ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

A High-Frequency Measure of Income Inequality

Authors	Marie Hogan, Laura E. Jackson, and Michael T. Owyang
Working Paper Number	2024-021A
Creation Date	September 2024
Citable Link	https://doi.org/10.20955/wp.2024.021
Suggested Citation	Hogan, M., Jackson, L.E., Owyang, M.T., 2024; A High-Frequency Measure of Income Inequality, Federal Reserve Bank of St. Louis Working Paper 2024-021. URL https://doi.org/10.20955/wp.2024.021

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

The views expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment.

A High-Frequency Measure of Income Inequality*

Marie Hogan†, Laura E. Jackson‡and Michael T. Owyang†

keywords: temporal disaggregation, productivity shocks, house price shocks

August 30, 2024

Abstract

To identify shocks in VARs using short-run sign or exclusion restrictions, the highest-frequency data possible is usually preferred. For income inequality, there is tension between high frequency and high quality. Annual datasets that survey large numbers of people provide high-quality estimates of income. Higher frequency surveys generally provide a sparser sampling of individual income. Some previous studies have used the the higher frequency data, presumably to match the frequency necessary to identify the shock. Using data obtained from the higher frequency, lower respondent surveys might result in misleading conclusions. We combine the two surveys and construct a set of quarterly-frequency income quantiles that are scaled to the annual data but fluctuate according to the high-frequency survey. We show that using these data yields very different economic conclusions than simply using the raw high-frequency income series. In particular, we show in two simple applications to technology shocks and house price shocks that one obtains different conclusions about the permanence and/or the direction of the responses of income inequality.

JEL Codes: C32, E00, E32

*Brooke Hathhorn, Hoang Le and Ashley H. Stewart provided research assistance. The authors benefitted from discussions with Neville Francis and Mike McCracken and thank Gene Kindberg-Hanlon for code. The views expressed here are not the official opinions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

†Research Division. Federal Reserve Bank of St. Louis.

‡Corresponding Author: Department of Economics, Bentley University. Email: ljackson@bentley.edu

1 Introduction

Identifying a shock's effect on inequality in a VAR can rely on short-run timing or sign restrictions that are complicated by the fact that many income inequality measures are observed only annually. For example, the March Supplement of the Current Population Survey (CPS) conducted by the Census is one source of measures of income inequality. Higher frequency measures of income inequality do exist. The Consumer Expenditure Survey (CEX) provides a quarterly measure that has been used by a number of studies [e.g., Coibion, Gorodnichenko, Kueng, and Silvia (2017) and others]. However, the number of survey participants in the CEX is smaller than the annual surveys and the incidence of nonresponse is higher. These issues may affect the measured distribution of income for the higher-frequency surveys.[1] If incomes are mismeasured, the subsequent impulse responses may lead to false economic conclusions about the effects of the shocks.

We construct high frequency measures of income inequality. To balance the natural tension between the number of respondents needed to fully capture the density of income and the frequency of the survey, we use the annual CPS as our reference measure for the income quantiles and create a high-frequency interpolation using the income quantiles of the quarterly CEX. We perform the interpolation using a Bayesian version of the Chow-Lin temporal disaggregation [see Chow and Lin (1971)]. While it is not explicitly necessary to use a Bayesian algorithm, the rejection sampler helps solve a quantile crossing problem that can arise when obtaining the Chow-Lin regression parameters independently for each quantile.

We use our quarterly CPS income data to estimate the effects of technology and housing price shocks on inequality and compare our results to those using the raw, unadjusted CEX.[2] In the first application, the technology shock is identified using the Max Share algorithm

[1] Very recently, Blanchet, Saez, and Zucman (2023) also produced a high-frequency estimate of the income distribution in the U.S. using data from monthly household and employment surveys, quarterly censuses of employment and wages, and monthly and quarterly national accounts statistics.

[2] We are unable to perform a similar exercise using the inequality measures of Blanchet, Saez, and Zucman (2023). Our series are constructed as the income levels at various quantiles. The public-facing data of Blanchet, Saez, and Zucman (2023) are transformed in such a way that doesn't allow for a direct comparison. For instance, we produce the income level at the 99th percentile while Blanchet, Saez, Zucman (2023) produce the income *of the top 1%*. Similarly, we produce an income level at the 50th percentile while they produce the income *of the bottom 50%*. However, we do attempt to adjust their measure to be more aligned with ours and compare the results in the Appendix.

[Francis, Owyang, Roush, and DiCecio (2014)] as the shock that maximizes the forecast error variance share of labor productivity at a long but finite horizon. We find that technology shocks raise income inequality. Moreover, we find that when using our quarterly CPS measures, the effect of technology on income inequality is permanent, possibly suggesting either a heterogeneous producivity response or an amplification mechanism resulting from, say, heterogeneous ownership of capital [as in Amberg, Jansson, Klein, and Rogantini Picco (2022) or Jackson, Otrok, Owyang (2022)]. On the other hand, using the CEX data, the effect of technology on inequality can be transitory, depending on how inequality is measured.

In our second application, we identify house price shocks using short-run exclusion restrictions and find that house price increases raise income inequality in the short run. However, the results at longer horizons differ across the datasets. For our quarterly CPS measures, the effect of house price shocks on income inequality is negligible in the long run. On the other hand, using the CEX measures, income inequality falls in the long run. These results lead us to conclude that using different inequality data can lead to very different economic conclusions about how shocks affect income disparities.

The remainder of the paper is laid out as follows: Section 2 outlines the income data used to construct our high-frequency inequality measures. Section 3 describes the methods used for the interpolation. Section 4 compares our high-frequency CPS measure to some of the available income inequality data. Section 5 contains two applications: Section 5.1 shows the effects of technology shocks on income inequality and Section 5.2 shows the effects of house price shocks on income inequality. In these two applications, we pay particular attention to how the results would differ using the raw CEX inequality data versus our quarterly CPS data. Section 6 concludes.

2 Inequality Data

A number of sources of income data can be used to form measures of income inequality. For example, the Internal Revenue Service's Statistics of Income program (SOI) uses information from taxpayers or the Census Bureau's CPS Annual Social and Economic Supplement (ASEC)

uses annual survey data of 75,000 households.[3] Unfortunately, the frequency of these datasets are undesirably low for VAR analysis, as many of the shock identification schemes require higher-than-annual frequency data. The CEX offers a higher-frequency alternative source of income data but offers a smaller cross-sectional sample with a higher rate of nonresponse.

Our objective is to combine the higher sampling density of the CPS with the high frequency fluctuations of the CEX. Thus, we sum CPS income variables at the ASEC-defined household level and follow Coibion, Gorodnichenko, Kueng, and Silvia (2017) to scale household income by household size using the OECD household equivalency scale. We then convert to real dollars. For each year, we calculate tenth through ninth deciles, 95th, and 99th percentile incomes at the national level.

To construct quarter-year income quantiles from the CEX, we use two main files: FMLI, which reports household level summary variables for demographics, income, and spending; and MEMI, which reports more detailed household member level income variables. We merge the FMLI, MEMI, and MTBI files at the household level, summing within household to construct our measure of pre-transfer income. We again scale household income by household size using the OECD equivalency scale. We impute income of nonreporters by category at the household level for all years prior to 2004 similar to Coibion, Gorodnichenko, Kueng, and Silvia (2017).[4]

We include income from the following (keeping definitions as consistent as possible throughout the same period): Wages and salaries (including tips); Social security (since it can be taxable); Own-business income and self-employment income; Farming income; Income from rents; Financial income (i.e., dividends, interest); and Royalties. We exclude government transfers except for social security.[5]

[3] The SOI data better captures the top of the income distribution because some high incomes are top-coded in the CPS. On the other hand, low-income households sometimes do not file income taxes. Thus, the CPS is often thought to be better than the SOI for low- and middle-income data.

[4] From 2004 onward, the BLS imputes income. In those years, we follow their recommendation and take the mean of the five iterations of BLS imputations. We try to replicate this process for earlier years and regress reported income on householder education level, age (quadratic), race, number of weeks worked per year, family size, and family age distribution using complete reporters within that year. We use this regression to predict income for nonreporting households, then add random shocks drawn with replacement from the regression's residuals. Finally, we trim our imputed values to fit within the maximum and minimum values reported and add the imputed values back into the full dataset.

[5] There are definitional changes in subcategories over the years and the subcategories used by CEX and CPS do not match exactly in their delineation between sources of non-wage income.

3 Data Methodology

To obtain a high-quality, high-frequency time series of income quantiles, we calibrate the level of our series to the annual CPS but infer the high-frequency movements from the quarterly CEX. Because of differences in how the questions in each survey are worded, interpolating the CPS income quantiles on the CEX income quantiles requires a two-step procedure described in this section.

First, we temporally disaggregate the income quantiles from the CEX. The question in the CEX survey asks the respondent about their income *over the last year*, which presents two problems: (i) the quarterly-sampled annual CEX income may be serially correlated and (ii) the quarterly-sampled values should not—even hypothetically—sum to the aggregate CPS value. Thus, we must first disaggregate the quarterly-sampled CEX annual income (QA CEX) into a measure of quarterly-sampled quarterly income (QQ CEX). Second, we interpolate the quarterly-sampled quarterly CPS income (QQ CPS) from the QQ CEX and the annually-sampled CPS annual income (A CPS). As part of this procedure, we impose restrictions on the QQ CPS income to prevent quantile crossings.

3.1 Temporal Disaggregation

In this section, we describe our disaggregation procedure for each income quantile separately, suppressing the income quantile notation. Define the quarterly time index as $t = 1, ..., T$. At each t, we observe from the CEX x_t, the total income over the last 4 quarters (QA CEX). Define ξ_t as the (latent) time$-t$ quarterly income (QQ CEX) so that:

$$x_t = \xi_t + \xi_{t-1} + \xi_{t-2} + \xi_{t-3} + u_t^x, \tag{1}$$

where $u_t^x \sim N\left(0, \omega^2\right)$ reflects some normally-distributed observation error. Notice that ξ_t contains information about $x_t, ..., x_{t+3}$ and vice versa. To obtain an estimate of ξ_t, we need to assume (i) a process for the underlying trend in income; (ii) a stationary propagation process

for quarterly income; and (iii) some initial values, ξ_0.[6]

We assume an AR(4) for the latent stationary component and a unit root trend, τ_t, so that:

$$\xi_t - \tau_t = \sum_{i=1}^{4} \phi_i \left(\xi_{t-i} - \tau_{t-i}\right) + u_t^{\xi} \tag{2}$$

and

$$\tau_t = \tau_{t-1} + u_t^{\tau}, \tag{3}$$

where u_t^{ξ} and u_t^{τ} are iid Normals. Based on (1), (2), and (3), we can then construct a state-space model and obtain the latent $\{\xi_t\}_{t=1}^{T}$, conditional on ξ_0.[7] We assume $\xi_0 \sim N\left(\frac{\Sigma_{t=1}^{4} x_t}{16}, \omega_{\xi}^2\right)$, where the mean $\left(\frac{\Sigma_{t=1}^{4} x_t}{16}\right)$ reflects the average quarterly income for the first year.

3.2 Interpolation

Given our estimates of QQ CEX, ξ_t, we can now interpolate QQ CPS, z_t. Let m represent the sampling multiplier—that is, the high-frequency variable is sampled m times for every low frequency sample (in our case, $m = 4$). The interpolation process assumes that the unobserved high-frequency variable, z_t, is related to the observed high-frequency variable, ξ_t, via the linear relationship:

$$z_t = \xi_t \beta + e_t, \tag{4}$$

where the innovation process is assumed to follow a stationary AR(1), $e_t = \rho e_{t-1} + v_t$, with $v_t \sim N\left(0, \omega_v^2\right)$ is a high-frequency innovation. Eq. (4) implies a similar low-frequency relationship:

$$Z_{\tau} = \Xi_{\tau} \beta + E_{\tau}, \tag{5}$$

[6] We perform the disaggregation with log income (i.e., $x_t = \log(CEX_t)$) to control the scale of income in the filter. To transform the latent ξ_t back into levels, we use the approximation, $QuarterlyIncome_t = \frac{\exp(4\xi_t)}{4}$, which requires the addition of the error term in eq. (1).

[7] Details on the state space are provided in the Appendix.

where Z_τ, Ξ_τ, and E_τ are the low-frequency analogs of z_t, ξ_t, and e_t and, importantly, Z_τ is observed. Note that the coefficient, β, is the same in both the high- and low-frequency regressions, implying a relationship between eqs. (4) and (5). The interpolation algorithm introduced in Chow and Lin (1971) estimates β by GLS, exploiting assumptions about the temporal aggregation of both variables and the process for the innovations. Additional details are provided in the Appendix.

3.3 Quantile Crossing

Often, temporal aggregation is a univariate problem—that is, each low-frequency variable is interpolated separately and independently. The wrinkle in our problem is that the quantiles of the income distribution cannot cross. For example, the (interpolated) high-frequency 60th quantile cannot take a value below the median, etc.

To avoid quantile crossing problems, we transform the income data so that, at each quantile, we draw the income value of that quantile relative to the income at the next quantile closest to the median. Instead of drawing $z_t = [z_{10t}, ...z_{40t}, z_{50t}, z_{60t}, ..., z_{99t}]'$ each independently, we first draw z_{50t}. Then, based on z_{50t}, we can obtain z_{40t} and z_{60t} by drawing $z_{50t} - z_{40t} > 0$ and $z_{60t} - z_{50t} > 0$. We proceed in steps away from the median until we have drawn all of the quantiles of interest.

4 The Interpolated Quarterly CPS Series

Figure 1 shows our quarterly CPS series (black line), interpolated as described above, for the first through ninth deciles, the 95th, and 99th percentiles. Figure 1 also depicts the income values for each quantile constructed from the raw CEX (blue line). Each series is scaled relative to its value at the beginning of the sample (1980Q1) to illustrate how it evolves over time. To facilitate comparison, we maintain the scale of (relative) income across the panels in each row.[8]

[8] The Appendix contains figures that show the income quantiles for our constructed quarterly CPS income quantiles and a comparison of the four quarter sum of our quarterly CPS and the annual CPS.

As discussed in Section A.2 of the Appendix, there were several procedural changes in the survey questioning and data reporting for the CEX between 2001 and 2004. First, Figure 1 treats the 2001Q1 observations of the CEX as missing to avoid distorting the figure with outlier values. Second, the procedural changes appear to have a more pronounced effect for incomes at lower quantiles, reflected by the apparent structural break in the series. We account for this in the applications by incorporating a break into the VAR when using CEX data.

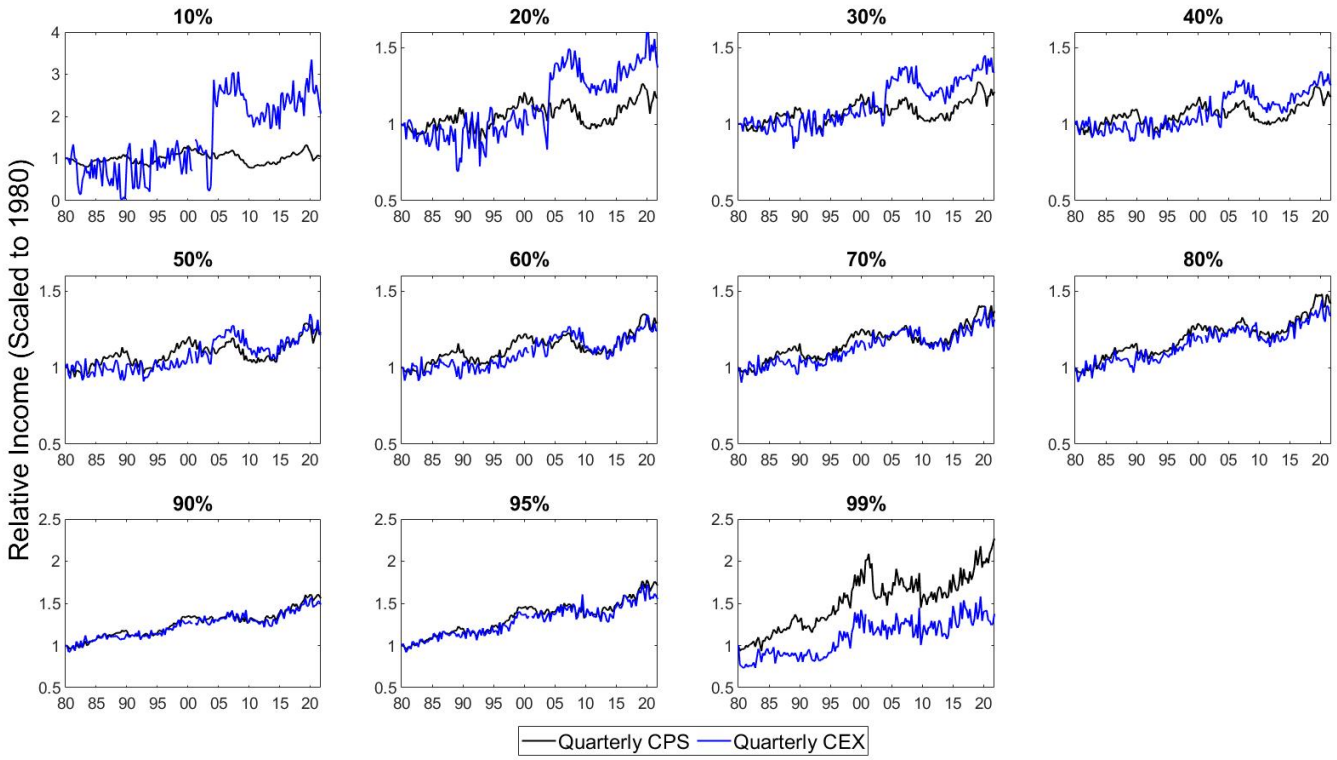

Figure 1: Quarterly CPS and Quarterly raw CEX. The Quarterly CPS series are generated via the Chow-Lin procedure outlined in Section 3.2. All series are scaled relative to their values at the beginning of the sample, 1980Q1, to illustrate the interpolation from low-frequency (annual) CPS to high-frequency (quarterly) CPS using high-frequency (quarterly) fluctuations in CEX. Note that the 2001Q1 observations of the raw CEX are treated as missing in the figure due to changes in the questionnaire. The scale on the y-axis is consistent across rows, except for the top left subplot which has a unique axis to account for the larger magnitude of the change at the 10% quantile.

To get a sense of how the landscape of inequality has evolved over time, we first compute several commonly studied measures using our new income levels. We look at the difference in the log levels of income at the following percentiles: 95th-50th, 99th-50th, 95th-90th, 99th-90th, and 90th-10th. We also compute the Gini index based on our income distribution. Figure 2 plots these series, scaled relative to the value at the beginning of our sample in 1980, along with real GDP. It is immediately apparent that these measures of inequality have risen considerably over time, especially so considering the divergence at the top end of the distribution. Aggregate measures of the economy, like GDP, show consistent growth over time but mask this disaggregate variation.

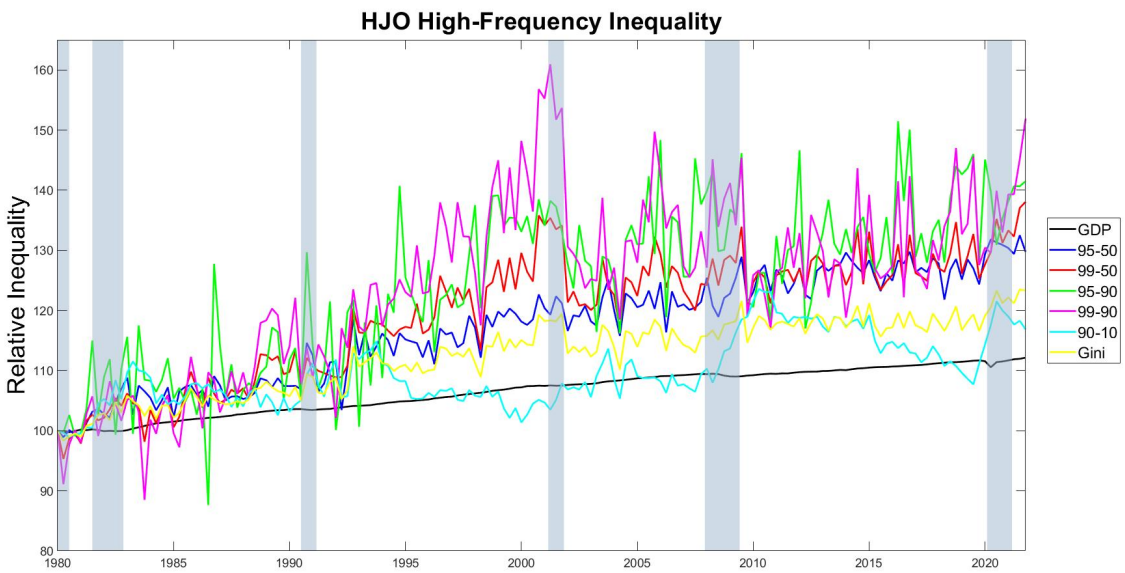

Figure 2: Relative inequality measures of quarterly interpolated CPS incomes. The inequality measures are computed the difference in the log level of incomes at each of the following percentiles: 95-50, 99-50, 95-90, 99-90, 90-10. We also compute the Gini index based on the income distibution suggested by our series. Log level of real GDP is shown for comparison. All series are scaled relative to the value in 1980.

5 Applications

To demonstrate the use of the quarterly inequality data, we explore two applications. First, we evaluate the effect of a technology shock on income inequality. Second, we explore the effect

of house price shocks on income inequality. In each case, our baseline model is a reduced-form VAR:

$$Y_t = \Phi(L) Y_{t-1} + \varepsilon_t, \tag{6}$$

where Y_t is defined below for each application and $\varepsilon_t \sim N(0, \Omega)$. The reduced-form VARs are estimated using Bayesian methods. The specific prior used for estimation and the method for identifying the structural shocks are discussed in the respective subsection.

5.1 Application 1: Technology Shocks

Our first application considers the effect of shocks to technological progress on income inequality. Several studies have considered the effects of technology (or productivity) shocks on macroeconomic variables such as output, hours [Gali (1999); Francis and Ramey (2005); and Christiano, Eichenbaum, and Viggfusson (2004); among many others]. Few others have investigated how inequality responds to aggregate productivity shocks in a VAR setting. For example, Faia and Shabalina (2023) investigate job reallocation that produces heterogeneous income effects in response to TFP shocks.

5.1.1 Data and Identification

Gali (1999) and the others mentioned above identify technology shocks by assuming that they account for all of the long-run variance in labor productivity. We identify technology shocks using the Max Share methodology of Francis, Owyang, Roush, and DiCecio (2014), which associates the shock with explaining the largest share of the forecast-error variance (FEV) of labor productivity at some long, finite horizon.[9] The full VAR is consistent with Dieppe, Francis, and Kindberg-Hanlon (2021), but adds a measure of income inequality. The VAR consists of log labor productivity, log total hours worked, log investment share of GDP, log consumption share of GDP, log difference of the PCE deflator, the 10-year Treasury bond

[9] This identification is the finite horizon analogue of the Gali assumption. Francis, Owyang, Roush, and DiCecio (2014) consider horizons h between 2.5 and 20 years. We identify the shock associated with the maximum share of the FEV of labor productivity at the 40-quarter horizon, ordering this series first in the VAR.

yield, and a measure of inequality. We estimate the VAR separately for each measure of inequality.

To identify the shock, we transform eq. (6) into the moving-average representation:

$$Y_t = \Theta\left(L\right)\Psi_0^{-1}\Psi_0\varepsilon_t$$

where we can identify the structural shocks as $\mu_t = \Psi_0\varepsilon_t$. From here, we can express the h-step-ahead forecast error for Y as the difference between observed data and the h-step-ahead conditional forecast:

$$Y_{t+h} - \hat{Y}_{t+h} = \sum_{\tau=0}^{h-1}\Theta_\tau\varepsilon_{t+h-\tau}.$$

We can then decompose the h-step-ahead FEV of variable i into the share attributable to shock j:

$$FEV_{ij}^h = \frac{e_i'\left[\sum_{\tau=0}^{h-1}\Theta_\tau\alpha\alpha\Theta_\tau'\right]e_i}{e_i'\left[\sum_{\tau=0}^{h-1}\Theta_\tau\Omega_\mu\Theta_\tau'\right]e_i}, \tag{7}$$

where $\alpha = De_j$, D is an orthonormal matrix, and e_i is an $n \times 1$ indicator vector that selects the ith column. The technology shock is identified as the shock that maximizes eq. (7).

5.1.2 Results

Figure 3 shows the responses of the macro aggregates to the technology shock. These results are consistent with previous work (for example, Francis and Ramey, 2005). Productivity permanently rises in response to the technology shock; this increase in productivity is partially offset by a corresponding decline in hours worked, albeit with some delay. As is relatively common in the literature, technology's effect on hours is either permanent or at least persistent across the response horizon.

The question of interest here is how the technology shock affects income inequality. The technology shock could raise inequality by increasing income proportionately across the board. On the other hand, the shock could produce a countervailing effect via its effect on labor markets. An increase in labor productivity disproportionately at lower incomes raises average

productivity but also compresses the productivity distribution and, thus, the income distribution. Figure 4 shows the effect of the technology shock on three measures of inequality: (i) the difference between the log of the 95th and 50th percentiles; (i) the difference between the log of the 99th and 50th percentiles, and (iii) the Gini index.[10] The first two inequality measures represent the difference between the top end of the income distribution and its center; the third inequality measure provides some insight over the whole income distribution. The left column of the figure shows income inequality measures based on our quarterly-interpolated CPS income. The middle column shows the income inequality measures based on the quarterly-disaggregated CEX (QQ CEX) income. The right column shows the income inequality measures based on raw quarterly-sampled annual CEX (QA CEX) income.

The left column of the figure shows the results using our data: Income inequality rises on impact across all parts of the distribution (bottom row) but, in particular, at the high end (top and middle rows). In the long run, the technology shock leads to a permanent positive increase in the growth rate of income inequality. The shock raises aggregate labor productivity, causing a permanent shift in the growth rate of income that is larger at the top end than at the bottom.

While the patterns of the responses are similar across the three datasets, the long-run effects of the technology shock on income inequality vary depending on how inequality is measured. For inequality measures generated by QQ CEX (middle column), technology's effect on income inequality is temporary. For inequality measures generated by QA CEX (right column), technology's effect on inequality is mixed: The effects on overall inequality and on the inequality with respect to the highest income levels in our sample are permanent but the effect on 95-50 is temporary. Thus, using different inequality data can lead to fundamentally different economic conclusions about how technology shocks affect income inequality. Scaling the income data to the annual survey, with a larger number of respondents, produces more consistent conclusions across measures of inequality at different parts of the distribution, thus providing us with greater confidence in our results from the left column.

[10] Given the interpolated CPS-CEX Series, we approximate a high-frequency Gini coefficient by integrating over discrete blocks centered at each quantile for which income is calculated.

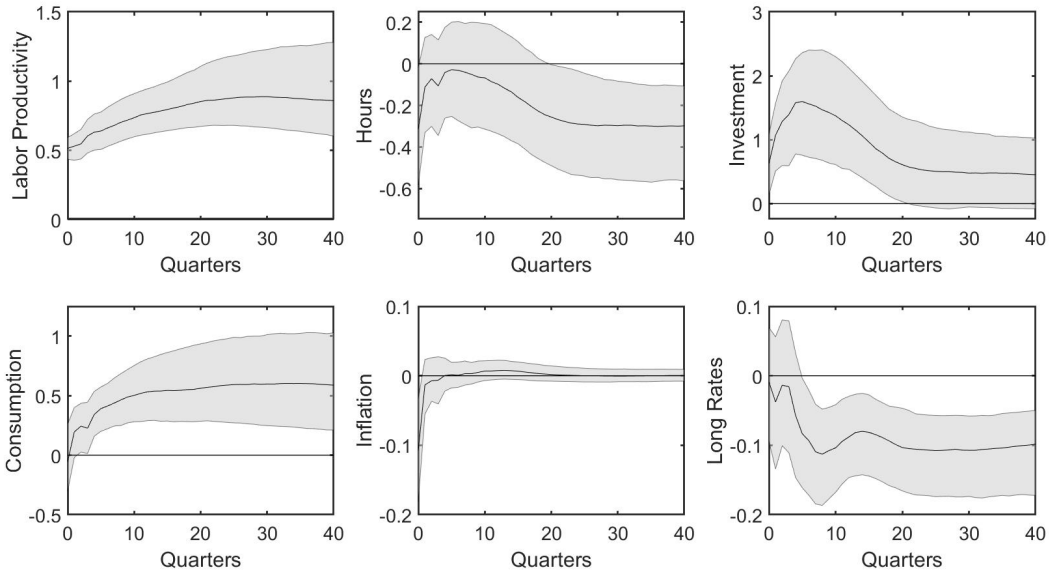

Figure 3: Impulse responses of macro variables to technology shock identified using the Max Share methodology of Francis, Owyang, Roush, and DiCecio (2014). The shock is associated with explaining the largest share of the forecast-error variance (FEV) of labor productivity at the 40-quarter horizon, ordering this series first in the VAR. The full VAR consists of log labor productivity, log total hours worked, log investment share of GDP, log consumption share of GDP, log difference of the PCE deflator, the 10-year Treasury bond yield, and a measure of inequality.

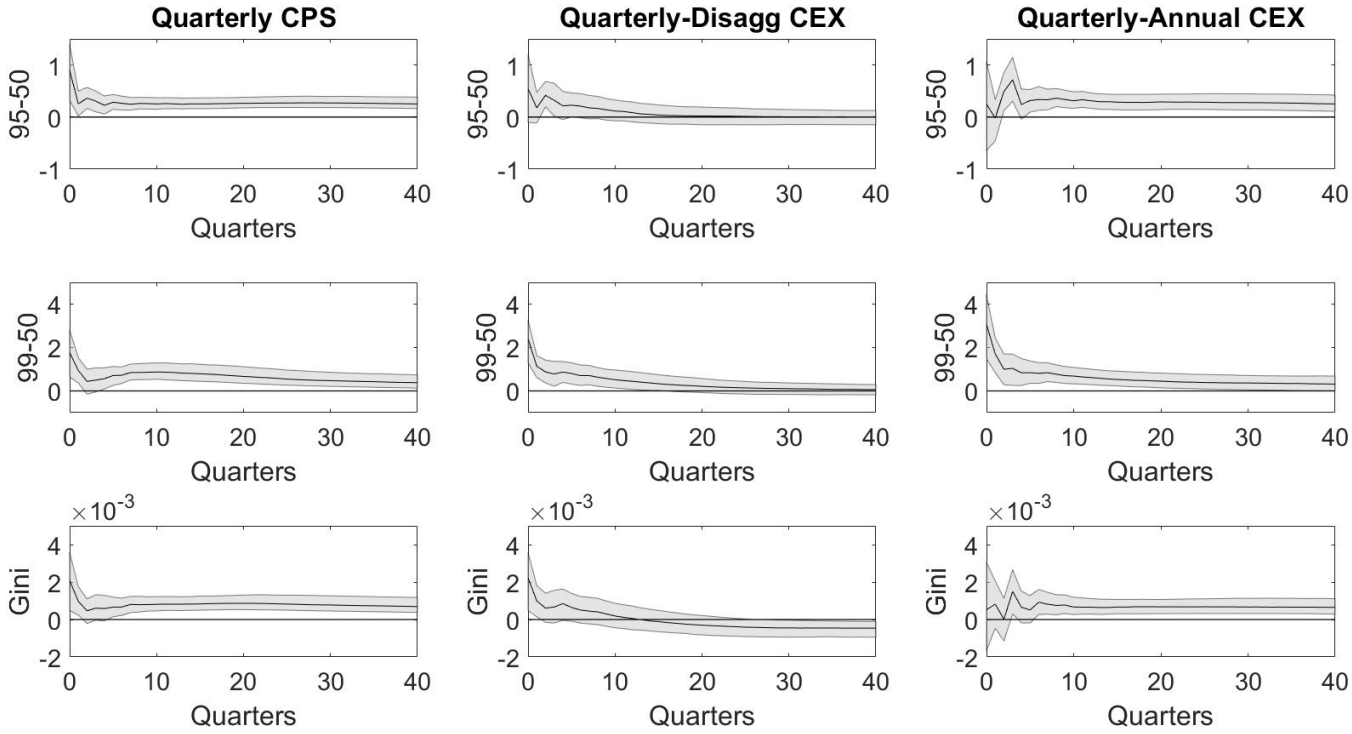

Figure 4: Impulse responses of inequality measures to to technology shock identified using the Max Share methodology of Francis, Owyang, Roush, and DiCecio (2014). The shock is associated with explaining the largest share of the forecast-error variance (FEV) of labor productivity at the 40-quarter horizon, ordering this series first in the VAR. The full VAR consists of log labor productivity, log total hours worked, log investment share of GDP, log consumption share of GDP, log difference of the PCE deflator, the 10-year Treasury bond yield, and a measure of inequality. In the first column, the inequality measures come from our quarterly-interpolated CPS income levels. In the second column, the inequality measures come from the quarterly-disaggregated CEX. In the third column, the inequality measures come from the raw quarterly-annual CEX.

5.2 Application 2: House Price Shocks

Our second application explores the channel through which house price shocks propagate through the economy, potentially affecting the distribution of income. Housing has been shown to be a special commodity that acts both as an asset and as a consumption good. Previous VAR studies of housing have studied various roles for housing in the economy including its important role in the propagation of monetary policy [e.g., Iacoviello (2005) and others]. Generally, the literature has found a positive relationship between housing prices and other macroeconomic variables such as income [Haurin and Rosenthal 2007; and Miller, Peng, and Sklarz 2011] and consumption [Case, Shiller, and Quigley 2005].

While the shelter consumption component of housing is relevant for all agents, housing serves as an asset only for those who own. Since homeownership rates vary by income, this asymmetry may then affect different parts of the income distribution disproportionately. At the top end, the rise in house prices appreciates the asset, while the passthrough of house price shocks to rental prices may affect the bottom end in a (potentially) opposite manner.

5.2.1 Data and Identification

Similarly to Iacoviello (2005), we identify the structural shocks from a Cholesky decomposition of the reduced-form VAR, eq. (6), with Y_t ordered as follows: output, house prices, price level, monetary policy rate, and inequality.[11] Thus, inequality is the most endogenous variable, sensitive to fluctuations in the real economy, nominal price pressures, and monetary policy. As is standard in the literature, monetary policy responds to aggregate economic conditions (including house prices) contemporaneously but not to inequality. It might be relevant to note that the house price shocks that we will investigate here are orthogonal to changes in house prices induced by policy or growth. We think of these shocks as exogenous housing preference shocks, since housing supply is assumed to evolve slowly.

In order to estimate the VAR, we include quarterly data on log real GDP, log of the Case-Shiller U.S. National Home Price Index (monthly, averaged over the quarter), log of the

[11] The VAR(2) includes a trend. We estimate the model using a Minnesota prior.

PCE deflator, the 1-year Treasury yield, and a measure of inequality.[12] We estimate the VAR separately for each measure of inequality. The sample period is consistent with that of our first application, 1980:Q1-2021:Q4, to match with the availability of the inequality data.

5.2.2 Results

Figure 5 shows the results of the house price shock on four of the aggregate variables. These responses are in line with Iacoviello's (2005) housing demand shock: output and prices rise in the short term and only begin to steady after about 6 quarters as the Fed responds by increasing interest rates. House prices, in particular, rise on impact and continue to rise over the response horizon (albeit at a slower rate), despite intervention by the Fed.

Figure 6 shows the effect of the house price shock on three measures of inequality: (i) the difference between the log of the 95th and 50th percentiles; (i) the difference between the log of the 99th and 50th percentiles, and (iii) the Gini index. The first two inequality measures represent the difference between the top end of the income distribution and its center; the third inequality measure provides some insight over the whole income distribution. The left column of the figure shows income inequality measures based on our quarterly-interpolated CPS (QQ CPS) income. The middle column shows the income inequality measures based on the quarterly-disaggregated CEX (QQ CEX) income. The right column shows the income inequality measures based on raw quarterly-sampled annual CEX (QA CEX) income.

The figure shows that, on impact, income inequality rises across all parts of the distribution (bottom row) but, in particular, at the high end of the income distribution (top and middle rows). The rise in income inequality is transitory. Generally, the pattern of the responses are similar across the three datasets; however, the exercise highlights important differences that may be the result of differences in our scaling to the CPS. For inequality measures generated by the CEX (middle and right columns), inequality falls at longer response horizons. Moreover, the response over the entire horizon using CEX data suggests that a positive house price shock

[12] We include the 1-year rate as the dates span several years in which conventional monetary policy was constrained at the effective lower bound. There are several strategies in the literature to account for this, and we choose the 1-year rate as it incorporates effects of changes in the short-term policy instrument as well as market expectations of near-term monetary policy actions.

yields a net *decline* in income inequality. This result is not apparent in our quarterly CPS data, where we conclude that house price shocks significantly increase income inequality, a result that is consistent with the trickle-up story in models such as Jackson, Otrok, Owyang (2022) and Amberg, Jansson, Klein, and Rogantini Picco (2022). In those models, an expansionary shock (tax progressivity in the former and easing monetary policy in the latter) leads to a disproportionate increase in income at the higher end of the distribution.

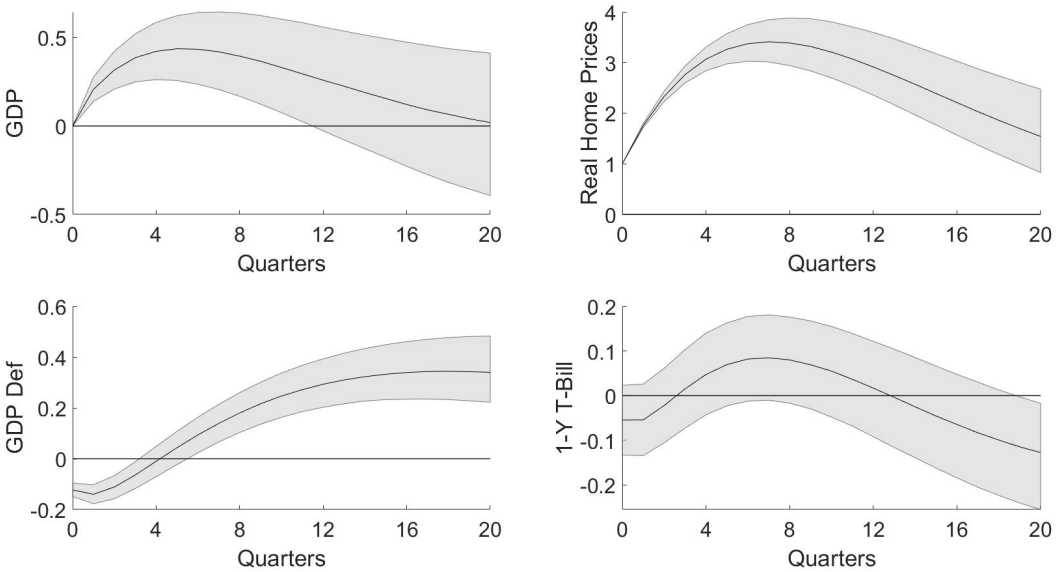

Figure 5: Impulse responses of macro variables to real home price shock identified from a Cholesky decomposition of the reduced-form VAR(2). The ordering of data in the quarterly VAR is as follows: log real GDP, log of the Case-Shiller U.S. National Home Price Index (monthly, averaged over the quarter), log of the PCE deflator, the 1-year Treasury yield, and a measure of inequality. We estimate the model using a Minnesota prior and include a trend.

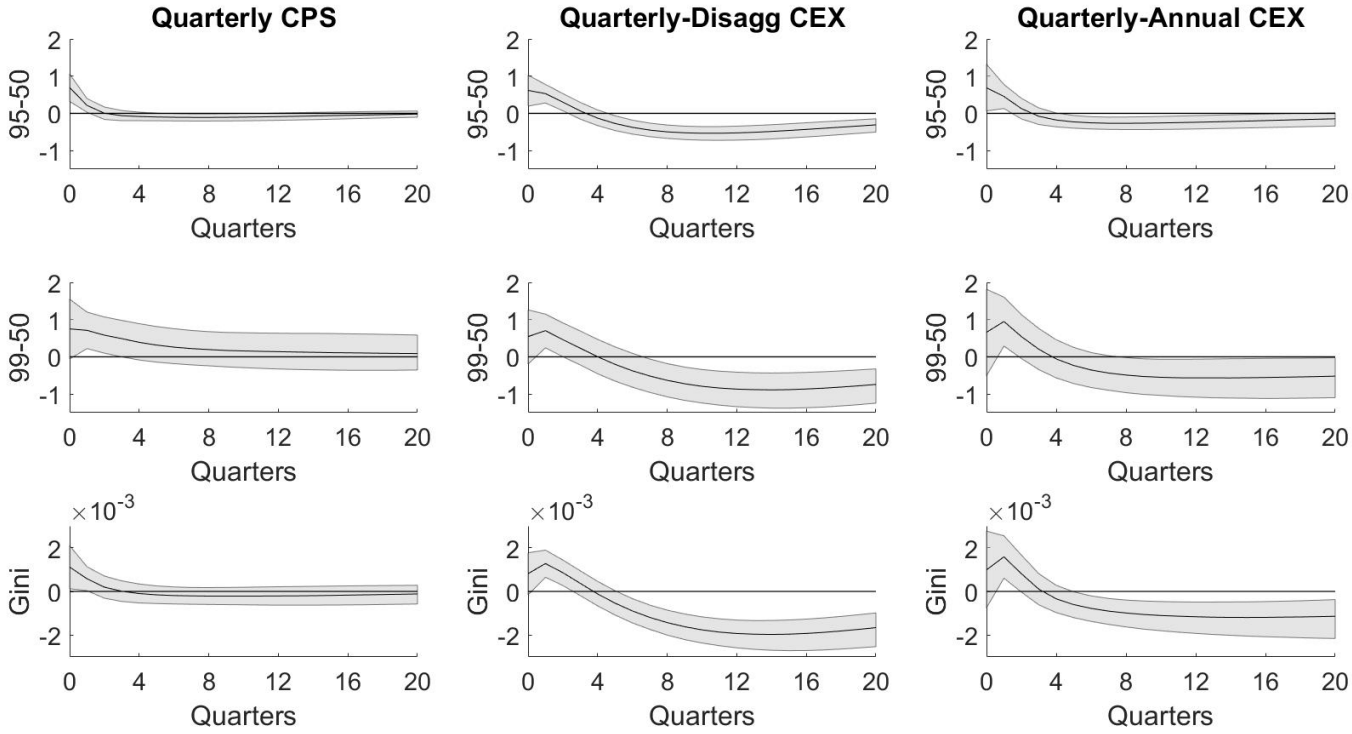

Figure 6: Impulse responses of inequality measures to real home price shock identified from a Cholesky decomposition of the reduced-form VAR(2). The ordering of data in the quarterly VAR is as follows: log real GDP, log of the Case-Shiller U.S. National Home Price Index (monthly, averaged over the quarter), log of the PCE deflator, the 1-year Treasury yield, and a measure of inequality. We estimate the model using a Minnesota prior and include a trend. In the first column, the inequality measures come from our quarterly-interpolated CPS income levels. In the second column, the inequality measures come from the quarterly-disaggregated CEX. In the third column, the inequality measures come from the raw quarterly-annual CEX.

6 Conclusion

We construct quarterly income quantiles that are tied to quarterly fluctuations from the CEX but are scaled to a temporal disaggregation of the annual CPS. Because the ASEC from the CPS surveys both more households and has a higher response rate, we feel that it more accurately represents the actual levels of the income quantiles. However, because higher frequency fluctuations are often necessary to identify VARs, we use the raw CEX quantiles to capture shorter-run fluctuations.

We compare our quarterly CPS inequality measures to the raw quarterly CEX measures in two applications. The applications use VARs to compute the impulse responses of income inequality to technology shocks and to house price shocks. Using our quarterly CPS inequality measures, technology shocks permanently raise income inequality, while house price shocks raise inequality temporarily but are neutral in the medium run. On the other hand, using the raw CEX measures leads to different economic conclusions about either the permanence of technology's effect or the sign of housing's effect on income inequality.

References

[1] Amberg, Niklas, Thomas Jansson, Mathias Klein, and Anna Rogantini Picco. 2022. Five Facts about the Distributional Income Effects of Monetary Policy Shocks, *American Economic Review: Insights*, 4(3). pp. 289–304.

[2] Blanchet, Thomas, Saez, Emmanuel, and Zucman, Gabriel, 2023. Real-Time Inequality, *NBER Working Paper Series*, 30229.

[3] Case, Karl, Quigley, John and Shiller, Robert, 2005. Comparing Wealth Effects: The Stock Market versus the Housing Market, *The B.E. Journal of Macroeconomics*, 5(1), pp. 1-34.

[4] Chow, G. C., and Lin, A., 1971. Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. *The Review of Economics and Statistics*, 53(4), pp. 372–375. DOI: https://doi.org/10.2307/1928739.

[5] Christiano, L. J., Eichenbaum, M., and Vigfusbson, R., 2004. The Response of Hours to a Technology Shock: Evidence Based on Direct Measures of Technology, *Journal of the European Economic Association*, 2(2/3), pp. 381–395.

[6] Coibion, Olivier, Gorodnichenko, Yuriy, Kueng, Lorenz, and Silvia, John, 2017. Innocent Bystanders? Monetary policy and inequality, *Journal of Monetary Economics*, 88, pp. 70-89.

[7] Dieppe, Alistair, Francis, Neville, and Kindberg-Hanlon, Gene, 2021. Technological and non-technological drivers of productivity dynamics in developed and emerging market economies, *Journal of Economic Dynamics and Control*, 131, 104216, DOI: https://doi.org/10.1016/j.jedc.2021.104216.

[8] Faia, E and Shabalina, E., 2023. Cyclical Move to Opportunities, *CEPR Discussion Paper,* 18546. CEPR Press, Paris & London. https://cepr.org/publications/dp18546.

[9] Francis, Neville and Ramey, Valerie A., 2005. Is the technology-driven real business cycle hypothesis dead? Shocks and aggregate fluctuations revisited, *Journal of Monetary Economics*,52(8), pp. 1379-1399, ISSN 0304-3932, DOI: https://doi.org/10.1016/j.jmoneco.2004.08.009.

[10] Francis, N., Owyang, M. T., Roush, J. E., and DiCecio, R., 2014. A Flexible Finite-Horizon Alternative to Long-Run Restrictions with an Application to Technology Shocks, *The Review of Economics and Statistics*, 96(4), pp. 638–647. http://www.jstor.org/stable/43554945.

[11] Galí, Jordi, 1999. Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?, *American Economic Review*, 89 (1), pp. 249–271, DOI: 10.1257/aer.89.1.249

[12] Haurin, Donald R. and Herbert, Christopher E. and Rosenthal, Stuart S.,2007. Homeownership Gaps Among Low-Income and Minority Households. *Cityscape*, 9(2).

[13] Iacoviello, Matteo, 2005. House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle. *American Economic Review*, 95(3), pp. 739-764.

[14] Jackson, Laura E., Otrok, Christopher, and Owyang, Michael, 2022. Tax Progressivity, Economic Booms, and Trickle-Up Economics, *Federal Reserve Bank of St. Louis Working Paper,* 2019-034. DOI: https://doi.org/10.20955/wp.2019.034.

[15] Miller, Norman G. and Peng, Liang and Sklarz, Michael, 2011. House Prices and Economic Growth. *Journal of Real Estate Finance and Economics*, 42(4), 2011.

A Appendix

A.1 CPS Data

We download annual Current Population Survey March Supplement microdata 1980 to present as a DAT file off of the IPUMS website, including variables related to income, geography, tax filing status, and demographic characteristics. We import the DAT file into STATA using IPUMS generated code and use STATA for all subsequent ASEC data analysis.

We define income as in the Consumer Expenditure Survey, including income from wages, rents, and capital but excluding government and nonwork transfer payments. Following the calculation of factor income of Blanchet, Saez, Zucman (2022), we include both old and redesign sub-samples in the 2014 ASEC, reweighting by factors of 3/8 and 5/8 respectively to correct for the oversample.

Exact income variables vary over by sample year due to updates in surveying methodology. We implement CPS provided rank proximity swap values for all top-coded income variables, which helps to address but does not resolve income top coding.

We sum income variables at the ASEC defined household level and scale according to the OECD equivalency scale and convert to real dollars. For each year, we calculate tenth through ninth deciles, 95th, and 99th percentile incomes at the national level.

A.2 CEX Data

We use Consumer Expenditure Interview Survey Microdata for each quarter-year available 1984-2022. Until 1990, microdata is only available in ASCII files from the ICSPR, which we convert to Stata using year specific data dictionaries. For all other years, we download microdata as CSV files off from the BLS website directly. Data analysis occurs in R, where it is broken down by year. We use the Tidyverse libraries for most analysis.

To construct quarter-year income quantiles we use two main files: FMLI, which reports household level summary variables for demographics, income, and spending; and MEMI, which reports more detailed household member level income variables.

We merge the FMLI, MEMI, and MTBI files at the household level, summing within household to construct our measure of pre-transfer income. We follow Coibion, Gorodnichenko, Kueng, and Silvia (2017) and scale household income by household size using the OECD equivalency scale. Over all samples, we consider the following sources under our definition of income: dividends, estates, trusts, royalties, and interest income and losses; rental income; workers' comp and veteran's payments; wages and salaries; self-employment, nonfarm and farm business income; and social security payments. Because the Interview Survey's income data structure has changed multiple times since the survey began, the exact income variables covered by MEMI and FMLI files varies by quarter-year. We use a variety of variables over the years to keep our definition of income consistent.

A major limitation of Consumer Expenditure Microdata is its high rate of income non-response and incomplete response. Prior to 2001, respondents could either report their actual income for each income category or decline to respond. Starting in 2001, respondents could report their actual income, report a bracketed income range, or decline to respond. Starting in mid-2004, the BLS began imputing income for non-reporters.

We use BLS-imputed income in all quarters after it becomes available. Between 2001 and 2004, we take actual income whenever available; when it is not, we substitute the median

of the bracketed range whenever that value is available. If neither is reported, we treat the household as a non-reporter for that income category. Prior to 2001, we take actual income whenever available, and treat the household as a non-reporter whenever that income category is missing for at least one household member.

Following Coibion, Gorodnichenko, Kueng, and Silvia (2017), we impute income by category at the household level for all years prior to 2004, trying to replicate BLS income imputation as closely as possible. We regress reported income on householder education level, age (quadratic), race, number of weeks worked per year, family size, and family age distribution using complete reporters within that year. We use this regression to predict income for non-reporting households, then add random shocks drawn with replacement from the regression's residuals. Finally, we trim our imputed values to fit within the maximum and minimum values reported and add the imputed values back into the full dataset.

Because the BLS only asks households about their income during the first and fourth of four interviews, we drop all households' second and third interviews from subsequent income samples to avoid reusing outdated information. However, we keep households from all interviews in samples when calculating contemporaneous consumption quantiles.

For each quarter-year, we calculate tenth through ninth deciles, 95th, and 99th percentiles of pre-transfer income at the national level. We calculate the same statistic within each month-year at the national level. Finally, we use quarterly national CPI data from the BLS to covert incomes from nominal to real.

Although the Consumer Expenditure Survey is quarterly, it asks respondents to describe their income over the previous year, which means that the income they report is actually an annual value corresponding to the 4-quarter period immediately prior to the interview date. We convert annual income to quarterly values using the temporal disaggregation procedure described in Section 3.1. Details of the state space are provided in the next section.

A.3 The disaggregation state space

Define the quarterly time index as $t = 1, ..., T$. At each t, we observe from the CEX x_t, the total income over the last 4 quarters (QA CEX). Define ξ_t as the (latent) time$-t$ quarterly income (QQ CEX) so that:

$$x_t = \xi_t + \xi_{t-1} + \xi_{t-2} + \xi_{t-3} + u_t^x, \tag{8}$$

where $u_t^x \sim N\left(0, \omega^2\right)$ reflects some normally-distributed observation error. To obtain an estimate of ξ_t, we assume (i) a process for the underlying trend in income; (ii) a stationary propagation process for quarterly income; and (iii) some initial values, ξ_0.

We assume an AR(4) for the latent stationary component and a unit root trend, τ_t, so that:

$$\xi_t - \tau_t = \sum_{i=1}^{4} \phi_i \left(\xi_{t-i} - \tau_{t-i}\right) + u_t^{\xi} \tag{9}$$

and

$$\tau_t = \tau_{t-1} + u_t^{\tau}. \tag{10}$$

We obtain the latent $\{\xi_t\}_{t=1}^T$, conditional on ξ_0. To construct the state space, we define $\zeta_t = \left[\xi_t, \xi_{t-1}, \xi_{t-2}, \xi_{t-3}\right]'$ and $H = [1, 1, 1, 1]$. Thus,

$$\zeta_t = \begin{bmatrix} \xi_t \\ \tau_t \\ \xi_{t-1} \\ \tau_{t-1} \\ \xi_{t-2} \\ \tau_{t-2} \\ \xi_{t-3} \\ \tau_{t-3} \end{bmatrix} = \begin{bmatrix} \phi_1 & -\phi_1 & \phi_2 & -\phi_2 & \phi_3 & -\phi_3 & \phi_4 & -\phi_4 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_{t-1} \\ \tau_{t-1} \\ \xi_{t-2} \\ \tau_{t-2} \\ \xi_{t-3} \\ \tau_{t-3} \\ \xi_{t-4} \\ \tau_{t-4} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ e_t \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

We assume $\xi_0 \sim N\left(\frac{\Sigma_{t=1}^4 x_t}{16}, \omega_\xi^2\right)$, where the mean $\left(\frac{\Sigma_{t=1}^4 x_t}{16}\right)$ reflects the average quarterly income for the first year.

A.4 Chow-Lin Interpolation

To fix ideas, our objective is to obtain an estimate of the high-frequency variable, z_t, which is observed only at a lower frequency, denoted Z_τ. Let m represent the sampling multiplier—that is, the high frequency is sampled m times for every low frequency sample (in our case, $m = 4$). Implementation of Chow-Lin assumes that Y_t is related to another (observed) high-frequency variable, X_t, via the linear regression:

$$z_t = \xi_t \beta + e_t, \tag{11}$$

where the innovation process is assumed to follow a stationary AR(1), $e_t = \rho e_{t-1} + v_t$, with $v_t \sim N\left(0, \omega_v^2\right)$ is a high frequency innovation.

Eq. (4) implies a similar low-frequency relationship:

$$Z_\tau = \Xi_\tau \beta + E_\tau, \tag{12}$$

where Z_τ, Ξ_τ, and $E_\tau$are the low frequency analogs of z_t, ξ_t, and e_t and, importantly, Z_τ is observed. Note that the coefficient, β, is the same in both the high- and low-frequency regressions, implying a relationship between eqs. (11) and (12).

Define the aggregator matrix

$$C_m = \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m & \cdots & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{1}_m & & \mathbf{0}_m \\ \vdots & & \ddots & \vdots \\ \mathbf{0}_m & \mathbf{0}_m & \cdots & \mathbf{1}_m \end{bmatrix},$$

where $\mathbf{1}_m$ and $\mathbf{0}_m$ are $(m \times 1)$ vectors of ones and zeroes, respectively. This aggregator identifies the relationship between the high- and low-frequency variables: $Z_\tau = C_m z_t$, $\Xi_t = C_m \xi_t$, and $V_t = C_m v_t$.

$$\beta = \left(\Xi_t' \Sigma_l^{-1} \Xi_t\right)^{-1} \Xi_t' \Sigma_l^{-1} Z_t, \tag{13}$$

where

$$\Sigma_l = C_m \Sigma_h C_m',$$

Σ_l is the autocovariance of the low frequency innovations and Σ_h is the autocovariance of the high frequency innovations: $\Sigma_h = A\frac{\omega^2}{1-\rho^2}$ and

$$A = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{mT-1} \\ \rho & 1 & \rho & \cdots & \rho^{mT-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{mT-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{mT-1} & \rho^{mT-2} & \rho^{mT-3} & \cdots & 1 \end{bmatrix}, \tag{14}$$

by assumption.

To execute Chow-Lin, we obtain an estimate of $\rho = \sqrt[12]{\lambda}$, where λ is the AR(1) coefficient from a regression of the annual residuals on their lags. Because the residuals are not available to initialize the algorithm, we start with some initial value for ρ. Conditional on ρ, we obtain A (and, equivalently Σ_l) and an estimate of β from eq. (13).[13] Next, conditional on β, we can obtain the *low-frequency* residuals

$$\widehat{V} = Z - \beta\Xi.$$

Then, we obtain the high-frequency interpolation as a combination of the projection onto ξ and rescaled high frequency transformations of the low frequency residuals:

$$z = \xi\beta + AC'\left(CAC'\right)^{-1}\widehat{v},$$

where A is defined in eq. (14) above. We also run the AR(1) regression of $\widehat{v}_t$ on its lag to obtain an estimate of ρ. We compare this new estimate of ρ to the previous estimate. If $\rho' \approx \rho$, we stop; otherwise, we iterate.

A.5 Comparing Quarterly to Annual CPS

Figure A1 shows the income quantiles for our constructed quarterly CPS. Figure A2 plots a comparison of the four quarter sum of our quarterly CPS and the annual CPS.

[13] Note that the estimate of β does not depend on an estimate of ω.

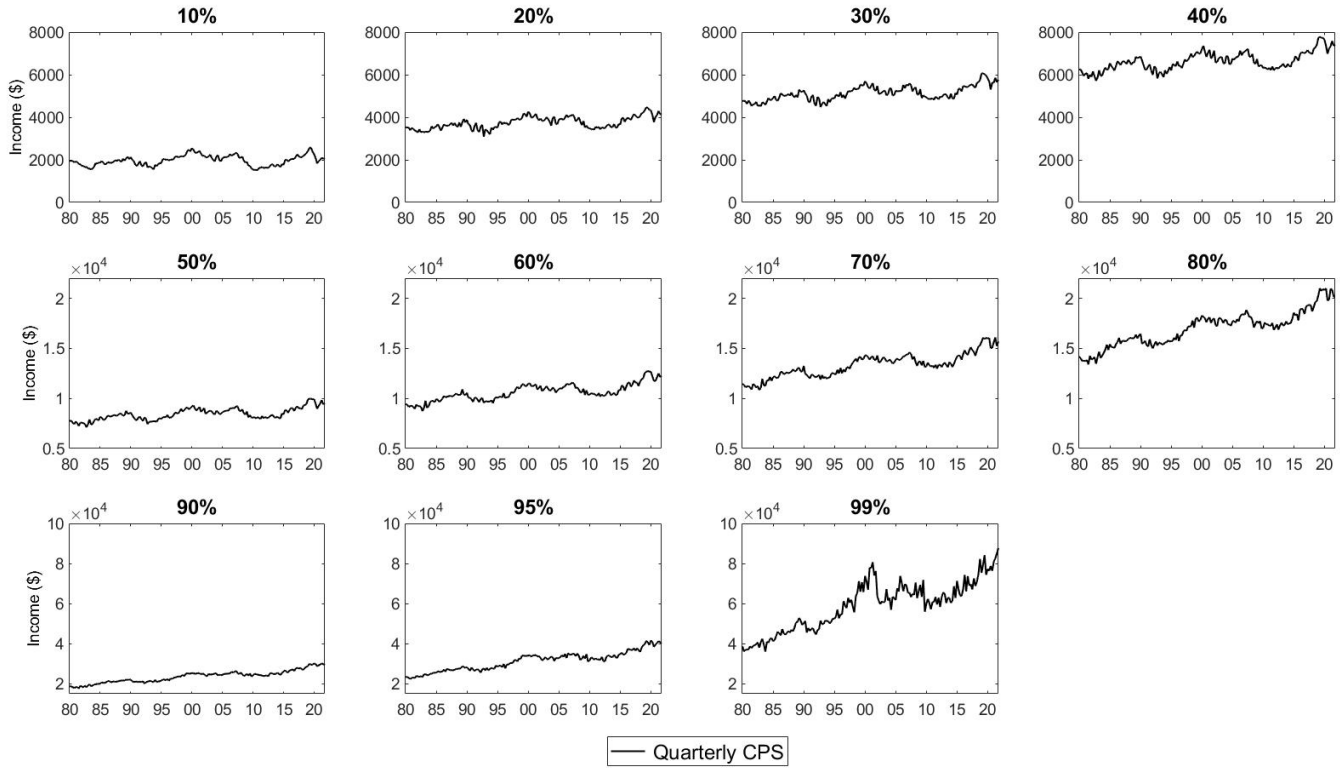

Figure 7: Quarterly CPS series are generated via the Chow-Lin procedure outlined in Section 3.2. The scale on the y-axis is consistent across rows.

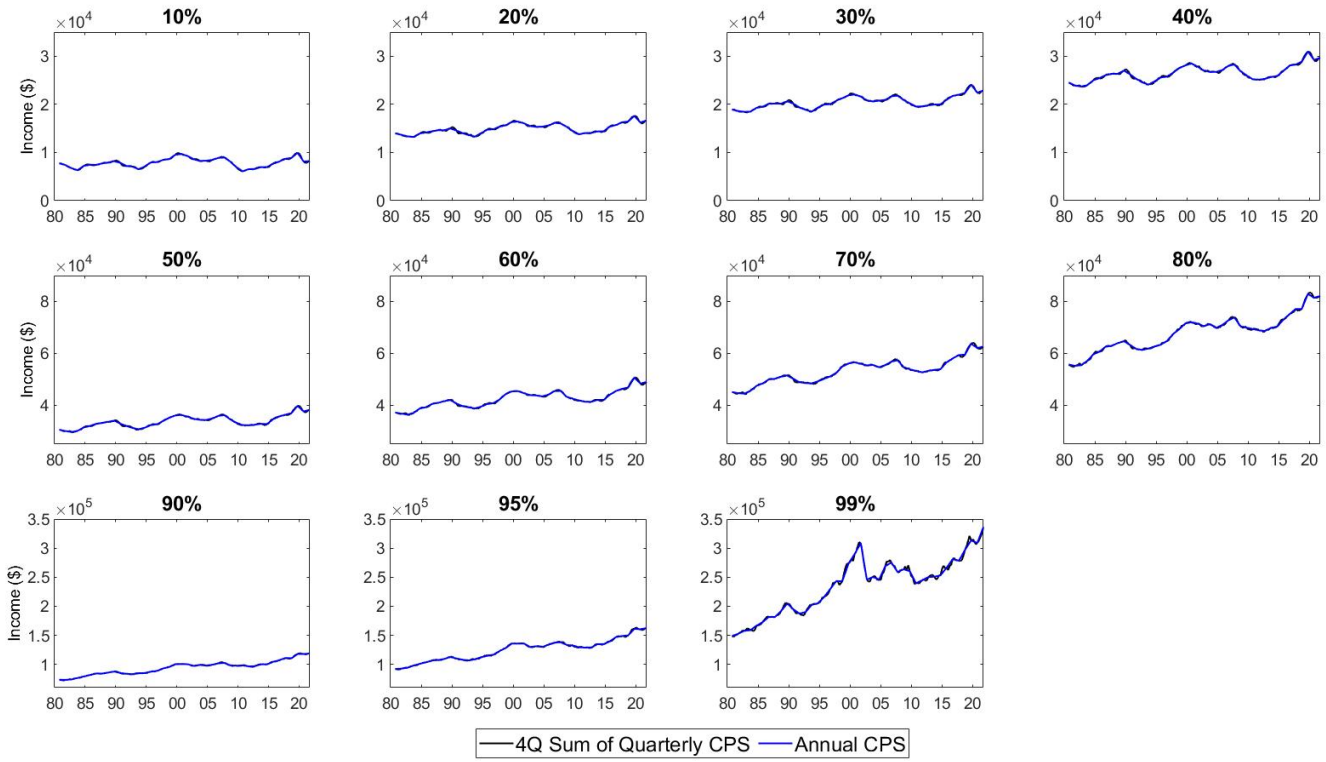

Figure 8: 4Q sum of Quarterly CPS and Annual CPS. Quarterly CPS series are generated via the Chow-Lin procedure outlined inSection 3.2. The quarterly CPS series represent quarterly income thus are transformed and plotted as the trailing 4-quarter sum in order to be comparable in scale to the annual CPS data. The scale on the y-axis is consistent across rows.

A.6 Comparison to BSZ

For comparison, the right panel of Figure A3 displays three relative inequality series based on the measures of total factor income shared in the Realtime Inequality dataset of Blanchet, Saez, and Zucman (2023) (henceforth BSZ). Note that the BSZ series are not directly comparable to ours. For instance, we produce the income level at the 99th percentile while BSZ produce the income of the top 1%. Similarly, we produce an income level at the 50th percentile while they produce the income of the bottom 50%. To compare measures as similar as possible, we include the BSZ measures of "per-unit" income of the top 1%, the top 10%, and the bottom 50%. These amount to per-household incomes, on average, in each of the top 1%, top 10%, and bottom 50%. Similarly to our measures, we compute the difference in the log level of each of these average incomes. Finally, we also borrow the BSZ measure of the top 1% share of total factor income.

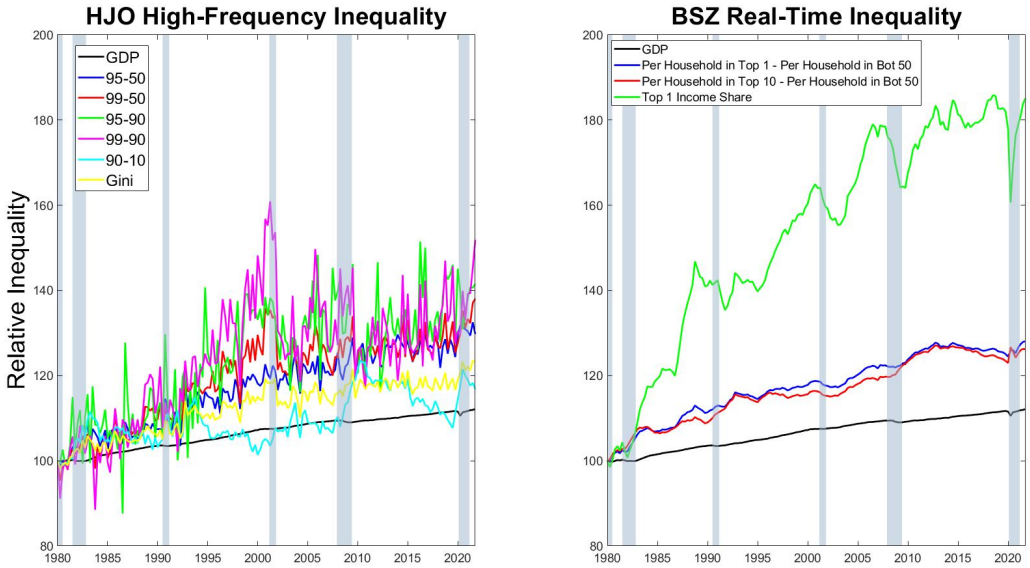

Figure 9: Relative inequality measures of quarterly interpolated CPS incomes and Blanchet, Saez, Bucman (2023) Realtime Inequality incomes. For the quarterly CPS, the inequality measures are computed as the difference in the log level of incomes at each of the following percentiles: 95-50, 99-50, 95-90, 99-90, 90-10. We also compute the Gini index based on the income distibution suggested by our series. For BSZ, the inequality measures are computed as the difference in the log level of income per-unit, on average in each portion of the distribution (i.e., the average income per household in the top 1%, in the top 10%, or in the bottom 50%). The BSZ series also includes income at the top 1% share of total factor income. Log level of real GDP is shown for comparison. All series are scaled relative to the value in 1980.