta_{D,t}^R$ increases in s_t^K. This implies that financial assets are concentrated among fewer people (high $\theta_{D,t}^R$) in a recession when more agents are credit constrained (high s_t^K). Indeed, the correlation between the HP filtered top 10% wealth share of Saez and Zucman (2016) and log real GDP per capita is -0.26. Finally, I assume that a condition $\theta_D^K \leq \theta_{D,t}^R$ holds in all cases I study.

There is a continuum of agents in both families supplying different types of labor in a monopolistically competitive way. Subject to quadratic wage adjustment costs in a nominal wage inflation $\pi_{l,t}^W$, a Keynesian worker has the following budget constraint:

$$P_t C_{l,t}^K = Z_t^K W_{l,t} N_{l,t}^K - \frac{\psi_W}{2} \left(\pi_{l,t}^W\right)^2 Z_t^K W_t N_t^K + \theta_D^K D_t. \tag{3.21}$$

The decisions on the wage rate and the hours are relegated to the type l labor union. The consumption utility $u\left(C_{l,t}^K - b^K C_{t-1}^K\right)$ features an external habit. The (negative) elasticity of the marginal utility function at the steady state is denoted by $\gamma^K = -\frac{u_{CC}(\bar{C}^K - b^K \bar{C}^K) \times (\bar{C}^K - b^K \bar{C}^K)}{u_C(\bar{C}^K - b^K \bar{C}^K)}$.

When a Ricardian becomes a Keynesian, one brings θ_D^K share of the stocks and leave all the other assets to the Ricardian family. On the other hand, when a Keynesian becomes a Ricardian, one carries all the wealth to the new family. Those assumptions make each Keynesian hold θ_D^K share while the number of Keynesians is not a constant. For equalizing financial resources available to the new and continuing Ricardians, an intra-family redistribution occurs, which is in a lump sum. A budget constraint for a type l Ricardian worker is given by

$$P_t C_{l,t}^R + \frac{B_{l,t}^R}{1+i_t} = B_{l,t-1}^R + Z_t^R W_{l,t} N_{l,t}^R - \frac{\psi_W}{2} \left(\pi_{l,t}^W\right)^2 Z_t^R W_t N_t^R + \theta_{D,t}^R D_t - T_t + R_t, \tag{3.22}$$

where R_t denotes the lump sum redistribution inside the Ricardian family.

I define b^R and γ^R similar to their Keynesian counterparts. Note that b^R and γ^R are not necessarily equal to b^K and γ^K. I instead consider decreasing relative risk aversion preferences in accordance with the empirical results of Calvet, Campbell and Sodini (2009, Section IV.C). In the model, at the steady state, this corresponds to a condition that $\frac{\gamma^K}{1-b^K} \geq \frac{\gamma^R}{1-b^R}$. Agents are more relative risk averse in the Keynesian family where they consume less than in the Ricardian family.

Labor Market Next, I turn to the labor market. I introduce labor unions whose preferences are based only on aggregate variables. By minimizing the effects of individual heterogeneity on the decision of labor unions, I can make the resulting wage Phillips curve similar to the RANK counterpart. This allows me to focus on the new features in the demand block, while reducing deviations in the supply block from the RANK model. I also show that a competitive labor market induces negatively correlated earnings inequality and consumption inequality, which is at odds with data. This finding further necessitates a non-competitive labor market institution such as labor unions.

A labor union for type l workers chooses a nominal wage rate $W_{l,t}$ and supplies $N_{l,t} = \left(\frac{W_{l,t}}{W_t}\right)^{-\epsilon_W} N_t$ in efficiency units, while taking W_t and $N_t = \left(\int_0^1 N_{l,t}^{\frac{\epsilon_W - 1}{\epsilon_W}} dl\right)^{\frac{\epsilon_W}{\epsilon_W - 1}}$ as given. The union should determine how to allocate the total labor $N_{l,t}$ to the Keynesian and the Ricardian workers. I assume that the union makes working hours of each agent be in proportion to their steady state values. Thus, $N_{l,t} = s_t^K Z_t^K N_{l,t}^K + s_t^R Z_t^R N_{l,t}^R$ where $\frac{N_{l,t}^K}{\bar{N}^K} = \frac{N_{l,t}^R}{\bar{N}^R}$ for all l and t. For example, when both agents work the same number of hours at the steady state ($\bar{N}^K = \bar{N}^R$), the type l union always assigns common labor hours to both Keynesians and Ricardians workers ($N_{l,t}^K = N_{l,t}^R$).

The utility function of the union l is denoted by U_l^L. Following Pencavel (1984), it is based on the total real earnings $e_{l,t}$ and the total labor supply $N_{l,t}$: $U_{l,t}^L = u^L\left(e_{l,t} - b^L e_{t-1}\right) - v^L\left(N_{l,t}\right)$ where $e_{l,t} = \left(W_{l,t}N_{l,t} - \frac{\psi_W}{2}\left(\pi_{l,t}^W\right)^2 W_t N_t\right)/P_t$ and $e_t = \left(W_t N_t - \frac{\psi_W}{2}\left(\pi_t^W\right)^2 W_t N_t\right)/P_t$.[18] Note that the earnings are subject to quadratic nominal wage adjustment costs, which induce sticky wages. The utility function u^L features an external habit and the elasticities of the marginal utilities u_e^L and v_N^L with respect to their inputs at the steady state are denoted by

[18] An alternative setup is to make labor unions maximize the average utility of members $E_t\left[\sum_{\tau=0}^{\infty} \beta^\tau (s_{t+\tau}^K U_{l,t+\tau}^K + s_{t+\tau}^R U_{l,t+\tau}^R)\right]$ where $U_{l,t}^\iota = u\left(C_{l,t}^\iota - b^\iota C_{t-1}^\iota\right) - v\left(N_{l,t}^\iota\right)$ for all ι, l, and t. However, the model impulse responses under this setup are similar to what are obtained in the benchmark case. See Appendix D.2.2.2.

γ^L and φ, respectively. The first-order condition at a symmetric equilibrium for maximizing $E_t\left[\sum_{\tau=0}^{\infty}\beta^\tau U^L_{l,t+\tau}\right]$ yields a standard wage Phillips curve, where the wage markup equals to $\frac{w_t}{v^L_N/u^L_e}$. Finally, the steady state wage markup is denoted by $\mathcal{M}_W = \frac{\epsilon_W}{\epsilon_W-1}$.

One may instead consider a competitive labor market. For a simple exposition, suppose that $\gamma^K = \gamma^R = \gamma$ and $b^K = b^R = 0$. Then individual labor supply schedule becomes Equation (8) and (10), or $\check{Z}^\iota_t + \check{w}_t = \varphi\check{N}^\iota_t + \gamma\check{C}^\iota_t$ for $\iota \in \{K, R\}$ in log-linearization. For structural shocks not affecting $\check{Z}^R_t$ or $\check{Z}^K_t$ directly, we have $\gamma(\check{C}^R_t - \check{C}^K_t) = -\varphi(\check{N}^R_t - \check{N}^K_t)$. This implies that consumption inequality $\log\left(\frac{C^R_t}{C^K_t}\right)$ and earnings inequality $\log\left(\frac{Z^R_t N^R_t}{Z^K_t N^K_t}\right)$ are negatively correlated. Such prediction contradicts empirical findings of Coibion et al. (2017) that consumption inequality increases in response to a contractionary monetary policy shock, whereas earnings inequality is unresponsive. Therefore, I rule out a competitive labor market and take another widely used setup, labor unions and sticky wages.

Firms, Monetary Policy, and Aggregate Resource Constraint Here the production block is briefly discussed because it is similar to that of the RANK, where the details are relegated to Appendix D.2. A firm j selects four variables: a price $P_{j,t}$, a labor input $N_{j,t}$, investment $I_{j,t}$, and a capital utilization rate $\nu_{j,t}$ given $P_{j,t-1}$, $I_{j,t-1}$, and a firm-specific physical capital stock $\kappa_{j,t}$. A nominal profit, which is paid out as dividends, is given by:

$$D_{j,t} = P_{j,t}Y_{j,t} - W_t N_{j,t} - \frac{\psi_P}{2}\left(\pi^P_{j,t}\right)^2 P_t Y_t - P_t I_{j,t} - \Phi^\nu(\nu_{j,t})P_t\kappa_{j,t}, \tag{3.23}$$

where the demand for good j is given by $Y_{j,t} = \left(\frac{P_{j,t}}{P_t}\right)^{-\epsilon_P} Y_t$. The third term represents nominal price adjustment costs, which make prices sticky. There are investment adjustment costs and the law of motion for the physical capital stock is $\kappa_{j,t+1} = (1-\delta)\kappa_{j,t} + \left(1 - \Phi^I\left(\frac{I_{j,t}}{I_{j,t-1}}\right)\right) I_{j,t}$. The last term in Equation (23) is for adjustment costs to capital utilization. Capital service $K_{j,t}$ is determined by the utilization rate and the physical capital stock: $K_{j,t} = \nu_{j,t}\kappa_{j,t}$. Finally, a Cobb-Douglas production function is assumed: $Y_{j,t} = A_t K^{1-\alpha}_{j,t} N^\alpha_{j,t}$.

The firm j maximizes the discounted dividends $E_t\left[\sum_{\tau=0}^{\infty} Q^D_{t,t+\tau} D_{j,t+\tau}\right]$ subject to the constraints above. The stochastic discount factor $Q^D_{t,t+\tau}$ is based on the marginal consumption utilities weighted by the time-varying population and equity shares:

$$Q^D_{t,t+\tau} = \beta^\tau \frac{s^K_{t+\tau}\theta^K_D u^K_{C,t+\tau} + s^R_{t+\tau}\theta^R_{D,t+\tau}u^R_{C,t+\tau}}{s^K_t\theta^K_D u^K_{C,t} + s^R_t\theta^R_{D,t}u^R_{C,t}} \frac{P_t}{P_{t+\tau}}. \tag{3.24}$$

Finally, an aggregate resource constraint and a policy rule for the central bank are as follows:

$$Y_t = C_t + I_t + G_t + \frac{\psi_P}{2}\left(\pi_t^P\right)^2 Y_t + \frac{\psi_W}{2}\left(\pi_t^W\right)^2 N_t w_t + \Phi^\nu(\nu_t)\kappa_t, \tag{3.25}$$

$$i_t = (1-\rho_i)\bar{i} + \rho_i i_{t-1} + (1-\rho_i)(\zeta_\pi \pi_t^P + \zeta_Y \check{Y}_t) + \sigma_i u_t^i. \tag{3.26}$$

Calibration and Estimation

The THINK model has many parameters. This subsection discusses how those parameters are calibrated and estimated. I calibrate parameters when there are commonly used values or when the empirical impulse responses are less informative about them. This sharpens identification of the estimator by reducing the number of parameters to be estimated substantially. Then I estimate the other parameters including newly introduced ones such as η_Y, η_s, γ^K, and the ratio of the marginal consumption utilities $\bar{u}_C^K/\bar{u}_C^R$ at the steady state, where $\bar{u}^\iota \equiv u(\bar{C}^\iota - b^\iota \bar{C}^\iota)$ for $\iota \in \{K, R\}$. For estimation, I match the empirical and the model impulse responses in a Bayesian framework following Christiano, Trabandt and Walentin (2010). This limited information approach allows me to focus on several shocks of interests, while not being specific about the remaining part of the data generating process.

A list of parameters and their calibrated values can be found in Table 3.2. For example, $\beta = 0.99$ is determined by the steady state investment share $\phi_I \equiv \bar{I}/\bar{Y}$. Following Debortoli and Galí (2017), I assume that one-fifth of the population are Keynesian in the steady state. The consumption and earnings share of the Keynesians are based on those of the bottom quintile households sorted by wealth.[19] For the transition probability from the Ricardian family to the Keynesian family at the steady state $\bar{q}^{RK}$, I note that Equation (19) and (20) lead to

$$\check{s}_t^K = (\bar{q}^{KK} - \bar{q}^{KK}\eta_s - \bar{q}^{RK})\check{s}_{t-1}^K - \bar{q}^{KK}\eta_Y \check{Y}_t, \tag{3.27}$$

where $\bar{q}^{KK} = 1 - \bar{q}^{KR} = 1 - \frac{\bar{s}^R}{\bar{s}^K}\bar{q}^{RK}$. Because η_Y and η_s appear only in the above equation in the log-linearized system, I fix $\bar{q}^{RK}$ and estimate η_Y and η_s, where η_Y and η_s govern the cyclicality and the persistence of $\check{s}_t^K$, respectively. Assuming $\bar{q}^{RK} = 0.0025$ implies that 4.5 percent of the Ricardians will transition to the Keynesians after 5 years at the steady state transition rates. This is similar to the transition probability from positive to strictly negative

[19] In light of the wealthy hand-to-mouth agents of Kaplan and Violante (2014), I later consider a case with higher $\bar{s}_C^K$ and $\bar{s}_N^K$. With different parameter estimates, the model generates similar dynamics of aggregate variables. (see Appendix D.2.2).

net worth in the Panel Study of Income Dynamics. For example, between 1984 and 1989 (1989 and 1994), the transition probability was 4.4 (4.7) percent in the data.

For adjustment cost parameters ψ_P and ψ_W, I make the slopes of the price and wage Phillips curves equal those implied by 75 percent of Calvo (1983) probabilities of non-adjustment. Finally, using $\bar{s}^K \bar{C}^K = \bar{s}_C^K \phi_C \bar{Y}$, $\bar{s}^K \bar{Z}^K \bar{N}^K \bar{w} = \bar{s}_N^K \mathcal{M}_P^{-1} \alpha \bar{Y}$, and Equation (21), one can show that $\theta_D^K = 0.44$. Because $\bar{s}^K = 0.2$, this means that about 9 percent of the total financial income goes to the Keynesians. This may reflect the presence of wealthy hand-to-mouth agents among the Keynesians, financial income taxation and government transfers, or pensions.

Next, I explain how the remaining parameters are estimated. Suppose that Ψ is a vector of impulse response coefficients of interests and an estimator $\hat{\Psi}$ has an asymptotic normal distribution $N(\Psi, V/T)$, where T is a sample size. The model parameters and the model-implied impulse responses are denoted by Θ and $\Psi(\Theta)$, respectively. A limited information likelihood $p(\hat{\Psi}|\Theta)$ of Kim (2002) is based on the distribution that $\hat{\Psi} \sim N(\Psi(\Theta), \hat{V}/T)$, where $\hat{V}$ is a diagonal matrix following Christiano, Eichenbaum and Evans (2005) and Christiano, Trabandt and Walentin (2010). Given the likelihood and a prior on Θ, we can think of a posterior $p(\Theta|\hat{\Psi}) = p(\hat{\Psi}|\Theta)p(\Theta)/p(\hat{\Psi})$.

Ψ consists of responses of real GDP, consumption, investment, the GDP deflator, and EFFR to a one standard deviation shock to the inequality, MP, and TFP. I estimate the responsesof major macroeconomic variables to the MP and TFP shocks using local projections similar to Equation (5), where contemporaneous variables in $\mathbf{Z}_t^{(m)}$ are included only when the minimum delay assumption is relevant for the identification. Lags of the impulse response functions that are included in Ψ might matter for the estimation of Θ. Because the minimum delay assumptions are imposed for the inequality and MP shocks in the data, the contemporaneous responses are nil by construction. Obviously, this assumption may have further effects at short lags. On the other hand, the TFP shocks are free of such concerns, and the responses at short lags may provide useful information on the short-run dynamics. However, including all the lags only for the TFP shocks might overweight the impulse responses induced by the TFP shocks. Given these consideration, I use the responses at lags 0, 4, 5, $\ldots$, 12 and the initial responses are dropped when the minimum delay assumption is used for the identification. Because there are five variables and three shocks, Ψ includes $5 \times (13 - 3) \times 3 - 9 = 141$ moments in total. To simulate the posterior, I draw 200,000 observations using a random walk Metropolis-Hastings algorithm and drop the first 50,000 observations (see Herbst and Schorfheide, 2015). The acceptance rate of the Markov chain is 34 percent.

The results are summarized in Table 3.3. The Keynesian consumption habit parameter b^K is assumed to be 0 because pre-MCMC numerical optimizations assign 0 to b^K when it is also estimated. γ^K is 8.39 at the posterior mode, which is much greater than $\gamma^R = 2$. This is consistent with the view that agents become more risk averse when consuming less and being credit constrained. In line with this view and my results, Guiso and Sodini (2013) report that the 90th percentile of a cross-section of coefficients of RRA in the U.S. is 16.4. Given that the population share of the Keynesians is 20 percent and coefficients of RRA decrease in wealth, my estimate seems reasonable. Note also that '$\gamma^K - 1$' follows a gamma distribution a priori, and so γ^K estimates are always larger than 1.

$\Phi^{\nu}_{\nu\nu}(\bar{\nu})$ denotes the second derivative of the utilization adjustment cost function Φ^{ν} that is evaluated at the steady state. Christiano, Eichenbaum and Evans (2005) fix it at 0.000457, while the estimates of Justiniano, Primiceri and Tambalotti (2010, 2011) are about 0.15.[20] My estimate is 0.23, closer to that of Justiniano, Primiceri and Tambalotti. The second derivative of the investment adjustment costs $\Phi^{I}_{II}(1)$ is 1.64, much smaller than 2.48 in Christiano, Eichenbaum and Evans or 3.14 in Justiniano, Primiceri and Tambalotti (2011). Relatedly, the estimated process of the technology shock is less persistent than usual. For example, a half-life of the technology shock is 3 quarters given $\rho_A = 0.77$, while it takes more than 3 years given a standard estimate around 0.95. Thus, the THINK model can generate hump-shaped and persistent responses successfully with small investment adjustment costs and less persistent inputs.

The estimated ratio of the marginal consumption utilities at the steady state $\bar{u}^K_C/\bar{u}^R_C$ is 3.99. This means that the Keynesian values a marginal consumption good much more than the Ricardian. The support of η_s is $(-1, \infty)$ including negative values because the sign of η_s is ambiguous a priori as discussed before. Given the estimates for η_Y and η_s, Equation (27) becomes $\check{s}^K_t = 0.48\check{s}^K_{t-1} - 4.27\check{Y}_t$. Thus, a recession with a 1 percent decline in output increases the number of the Keynesians by $4.27 \times \bar{s}^K = 0.85$ percentage points.

My interpretation of the magnitude of σ_Z, which represents how much $\check{Z}^K_t$ decreases concurrently in response to a one standard deviation inequality shock, is as follows. Sup-

[20] $\Phi^{\nu}_{\nu}(\bar{\nu})$ and $\frac{\Phi^{\nu}_{\nu\nu}(\bar{\nu})}{\Phi^{\nu}_{\nu}(\bar{\nu})}$ in Justiniano, Primiceri and Tambalotti (2010) are about 0.03 and 5, respectively, implying that $\Phi^{\nu}_{\nu\nu}(\bar{\nu})$ is about 0.15, where Φ^{ν}_{ν} is the first derivative of Φ^{ν}. Phaneuf, Sims and Victor (2018) do a similar algebra for estimates in Christiano, Eichenbaum and Evans (2005) and obtain 0.000457.

pose that there exists a continuum of agents whose idiosyncratic log labor productivity is denoted by $z_{i,t}$. Let $\sigma(z_{i,t})$ be its cross-sectional standard deviation and y_t be the log P90/P10 index. If $z_{i,t}$ is normally distributed, $\sigma(z_{i,t}) = \frac{y_t}{2N^{-1}(0.9)}$, where $N^{-1}(\cdot)$ is the inverse cumulative distribution function of the standard normal distribution. It is because $\log(P90/P50)$ and $\log(P50/P10)$ equal $N^{-1}(0.9)$ times the standard deviation for a normally distributed random variable. Furthermore, the population share of the Keynesians is 0.2, and so their productivity can be represented by the 10th percentile of $z_{i,t}$. This implies that $\log(Z_t^K) \approx N^{-1}(0.1)\sigma(z_{i,t})$.[21] By combining these two equations, I obtain an expression for Z_t^K and y_t that

$$\log(Z_t^K) \approx -\frac{y_t}{2}. \tag{3.28}$$

As shown in Figure 3.6, y_t increases by 2 log points (annualized) or 0.5 log points at the quarterly frequency in response to a one standard deviation inequality shock. This translates into a 25 log basis points decrease in $\log(Z_t^K)$ according to Equation (28), which is similar to the posterior mode of σ_Z, 30 log basis points.

Figure 3.8 illustrates how major macroeconomic variables respond to a one standard deviation inequality shock in the model, when Θ equals the posterior mode. For the earnings inequality y_t in the bottom right panel, I use Equation (28) and plot $-2E_t\left[\check{Z}_{t+h}^K\right] \times 400$. The fit of the model is reasonably good in the sense that the peak effects and the shapes are similar. Although the model impulse responses have the peaks earlier than the empirical impulse responses, it is well known that it is hard to generate much delayed responses with purely forward-looking Phillips curves. Furthermore, the estimated model is good at replicating the empirical responses to the MP and TFP shocks (see Appendix D.2.2).

Dupor, Han and Tsai (2009) point out that estimated parameters by matching impulse responses heavily depend on which shock is studied. Similarly, I find that the estimates in Table 3.4 vary when impulse responses to each shock are used separately to estimate Θ. For example, γ^K based on the TFP shock is 13.28, much greater than that based on either the inequality shock (6.14) or the MP shock (8.14). On the other hand, using all three shocks gives a moderate estimate of 8.39. I also review a case where the Keynesians consume and earn more than the benchmark calibration in light of the wealthy hand-to-mouth agents in the Keynesian family. When I increase $\bar{s}_C^K$ and $\bar{s}_N^K$ by 4 percentage points, $\frac{\gamma^K}{1-b^K}$ and $\frac{\bar{u}_C^K}{\bar{u}_C^R}$ decrease and become closer to $\frac{\gamma^R}{1-b^R}$ and 1. This reduces heterogeneity in preferences between the two

[21] Here I ignore the mean of $z_{i,t}$ for simplicity, because I only consider mean-preserving spreads.

agents by making the coefficients of RRA and the marginal consumption utility alike. This change seems reasonable because idiosyncratic labor productivity and the level of individual consumption become less diverging between the Keynesians and Ricardians in this case.[22]

It is worth mentioning that the model impulse responses are robust to the different parameter estimates. For example, the estimates in 'All shocks', 'Inequality shock', and 'High $\bar{s}_C^K$ and $\bar{s}_N^K$' columns induce almost identical responses of key macroeconomic variables to the inequality shock (Appendix D.2.2). In other words, it is a robust prediction of the THINK model that the inequality shock reduces aggregate demand substantially. This calls for a further inspection on determinants of aggregate demand in the model, which is the topic for the next subsection.

Aggregate Demand in the THINK Model

This subsection studies amplification and propagation mechanisms of the inequality shock in the THINK model with a focus on aggregate demand. I investigate where the large, U-shaped decline in aggregate demand comes from and how they are related to the new features in the model. Below I discuss C_t^K, C_t^R, I_t, and C_t in order.

I begin with the consumption of the Keynesians (C_t^K). They are hand-to-mouth and their consumption is determined by income as in Equation (21). When the inequality shock lowers the labor productivity of the Keynesians (Z_t^K), they lose earnings and reduce consumption. However, this 'direct effect' may be important only for the first few quarters for three reasons. First, the population share of the Keynesians (s_t^K) is only about 20 percent, and their consumption share is even less. Second, Z_t^K is not very persistent. Its half-life is about 3 quarters given $\rho_Z = 0.77$. Lastly, the dividend income plays a role of countercyclical transfers. It is because the price markup is countercyclical conditional on the inequality shock, and so are the dividends. This more or less offsets the effects of the decline in labor productivity on consumption.

[22]Interpretation of σ_Z also differs in this specification. Suppose that $z_{i,t} - \mu = \rho(z_{i,t-1} - \mu) + \epsilon_{i,t}$ where $\epsilon_{i,t} \sim N(0, \sigma_t^2)$ is independent across individual and time. In one extreme, the productivity distribution of the hand-to-mouth agents is ex ante similar to that of the others, but they receive a large idiosyncratic shock $\epsilon_{i,t}$ which makes their credit constraints bind. In this framework, shocks to $\log(Z_t^K)$ is more tightly related to the dispersion of $\epsilon_{i,t}$, not $z_{i,t}$, and, an inequality shock originates from an increase in σ_t^2. Given $\rho = 0.966$ and $\sigma_{t-1}^2 = \sigma_{t-2}^2 = \cdots = \bar{\sigma}^2 = 0.017$ from McKay, Nakamura and Steinsson (2016), a rise in the log P90/P10 index y_t by 2 log points (annualized) from its steady state value is induced by $\sigma_t^2 = 0.019 > \bar{\sigma}^2$. Then a shock to $\log(Z_t^K)$ is approximated by $N^{-1}(0.9) \times (\sigma_t - \bar{\sigma}) = 0.0094$. This is based on the extreme assumption of ex ante identical distributions, and the estimated σ_Z in the high $\bar{s}_C^K$ and $\bar{s}_N^K$ scenario is 0.0061, less than 0.0094.

The Ricardians are aware that there is a chance of receiving large negative idiosyncratic shocks and being credit constrained in the next period. Therefore, the Euler equation for the Ricardians becomes

$$1 = E_t \left[\beta \frac{q^{RR}_{t+1} u^R_{C,t+1} + q^{RK}_{t+1} u^K_{C,t+1}}{u^R_{C,t}} \frac{1+i_t}{1+\pi^P_{t+1}} \right], \tag{3.29}$$

where $u^\iota_{C,t} \equiv u'\left(C^\iota_t - b^\iota C^\iota_{t-1}\right)$ for $\iota \in \{K, R\}$. Note that there is a new element in the stochastic discount factor reflecting precautionary motivations. This raises the Ricardian's propensity to save and lowers the interest rate. For example, at the steady state, $1 + \bar{i} = \beta^{-1}\left[1 + \bar{q}^{RK}\left(\frac{\bar{u}^K_C}{\bar{u}^R_C} - 1\right)\right]^{-1} < \beta^{-1}$, because $\bar{u}^K_C > \bar{u}^R_C$. As a result, the benchmark value of $\bar{i}$ is only 0.0022, while β is 0.99. When log-linearized around the steady state, the Euler equation becomes

$$\check{u}^R_{C,t} = \beta(1+\bar{i}) \left(\bar{q}^{RR} E_t \left[\check{u}^R_{C,t+1} \right] + \bar{q}^{RK} \frac{\bar{u}^K_C}{\bar{u}^R_C} E_t \left[\check{u}^K_{C,t+1} \right] \right) + \left(\check{i}_t - E_t \left[\pi^P_{t+1} \right] \right), \tag{3.30}$$

where $\check{i}_t \equiv \log(\frac{1+i_t}{1+\bar{i}}$. The two non-standard aspects in Equation (30) reflects precautionary motivation in the THINK model. First, $\beta(1+\bar{i}) < 1$, and so Equation (30) resembles the discounted Euler equation of McKay, Nakamura and Steinsson (2017). Second, the Ricardians care for not only $\check{u}^R_{C,t+1}$, but also $\check{u}^K_{C,t+1}$, because of the uninsurable idiosyncratic risk.

The inequality shock is a positive productivity shock to the Ricardians. Thus, they will increase their consumption in response, but with some endogenous delay. There are two reasons for the delay. First, $u^R_{C,t}$ features consumption habits. Second, the inequality shock reduces $E_t\left[\check{C}^K_{t+1}\right]$, or equivalently increases $E_t\left[\check{u}^K_{C,t+1}\right]$. When being credit constrained is a more unpleasant experience than usual, the Ricardians become more cautious (save more for the future and consume less today). In Equation (30), an increase in $E_t\left[\check{u}^K_{C,t+1}\right]$ leads to an increase in $\check{u}^R_{C,t}$, corresponding to a decrease in C^R_t. The fact that the coefficients of RRA decrease, (*i.e.*, γ^K is greater than $\frac{\gamma^R}{1-b^R}$), further amplifies this effect, because $\check{u}^K_{C,t+1} = -\gamma^K \check{C}^K_{t+1}$ and $\check{u}^R_{C,t} = -\frac{\gamma^R}{1-b^R}(\check{C}^R_t - b^R \check{C}^R_{t-1})$. When the Keynesians reduce consumption, the marginal consumption utility increases faster than the Ricardians, and therefore $\check{C}^R_t$ should decrease more. For these reasons, $\check{C}^R_t$ responds to the inequality shock in a hump-shaped manner, and the initial increase in $\check{C}^R_t$ is rather muted. This helps the concurrent decline in $\check{C}^K_t$ to propagate and to reduce aggregate demand. However, it is clear that the 'direct effect' on $\check{C}_t^R$ cannot drive a recession in response to the inequality shock, because the Ricardians increase their consumption.

Another component of aggregate demand is investment. The two-agent structure adds a new dynamic to it through the discount factor $Q_{t,t+\tau}^D = \beta^\tau \frac{s_{t+\tau}^K \theta_D^K u_{C,t+\tau}^K + s_{t+\tau}^R \theta_{D,t+\tau}^R u_{C,t+\tau}^R}{s_t^K \theta_D^K u_{C,t}^K + s_t^R \theta_{D,t}^R u_{C,t}^R} \frac{P_t}{P_{t+\tau}}$ in Equation (24). In response to the inequality shock, $u_{C,t}^K$ increases a lot, because C_t^K decreases and γ^K is high. In other words, one more unit of financial income becomes much valuable when constrained agents have to reduce their consumption significantly. As a result, the utility value of the current marginal profit $s_t^K \theta_D^K u_{C,t}^K + s_t^R \theta_{D,t}^R u_{C,t}^R$ increases and $Q_{t,t+\tau}^D$ decreases. While the other terms in $Q_{t,t+\tau}^D$ may vary, the marginal utilities are the most important driver of $Q_{t,t+\tau}^D$ quantitatively (see Appendix D.2.5).

A decline in $Q_{t,t+\tau}^D$ leads to a lower value of a physical capital today. Firm's optimality condition related to the shadow value of a one unit of physical capital, denoted by q_t^κ, is as follows:

$$q_t^\kappa = E_t \left\{ \sum_{\tau=1}^{\infty} Q_{t,t+\tau}^D \frac{P_{t+\tau}}{P_t} (1-\delta)^{\tau-1} \left[r_{t+\tau}^K \nu_{t+\tau} - \Phi^\nu(\nu_{t+\tau}) \right] \right\}, \tag{3.31}$$

where r_t^K is the shadow value of a one unit of capital service K_t. Therefore, q_t^κ decreases as $Q_{t,t+\tau}^D$ lowers the current value of a physical capital, and so firms reduce investments accordingly.

Now I combine the discussions so far and look at aggregate consumption in detail. The following decomposition of aggregate consumption into several pieces highlights a key, new characteristic of the THINK model. By log-linearizing $C_t = s_t^K C_t^K + s_t^R C_t^R$, one obtains:

$$\check{C}_t = \bar{s}_C^K \check{C}_t^K + \bar{s}_C^R \check{C}_t^R + \left(\bar{s}_C^K - \bar{s}_C^R \frac{\bar{s}^K}{\bar{s}^R} \right) \check{s}_t^K. \tag{3.32}$$

The first and second terms represent the direct effects. When the inequality shock reduces $\check{Z}_t^K$, consumption of the Keynesians decrease while the opposite holds for the Ricardians. Those direct effects constitute all of aggregate consumption responses in other TANK models where agents' types are fixed. However, my model features an additional channel of distributional effects. In the THINK model, higher $\check{s}_t^K$ leads to lower aggregate consumption, because individuals who become credit constrained reduce their consumption substantially. The last term in Equation (32) represents this channel, where the coefficient on $\check{s}_t^K$ is negative when the consumption share of the Keynesian is less than their population share at the steady state ($\bar{s}_C^K < \bar{s}^K$), or equivalently, $\bar{C}^K < \bar{C}^R$. Because $\check{s}_t^K$ is countercyclical as illustrated in Equation (27), this channel can amplify aggregate fluctuations.

How the three terms in Equation (32) and aggregate consumption react to a one standard deviation inequality shock is depicted in Figure 3.9. The left panel is based on the benchmark parameters and the right one is for the high $\bar{s}_C^K$ and $\bar{s}_N^K$ case. As discussed above, the direct effects to the Keynesians contribute little to the responses of aggregate consumption after a few quarters. While it is more important in the high $\bar{s}_C^K$ and $\bar{s}_N^K$ case, it is not the most important component of $\check{C}_t$ at least after a year. Besides, the direct effects to the Ricardians are positive. Thus, the negative and U-shaped responses of aggregate consumption are mostly driven by the distributional effects.

Recall that the inequality shock decreases investment significantly. This negative effect to aggregate demand, combined with the direct effects to the Keynesian consumption, lowers economic output. When the economy turns into a recession, idiosyncratic earnings risk increases and some Ricardians become Keynesians. Then their consumption decreases, which further reduces aggregate demand. As a result, output decreases, more Ricardians become Keynesians, and so on. This aggregate demand spiral amplifies the distributional effects substantially and makes it the most important determinant of aggregate consumption.[23]

It is clear from equations (27) and (32) that the value of η_Y is crucial for determining the magnitude of the distributional effects. However, it is hard to estimate η_Y directly from (27), because there is no available quarterly time series data of s_t^K. Therefore, I take an indirect route to supplement the discussion and make three points on η_Y. First, the Great Recession was a period when access to credit was limited and as a result many people were credit constrained (Mian, Rao and Sufi, 2013; Mian and Sufi, 2015). This is consistent with an implication of a positive η_Y in the model that the number of credit constrained agents increases in recessions. Second, unemployment risk may contribute to the countercyclical variations in $\check{s}_t^K$ significantly. Given $\eta_Y = 4.32$ and other parameters, Equation (27) becomes $\check{s}_t^K = 0.48\check{s}_{t-1}^K - 4.27\check{Y}_t$. Thus, the semi-elasticity of s_t^K with respect to output is $\frac{\partial s_t^K}{\partial \check{Y}_t} = -4.27 \times \bar{s}^K = -0.85$, meaning that the population share of the Keynesian

[23] Auclert and Rognlie (2018) also study how an earnings inequality shock affects economic output in their HANK model and find little aggregate effects. However, the model of Auclert and Rognlie features a CRRA preference, flexible prices, and no autoregressive term in the monetary policy rule, unlike my THINK model. When I make the THINK model similar to the model of Auclert and Rognlie by changing parameters as $\gamma^K = \gamma^R$, $b^K = b^R = 0$, $\mathcal{M}_P = 1$, $\psi_P = 0$, and $\rho_i = 0$, the THINK model also predicts little effects of an inequality shock on real variables. Furthermore, each of the factors above is important for rationalizing the large, U-shaped, estimated responses in Section 3.4. For example, when I fix ρ_i at 0 while not changing the other parameters, the peak effect of an inequality shock on real GDP becomes less than half of the benchmark case (see Appendix D.2.4).

family increases by 0.85 percentage points when output decreases by 1 percent. A similar semi-elasticity of unemployment rate with respect to real GDP per capita in the U.S. is -0.44 based on the HP filtered quarterly series. Because there also exists earnings risk conditional on being employed (Guvenen, Ozkan and Song, 2014), the implied sensitivity of $\check{s}_t^K$ to $\check{Y}_t$ in the model may not be unreasonably large. Finally, one may rely on the micro-foundation in Section 3.5.2 and infer η_Y from the left-tail of the earnings distribution. The micro-foundation assumes that the left-tail of the earnings distribution can be approximated by an inverse Pareto random variable $\upsilon_{i,t}^{-1} Y_t$, where $\upsilon_{i,t} \sim \text{Pareto}(\eta_Y)$ for $\upsilon_{i,t} \geq \upsilon_m$. It follows that the log cumulative distribution function is linear in the log earnings with a slope η_Y in the left-tail, and this can be estimated from the QCEW. For example, the estimate is 5.04 based on the data in the first quarter of 2000. One need to exercise caution when relating this estimate to η_Y in the model because of measurement errors, minimum wages, and unemployment risk. Nevertheless, the estimate for η_Y (4.32) may not be unrealistic in light of the micro-founded slope coefficient (5.04). See Appendix D.2.3 for details on this estimation.

So far, I have shown how inequality can be a source of demand-driven business cycles. I introduce three new features to the THINK model, an extensive margin of being credit constrained, DRRA preferences, and a small amount of financial income for the Keynesians. Equipped with the new channels, I illustrate how they can rationalize the large, U-shaped, empirical impulse responses of aggregate variables. In doing so, I use solution and estimation methods developed for linear systems. However, inequality may have a non-linear effect on an economy by altering how it responds to policies and other structural shocks. An analysis of those effects requires a separate approach because of its non-linear nature.

3.6 Inequality and the Power of Stabilization Policies

This section covers policy implications of rising inequality in the U.S. based on the non-linear dynamics of the THINK model. Intuitively, there are more earnings- or wealth-poor people in an economy when the level of inequality is higher. They have higher MPCs and an interaction effect between more people and higher MPCs can make aggregate consumption demand more sensitive to economic conditions and policies. Consistent with the intuition, the THINK model predicts that the power of stabilization policies increases in the level of inequality. Empirical evidence based on various datasets is also in line with the prediction.

On top of my findings that inequality and redistributive factors can drive macroeconomic fluctuations, this policy implication provides another reason why understanding inequality is important for policymakers.

3.6.1 Policy Implications of the THINK Model

For understanding the relationship between the power of stabilization policies and the level of inequality, the following decomposition of aggregate consumption in the THINK model is useful. Let $ds_t^K \equiv s_t^K - \bar{s}^K$ and other linear deviations be denoted similarly. From $C_t = s_t^K C_t^K + s_t^R C_t^R$, it follows that

$$dC_t = \bar{s}^K dC_t^K + \bar{s}^R dC_t^R + (\bar{C}^K ds_t^K + \bar{C}^R ds_t^R) + (ds_t^K dC_t^K + ds_t^R dC_t^R). \tag{3.33}$$

Note that this is an exact equation, not an approximation. When compared to the log-linear approximation in Equation (32), it is clear that the first three terms in Equation (33) correspond to the direct effects to the Keynesian consumption, the direct effects to the Ricardian consumption, and the distributional effects in Equation (32). However, there exists an additional term representing the interaction effect between distribution (ds_t^K and ds_t^R) and marginal propensities to consume (dC_t^K and dC_t^R). Thus, aggregate consumption demand may become more sensitive to policies when there are more agents with higher MPCs. If the same mechanism applies to other structural shocks, aggregate fluctuations may become larger and macroeconomic volatility in general may be elevated. On the bright side, however, stabilization policies become more powerful too.

Note that high inequality corresponds to high s_t^K in the discussion above. This is because more earnings or income inequality implies that there are more people with higher MPCs at the bottom of wealth distribution. For example, Wolff (2017) reports that the share of households holding non-positive (less than $5,000 constant 1995 dollars) net worth increased by about 6 (13) percentage points from 1969 to 2013 in the U.S.

I consider two initial states of the model economy: high and low inequality. In the high (low) inequality state, s_{t-1}^K is 0.25 (0.15) and all the other variables equal their steady state values, where the range of 10 pernetage points is about the midpoint between 6 and 13 percentage points of Wolff (2017). To evaluate the non-linear dynamics of the THINK model, I employ a third-order pruned state-space system approach of Andreasen, Fernández-Villaverde and Rubio-Ramírez (2017) and the generalized impulse response functions of

Koop, Pesaran and Potter (1996). The responses of aggregate consumption conditional on both states are plotted in Figure 3.10. The left panel illustrates the generalized impulse response functions to a one standard deviation contractionary monetary policy shock, while the right one is for an expansionary fiscal policy shock. It is clear from both panels that aggregate consumption reacts more strongly to the policy shocks when the level of inequality is higher. This implies more powerful stabilization policies conditional on the higher level of inequality, and the results for other variables are similar (Appendix E.3).

The discussion so far illustrates a mechanism through which the level of inequality can affect propagation of structural shocks. Among many structural shocks, I concentrate on monetary and fiscal policy shocks and derive novel policy implications. In the next subsection, I investigate several datasets to test this theoretical prediction and find qualitatively consistent results.

3.6.2 Empirical Evidence

Here I test the theoretical prediction above empirically using U.S. data. The main idea is to include an interaction term between an inequality measure and a structural shock in local projections. If the coefficient is statistically and economically significant, I would conclude that inequality matters for the propagation and amplification of stabilization policies.

I examine three different datasets for robustness of the results. The first dataset consists of quarterly observations of earnings inequality, various aggregate variables, and several structural shocks in recent decades. The second dataset includes an annual but long history of income inequality, some aggregate variables, and a military news shock. The last one is based on state-level annual series of income inequality, real GDP, and military procurement spending since the 1960s. For identification, I exploit time series variation in the first two datasets and variation across states and time in the last dataset. For all the data, shocks, and specifications, the results consistently imply that more inequality leads to larger responses to policy shocks of the same size, consistent with the theoretical policy implications of my THINK model.

Recent Data

The first dataset consists of quarterly observations including the MP and FP shocks in Section 3.3, my log P90/P10 index based on the QCEW, and key macroeconomic variables in Section 3.4.[24] I consider the following local projections with an interaction term:

$$m_{t+h} - m_{t-1} = \beta_h x_t + \gamma_h x_t y_{t-1} + \Gamma'_{xy,h}\mathbf{Z}^{(xy)}_{t-1} + u^{(xy)}_{t,h}. \qquad (3.34)$$

Given an impulse of a unit structural shock x_t, a macroeconomic variable m_t responds by $\beta_h + \gamma_h y_{t-1}$ after h periods, where y_{t-1} is the inequality index in the previous quarter. The way m_t reacts to the shock x_t depends on the state of inequality y_{t-1}, and the dependence is parametrized by γ_h. $\mathbf{Z}^{(xy)}_{t-1}$ includes an intercept and four lags of x_t, y_t, $x_t y_{t-1}$, Δm_t, and $\Delta m_t y_t$. The sample period is from 1978:Q1 to 2008:Q4.

The left panel of Figure 3.11 shows the results for real consumption in response to a one standard deviation contractionary MP shock conditional on y_{t-1} being plus or minus one standard deviation from the average. It is clear that the contractionary effects of the MP shock is much stronger when earnings inequality is higher. For example, the t-statistic for the null hypothesis that $\gamma_{14} = 0$ is -5.15, and so $\hat{\gamma}_{14}$ is statistical significant at the 1% level. The right panel is for a one standard deviation expansionary FP shock conditional on the same values of y_{t-1}. Similarly, consumption increases more when earnings are more unequally distributed. The t-statistic for $\gamma_8 = 0$ is 2.91 and the $p-$value is only 0.002. Thus, I conclude that high earnings inequality makes contractionary MP shocks more contractionary and expansionary FP shocks more expansionary. The estimates for other variables such as real GDP, investment, price level, and unemployment rate are in line with the findings here (Appendix E.1).

Historical Data

Although the results above are intriguing, one may worry about a rising trend in inequality during the sample period. In the worst case, earnings inequality might be just capturing a trend in the U.S. economy becoming more volatile due to some other reasons.

To address this concern, I look at a long history of inequality and economic growth in the U.S. throughout the 20th century. The top 10% income share of Piketty and Saez (2003) serves the purpose well because it starts from 1917. Importantly for my identification, it follows a U-shaped pattern instead of an upward trend.

The cost of extending the sample backward is that there are not many reliable identified shock series available. A narrative measure of military news shock constructed by Ramey

[24]See Appendix E.1 for the TFP shocks.

and Zubairy (2018) is an exception. It dates back to 1889. Ramey and Zubairy also provide real GDP, price level, and unemployment rates in addition to the military news shock. By combining the two sources, the sample spans from 1917 to 2015.

I estimate Equation (34) using the historical data, where the dependent variable is the real GDP per capita. As illustrated in Figure 3.12, the U.S. economy responds more strongly to the military news shocks when the top 10 percent takes more income. For example, a military news shock whose present-discounted value is 10 percent of the trend GDP increases real GDP per capita by 4 percent after 3 years when the top 10% income share is 43.9 percent. However, the same shock raises real GDP per capita only by 1.5 percent when the top 10% holds 32.6 percent of income. Also, the t-statistic for $\gamma_3 = 0$ is 4.81 and the null is rejected at the 1 percent level. Similarly, GDP deflator and the unemployment rate react more strongly conditional on higher inequality (Appendix E.2).

State-level Data

Finally, I compare states with different levels of inequality. For inequality, I employ Frank-Sommeiller-Price series for the top 10% income share by state (Frank et al., 2015). This series is constructed by applying methods similar to Piketty and Saez (2003) at the state-level. For real state GDP, price level, employment, population, and most importantly, military procurement spending, I use the data from Nakamura and Steinsson (2014). The sample period is from 1969 to 2008.

Let $m_{i,t}$, $g_{i,t}$, and $y_{i,t}$ be real GDP per capita, real military procurement spending per capita, and top 10% income share in state i in year t. m_t and g_t without subscript i refer to the same variables at the U.S. level. Instrumental variables $D_i \cdot \frac{g_t - g_{t-1}}{m_{t-1}}$ for all i are used for the first two regressors in Equation (35), where D_i is a dummy variable for state i:

$$\frac{m_{i,t+h} - m_{i,t-1}}{m_{i,t-1}} = \beta_h \frac{g_{i,t} - g_{i,t-1}}{m_{i,t-1}} + \gamma_h \frac{g_{i,t} - g_{i,t-1}}{m_{i,t-1}} \cdot y_{i,t-1} + \Gamma'_{i,t,h} \mathbf{Z}_{i,t} + u_{i,t,h}. \tag{3.35}$$

$\mathbf{Z}_{i,t}$ includes time and state fixed-effects, $\frac{m_{i,t-1} - m_{i,t-2}}{m_{i,t-2}}$, and $y_{i,t-1}$. Standard errors are clustered by state.

When the aggregate military expenditures increase, some states receive more military spending or have higher income inequality on top of that. My identifying assumption, similar to that of Nakamura and Steinsson, is that the U.S. does not engage in aggregate military buildups because these states are experiencing or expected to suffer from sluggish growth relative to the others.

The estimated γ_h in Table 3.5 is positive and statistically significant at the 1 percent level for $h = 1, 2$, and 3. To fix the idea, consider $h = 2$, and a military spending shock amounts to 1% of real state GDP, *i.e.*, $\frac{g_{i,t}-g_{i,t-1}}{m_{i,t-1}} = 0.01$. Real state GDP per capita responds insignificantly when the top 10% share is only 30 percent. However, when $y_{i,t-1}$ is 40 (50) percent, the response becomes 5.03 (11.89) percent and statistically significant at the 1 percent level. In other words, fiscal expansion becomes more powerful in states where income inequality is higher.

In summary, I look at the three datasets so far and rely on several variation to identify the effects of the level of inequality on the propagation of monetary and fiscal policies. Those results provide a set of extensive empirical facts consistent with the policy implication of the THINK model that the power of stabilization policies increases in the level of inequality.

3.7 Conclusion

The Great Recession stimulated interest in how inequality, aggregate fluctuations, and stabilization policies are related. For example, policymakers have become concerned about the distributional effects of stabilization policies in addition to their aggregate effects. Another important issue is to understand the direction from inequality to business cycles. If redistribution and inequality affect aggregate demand and how shocks propagate, policymakers should incorporate such considerations when they design policies to stabilize the economy. This paper explored these important relationships both empirically and theoretically.

Using a new quarterly measure of earnings inequality based on high-quality administrative data, I illustrate that inequality matters for policymakers in three aspects. First, earnings inequality reacts to shocks to fiscal policy and total factor productivity at business cycle frequencies. Second, unanticipated increases in earnings inequality induce recessions by reducing aggregate demand. Lastly, high levels of inequality make stabilization policies more powerful.

I further develop a new, tractable theoretical framework for studying the interaction between inequality, business cycles, and stabilization policies. This framework rationalizes my empirical findings and provides novel insights on the mechanisms through which inequality affects aggregate demand and the power of stabilization policies. The simplicity of the approach can help researchers easily link models to data, and thus stimulate further research in this area which historically relied on computationally intensive heterogeneous agent models.

My results may have implications for other macroeconomic phenomena. For example, inequality shocks may have contributed to slow recoveries from the Great Recession. The unanticipated inequality shocks are mostly positive between 2006:Q3 and 2008:Q4. This may be related to the large, prolonged decline in economic activity afterwards, because the impulse responses of real GDP to inequality shocks are persistent. Furthermore, my finding may provide a new interpretation of (a possible end of) the Great Moderation based on an upward trend in inequality. As discussed in Section 3.6, aggregate consumption demand can be more sensitive to economic condition when the level of inequality is higher. Because inequality has been rising in recent decades, this implies the level of cyclical volatility has been also rising, which may mean an end of the Great Moderation. Finally, my THINK model predicts that fiscal multipliers can increase in recessions. In the model, the number of credit constrained agents is countercyclical, and therefore there are more agents with higher MPCs during economic downturns. This makes aggregate consumption demand sensitive to shocks, which leads to large fiscal multipliers. Rigorous investigation of these hypotheses is left for future research.

Incorporating new features introduced in the THINK model into a heterogeneous agent framework is another topic for future research. Fully-specified HANK models where agents are endogenously credit constrained may provide an useful laboratory for studying the interaction between inequality and business cycles.

Table 3.1: Summary statistics.

	1975:Q1	1984:Q1	2001:Q1	2014:Q1
Industrial classification	SIC 2-digit	SIC 4-digit	NAICS 6-digit	NAICS 6-digit
Number of cells	105,026	219,300	265,805	268,875
Total number of workers, million	59.9	64.8	89.0	96.3
Total earnings, USD billion	145.5	297.6	846.4	1,303.4
Average earnings, USD	2,430	4,590	9,514	13,540
Distribution of the number of workers in a cell				
P1	1	1	1	1
P25	24	18	23	23
P50	78	51	64	66
P75	280	167	207	214
P99	8,750	4,109	4,543	4,948

Notes: A cell means an industry/county/ownership-type combination in the QCEW. In the data, the number of workers is counted in each month, while the earnings are available only quarterly. Thus, I use the average number of workers over three months in each cell, which may not be an integer. The fractional parts are rounded in the table. For example, there are about 66 workers in a median-sized cell in the first quarter of 2014.

Table 3.2: Model parameters.

Parameter	Value	Description, Source, and Comment
β	0.99	Time preference. In the SS, $1/\beta = 1 - \delta + \delta(1-\alpha)/(\phi_I \mathcal{M}_P)$.
γ^K	8.39^e	Negative elasticity of u_C^K at the SS
b^K	0	Consumption habits for the Keynesian, Pre-MCMC numerical optimization gives 0.
γ^R	2	Negative elasticity of u_C^R at the SS
b^R	0.7	Consumption habits for the Ricardian
γ^L	2	Negative elasticity of u_e^L at the SS
b^L	0.7	Earnings habits for the labor unions
$\bar{u}_C^K/\bar{u}_C^R$	3.99^e	Ratio of the marginal consumption utilities at the SS
$\bar{s}^K$	0.2	Population share of the Keynesian family at the SS, Debortoli and Galí (2017).
$\bar{s}_C^K$	0.11	Consumption share of the Keynesian family at the SS, Krueger, Mitman and Perri (2016).
$\bar{s}_N^K$	0.08	Labor share of the Keynesian family at the SS, Kuhn and Rios-Rull (2016).
q^{RK}	0.0025	Transition probability from the Ricardian to the Keynesian family, see Section 3.5.2.
η_Y	4.32^e	Negative elasticity of q_t^{KK} to Y_t
η_s	0.51^e	Negative elasticity of q_t^{KK} to s_{t-1}^K
φ	1/0.54	Elasticity of labor disutility v^L at the SS. Chetty et al. (2011).
δ	0.025	Capital depreciation rate
α	2/3	Production function: $Y = AK^{1-\alpha}N^{\alpha}$.
$\Phi_{\nu\nu}^{\nu}(\bar{\nu})$	0.23^e	Second derivative of the capital utilization costs at the SS. $\Phi_{\nu}^{\nu}(\bar{\nu}) = 0.035$ is chosen to make $\bar{\nu} = 1$.
$\Phi_{II}^{I}(1)$	1.64^e	Second derivative of the investment adjustment costs at the SS

Continued on the next page.

Notes: e: estimated, SS: steady state.

Table 3.2: Model parameters, continued.

Parameter	Value	Description, Source, and Comment
$\mathcal{M}_P$	1.2	Gross price markup at the SS, Rotemberg and Woodford (1997). Equivalent to $\epsilon_P = 6$.
ψ_P	233.3	Price adjustment costs. Equivalent to the Calvo probability of 0.75.
$\mathcal{M}_W$	1.2	Gross wage markup at the SS, Huang and Liu (2002), Griffin (1992). Equivalent to $\epsilon_W = 6$.
ψ_W	706.3	Wage adjustment costs. Equivalent to the Calvo probability of 0.75.
ϕ_C	0.6	$\bar{C}/\bar{Y}$.
ϕ_I	0.2	$\bar{I}/\bar{Y}$. $\phi_G \equiv \bar{G}/\bar{Y} = 0.2$.
ρ_i	0.9	Monetary policy: interest rate smoothing
ζ_π	2	Monetary policy: responsiveness to price inflation
ζ_Y	0.15	Monetary policy: responsiveness to output
ρ_Z	0.77^e	Persistence of inequality shocks
ρ_A	0.77^e	Persistence of productivity shocks
ρ_G	0.97	Persistence of government expenditure shocks, Smets and Wouters (2007).
σ_Z	0.0030^e	Standard deviation of inequality shocks
σ_i	0.0008^e	Standard deviation of monetary policy shocks
σ_A	0.0101^e	Standard deviation of productivity shocks
σ_G	0.0050	Standard deviation of government expenditure shocks, Smets and Wouters (2007).

Notes: e: estimated, SS: steady state.

Table 3.3: Parameter estimation, the benchmark case.

	Prior			Posterior			
Parameter	Distribution	Mean	St. Dev.	Mode	Mean	P5	P95
$\gamma^K - 1$	Gamma	10	5	7.39	7.22	6.98	7.49
$\Phi^{\nu}_{\nu\nu}(\bar{\nu})$	Gamma	0.1	0.05	0.23	0.18	0.15	0.24
$\Phi^{I}_{II}(1)$	Gamma	3	1	1.64	1.74	1.61	1.86
$\bar{u}^K_C/\bar{u}^R_C - 1$	Gamma	1	1	2.99	3.25	2.93	3.60
η_Y	Gamma	3	1.5	4.32	4.07	3.79	4.37
$\eta_s + 1$	Gamma	1.5	0.5	1.51	1.52	1.51	1.53
ρ_Z	Beta	0.8	0.05	0.770	0.777	0.763	0.789
ρ_A	Beta	0.9	0.05	0.772	0.800	0.767	0.818
σ_Z	Inv. Gam.	0.0025	0.0010	0.0030	0.0030	0.0028	0.0032
σ_i	Inv. Gam.	0.0013	0.0003	0.0008	0.0009	0.0008	0.0009
σ_A	Inv. Gam.	0.0081	0.0016	0.0101	0.0097	0.0088	0.0112

Notes: The supports of priors are not $(-\infty, \infty)$. For a gamma or an inverse gamma distribution, the support is $(0, \infty)$, and a beta distribution is defined on $(0, 1)$. This might incur problems near the boundary when a random walk algorithm suggests a value outside the support. Thus, I re-parameterize Θ. For a beta random variable, I use $f(x) = \tan(\pi x - \pi/2)$, and $\log(\cdot)$ function is employed for the others. I write the transformation as $\tilde{\Theta} = F(\Theta)$. I work with $\tilde{\Theta}$ throughout the estimation and inverse the chained samples to Θ at the last step. I simulate a Markov chain whose length is 200,000, and the first 50,000 observations are dropped. The acceptance rate is 34%.

Table 3.4: Parameter estimation, robustness check.

Parameter	All shocks	Inequality shock	Monetary shock	TFP shock	High $\bar{s}_C^K$ and $\bar{s}_N^K$
γ^K	8.39 (0.15)	6.14 (0.44)	8.14 (0.53)	13.28 (1.35)	7.11 (0.88)
$\Phi_{\nu\nu}^{\nu}(\bar{\nu})$	0.23 (0.027)	0.06 (0.004)	0.05 (0.013)	0.19 (0.010)	0.17 (0.003)
$\Phi_{II}^{I}(1)$	1.64 (0.08)	1.69 (0.07)	1.89 (0.19)	2.25 (0.25)	2.23 (0.13)
$\bar{u}_C^K/\bar{u}_C^R$	3.99 (0.18)	4.90 (0.19)	3.82 (0.21)	1.75 (0.24)	2.01 (0.10)
η_Y	4.32 (0.17)	4.86 (0.26)	5.55 (0.37)	4.67 (0.13)	4.57 (0.17)
η_s	0.51 (0.01)	0.82 (0.04)	0.58 (0.02)	0.39 (0.03)	0.38 (0.02)
ρ_Z	0.77 (0.007)	0.78 (0.005)	-	-	0.79 (0.009)
ρ_A	0.77 (0.016)	-	-	0.85 (0.006)	0.87 (0.004)
σ_Z	0.0030 (0.0001)	0.0021 (0.0002)	-	-	0.0061 (0.0007)
σ_i	0.0008 (0.00001)	-	0.0008 (0.00002)	-	0.0009 (0.00002)
σ_A	0.0101 (0.0007)	-	-	0.0069 (0.0002)	0.0061 (0.0002)
Log Posterior	-70.0	-234.6	-134.6	-132.3	-88.0

Notes: This table displays the posterior mode with the standard deviation in parentheses. The column labeled as all shocks is based on the benchmark results in Table 3. The next three columns are for restricted moment conditions. For example, the estimates in the inequality shock column are obtained by matching responses to a one standard deviation inequality shock only. The last column is for a case where $\bar{s}_C^K = 0.15$ and $\bar{s}_N^K = 0.12$ in view of the wealthy hand-to-mouth agents. The benchmark values are 0.11 and 0.08, respectively, which are based on the bottom 20 percent of the wealth distribution in the U.S. I report the value of log posterior based on the posterior distribution in the benchmark case for comparison. For parameters not estimated, *e.g.,* ρ_A in the inequality shock column, I use the prior mean in Table 3 when evaluating the log posterior at the mode.

Table 3.5: Responses of state real GDP per capita to government spending shocks conditional on inequality.

	h				
	0	1	2	3	4
β_h	2.48	-10.24**	-22.40***	-21.12**	-26.78
	(6.54)	(4.41)	(7.80)	(8.46)	(12.09)
γ_h	-5.78	33.31***	68.58***	67.06***	83.43
	(18.48)	(12.51)	(21.92)	(24.14)	(33.67)
Observations	1,746	1,740	1,734	1,684	1,634
$\beta_h + \gamma_h \cdot 0.3$	0.74	-0.25	-1.83	-1.00	-1.75
$\beta_h + \gamma_h \cdot 0.4$	0.16	3.08***	5.03***	5.70***	6.59**
$\beta_h + \gamma_h \cdot 0.5$	-0.41	6.41**	11.89***	12.41***	14.93**

Notes: This table is based on Equation (35). As is clear from the bottom half of the table, the response of real state GDP to state fiscal expenditure shocks depends on the state top 10% income shares significantly. Standard errors are in parentheses, which are clustered by state. The number of asterisks denotes statistical significance of the estimate. *: 10%, **: 5%, ***: 1%.

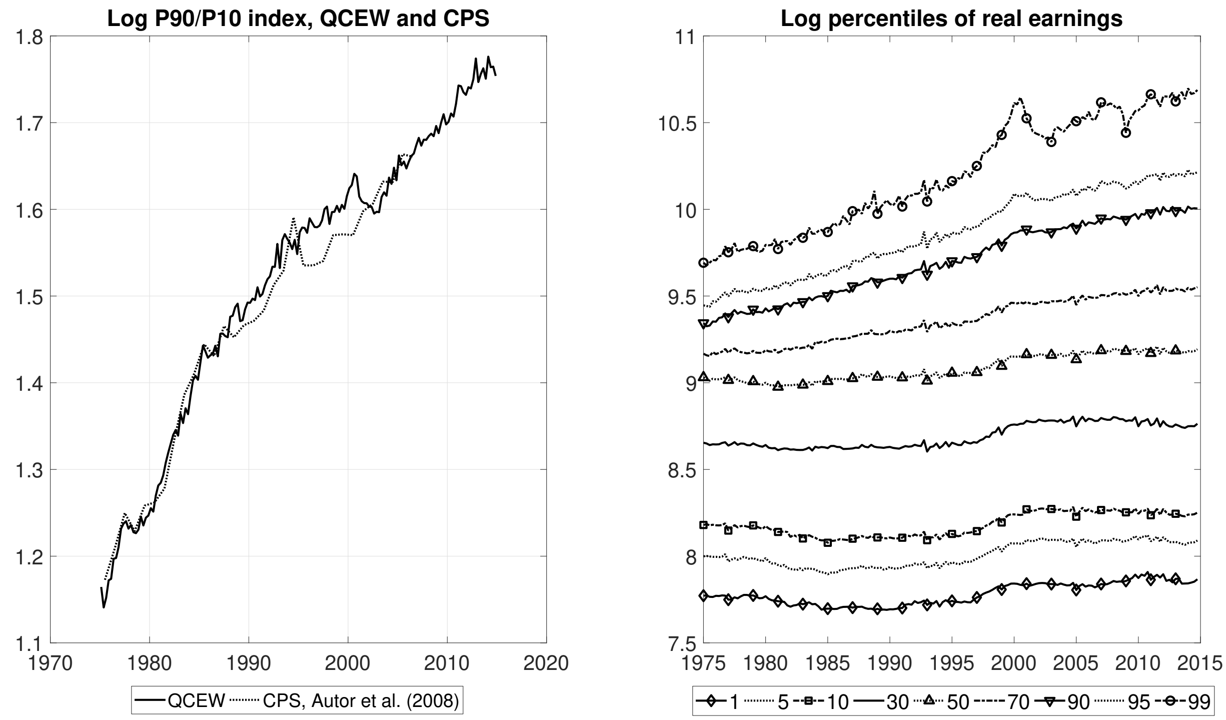

Figure 3.1: New inequality index and log percentiles.

Notes: The left panel depicts my new inequality index in comparison with the CPS-based measure, where the log percentiles of the real earnings distribution from the QCEW are shown in the right panel. Autor, Katz and Kearney (2008) construct the log P90/P10 index from the March annual demographic survey in the CPS. They focus on male respondents and derive weakly earnings from annual earnings and number of working weeks. On the other hand, the QCEW covers both male and female workers. Each percentile is deflated using the GDP implicit deflator.

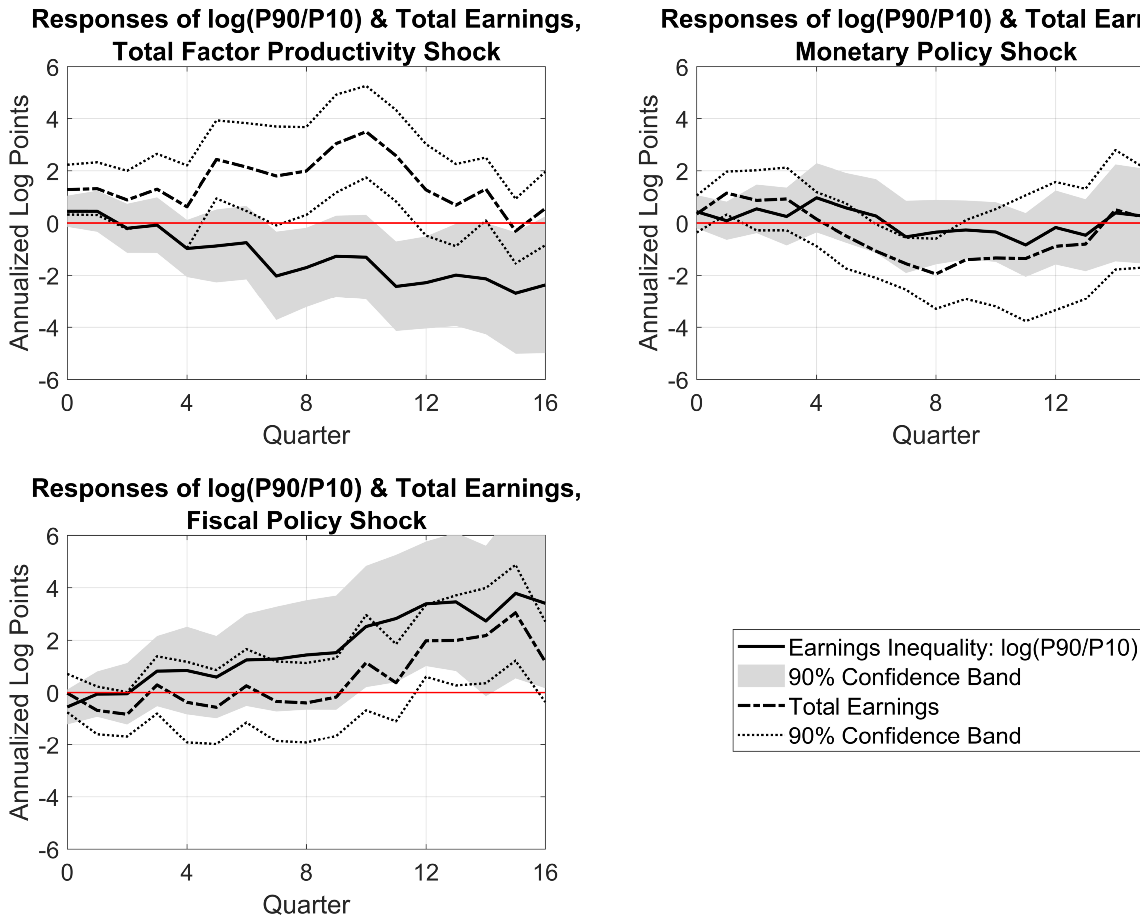

Figure 3.2: Effects of structural shocks on the inequality index and aggregate real earnings.

Notes: The solid lines and the shaded area represent how the inequality index responds to a one standard deviation TFP, MP, and FP shock with the 90% confidence bands. The dash-dot lines and the dotted lines are for aggregate real earnings from the QCEW. Units are annualized percent and annualized log points. The bandwidth for the Newey-West variance estimator increases in the horizon of local projections one for one. For the inequality index, the Phillips-Perron test does not reject the null of a unit root, while the KPSS test rejects the null of trend-stationarity at the 1 percent level (Kwiatkowski et al., 1992; Phillips and Perron, 1988). On the other hand, both tests do not reject own nulls for aggregate earnings. I assume a trend-stationary model in this case and include $d_h t$ term and use y_{t-i} in place of Δy_{t-i} in Equation (2). The benchmark result is not sensitive to the specification details. The results based on other inequality measures, specifications controlling for the early Volcker period, a model with an oil supply shock of Kilian (2008), or the impulse response functions estimated in a shock by shock manner can be found in Appendix B.3.

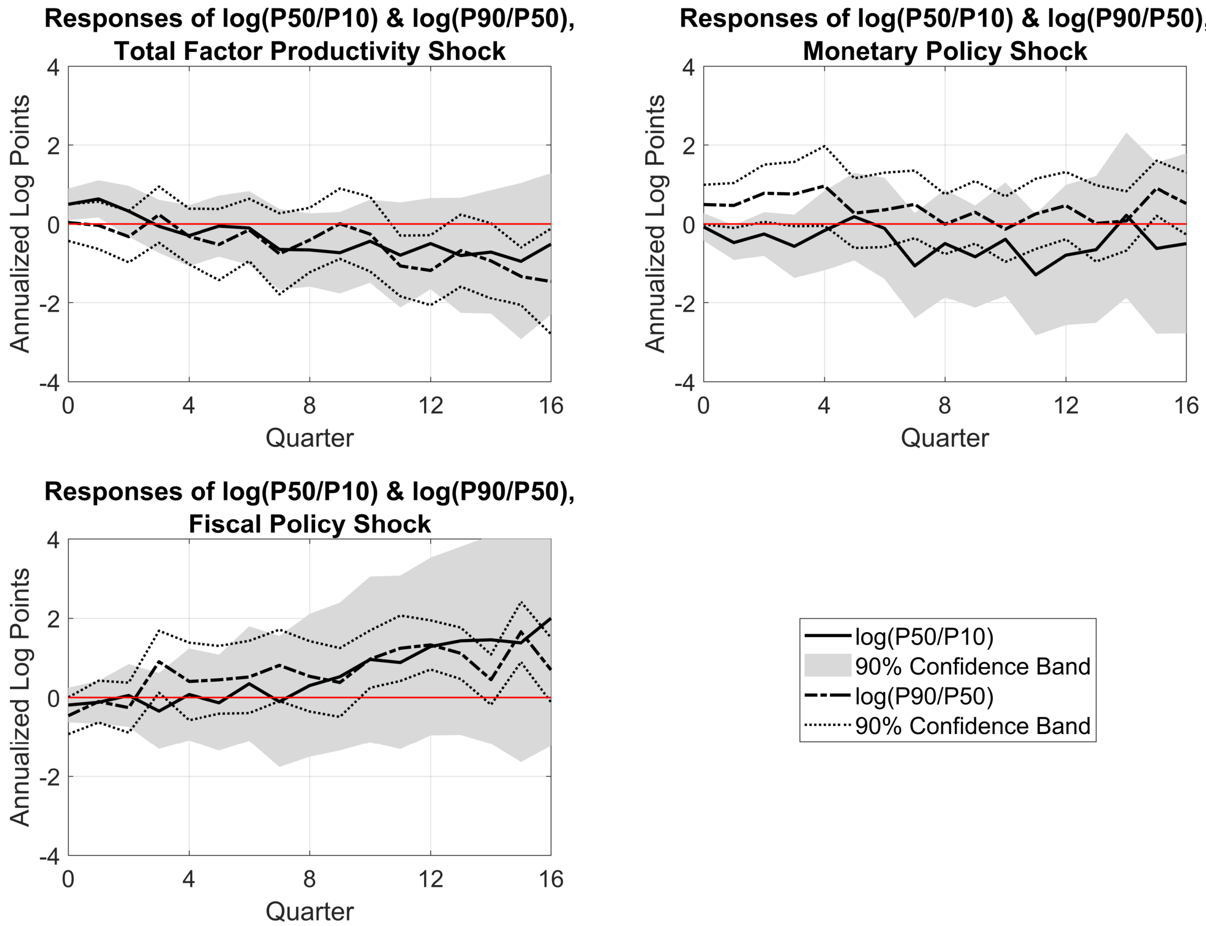

Figure 3.3: Decomposing the responses of the inequality index.

Notes: The solid lines and the shaded area represent how the log P50/P10 index, which represents the dispersion among the bottom half of earnings distribution, responds to a one standard TFP, MP, and FP shock. The dash-dot lines and the dotted lines are for the log P90/P50 index. The results are similar when a dummy variable for the early Volcker period is added or sample after the period is used (see Appendix B.4). I use Equation (2) and assume that both series have a unit root based on the results of the Phillips-Perron test and the KPSS test (Kwiatkowski et al., 1992; Phillips and Perron, 1988).

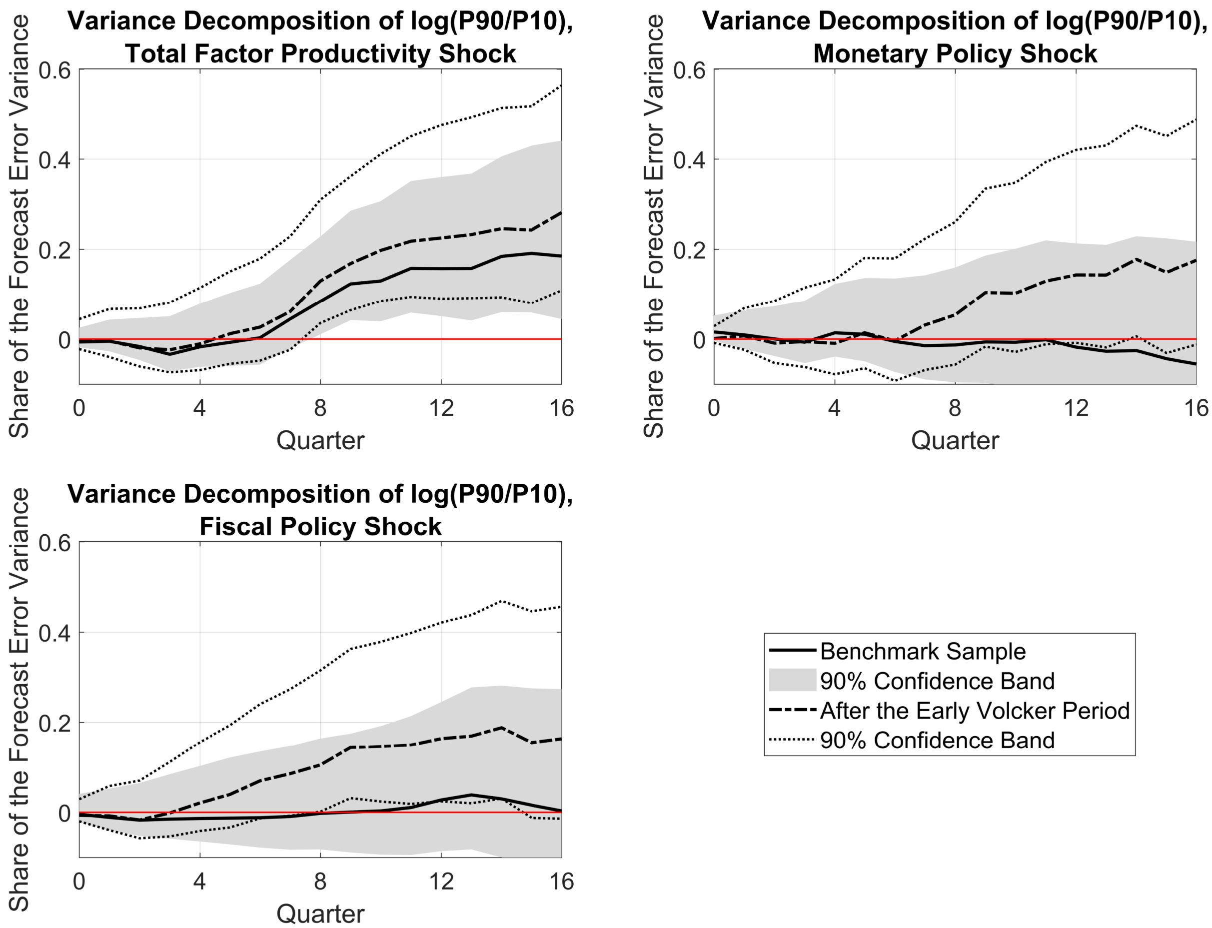

Figure 3.4: Variance decomposition of the inequality index.

Notes: The solid lines denote the estimated forecast error variance decompositions (FEVD) based on the benchmark sample with the shaded area being the bootstrapped 90 percent confidence bands. The benchmark sample spans from 1978:Q1 to 2008:Q4. The other sample begins in 1983:Q1 after the early Volcker period when the Fed targeted the amount of non-borrowed reserves. The results using the second sample are represented by the dash-dot and the dotted lines. I use the bias-corrected R^2 estimator of Gorodnichenko and Lee (2017) to estimate the FEVDs. Other lag length, the inclusion of the oil supply shocks of either Kilian (2008) or Kilian (2009), and considering a smaller information set by letting an information set contain only one shock each time do not change the results significantly (see Appendix B.6).

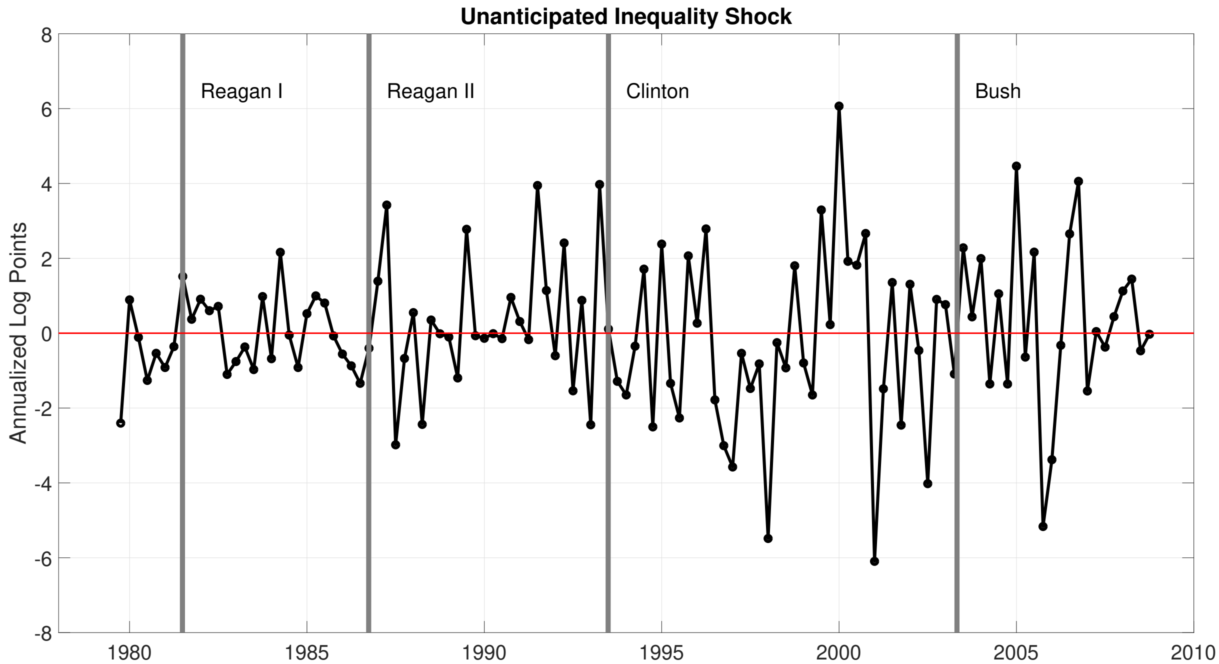

Figure 3.5: Identified unanticipated innovations in earnings inequality.

Notes: This figure plots the unanticipated innovations in earnings inequality ($x_{t,ineq}$) in Equation (4). The units are annualized log points. The grey bars depict when major tax reforms in the U.S. were signed into laws with the name of president who signed: (i) the Economic Recovery Tax Act of 1981 (ERTA 81), Ronald Reagan, August 13, 1981, (ii) the Tax Reform Act of 1986 (TRA 86), Ronald Reagan, October 22, 1986, (iii) the Omnibus Budeget Reconciliation Act of 1993 (OBRA 93), Bill Clinton, August 10, 1993, and (iv) the Jobs and Growth Tax Relief Reconciliation Act of 2003 (JGTRRA 03), George W. Bush, May 28, 2003. The OBRA 93 raised the top marginal income tax rates while the others did the opposite. Considering the dates when the new tax rates became effective for the first time does not change the implication from the figure. The first tax cut following the ERTA 81 happened on October 1, 1981, and $x_{t,ineq}$ is positive in 1981:Q4. The income tax rates defined on the TRA 86 were applicable from the tax year 1987, and similarly the identified shock in 1987:Q1 is positive. The new rates from the OBRA 93 were applied since the tax year 1993. Because it was signed in the middle of 1993:Q3, it is natural that 1993:Q4 is the first quarter under the new tax rates. And the identified shock is negative in 1993:Q4. Similarly, the tax rates from the JGTRRA 03 were valid for the tax year 2003 and it was signed in the middle of 2003:Q2. Consistently, the identified shock in 2003:Q3 is positive.

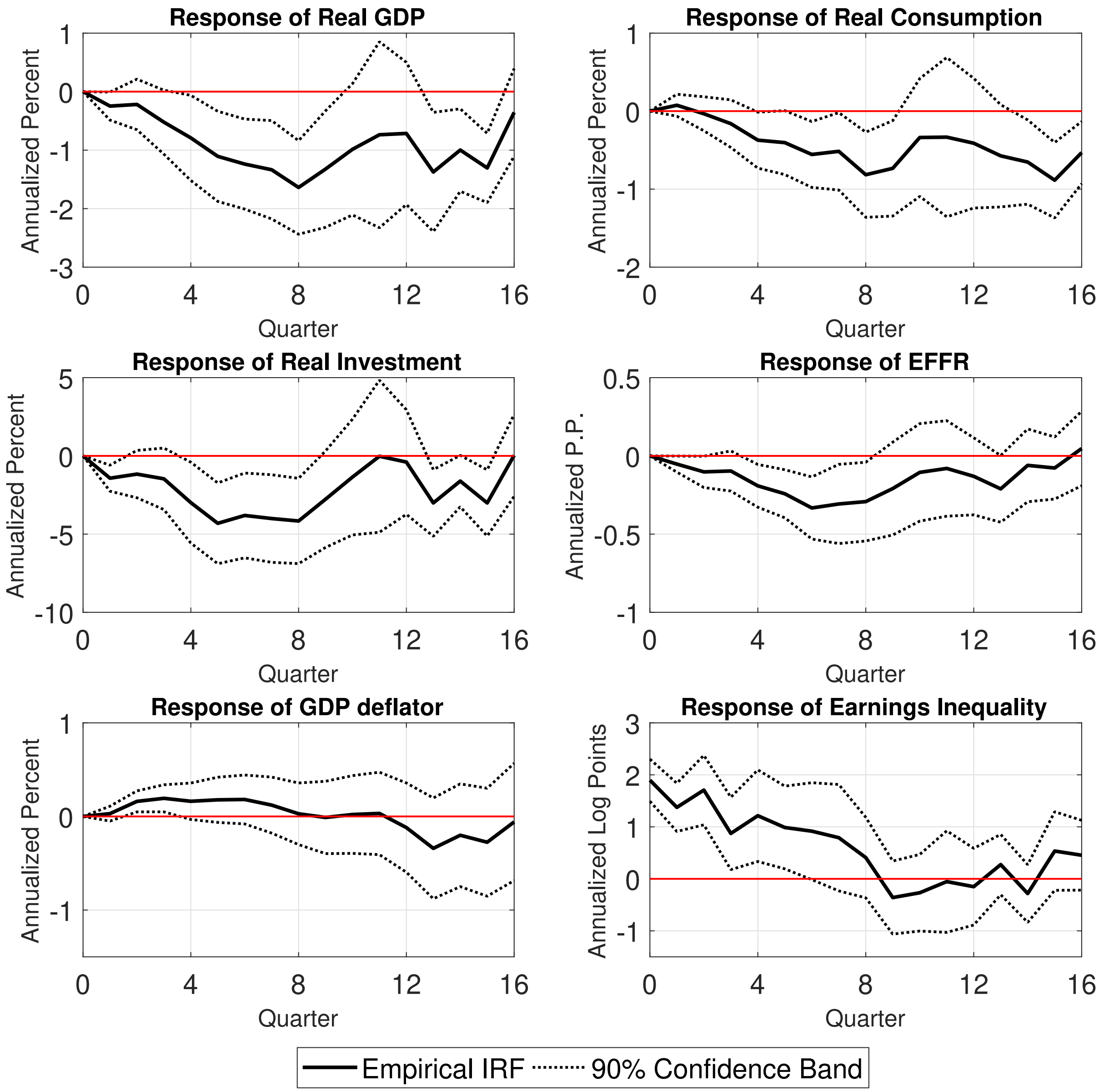

Figure 3.6: Responses of macroeconomic variables to an inequality shock.

Notes: The impulse responses are estimated by local projections in Equation (5). The dotted lines denote the 90 percent confidence bands where the bandwidth for the Newey-West variance estimator increases in the horizon of local projections one for one. All responses are either in annualized percent, annualized percentage points, or annualized log points. The bottom right panel illustrates the response of the log P90/P10 index. The results here are robust to various specification details (see Appendix C.2).

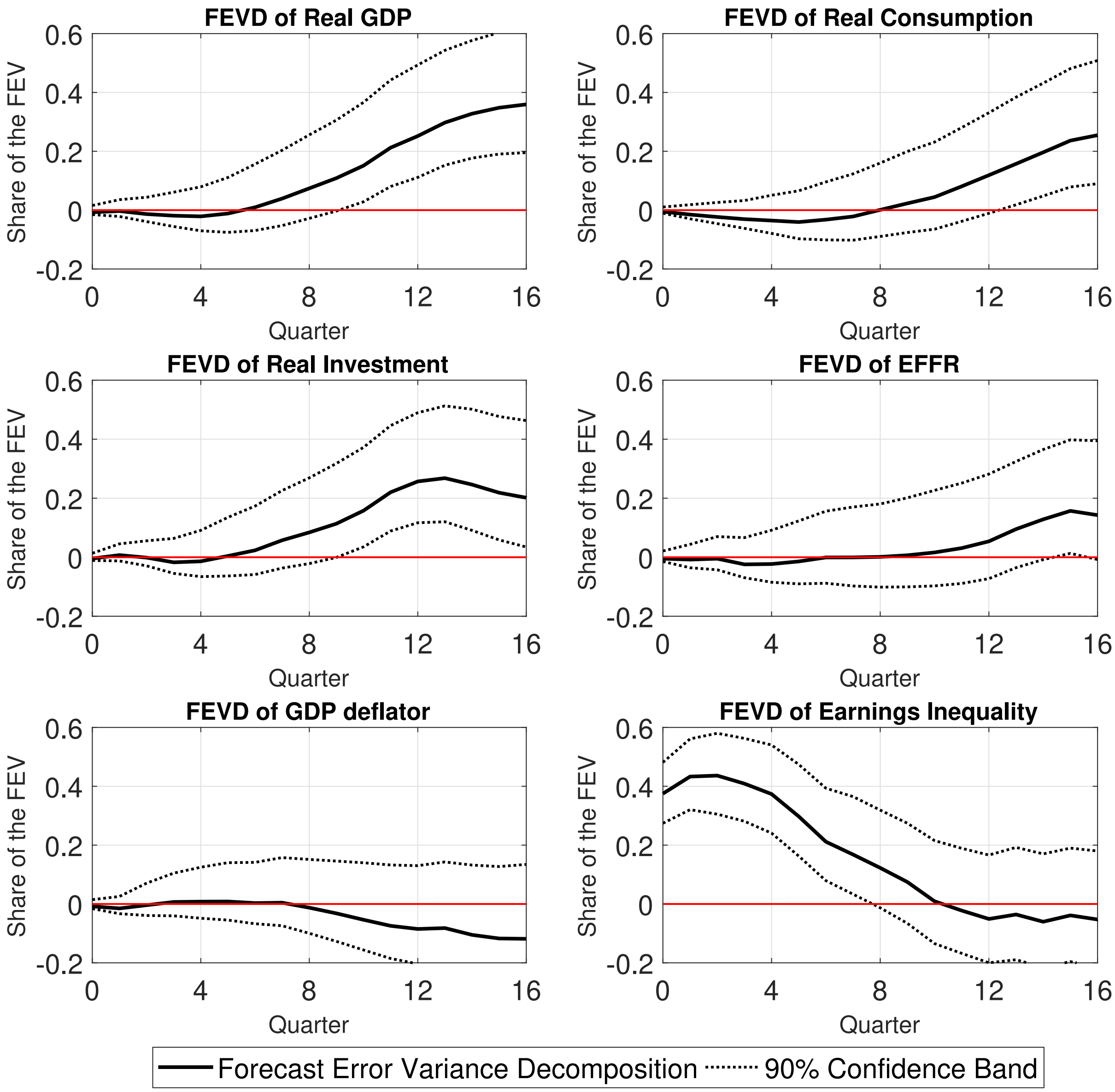

Figure 3.7: FEVDs, unanticipated innovations in inequality.

Notes: I employ the bias-corrected R^2 estimator of Gorodnichenko and Lee (2017) to estimate the forecast error variance decompositions, where the unanticipated innovations in inequality is from Section 3.4.2. The bias-correction uses bootstrapped samples from a vector autoregression model for variables in $\mathbf{Z}_t^{(m)}$ and the inequality shock. The 90 percent bootstrapped confidence bands are denoted by the dotted lines. The results are robust to specification details (see Appendix C.3).

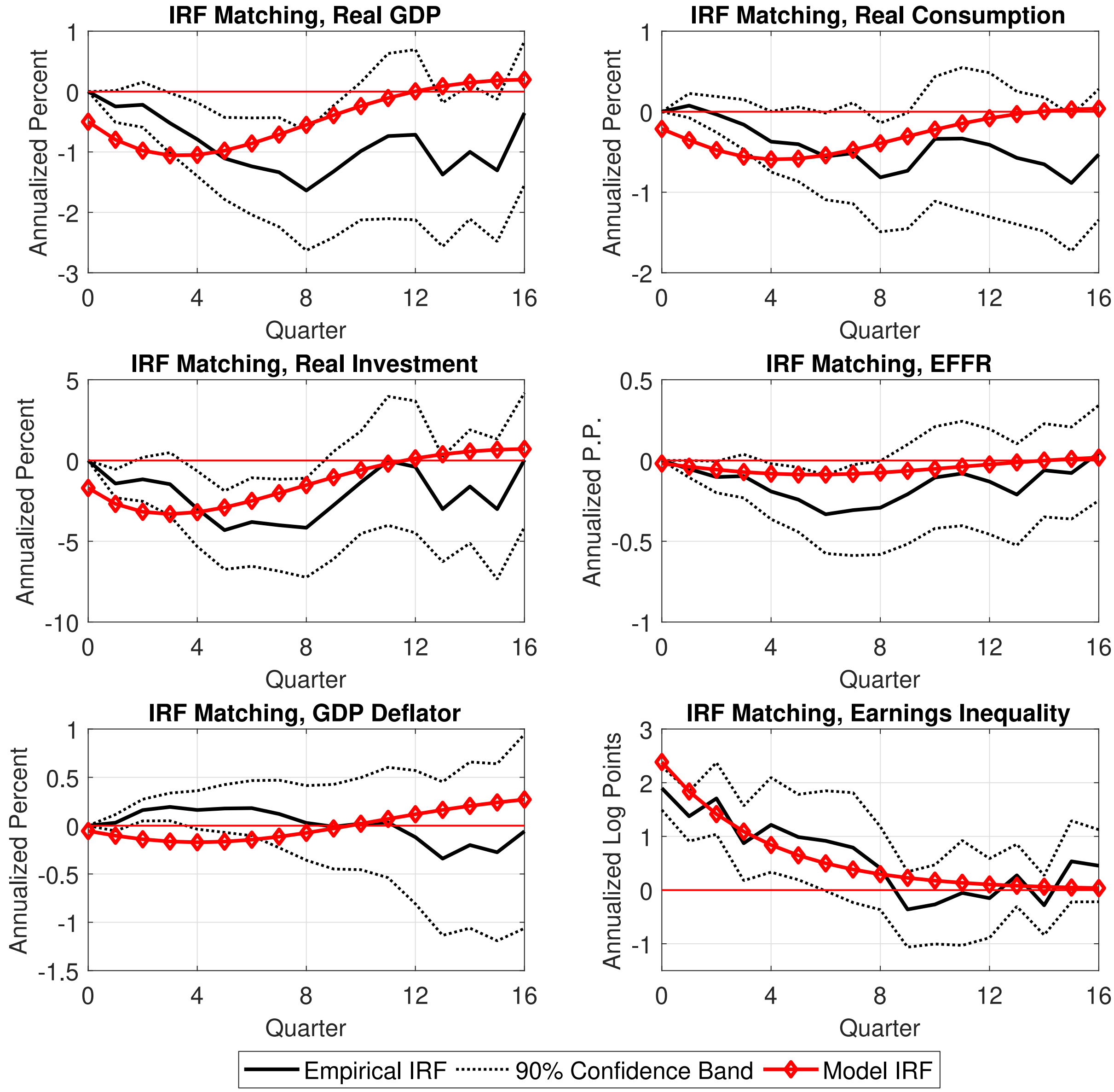

Figure 3.8: Matching impulse response functions.

Notes: The empirical impulse response functions and the confidence bands are from Figure 6. The model is evaluated at the posterior mode and the corresponding impulse response functions are illustrated by the solid lines with diamonds. For the calibrated parameters and the posterior mode, see Table 2 and 3. The peak effects of both the empirical and model responses are similar. Furthermore, the model responses are in the 90 percent confidence bands at most lags. While this is not the case for small h's, I include the moment conditions related to the inequality shocks only for $4 \leq h \leq 12$ when evaluating the posterior. This is because the empirical responses for the variables other than the inequality index at $h = 0$ are zero by construction, and this obviously affects estimates for small h's too.

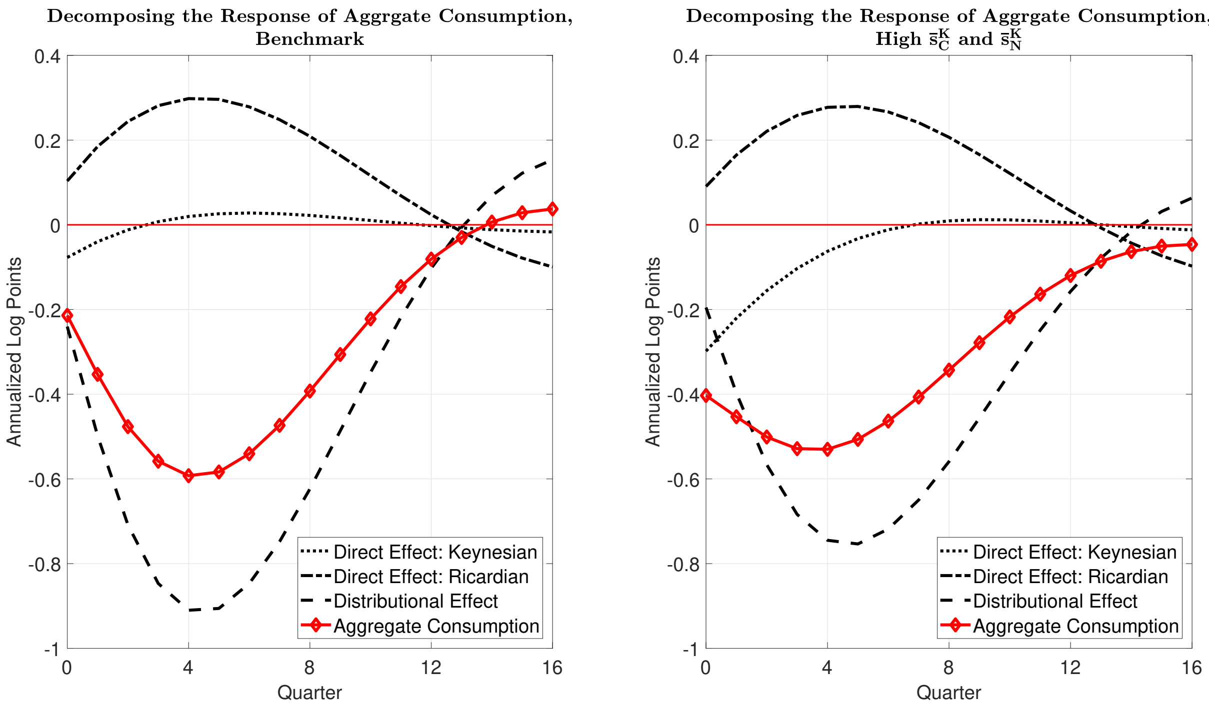

Figure 3.9: Decomposition of aggregate consumption responses.

Notes: The dotted line represents $E_t\left[\bar{s}_C^K \check{C}_{t+\tau}^K\right]$, the direct effect of a one standard deviation inequality shock to the Keynesians' consumption. The dash-dot line is for $E_t\left[\bar{s}_C^R \check{C}_{t+\tau}^R\right]$, the direct effect to the Ricardians' consumption. The distributional effect, $E_t\left[\left(\bar{s}_C^K - \frac{\bar{s}^K}{\bar{s}^R}\bar{s}_C^R\right)\check{s}_{t+\tau}^K\right]$, is illustrated by the dashed line. The total aggregate consumption response, $E_t\left[\check{C}_{t+\tau}\right]$, is shown by the red solid line with diamonds. The units are annualized percent obtained by multiplying 400 to the model outcome. The left panel is based on the benchmark parameter estimates, while the right panel is based on the posterior mode when $\bar{s}_C^K$ and $\bar{s}_N^K$ are calibrated at 0.15 and 0.12, respectively.

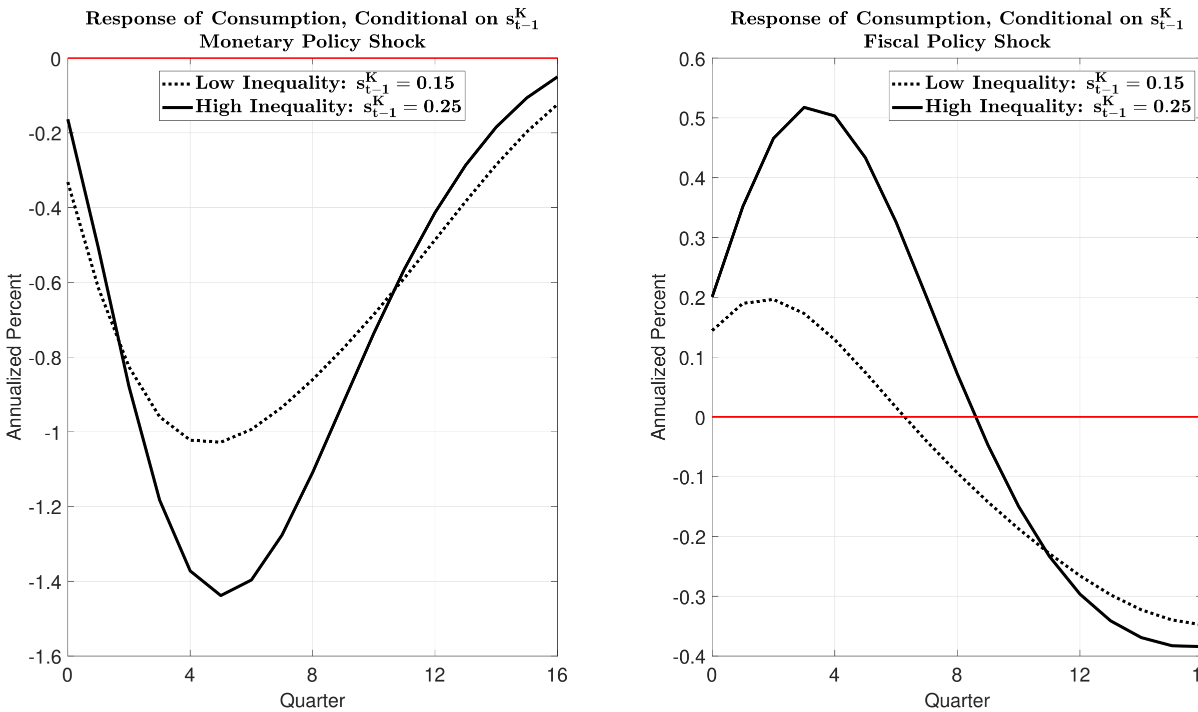

Figure 3.10: Responses of consumption conditional on the level of inequality, model.

Notes: The generalized impulse responses are based on the third order pruned state-space system in Andreasen, Fernández-Villaverde and Rubio-Ramírez (2017). In the high inequality state, $s^K_{t-1} = 0.25$ and all the other variables equal their steady-state values. The low inequality state is based on $s^K_{t-1} = 0.15$. Inputs are one standard deviation contractionary monetary policy shocks and expansionary fiscal policy shocks. The units are annualized percent. The results for other variables are in Appendix E.3.

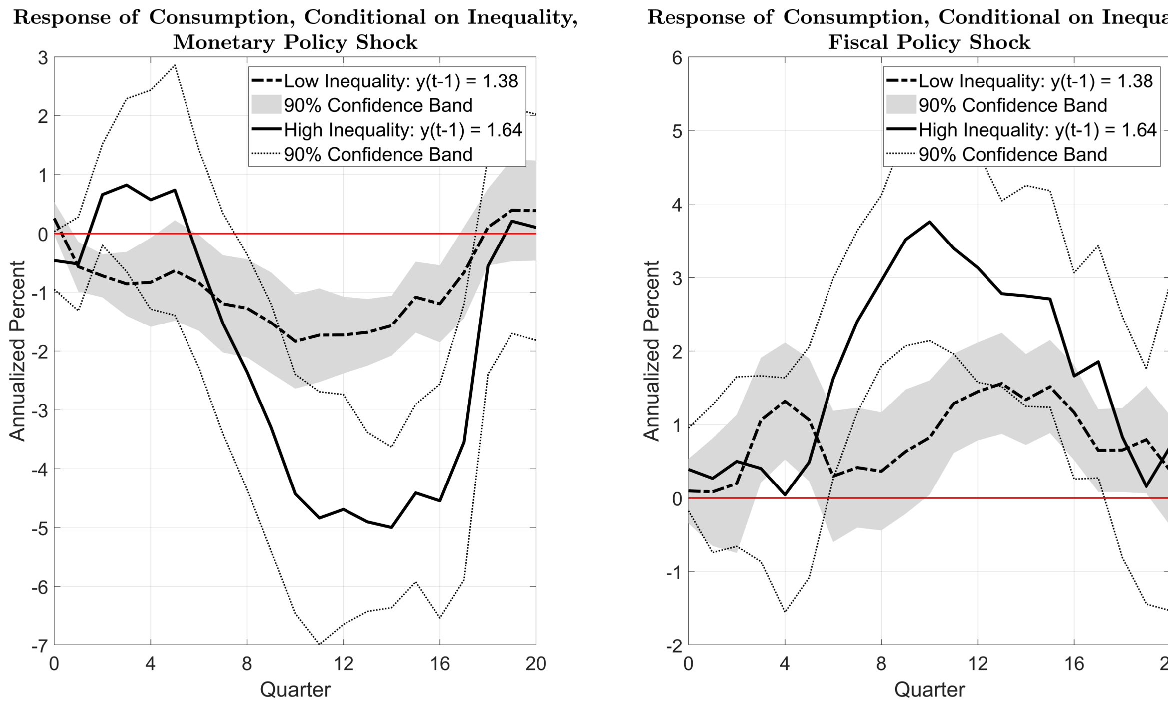

Figure 3.11: Responses of consumption conditional on the level of inequality, recent data.

Notes: Two panels depict the impulse responses of consumption given a one standard deviation contractionary monetary policy shock and an expansionary fiscal policy shock, respectively. Each panel plots two sets of results: the impulse responses conditional on the inequality index being one standard deviation below or above the mean, 1.38 and 1.64, respectively. The results are based on Equation (34), a local projection with an interaction term between the lagged inequality index and a shock. The bandwidth for the Newey-West variance estimator increases in the horizon of local projections one for one. For responses of other macroeconomic variables and the results based on total factor productivity shocks, see Appendix E.1.

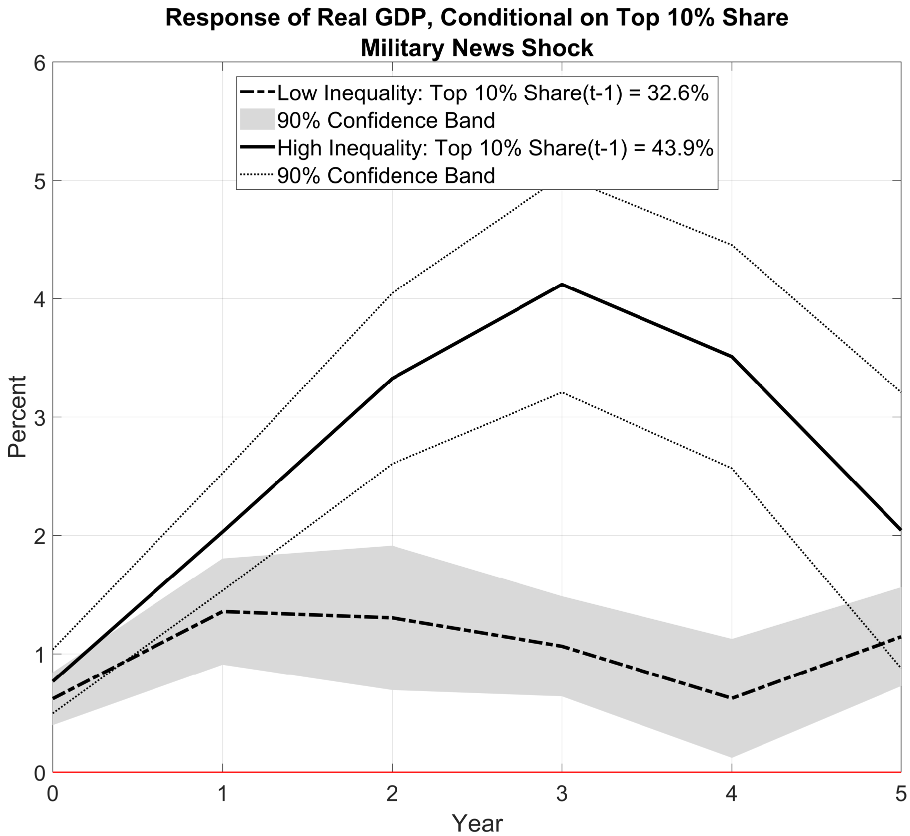

Figure 3.12: Responses of real GDP conditional on the level of inequality, historical data.

Notes: The results in this figure is based on the long historical data of Piketty and Saez (2003) and Ramey and Zubairy (2018). The top 10% income share series of Piketty and Saez is annual, and therefore the unit for the horizontal axis is a year. The sample period is from 1917 to 2015. The top 10% income share displays a U-shaped pattern during the sample periods. The results are based on Equation (34), where the bandwidth for the Newey-West variance estimator increases in the horizon of local projections one for one. The input is a military news shock whose present discounted value amounts to 10 percent of the trend GDP. The result for the GDP deflator and unemployment rate, and further robustness checks are in Appendix E.2.

Bibliography

Acharya, Sushant, and Keshav Dogra. 2018. "Understanding HANK: Insights from a PRANK." *Manuscript.*

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

Alesina, Alberto, and Roberto Perotti. 1996. "Income distribution, political instability, and investment." *European Economic Review* 40(6): 1203–1228.

Andreasen, Martin M., Jesús Fernández-Villaverde, and Juan F. Rubio-Ramírez. 2017. "The pruned state-space system for non-linear DSGE models: Theory and empirical applications." *Review of Economic Studies* 85(1): 1–49.

Andrews, Donald W.K., and J. Christopher Monahan. 1992. "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator." *Econometrica* 60(4): 953–966.

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.

Auclert, Adrien. 2017. "Monetary Policy and the Redistribution Channel." National Bureau of Economic Research Working Paper 23451.

Auclert, Adrien, and Matthew Rognlie. 2018. "Inequality and aggregate demand." National Bureau of Economic Research Working Paper 24280.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2012. "Measuring the output responses to fiscal policy." *American Economic Journal: Economic Policy* 4(2): 1–27.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in US wage inequality: Revising the revisionists." *Review of Economics and Statistics* 90(2): 300–323.

Baek, ChaeWon, and Byoungchan Lee. 2019. "ABCs of understanding single equation regressions." *manuscript.*

Barakchian, S. Mahdi, and Christopher Crowe. 2013. "Monetary policy matters: Evidence from new shocks data." *Journal of Monetary Economics* 60(8): 950–966.

Barsky, Robert B., and Eric R. Sims. 2012. "Information, animal spirits, and the meaning of innovations in consumer confidence." *American Economic Review* 102(4): 1343–77.

Basu, Susanto, John G. Fernald, and Miles S. Kimball. 2006. "Are technology improvements contractionary?" *American Economic Review* 96(5): 1418–1448.

Beaudry, Paul, and Franck Portier. 2006. "Stock prices, news, and economic fluctuations." *American Economic Review* 96(4): 1293–1307.

Bilbiie, Florin O. 2008. "Limited asset markets participation, monetary policy and (inverted) aggregate demand logic." *Journal of Economic Theory* 140(1): 162–196.

Bilbiie, Florin O. 2017. "The New Keynesian cross: Understanding monetary policy with hand-to-mouth households." CEPR Discussion Paper No. DP11989.

Bilbiie, Florin O., André Meier, and Gernot J. Müller. 2008. "What accounts for the changes in US fiscal policy transmission?" *Journal of Money, Credit and Banking* 40(7): 1439–1470.

Blanchard, Olivier Jean, and Danny Quah. 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *American Economic Review* 79(4): 655–673.

Blank, Rebecca M., and Alan S. Blinder. 1986. "Macroeconomics, income distribution, and poverty." In *Fighting Poverty: What Works and What Does Not.* , ed. Sdon Danziger. Cambridge:Harvard University Press.

Blundell, Richard, and Ian Preston. 1998. "Consumption inequality and income uncertainty." *Quarterly Journal of Economics* 113(2): 603–640.

Bordo, Michael D., and Christopher M. Meissner. 2012. “Does inequality lead to a financial crisis?” *Journal of International Money and Finance* 31(8): 2147–2161.

Braun, R. Anton, and Tomoyuki Nakajima. 2012. “Uninsured countercyclical risk: an aggregation result and application to optimal monetary policy.” *Journal of the European Economic Association* 10(6): 1450–1474.

Byrne, David M., John G. Fernald, and Marshall B. Reinsdorf. 2016. “Does the United States have a productivity slowdown or a measurement problem?” *Brookings Papers on Economic Activity* 2016(1): 109–182.

Cairó, Isabel, and Jae Sim. 2017. “Income inequality, financial crisis and monetary policy.” *Manuscript.*

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini. 2009. “Fight or flight? Portfolio rebalancing by individual investors.” *Quarterly Journal of Economics* 124(1): 301–348.

Calvo, Guillermo A. 1983. “Staggered prices in a utility-maximizing framework.” *Journal of Monetary Economics* 12(3): 383–398.

Campbell, John Y., and N. Gregory Mankiw. 1989. “Consumption, income, and interest rates: Reinterpreting the time series evidence.” *NBER Macroeconomics Annual* 4: 185–216.

Chang, Minsu, Xiaohong Chen, and Frank Schorfheide. 2018. “Heterogeneity and Aggregate Fluctuations.” *Manuscript.*

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2005. “Nominal rigidities and the dynamic effects of a shock to monetary policy.” *Journal of Political Economy* 113(1): 1–45.

Christiano, Lawrence J., Mathias Trabandt, and Karl Walentin. 2010. “DSGE models for monetary policy analysis.” In *Handbook of Monetary Economics.* Vol. 3, 285–367. Elsevier.

Coibion, Olivier. 2012. “Are the effects of monetary policy shocks big or small?” *American Economic Journal: Macroeconomics* 4(2): 1–32.

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What can survey forecasts tell us about information rigidities?" *Journal of Political Economy* 120(1): 116–159.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia. 2017. "Innocent Bystanders? Monetary policy and inequality." *Journal of Monetary Economics* 88: 70–89.

Constantinides, George M., and Darrell Duffie. 1996. "Asset pricing with heterogeneous consumers." *Journal of Political Economy* 104(2): 219–240.

Cowell, Frank A., and Emmanuel Flachaire. 2015. "Statistical methods for distributional analysis." In *Handbook of Income Distribution.* Vol. 2, 359–465. Elsevier.

Cramer, Jan Solomon. 1987. "Mean and variance of R2 in small and moderate samples." *Journal of Econometrics* 35(2-3): 253–266.

Crook, Jonathan. 2001. "The demand for household debt in the USA: evidence from the 1995 Survey of Consumer Finance." *Applied Financial Economics* 11(1): 83–91.

Crook, Jonathan. 2006. "Household debt demand and supply: A cross-country comparison." In *The Economics of Consumer Credit.* , ed. Giuseppe Bertola, Richard Disney and Charles Grant, 63–92. MA:Mit Press Cambridge.

Debortoli, Davide, and Jordi Galí. 2017. "Monetary policy with heterogeneous agents: Insights from TANK models." *manuscript.*

Doepke, Matthias, and Martin Schneider. 2006. "Inflation and the redistribution of nominal wealth." *Journal of Political Economy* 114(6): 1069–1097.

Dolado, Juan, Gergo Motyovszki, and Evi Pappa. 2018. "Monetary policy and inequality under labor market frictions and capital-skill complementarity." *Manuscript.*

Dupor, Bill, Jing Han, and Yi-Chan Tsai. 2009. "What do technology shocks tell us about the New Keynesian paradigm?" *Journal of Monetary Economics* 56(4): 560–569.

Dynan, Karen E., Jonathan Skinner, and Stephen P. Zeldes. 2004. "Do the rich save more?" *Journal of Political Economy* 112(2): 397–444.

Elsby, Michael W.L., Bart Hobijn, and Ayşegül Şahin. 2013. "The decline of the U.S. labor share." *Brookings Papers on Economic Activity* 2013(2): 1–63.

Fernald, John. 2014. "A quarterly, utilization-adjusted series on total factor productivity." Federal Reserve Bank of San Francisco Working Paper 2012–19.

Fernald, John G. 2015. "Productivity and Potential Output before, during, and after the Great Recession." *NBER Macroeconomics Annual* 29(1): 1–51.

Floden, Martin, and Jesper Lindé. 2001. "Idiosyncratic risk in the United States and Sweden: Is there a role for government insurance?" *Review of Economic dynamics* 4(2): 406–437.

Frank, Mark, Estelle Sommeiller, Mark Price, and Emmanuel Saez. 2015. "Frank-Sommeiller-Price series for top income shares by US states since 1917." *WTID Methodological Notes.*

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. "The dynamics of inequality." *Econometrica* 84(6): 2071–2111.

Galí, Jordi. 2015. *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications.* Princeton University Press.

Galí, Jordi, J. David López-Salido, and Javier Vallés. 2004. "Rule-of-Thumb Consumers and the Design of Interest Rate Rules." *Journal of Money, Credit, and Banking* 36(4): 739–763.

Galí, Jordi, J. David López-Salido, and Javier Vallés. 2007. "Understanding the effects of government spending on consumption." *Journal of the European Economic Association* 5(1): 227–270.

Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit spreads and business cycle fluctuations." *American Economic Review* 102(4): 1692–1720.

Goldin, Claudia Dale, and Lawrence F. Katz. 2009. *The Race between Education and Technology.* Harvard University Press.

Gonçalves, Sılvia, and Lutz Kilian. 2004. "Bootstrapping autoregressions with conditional heteroskedasticity of unknown form." *Journal of Econometrics* 123(1): 89–120.

Gornemann, Nils, Keith Kuester, and Makoto Nakajima. 2016. "Doves for the rich, hawks for the poor? Distributional consequences of monetary policy." *Manuscript*.

Gorodnichenko, Yuriy, and Byoungchan Lee. 2017. "A Note on Variance Decomposition with Local Projections." National Bureau of Economic Research Working Paper 23998.

Guerrieri, Veronica, and Guido Lorenzoni. 2017. "Credit crises, precautionary savings, and the liquidity trap." *Quarterly Journal of Economics* 132(3): 1427–1467.

Guiso, Luigi, and Paolo Sodini. 2013. "Household finance: An emerging field." In *Handbook of the Economics of Finance*. Vol. 2, 1397–1532. Elsevier.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2015. "What do data on millions of US workers reveal about life-cycle earnings risk?" National Bureau of Economic Research Working Paper 20913.

Guvenen, Fatih, Sam Schulhofer-Wohl, Jae Song, and Motohiro Yogo. 2017. "Worker betas: Five facts about systematic earnings risk." *American Economic Review: Papers & Proceedings* 107(5): 398–403.

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

Hagedorn, Marcus, Iourii Manovskii, and Kurt Mitman. 2019. "The fiscal multiplier." National Bureau of Economic Research Working Paper 25571.

Hansen, Lars Peter. 1982. "Large sample properties of generalized method of moments estimators." *Econometrica* 50(4): 1029–1054.

Heer, Burkhard, and Christian Scharrer. 2016. "The Burden of Unanticipated Government Spending." CESifo Working Paper 5876.

Herbst, Edward P, and Frank Schorfheide. 2015. *Bayesian estimation of DSGE models.* Princeton University Press.

Hodrick, Robert J., and Edward C. Prescott. 1997. "Postwar U.S. business cycles: an empirical investigation." *Journal of Money, credit, and Banking* 1–16.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles. 2006. "Household expenditure and the income tax rebates of 2001." *American Economic Review* 96(5): 1589–1610.

Jordà, Òscar. 2005. "Estimation and inference of impulse responses by local projections." *American Economic Review* 95(1): 161–182.

Judd, Kenneth L. 1998. *Numerical methods in economics.* MIT press.

Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng. 2015. "Measuring uncertainty." *American Economic Review* 105(3): 1177–1216.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti. 2010. "Investment shocks and business cycles." *Journal of Monetary Economics* 57(2): 132–145.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti. 2011. "Investment shocks and the relative price of investment." *Review of Economic Dynamics* 14(1): 102–121.

Kaplan, Greg, and Giovanni L. Violante. 2014. "A model of the consumption response to fiscal stimulus payments." *Econometrica* 82(4): 1199–1239.

Kaplan, Greg, and Giovanni L. Violante. 2018. "Microeconomic Heterogeneity and Macroeconomic Shocks." *Journal of Economic Perspectives* 32(3): 167–194.

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

Kaplan, Greg, Giovanni L. Violante, and Justin Weidner. 2014. "The wealthy hand-to-mouth." *Brookings Papers on Economic Activity* 1: 77–138.

Karabarbounis, Loukas, and Brent Neiman. 2013. "The global decline of the labor share." *Quarterly Journal of Economics* 129(1): 61–103.

Kilian, Lutz. 2008. "Exogenous oil supply shocks: how big are they and how much do they matter for the US economy?" *Review of Economics and Statistics* 90(2): 216–240.

Kilian, Lutz. 2009. "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market." *American Economic Review* 99(3): 1053–69.

Kilian, Lutz, and Helmut Lütkepohl. 2017. *Structural vector autoregressive analysis.* Cambridge University Press.

Kilian, Lutz, and Yun Jung Kim. 2011. "How reliable are local projection estimators of impulse responses?" *Review of Economics and Statistics* 93(4): 1460–1466.

Kim, Jae-Young. 2002. "Limited information likelihood and Bayesian analysis." *Journal of Econometrics* 107(1-2): 175–193.

Koenker, Roger, and José A. F. Machado. 1999. "Goodness of fit and related inference processes for quantile regression." *Journal of the American Statistical Association* 94(448): 1296–1310.

Koh, Dongya, Raul Santaeulalia-Llopis, and Yu Zheng. 2018. "Labor Share Decline and Intellectual Property Products Capital." *Manuscript.*

Koop, Gary, M. Hashem Pesaran, and Simon M. Potter. 1996. "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics* 74(1): 119–147.

Koornwinder, Tom H., Roderick S. C. Wong, Roelof Koekoek, and René F. Swarttouw. 2010. "Orthogonal Polynomials." In *NIST Handbook of Mathematical Functions.* Cambridge:Cambridge University Press.

Kotz, Samuel, Tomasz J. Kozubowski, and Krzysztof Podgórski. 2001. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance.* Birkhäuser Boston.

Krueger, Dirk, Kurt Mitman, and Fabrizio Perri. 2016. "Macroeconomics and household heterogeneity." In *Handbook of Macroeconomics.* Vol. 2, 843–921. Elsevier.

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

Krusell, Per, Lee E. Ohanian, José-Víctor Ríos-Rull, and Giovanni L. Violante. 2000. "Capital-skill complementarity and inequality: A macroeconomic analysis." *Econometrica* 68(5): 1029–1053.

Kuhn, Moritz, and José-Victor Rios-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): 1–75.

Kumhof, Michael, Romain Rancière, and Pablo Winant. 2015. "Inequality, leverage, and crises." *American Economic Review* 105(3): 1217–45.

Kwiatkowski, Denis, Peter C.B. Phillips, Peter Schmidt, and Yongcheol Shin. 1992. "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of Econometrics* 54(1-3): 159–178.

Linden, Mikael. 2001. "A model for stock return distribution." *International Journal of Finance and Economics* 6(2): 159–169.

Lorenzoni, Guido. 2009. "A theory of demand shocks." *American Economic Review* 99(5): 2050–84.

McKay, Alisdair, and Ricardo Reis. 2016. "The role of automatic stabilizers in the US business cycle." *Econometrica* 84(1): 141–194.

McKay, Alisdair, Emi Nakamura, and Jón Steinsson. 2016. "The power of forward guidance revisited." *American Economic Review* 106(10): 3133–58.

McKay, Alisdair, Emi Nakamura, and Jón Steinsson. 2017. "The discounted euler equation: A note." *Economica* 84(336): 820–831.

Mian, Atif, and Amir Sufi. 2015. *House of debt: How they (and you) caused the Great Recession, and how we can prevent it from happening again.* University of Chicago Press.

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

Nakamura, Emi, and Jon Steinsson. 2014. "Fiscal stimulus in a monetary union: Evidence from US regions." *American Economic Review* 104(3): 753–92.

Pagan, Adrian. 1984. "Econometric issues in the analysis of regressions with generated regressors." *International Economic Review* 25(1): 221–247.

Parker, Jonathan A., and Annette Vissing-Jorgensen. 2009. "Who bears aggregate fluctuations and how?" *American Economic Review: Papers & Proceedings* 99(2): 399–405.

Parker, Jonathan A., Nicholas S. Souleles, David S. Johnson, and Robert McClelland. 2013. "Consumer spending and the economic stimulus payments of 2008." *American Economic Review* 103(6): 2530–53.

Pencavel, John H. 1984. "The tradeoff between wages and employment in trade union objectives." *Quarterly Journal of Economics* 99(2): 215–231.

Phaneuf, Louis, Eric Sims, and Jean Gardy Victor. 2018. "Inflation, output and markup dynamics with purely forward-looking wage and price setters." *European Economic Review* 105: 115–134.

Phillips, Peter C.B., and Pierre Perron. 1988. "Testing for a unit root in time series regression." *Biometrika* 75(2): 335–346.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income inequality in the United States, 1913–1998." *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." *Quarterly Journal of Economics* 133(2): 553–609.

Pistaferri, Luigi. 2001. "Superior information, income shocks, and the permanent income hypothesis." *Review of Economics and Statistics* 83(3): 465–476.

Plagborg-Møller, Mikkel, and Christian K. Wolf. 2018*a*. "Instrumental variable identification of dynamic variance decompositions." *manuscript*.

Plagborg-Møller, Mikkel, and Christian K. Wolf. 2018*b*. "Local Projections and VARs Estimate the Same Impulse Responses." *manuscript*.

Ragot, Xavier. 2018. "Heterogeneous agents in the macroeconomy: reduced-heterogeneity representations." In *Handbook of Computational Economics*. Vol. 4, 215–253. Elsevier.

Ramey, Valerie A. 2011. "Identifying government spending shocks: It's all in the timing." *Quarterly Journal of Economics* 126(1): 1–50.

Ramey, Valerie A., and Sarah Zubairy. 2018. "Government spending multipliers in good times and in bad: evidence from US historical data." *Journal of Political Economy* 126(2): 850–901.

Ravn, Morten O., and Harald Uhlig. 2002. "On adjusting the Hodrick-Prescott filter for the frequency of observations." *Review of Economics and Statistics* 84(2): 371–376.

Ravn, Morten O., and Vincent Sterk. 2017. "Job uncertainty and deep recessions." *Journal of Monetary Economics* 90: 125–141.

Ravn, Morten O., and Vincent Sterk. 2018. "Macroeconomic fluctuations with HANK & SAM: An analytical approach." *Manuscript*.

Romer, Christina D., and David H. Romer. 1999. "Monetary Policy and the Well-Being of the Poor." *Federal Reserve Bank of Kansas City Economic Review* 84(1): 21–49.

Romer, Christina D., and David H. Romer. 2004. "A new measure of monetary shocks: Derivation and implications." *American Economic Review* 94(4): 1055–1084.

Romer, Christina D., and David H. Romer. 2010. "The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks." *American Economic Review* 100(3): 763–801.

Rotemberg, Julio J. 1982. "Sticky prices in the United States." *Journal of Political Economy* 90(6): 1187–1211.

Rotemberg, Julio J., and Michael Woodford. 1997. "An optimization-based econometric framework for the evaluation of monetary policy." *NBER Macroeconomics Annual* 12: 297–346.

Rothschild, Michael, and Joseph E Stiglitz. 1970. "Increasing risk: I. A definition." *Journal of Economic Theory* 2(3): 225–243.

Rothschild, Michael, and Joseph E Stiglitz. 1971. "Increasing risk: II. Its economic consequences." *Journal of Economic Theory* 3(1): 66–84.

Rouwenhorst, K. Geert. 1995. "Asset Pricing Implications of Equilibrium Business Cycle Models." In *Frontiers of Business Cycle Research*. Princeton, New Jersey:Princeton University Press.

Rudin, Walter. 1987. *Real and complex analysis.* McGraw-Hill Book Company.

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

Sims, Christopher A. 1980. "Macroeconomics and reality." *Econometrica* 48(1): 1–48.

Smets, Frank, and Rafael Wouters. 2007. "Shocks and frictions in US business cycles: A Bayesian DSGE approach." *American Economic Review* 97(3): 586–606.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. 2018. "Firming up inequality." *Quarterly Journal of Economics* 134(1): 1–50.

Stock, James H., and Mark W. Watson. 2007. "Why has US inflation become harder to forecast?" *Journal of Money, Credit and Banking* 39(1): 3–33.

Stock, James H., and Mark W. Watson. 2011. "Dynamic factor models." In *Oxford Handbook on Economic Forecasting.* Oxford University Press.

Stock, James H., and Mark W. Watson. 2016. "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics." In *Handbook of Macroeconomics.* Vol. 2, 415–525. Elsevier.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron. 2004. "Cyclical dynamics in idiosyncratic labor market risk." *Journal of Political Economy* 112(3): 695–717.

Straub, Ludwig. 2018. "Consumption, Savings, and the Distribution of Permanent Income." *Manuscript.*

Tauchen, George. 1986. "Finite state markov-chain approximations to univariate and vector autoregressions." *Economics Letters* 20(2): 177–181.

Tauchen, George, and Robert Hussey. 1991. "Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models." *Econometrica* 371–396.

Taylor, James W. 2019. “Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution.” *Journal of Business and Economic Statistics* 37(1): 121–133.

Werning, Iván. 2015. “Incomplete markets and aggregate demand.” National Bureau of Economic Research Working Paper 21448.

Wolf, Christian K. 2017. “Masquerading Shocks in Sign-Restricted VARs.” *manuscript.*

Wolff, Edward N. 2017. “Household Wealth Trends in the United States, 1962 to 2016: Has Middle Class Wealth Recovered?” National Bureau of Economic Research Working Paper 24085.

Woodford, Michael. 2003. *Interest and prices: Foundations of a theory of monetary policy.* Princeton University Press.

Yellen, Janet L. 2016. “Macroeconomic research after the crisis.” *Speech given at the 60th Annual Economic Conference, Federal Reserve Bank of Boston,* Boston, October 14.

Zidar, Owen M. 2018. “Tax cuts for whom? Heterogeneous effects of income tax changes on growth and employment.” *Journal of Political Economy,* forthcoming.