Universitat Autònoma de Barcelona

ESSAYS IN APPLIED MACROECONOMICS

NICOLÒ MAFFEI FACCIOLI

Essays in Applied Macroeconomics

Nicolò Maffei Faccioli

September, 2020

Thesis supervisor: Luca Gambetti

ACKNOWLEDGEMENTS

I would like to thank Luca Gambetti for his valuable help and supervision on this project, and my family and friends that supported me over this period. A special thanks goes also to Drago Bergholt and Francesco Furlanetto. I gratefully acknowledge financial support from the "La Caixa-Severo Ochoa International Doctoral Fellowship".

TABLE OF CONTENTS

CHAPTER I

Monetary Policy Shocks and Inequality: Evidence From the US*

Nicolò Maffei Faccioli†

August, 2020

Abstract: This paper sheds new light on the effects of both conventional and unconventional monetary policy shocks on income and wealth inequality in the US. Measures of inequality and the distribution are constructed using income and wealth data from the CEX survey. Measurement issues are tackled by decomposing the measures of interest into two orthogonal components: one common to the macroeconomy and one idiosyncratic, which reflects measurement error and other series-specific factors unrelated to the economy. I find that a rise in the federal funds rate leads to a short-run increase in total, labor and financial income inequality. Financial wealth inequality, instead, rises more persistently. The effects appear sizeable, highly heterogeneous across different percentiles of the distribution, and characterized by a larger negative response of its lower tail. An increase in asset purchases leads to a fall in financial wealth, total and labor income inequality, while financial income inequality increases. These effects appear less heterogeneous and seem to benefit mostly the bottom of the distribution. The findings suggest that the transmission mechanisms of conventional and unconventional actions are potentially different.

JEL Classification: *E52, D31, E2, E3, E4.*
Keywords: *Monetary Policy, Inequality, SVAR, Local Projections.*

*I would like to thank Luca Gambetti for his valuable help and supervision on this project. This paper constitutes the first chapter of my PhD dissertation and has benefited from discussions with Knut Are Aastveit, Drago Bergholt, Ambrogio Cesa Bianchi, Luca Dedola, Francesco Furlanetto, Peter Karadi, Keith Kuester, Saskia Ter Ellen, Giulia Rivolta, Juan Rubio-Ramírez, all the professors of the Macro Club of Universitat Autònoma de Barcelona, and participants at various conferences and research seminars. I gratefully acknowledge financial support from the "La Caixa-Severo Ochoa International Doctoral Fellowship".
†Universitat Autònoma de Barcelona and Barcelona GSE.

1 INTRODUCTION

The well documented growth in income and wealth inequality in the US over the past four decades has been at the center of economic and political debates. Figure A.1 shows the evolution of different measures of inequality for various components of income and wealth over the last forty years.[1] While implying somewhat different trends, these series all display a strong and persistently increasing pattern.

Besides social concerns, continuously growing inequality has multiple turndowns for the economy: it weakens aggregate demand and growth, and is related to more recurrent and harsh boom-and-boost cycles (Stiglitz (2012)). Rising inequality has in fact been identified as one of the fault lines of the last financial crisis (Rajan (2010)), and a harmful phenomenon for economic stability (Kumhof, Rancière, and Winant (2015); Cairo and Sim (2018)). In the aftermath of the Great Recession, inequality has reached the attention of monetary policymakers, particularly after the raised concerns that monetary policy actions might have favored its growth, making rich richer and poor poorer. Several questions spontaneously arise: is tightening monetary policy increasing or decreasing inequality? Are these effects mainly short-run, or more persistent? Is there large heterogeneity in the response to monetary policy shocks? Do conventional and unconventional policies affect inequality and the distribution similarly? Are these effects considerable? The aim of this paper, in short, is to reply to such questions. First, I construct measures of inequality and the distribution which are purged from measurement errors. Then, I use two state-of-the-art identification strategies to measure the effects of conventional and unconventional monetary policies on inequality and the distribution.

This is not the first paper that evaluates the effects of monetary policy on inequality. A first step in this direction is the pioneer work of Coibion, Gorodnichenko, Kueng, and Silvia (2017), who assess empirically the effects of conventional monetary policy shocks on income and consumption inequality. The authors find that an unexpected increase in the short-term rate leads to medium-run increases in income and consumption inequality. Other studies point towards similar findings. Furceri, Loungani, and Zdzienicka (2018), using a panel of 32 advanced and emerging market economies, find an average increase in income inequality after a conventional tightening. Mumtaz and Theophilopoulou (2017) show that both income and consumption inequality increase in response to conventional monetary policy shocks in the UK. A smaller share of the literature attempted to assess the effects of unconventional monetary policies on the distribution of income and wealth. Montecino and Epstein (2015) find that the effects of QE were most likely dis-equalizing using income quintiles data from the SCF survey in the US. Lenza and Slacalek (2018) find that quantitative easing decreased income inequality and had no significant effect on wealth inequality in the Euro Area. Studies on the evolution of wealth inequality in response to unconventional monetary policy shocks in the US are, to the best of my knowledge, lacking.

The main challenges faced in these empirical studies relate to the measurement issues arising from the usage of survey data in the construction of inequality data, and to the identification strategy used to measure the effects of conventional and unconventional monetary policy shocks. In this study, I tackle both issues directly. The contribution is twofold: first, I propose a simple way of accounting for measurement issues arising from

[1] see Section 2 for details on the construction of the data

the construction of the measures of inequality and distribution of income and wealth from survey data. The idea is to have measures that are purged from measurement errors and reflect the macroeconomic dynamics of the series only. This methodology goes along the lines of De Giorgi and Gambetti (2017). Second, I attempt to give a complete overview of the effects of both conventional and unconventional monetary policy shocks on inequality and the distribution in the US.

To this end, I use data from the Consumption Expenditure Survey to construct measures of inequality and the distributions of income and wealth. The different measures are then decomposed into two orthogonal components: one common and one idiosyncratic. The former is common to the other measures and the macroeconomy, and thus represents the macroeconomic dynamics of a given measure. The latter captures series-specific factors, like measurement error, that are unrelated to aggregate economic conditions. We focus on the common components only for each measure of inequality and the distribution in the empirical analysis below, as the concern is to study aggregate shocks like monetary policy.

Then, I study the effects of both conventional and unconventional monetary policy shocks on these measures. Regarding the conventional side, I estimate a proxy SVAR along the lines of Gertler and Karadi (2015) using the quarterly monetary surprise measure proposed in Miranda-Agrippino (2016) as instrument. This measure is a reliable instrument as it is unpredictable by past information, independent of central bank information and not autocorrelated with its past. This approach is fundamentally different from the one of Coibion et al. (2017), which extends the narrative measure of monetary policy shocks of Romer and Romer (2004) and uses it in a local projection framework. This narrative measure suffers from a number of issues: it is autocorrelated with its past values, predicted by past information and it might include anticipated policy shifts. These issues are particularly concerning in the context of local projections and might lead to biased and puzzling results. On the unconventional side, I estimate a SVAR with sign restrictions following Weale and Wieladek (2016) and identify a monthly series of asset purchase announcements shocks. Notably, this series appears uncorrelated with its past and unpredictable by past information. Then, I aggregate this into a quarterly series by using a simple time sum, and project it onto the different measures of inequality and the distribution, one by one. The reason for using a simple local projection within this framework is purely driven by the small sample size of the data at hand for the period of unconventional monetary policies.

The main findings can be summarized as follows. An unexpected increase in the federal funds rate increases all measures of income inequality in the short-run, and financial wealth inequality quite persistently. These effects are sizeable: a 1% increase in the short-term rate leads to an increase by about $0.15 - 0.5\%$, depending on the measure considered. The increase in total and labor income inequality is largely driven by an important heterogenous response of the different percentiles of the distribution: while the top (99th percentile) experiences an increase by about 0.05%, the remaining percentiles of the distribution exhibit a substantial decrease, especially so the bottom (10th percentile), by about 0.2%. Differently from Coibion et al. (2017), the responses of income inequality appear significant and quantitatively relevant in the short-run only. The timing of these results are in line with those of the theoretical framework of Gornemann et al. (2016), both for income and wealth. An expansionary unconventional monetary policy shock, instead, decreases financial wealth, total and labor income inequality, while increasing financial income inequality for the bulk of the measures. Unlike conventional monetary policy, there is no large heterogeneity across the responses of the different percentiles. These all point in the same direction and, if anything, appear particular beneficial to the bottom of the distribution. These findings point towards significantly different channels of transmission between conventional and unconventional monetary policy shocks.

The remainder of the paper is organized as follows: section 2 discusses the data and the construction of the measures of inequality and the distribution of income and wealth. Section 3 lays out the econometric models used to assess the effects of both conventional and unconventional monetary policy shocks. Section 4 illustrates the main results. **??** and section 5 provide, respectively, the robustness checks and the main conclusions of this study.

2 Data

The measures of income and wealth inequality are all constructed from the Consumption Expenditure Survey (CEX), which is a US representative household survey conducted continuously by the Bureau of Labor Statistics since 1980, on a quarterly basis. I include waves from 1980Q1 up to 2014Q4, in order to cover both times of conventional (1980Q1-2008Q4) and unconventional (2009Q1-2014Q4) monetary policy actions. The CEX provides the richest dataset available on the buying habits of American consumers. In this paper, I focus on the Interview Survey only, which is designed to capture the bulk of US households' expenses and contains detailed data on different components of income and wealth. When constructing the data, I use the weights provided within the survey, exclude incomplete income reporters, assign an observation to the given quarter in which the interview is taken, and use non imputed data, when possible[2].

For a better understanding of the mechanisms through which monetary policy affects income inequality and its dynamics, I consider mainly three types of income: total income before tax, labor income and financial income. Financial income is defined as the sum of income received from interests on savings accounts and bonds, and the income received from dividends, royalties, estates, or trusts. Labor income is defined as wage and salary income before deductions obtained by the household. Total income before tax is given by the sum of the above plus income arising from other sources, as transfer income, business income, retirement benefits, pensions et cetera.

Even though not explicitly designed to measure wealth, the CEX survey contains data on multiple sources of financial wealth. Following Attanasio and Hoynes (2000), I construct a measure of financial wealth by summing the amount held by the household in four different components: checking accounts, savings accounts, U.S. saving bonds, and other bonds and equities.

Following Coibion et al. (2017), I consider three measures of inequality: the Gini coefficient, the cross sectional standard deviations of log-levels, and the difference between the 90th percentile and 10th percentile of log levels. Additionally, I consider the 99th,

[2]Since 2004, the CEX survey started imputing income data, with the aim of filling up blanks due to non response. For the years 2004 and 2005, only imputed data is available, whereas, from 2006 onwards, both imputed and non imputed data are available.

90th, 50th and 10th percentiles of the distribution. I construct these measures for all the definitions of income and wealth. Taken together, these are extremely valuable to have a complete overview of inequality, the distribution, and their dynamics.

As discussed in depth by the literature[3], household surveys are characterized by numerous measurement issues, the presence of which might affect importantly the constructed measures of inequality and of the percentiles of the distribution, and thus the estimates of the effects of monetary policy on these. This is certainly the case for the CEX data, which is characterised by sampling and measurement errors since it involves individuals responses, there is usually more measurement error than in administrative data and is top-coded. A frequent approach to tackle this issue is to eliminate outliers by cutting, for instance, the top and bottom 1% of the distributions. However, this strategy is largely arbitrary, doesn't solve any top-coding issue and, more importantly, the top and bottom of the distribution might contain relevant information, especially when considering the reactions to monetary policy. Rather than winsorizing, I propose a purely time series approach to deal with these measurement issues. Specifically, I model the inequality measures of interest and the percentiles of the distribution as noisy indicators of inequality and the distribution.

Let $z_{i,t}$ be the measure of interest. We can think of it as the sum of two orthogonal components: a common component and an idiosyncratic one. The former is common to the other measures $z_{j,t}$, for $j \neq i$, and the macroeconomy, and is driven by common macroeconomic shocks as, for instance, monetary policy. This component embodies the macroeconomic dynamics of the given measure $z_{i,t}$ and thus is our object of interest. The latter, instead, captures series-specific factors, like measurement error, that are unrelated to the macroeconomy and thus do not have pervasive effects on it. This decomposition might not fully address top coding, but it surely addresses more intuitively and directly the sampling and measurement errors mentioned above. Making a direct example, suppose, for instance, that the measure of interest is the 90th percentile of total income before tax. Then, the idiosyncratic component might include factors like health, luck, or measurement error that are specific to that measure.

Thus, we specify $z_{i,t}$ as follows:

$$z_{i,t} = \gamma_i f_t + \xi_{i,t}$$

where f_t is a rx1 vector of macroeconomic factors, γ_i is a 1xr vector of the individual responses of the measure of interest to these, and $\xi_{i,t}$ is a white noise shock which is orthogonal to f_t. This representation of the data is particularly convenient because it allows for different idiosyncratic components (and, thus, measurement errors) for each measure of interest. Moreover, the idiosyncratic components might be correlated among each other. For instance, measurement errors present in the CEX data that are common across the different measures $z_{i,t}$, but do not affect aggregate economic conditions, will be reflected in these components. Again, the idea is to have measures of inequality and the distribution that are purged from measurement error and other factors that are not related to the macroeconomy. This approach has been used, for instance, in De Giorgi and Gambetti (2017) to compute the common component of different percentiles of the

[3] Aguiar and Bils (2015), Attanasio (2003), Attanasio et al. (2004), Bee et al. (2015), Heathcote, Perri, and Violante (2010), Krueger et al. (2010), among others.

distribution of consumption expenditure from the CEX data and analyze the interaction between business cycles and the distribution of consumption.

In order to estimate the common components of the measures of interest, I propose the following steps. First, I merge the FRED-QD database developed by McCracken[4] with all the measures of inequality and the distribution of income and wealth described above and take the $r = 10$ principal components of this dataset, selected based on the test of Alessi, Barigozzi, and Capasso (2010). Let x_t be a vector containing all the macroeconomic variables of the FRED-QD dataset. Define the kx1 vector $\pi_t = [z_t, x_t]'$, where z_t is a vector containing all the measures of inequality and of the distribution, and assume the following structure:

$$\pi_t = \Gamma f_t + \xi_t$$

where $\Gamma = [\gamma_1', ..., \gamma_k']'$ is a kxr matrix and $\xi_t = [\xi_{1,t}, ..., \xi_{k,t}]'$ a rx1 vector. This way, we assume that each variable in π_t is given by the sum of a common and an idiosyncratic components.

Let $\hat{\Sigma}^{\pi}$ be the sample variance-covariance matrix of the data at hand. Let loadings $\hat{\Gamma} = (\hat{\gamma}_1', ..., \hat{\gamma}_k')'$ be the kxr matrix having on the columns the normalized eigenvectors corresponding to the first largest r eigenvalues of $\hat{\Sigma}^{\pi}$. Then, the factors are given by $\hat{f}_t = \hat{\Gamma}'(\pi_{1,t}, \pi_{2,t}, ..., \pi_{k,t})'$. The measure of interest $\hat{z}_{i,t}$ is then obtained as:

$$\hat{z}_{i,t} = \hat{\gamma}_i \hat{f}_t$$

Figure A.1 presents the evolution of the three main measures of inequality for total income before taxes, labor income, financial income and financial wealth. Consistently with the literature on US inequality growth and CEX data[5], all the measures of income and wealth inequality are increasing over time and show some cyclicality.

A spontaneous question is how these measures compare to the raw data and data from other studies. Table 1 shows the correlation between the measures constructed in this paper, the raw inequality data from the CEX obtained by simply excluding incomplete income reporters and using the population weights provided by the survey, and the ones of Coibion et al. (2017). While the estimated measures of inequality for total income are highly correlated to the raw CEX data, only the Gini coefficient appears correlated to the measures of Coibion et al. (2017). However, even if highly correlated, the measures of the gini index of total income have some important differences. To have a clear idea, in Figure A.2 I plot three different measures for the gini index of total income before tax: the estimated common component from this paper, the raw CEX data and the one of Coibion et al. (2017). Notwithstanding the high correlation between the different measures, the measurement errors appear far from negligible. For instance, both the raw CEX data and the measure of Coibion et al. (2017) appear to overstate inequality, especially after 1995.

[4]The dataset is available on the FRED website page: https://research.stlouisfed.org/econ/mccracken/fred-databases/. The original FRED-QD database includes 248 macroeconomic and financial variables that give a comprehensive view of the macroeconomy. Variables for which data is not available for the sample of interest, namely 1980Q1-2014Q4, are dropped. The final database includes 236 variables.

[5]Heathcote et al. (2010), Krueger and Pistaferri (2005), among others.

Table 1: Correlation among inequality measures

	Raw CEX Data	Coibion et al. (2017)
$\hat{z}_{Gini,t}$	0.97	0.85
$\hat{z}_{sd,t}$	0.94	0.5
$\hat{z}_{P9010,t}$	0.90	0.45

3 Econometric Methodology

3.1 Conventional Monetary Policy

To estimate the effects of conventional monetary policy shocks on inequality and the distribution, I estimate a proxy SVAR along the lines of Gertler and Karadi (2015). This methodology is a variation of the ones developed by Stock and Watson (2012) and Mertens and Ravn (2013), and is explained in detail hereinafter.
Consider the following reduced form VAR model:

$$Y_t = C + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + u_t \tag{1}$$

where Y_t is a n x 1 vector containing all the n endogenous variables, C is a n x 1 vector of constants, $A_1, ..., A_p$ are the n x n matrices of coefficients associated with the p lags of the dependent variable and $u_t \sim WN(0_n, \Omega)$[6] is the reduced form residual. Let the mapping from the reduced form residuals to the economically meaningful structural shocks be $u_t = S\epsilon_t$, where $\epsilon_t \sim WN(0_n, I_n)$ is a vector of structural white noise shocks with unit variance and S in a non-singular matrix with $SS' = \Omega$. Let **s** denote the column in matrix S corresponding to the impact of the structural monetary policy shock ϵ_t^p on each element of the vector of reduced form residuals u_t. To compute the impulse responses to a monetary policy shock only, (1) reduces to:

$$Y_t = C + \sum_{j=1}^{p} A_j Y_{t-p} + \mathbf{s}\epsilon_t^p \tag{2}$$

The objective is then to estimate **s** and we can make use of external instruments as an identification strategy. Let Z_t be a vector of instrumental variables and let ϵ_t^q be a vector of structural shocks other than the monetary policy one. For the instrumental variables to be valid, we need two conditions. First, Z_t must be correlated with the monetary policy shock ϵ_t^p:

$$E(Z_t \epsilon_t^{p'}) = \Phi \tag{3}$$

Second, Z_t must be orthogonal to the other variables' shocks ϵ_t^q:

$$E(Z_t \epsilon_t^{q'}) = 0 \tag{4}$$

Given these two conditions, the estimation can be carried out as follows. First, estimate the VAR(p) process in (1) by ordinary least squares and obtain the vector of reduced

[6]This indicates a white noise process with mean the n x 1 vector of zeros 0_n and with variance-covariance matrix Ω.

form residuals $\hat{u}_t$. Second, let $\hat{u}_t^p$ be the estimated residual in (1) for the monetary policy variable and let $\hat{u}_t^q$ be the estimated residual for the remaining variables in the system. In order to isolate the variation in $\hat{u}_t^p$ that is due to the structural monetary policy shock ϵ_t^p only, we can regress the former on the vector of instrumental variables Z_t and a constant:

$$\hat{u}_t^p = \alpha + \beta Z_t + \xi_t \tag{5}$$

The resulting fitted value $\hat{\hat{u}}_t^p$ can be then used to estimate consistently $\frac{\mathbf{s}^q}{s^p}$ as follows:

$$\hat{u}_t^q = \frac{\mathbf{s}^q}{s^p}\hat{\hat{u}}_t^p + \delta_t \tag{6}$$

Normalizing s^p to 1, then $\hat{\mathrm{s}} = (1, \hat{\mathrm{s}}^{q'})'$. In this way, we normalize the impact effect of the policy variable to a monetary policy shock to be of 1 percent.

The VAR in (1) is estimated using quarterly data that spans the period 1980Q1-2008Q4, in order to cover a time range of monetary policies that are conventional only. The set of endogenous variables Y_t includes five variables in levels: the effective federal funds rate, the inequality measure of interest $\hat{z}_{i,t}$, the log of real GDP per capita, the log of the GDP deflator and the excess bond premium of Gilchrist and Zakrajšek (2012). Thus, I estimate a different VAR for each measure $\hat{z}_{i,t}$. The lags are set to four, i.e. the average between the AIC, BIC and HQC criteria. This is also common in the time series literature, where it is usual to set a lag length that covers a year's worth of data.[7] Confidence bands for the impulse responses are computed using the bootstrap-after-bootstrap methodology of Kilian (1998) to correct for the bias and skewness of the small-sample distribution of the impulse response estimator.

The baseline specification of Gertler and Karadi (2015) uses changes in the three-month federal funds futures around a 30 minutes window of the FOMC announcement to construct a monthly series of monetary surprises that is used as instrument in a proxy SVAR. The instrument proposed, however, has two main flaws. First, it is predicted by past information and autocorrelated with its past. Second, there is a potentially important information problem due to the fact that central banks transfer information about future developments of the economy around policy announcements. Thus, it is difficult to disentangle a pure monetary policy surprise from one that arises, for instance, from central bank information, as illustrated in Jarocinski and Karadi (2018) and Nakamura and Steinsson (2018). These issues are potentially even more concerning once we aggregate the monthly measure into a quarterly one. For these reasons, I use the instrumental variable proposed by Miranda-Agrippino (2016), which projects the measure of Gertler and Karadi (2015) on central banks' forecasts and forecasts revisions of output, inflation and unemployment. This measure is unpredictable by past information, independent of central bank information and not autocorrelated with its past. This way we have a reliable instrument and thus identification of the effects of monetary policy shocks. The instrument is available at both monthly and quarterly frequencies from 1990Q1 onwards. Thus, the VAR in (1) is estimated on the full sample 1980Q1-2008Q4, whereas (5) and (6) are estimated on the sample 1990Q1-2008Q4. We present the impulse responses of the inequality measures as well as the remaining variables in the system to a conventional monetary policy shock in the next section.

[7] Results are very similar using different lag specification and are available upon request.

This approach is fundamentally different from the one of Coibion et al. (2017), which extends the narrative measure of monetary shocks of Romer and Romer (2004) and uses it in a local projection framework. This narrative measure, however, suffers from the same issues raised above: it is autocorrelated with its past values, predicted by past information and likely includes anticipated policy shifts. These issues are particularly concerning in the context of a local projection, in which the measure is included directly in the regression (and not as an instrument) and thus might lead to biased results, as shown in Miranda-Agrippino and Ricco (2017). Moreover, as shown in Ramey (2016), this measure produces puzzling results unless additional zero contemporaneous restrictions are imposed, for instance on output and inflation. These restrictions have been heavily criticized by the literature because not based on economic theory, thus not easily justifiable, and because these often produce responses to monetary policy shocks at odds with theoretical models. These observations motivated the usage of a proxy SVAR and of the measure of Miranda-Agrippino (2016) as identification strategy in this study.

3.2 Unconventional Monetary Policy

To study the effects of unconventional monetary policy shocks on inequality and the distributions of income and wealth, I proceed as follows. First, I estimate a SVAR with sign restrictions following Weale and Wieladek (2016) and identify a monthly series of asset purchase announcements shocks. Second, I aggregate these into quarterly shocks using a simple time sum and project these onto the different measures $\hat{z}_{i,t}$, one by one. The reason for using a simple local projection within this framework is purely driven by the small sample size of the data at hand for the period of unconventional monetary policies.

Weale and Wieladek (2016) estimate a SVAR with a combination of contemporaneous and sign restrictions to identify the shocks of the large-scale asset purchase program, undertook by the Federal Reserve in response to the financial crisis of 2008-2009, on the macroeconomy. Consider the following reduced form VAR model estimated on monthly data:

$$X_t = \Lambda + \sum_{j=1}^{p} A_j X_{t-j} + e_t, \quad e_t \sim N(0_n, \Omega) \tag{7}$$

where X_t is a jx1 vector containing the following endogenous variables: log of monthly GDP, log of CPI, asset purchase announcements in percentage of 2009Q1 GDP and log of real equity prices [8].

In keeping with the baseline specification of Weale and Wieladek (2016), the VAR in (7) is estimated with Bayesian methods and variables in levels. A flat prior is specified for the reduced form parameters so that the information in the likelihood is dominant. Let $e_t = A\eta_t$, where $\eta_t \sim N(0, I_n)$ is a n x 1 vector of structural disturbances and A is such that $AA' = \Sigma$. In order to identify all the shocks in the system we need at least $\frac{j(j-1)}{2}$ additional restrictions. The identification strategy corresponds to the second identification scheme of Weale and Wieladek (2016), which relies on the sign restrictions presented in Table 1 and is implemented using the QR algorithm of Rubio-Ramírez, Waggoner, and Zha (2010).

[8] Monthly averages of the daily $S\&P500$ Index deflated with CPI. Monthly GDP is taken from Macroeconomic Advisers.

Table 2: Sign restrictions

Sign Restrictions			
	Supply	Demand	AP
CPI	-	+	/
GDP	+	+	/
AP announcements	/	/	+
10yrs governement rate	+	+	-
SP500	+	+	+

Note: Sign restrictions on impulse responses in the SVAR model. The restrictions are imposed on impact and one period ahead.

Define $\hat{\eta}_t$ as the median structural shock, obtained from the estimation of the SVAR above. Consider the third shock in $\hat{\eta}_t$, i.e. the one corresponding to the asset purchase announcements variable, $\hat{\eta}_t^{UMP,m}$, where m indicates that the shock is monthly. To make sure that the series of shocks is unanticipated, i.e. orthogonal to other macroeconomic variables, I perform the sufficient information test proposed in Forni and Gambetti (2014). First, I take the FRED-MD dataset available on McCraken's website. I compute the first 8 principal components of this large dataset, again based on the test of Alessi et al. (2010). Then, I perform a Granger causality test to check whether the principal components Granger cause $\hat{\eta}_t^{UMP,m}$. Whether I include lags of $\hat{\eta}_t^{UMP,m}$ in the regression or not, or a multiplicity of lags of the factors, I don't reject the null of no granger causation, for all the specifications. Thus, this series doesn't appear to be predictable by the macroeconomy and past information.

Then, I aggregate this monthly shock into a quarterly shock, $\hat{\eta}_t^{UMP,q}$, by a simple time sum. Finally, to compute the effects of an unconventional monetary policy shock on inequality, I use a simple local projection à la Jordà (2005):

$$\Delta \hat{z}_{i,t+h} = \gamma_i^{(h)} + \beta_i^{(h)} \hat{\eta}_t^{UMP,q} + \psi_{i,t+h}, \quad h = 0, ..., H. \tag{8}$$

where $\gamma_i^{(h)}$ is a constant and $\psi_{i,t+h}$ is the residual.

The cumulated estimated coefficients $\sum_{h=0}^{h} \hat{\beta}_i^{(h)}$, for $h = 0, ..., H$, represent the effects of an unconventional monetary policy shock at time t on the inequality measure considered at time $t + h$. The specification of (8) allows for a contemporaneous effect of the unconventional monetary policy shock on the inequality measure of interest and it is particularly convenient given the small sample at hand. I don't include additional lags of the shock $\hat{\eta}_t^{UMP,q}$, as the sample autocorrelation function doesn't reveal significant correlation between different lags, both for the monthly and quarterly shock, and since the inclusion of these would imply dropping observations. Furthermore, (8) doesn't include lags of the dependent variable, which are included in the robustness checks section. The equation in (8) is estimated for each inequality measure of each variable. The impulse responses are presented in the next section with 1 and 1.65 standard deviation confidence bands, computed with Newey-West heteroskedasticity and autocorrelation robust standard errors, as, except for $h = 0$, the errors are serially correlated. I estimate (7) using data from 2008:09 to 2014:12, in order to capture the time horizon in which the asset purchases where announced and implemented, and using two lags, as in Weale and Wieladek (2016). This lag length choice coincides also with the average of the AIC, BIC and HQC criteria. Moreover, I estimate (8) including data from 2009Q1-2014Q4 and unconventional monetary policy shocks up to 10 quarters behind, i.e. $h = 0, ..., 10$.

4 Main Results

4.1 Conventional Monetary Policy

To assess the effects of conventional monetary policy shocks on inequality and the distributions of income and wealth, I estimate the proxy SVAR presented in Section 3.1 for each measure $\hat{z}_{i,t}$. In the following figures, the horizontal axis measures time from impact to 20 quarters after the innovation has occurred. The vertical axis represents the responses for each variable in percentage points.

Before presenting the results on inequality and the distribution, I perform the following exercise: estimate the proxy SVAR by excluding the measure of interest $\hat{z}_{i,t}$ from the system, and thus focusing on the responses of macroeconomic variables only to a conventional monetary policy shock. This constitutes a validating exercise for the identification strategy and it is presented in Figure B.1. In line with the bulk of the theoretical and empirical literature on conventional monetary policy shocks, an increase in the short-term interest rate leads to a persistent decrease in GDP and in the GDP deflator, and to a short-run rise in the excess bond premium. Moreover, the increase in the effective federal funds rate is rather short-lived and reverts after a couple of quarters.

Figure B.2 presents the impulse responses of the different measures of inequality of total income before tax, labor income and financial income to a contractionary monetary policy shock. An unexpected increase in the federal funds rate leads to a statistically significant short-run increase for all the measures considered. A couple of remarks are in place: first, the effects on total income and financial income appear, for the bulk of the measures, more persistent than the ones on labor income. Second, financial income inequality reverses after approximately 12 quarters when considering the cross-sectional standard deviation of log levels and the difference between the 90th and 10th percentiles. This could be linked to the decrease in the short-term interest rate after a couple of quarters, as observed in Figure B.1. Third, the increase in financial income inequality is of smaller magnitude than for the other two components of income. In fact, a 1% increase in the federal funds rate leads to an increase in total income inequality of about 0.25% for each measure, of around $0.25 - 0.5\%$ for labor income and of about 0.15% for financial income. Thus, monetary policy has sizeable effects on income inequality, especially so for labor income. Interestingly, most of the action takes place in the short-run. While the direction of the response to a monetary policy shock is the same as in Coibion et al. (2017), the timing is substantially different: in their framework inequality increases after approximately 2 years. The results of Figure B.2 are more in line, for instance, with the theoretical framework of Gornemann et al. (2016), in which the Gini coefficients of total income and labor income increase on impact and persistently for about 10 quarters after a contractionary monetary policy innovation has occurred.

In order to understand the underlying dynamics behind the responses of income inequality, we can focus our attention to the different percentiles of the distribution. Figure B.3 presents the impulse responses of the different percentiles of income considered: 99th, 90th, 50th and 10th. The main observation is that there is large heterogeneity in the responses across percentiles, for each income component. Starting with total and labor income, we observe a negative and persistent decrease for the 90th, 50th and 10th percentiles. The magnitude of the bottom 10th percentile is particularly relevant, since it more than doubles the decrease in the 50th percentile and it is five times stronger than the one of the 90th percentile. The effects are sizeable: a 1% increase in the policy rate decreases both total and labor income for the 10th percentile by about 0.2%. However, this decrease appears more short-lived than for the 90th and 50th percentiles. The only percentile experiencing an increase is the top 1%, by about $0.05 - 0.1\%$. While inequality between the 90th and 10th percentiles increases moderately, the distance between the top 1% and the bottom 10% increases substantially. These findings highlight a widening of the tails of the distribution after a tightening and a completely different transmission across agents with different levels of income.

Focusing on financial income, an unexpected rise in the interest rate has a positive and significant short-run effect on financial income for the 99th, 90th and 50th percentiles, whereas it affects negatively and significantly the 10th percentile. The increase in financial income is of much higher magnitude, on impact, for the 99th and 90th percentile, highlighting that these percentiles are the one benefiting and reacting the most to a monetary policy shock. A 1% increase in the short-term rate increases the 99th and 90th percentiles by about 0.2% and decrease the 10th percentile by about 0.1%. This finding seems to suggest a different portfolio composition between the agents that hold a high amount of financial income and the ones that hold a small one. More precisely, the top of the distribution seems to have an asset composition that is extremely positively affected by increases in the federal funds rate, whereas the bottom negatively. This result is in line with the findings of Doepke and Schneider (2006). Notice that, as the CEX is top coded, namely it doesn't include the top 1% of the population, these results could be a lower bound of the real increase in financial income, at least for the top percentiles of the distribution. Therefore, these effects could be of a potentially higher magnitude, indicating the importance of the channels related to this income component when considering monetary policy. However, the top-coding issue should be, at least in part, addressed by the decomposition proposed in Section 2.

To sum up, contractionary monetary policy seems to affect in a positive way mostly those agents that are connected to financial markets and hold assets that are positively correlated to changes in the short term rate. On the contrary, it affects negatively those agents who are more susceptible to earning risks, namely those at the lower tail of the income distribution.

Figure B.4 illustrates the impulse responses of the different measures of inequality and the distribution of financial wealth. Consistently with the findings of the theoretical model of Gornemann, Kuester, and Nakajima (2016), a contractionary monetary policy shock leads to a persistent increase in the Gini coefficient of financial wealth inequality, by about $0.1 - 0.2\%$ on impact. When focusing on the additional measures of inequality, the increase is on impact only. This is because wealth increases significantly for all the percentiles of the distribution. However, the rise of the 99th percentile is of a significantly higher magnitude than for the rest. In fact, the increase is relevant and persistent when taking the difference between the 99th and 10th percentiles, showing a widening of the upper and lower tails of the distribution, like in the case of income.

4.2 Unconventional Monetary Policy

As for the conventional case, we first set out the sign restricted SVAR to examine the behavior of the macroeconomy in response to asset purchase announcement shocks. Figure B.5 presents the results. An increase in asset purchases positively affects GDP, CPI and stock prices, while having a negative effect on the long term yield, in line with the findings of Weale and Wieladek (2016).

To assess the effects of unconventional monetary policy shocks on inequality and the distributions of income and wealth, I estimate (8) individually for each measure $z_{i,t}$. Figure B.6 presents the results for the measures of inequality. The first two columns exhibit a significant decrease in total and labor income inequality after approximately six quarters. The magnitude and dynamics are extremely similar for these two income components. Financial income inequality, on the other side, shows at first glance a quite puzzling behavior: while the Gini coefficient decreases strongly and persistently, the opposite is observed for the other two measures. The difference could be due to the fact that the Gini coefficient contains households with zero financial income.

In light of the foregoing, it is interesting to explore the dynamics of the different percentiles for the distribution in response to the expansionary unconventional monetary policy shock, which are presented in Figure B.7. Differently from the conventional case, there is no wide heterogeneity in the response of the different percentiles of total and labor income to an asset purchase announcement shock, and all the responses are of the same sign. If anything, the 10th percentile appears to be the one that benefits the most after a couple of quarters for both measures, while the magnitude of the responses of the remaining percentiles is nearly identical. The pattern is different if we focus on financial income, for which the highest increase is observed at the top of the distribution and the smallest at the bottom. Again, the sign of the responses is the same for each percentile. All in all, while the effects on inequality exhibit an opposite sign between the top and bottom of the distribution to a conventional monetary policy tightening, the same doesn't take place for unconventional monetary policy, highlighting potentially different channels of transmission.

The responses of the measures of inequality and of the distribution of financial wealth are presented in Figure B.8. Differently from the conventional case, there is an unambiguous and persistent decrease in each measure of inequality in response to an increase in asset purchases. This fall is mainly driven by the prompt rise in the 50th and 10th percentiles. The response is of higher magnitude for the latter. The responses appear largely insignificant for the 99th and 90th percentiles, for which, if anything, a slight decrease is observed after eight quarters. It's interesting to point out that, overall, an increase in financial wealth is observed, as for the case of a conventional monetary policy tightening. This result further highlights the presence of different transmission mechanisms between conventional and unconventional actions.

5 CONCLUSIONS

In this paper, I propose a simple methodology to purge the measures of inequality and the distributions of income and wealth from measurement errors, and then investigate the effects of both conventional and unconventional monetary policy shocks on these.

The analysis yields several interesting results. In response to a conventional monetary policy tightening, all the measures of inequality are increasing. While total, labor and financial income are affected mostly in the short-run, financial wealth inequality increase more persistently. The effects appear sizeable for each measure considered. The rise in inequality is largely driven by heterogeneous responses across the different percentiles of the distribution. While the 99th percentile of total and labor income exhibits a mild but persistent increase, all the remaining percentiles experience a fall, of particularly relevant magnitude for the bottom of the distribution. Financial wealth, on the other hand, increases for all the different percentiles, but especially so for the top 1%, reason for which inequality overall increases. When turning our attention to an increase in asset purchases, financial wealth, total and labor income inequality all decrease, whereas financial income inequality increases for the bulk of the measures. Like conventional monetary policy, the percentiles of financial wealth are all increasing. Unlike conventional monetary policy, there is not a wide heterogeneity across the responses of the different percentiles, and these all point in the same direction. These findings point towards different channels of transmission between conventional and unconventional monetary policy shocks.

REFERENCES

Aguiar, M. and M. Bils (2015). Has consumption inequality mirrored income inequality? *American Economic Review 105*(9), 2725–56.

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics Probability Letters 80*(23-24), 1806–1813.

Attanasio, O. P. and H. W. Hoynes (2000). Differential mortality and wealth accumulation. *The Journal of Human Resources 35*(1), 1–29.

Cairo, I. and J. W. Sim (2018, July). Income Inequality, Financial Crises, and Monetary Policy. Finance and Economics Discussion Series 2018-048, Board of Governors of the Federal Reserve System (US).

Coibion, O., Y. Gorodnichenko, L. Kueng, and J. Silvia (2017). Innocent bystanders? monetary policy and inequality. *Journal of Monetary Economics 88*, 70 – 89.

De Giorgi, G. and L. Gambetti (2017). Business Cycle Fluctuations and the Distribution of Consumption. *Review of Economic Dynamics 23*, 19–41.

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

Furceri, D., P. Loungani, and A. Zdzienicka (2018). The effects of monetary policy shocks on inequality. *Journal of International Money and Finance 85*(C), 168–186.

Gertler, M. and P. Karadi (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

Gilchrist, S. and E. Zakrajšek (2012). Credit spreads and business cycle fluctuations. *American Economic Review 102*(4), 1692–1720.

Gornemann, N., K. Kuester, and M. Nakajima (2016). Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy. CEPR Discussion Papers 11233, C.E.P.R. Discussion Papers.

Heathcote, J., F. Perri, and G. L. Violante (2010). Unequal we stand: An empirical analysis of economic inequality in the united states, 1967â2006. *Review of Economic Dynamics 13*(1), 15 – 51. Special issue: Cross-Sectional Facts for Macroeconomists.

Jarocinski, M. and P. Karadi (2018). Deconstructing Monetary Policy Surprises - The Role of Information Shocks. CEPR Discussion Papers 12765, C.E.P.R. Discussion Papers.

Kilian, L. (1998). Small-Sample Confidence Intervals For Impulse Response Functions. *The Review of Economics and Statistics 80*(2), 218–230.

Kumhof, M., R. Rancière, and P. Winant (2015). Inequality, Leverage, and Crises. *American Economic Review 105*(3), 1217–1245.

Lenza, M. and J. Slacalek (2018). How does monetary policy affect income and wealth inequality? Evidence from quantitative easing in the euro area. Working Paper Series 2190, European Central Bank.

Mertens, K. and M. O. Ravn (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review 103*(4), 1212–47.

Miranda-Agrippino, S. (2016). Unsurprising shocks: information, premia, and the monetary transmission. Bank of England working papers 626, Bank of England.

Miranda-Agrippino, S. and G. Ricco (2017). The transmission of monetary policy shocks. Bank of England working papers 657, Bank of England.

Montecino, J. A. and G. Epstein (2015). Did Quantitative Easing Increase Income Inequality? Working Papers Series 28, Institute for New Economic Thinking.

Mumtaz, H. and A. Theophilopoulou (2017). The impact of monetary policy on inequality in the UK. An empirical analysis. *European Economic Review 98*(C), 410–423.

Nakamura, E. and J. Steinsson (2018). High-Frequency Identification of Monetary Non-Neutrality: The Information Effect*. *The Quarterly Journal of Economics 133*(3), 1283–1330.

Rajan, R. G. (2010). *Fault Lines: How Hidden Fractures Still Threaten the World Economy*. Princeton University Press.

Ramey, V. (2016). *Macroeconomic Shocks and Their Propagation*, Volume 2 of *Handbook of Macroeconomics*, Chapter 0, pp. 71–162. Elsevier.

Romer, C. D. and D. H. Romer (2004). A New Measure of Monetary Shocks: Derivation and Implications. *American Economic Review 94*(4), 1055–1084.

Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010). Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies 77*(2), 665–696.

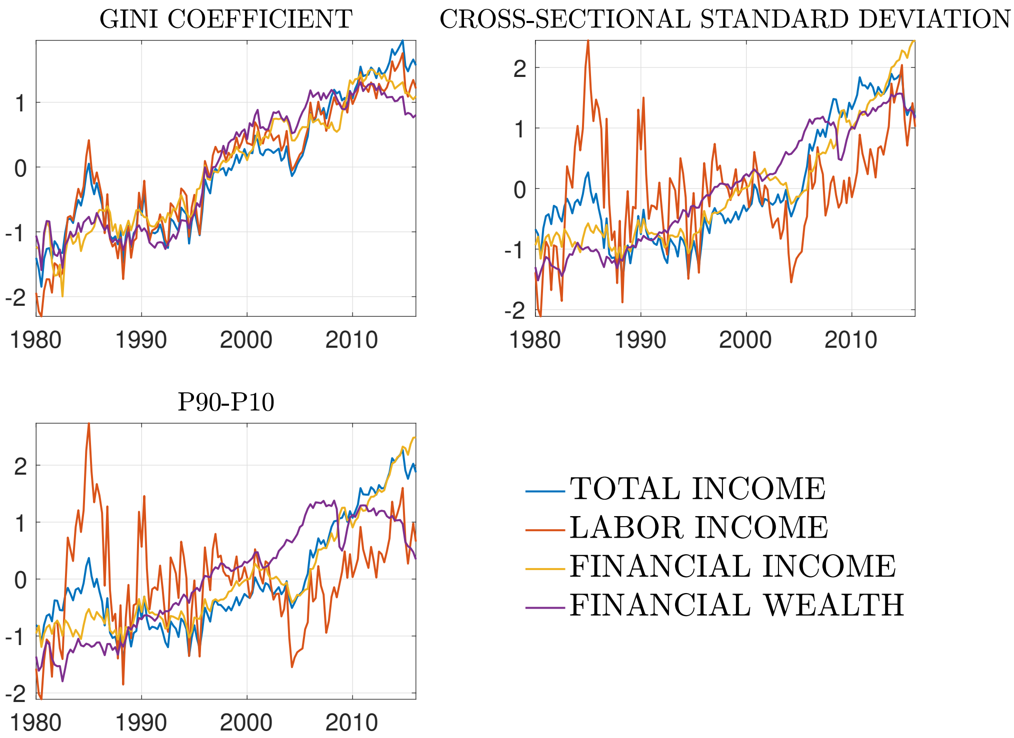
Stock, J. H. and M. W. Watson (2012). Disentangling the channels of the 2007-2009 recession. Working Paper 18094, National Bureau of Economic Research.

Weale, M. and T. Wieladek (2016). What are the macroeconomic effects of asset purchases? *Journal of Monetary Economics 79*(C), 81–93.

Appendix

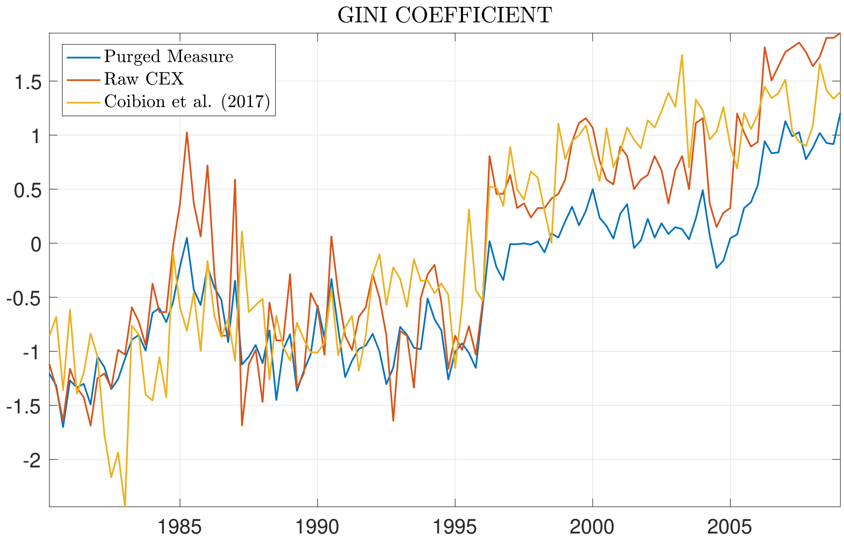
A Inequality measures

Figure A.1: Measures of income and wealth inequality



Source: Author's calculations using the CEX survey. Data is standardized. See Section 2 for details.

Figure A.2: Measures of income and wealth inequality

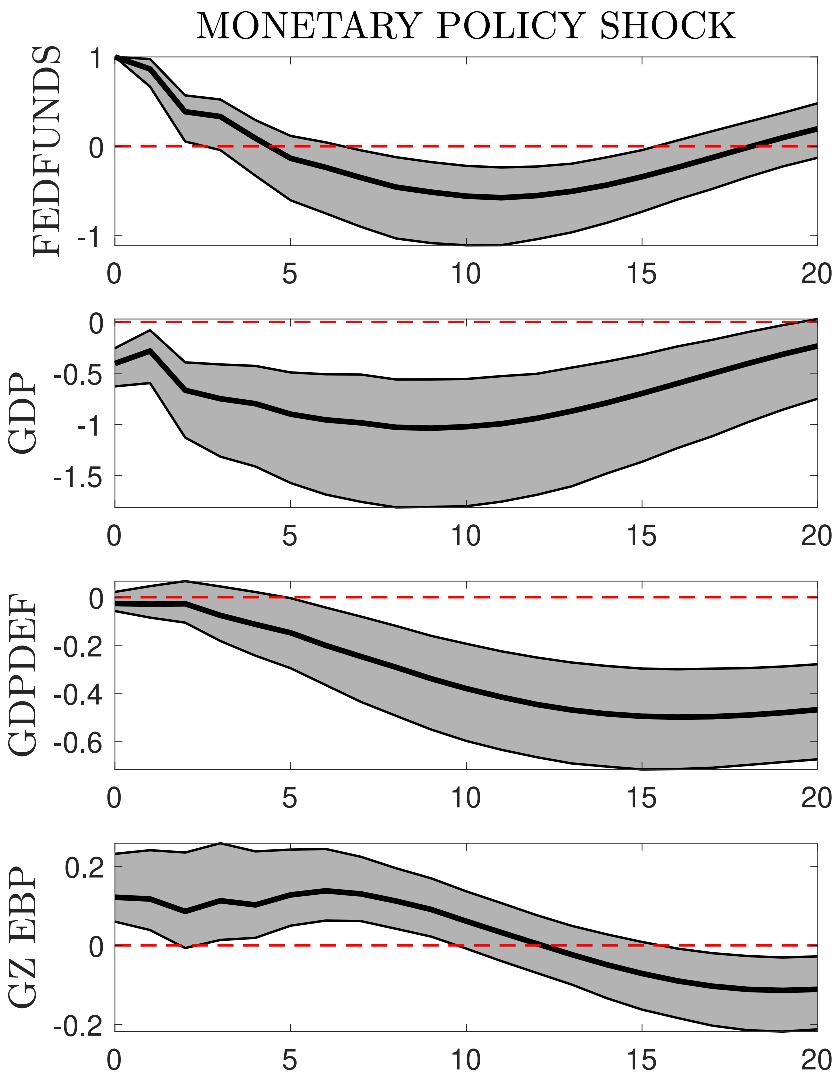


Source: Author's calculations using the CEX survey and data from Coibion et al. (2017). Data is standardized.

B IMPULSE RESPONSE FUNCTIONS

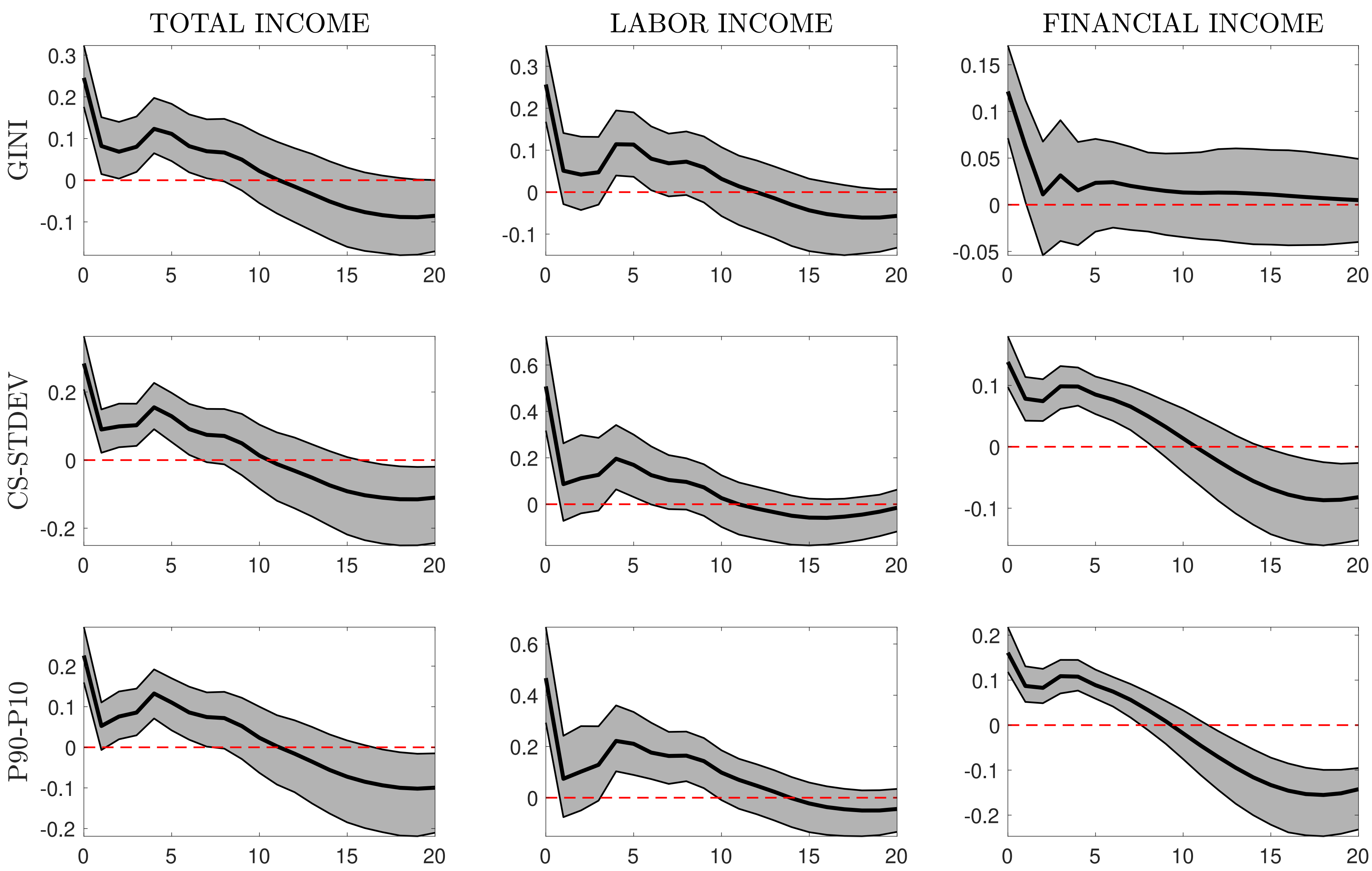
B.1 CONVENTIONAL MONETARY POLICY

Figure B.1: Impulse responses from the baseline Proxy SVAR model - Macroeconomic variables



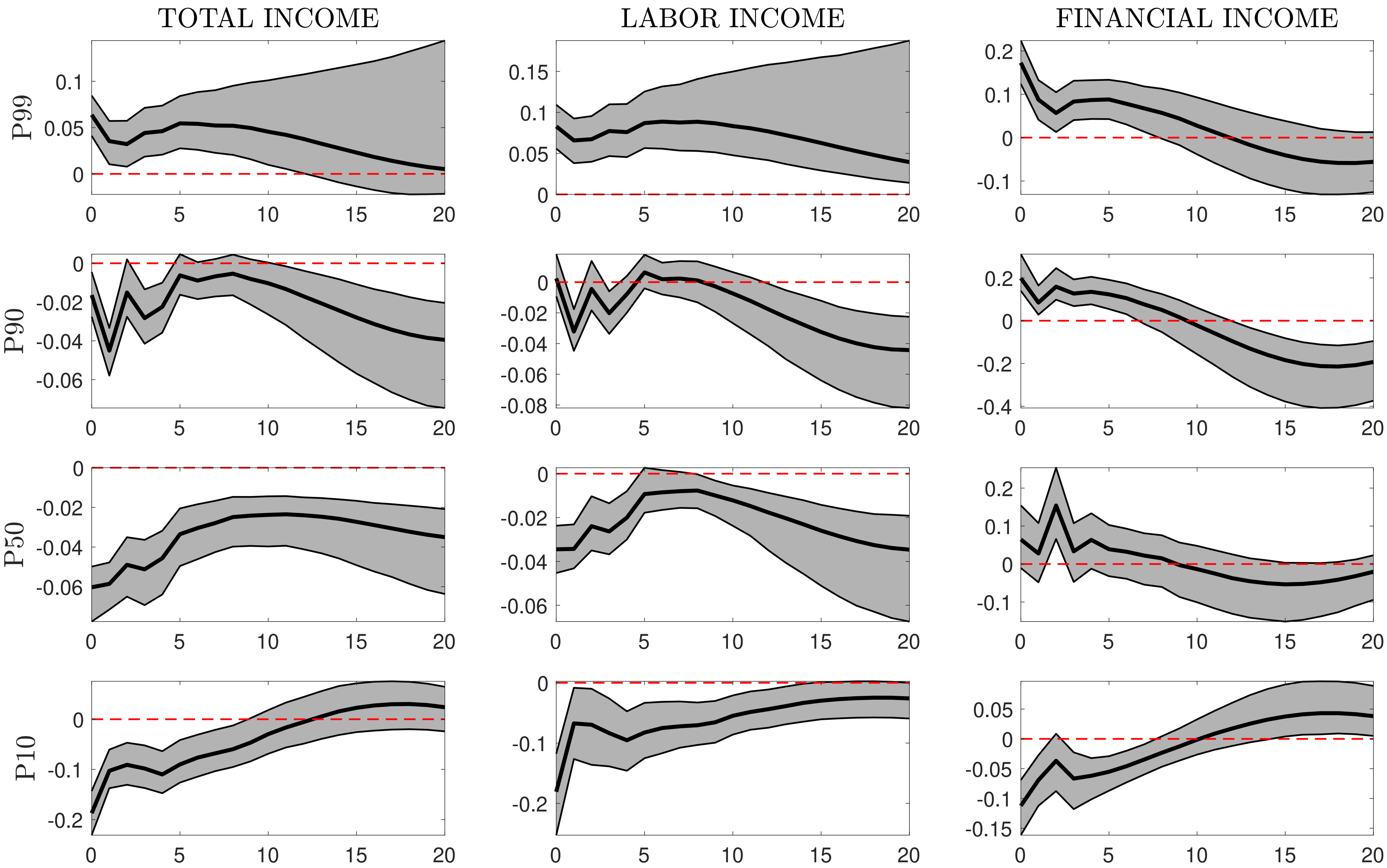
Note: Impulse responses of the different macroeconomic variables to a monetary policy shock using the baseline proxy SVAR model excluding the measure of interest $z_{i,t}$ from the system. The light-shaded areas are 68% bootstrap confidence bands.

Figure B.2: Impulse responses from the baseline Proxy SVAR model - Income inequality measures



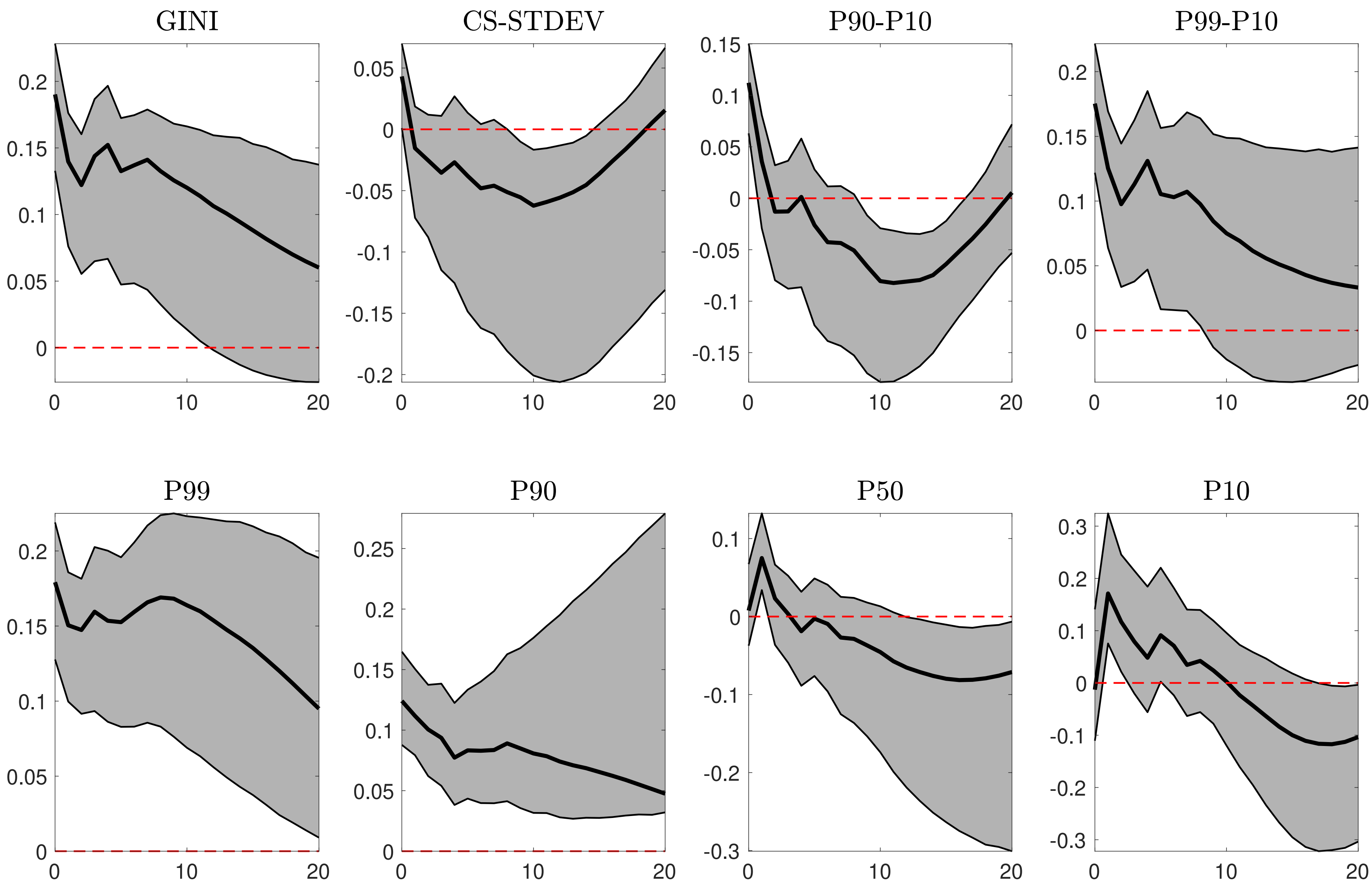
Note: Impulse responses of the different income inequality measures to a monetary policy shock using the baseline proxy SVAR model. The light-shaded areas are 68% bootstrap confidence bands.

Figure B.3: Impulse responses from the baseline Proxy SVAR model - Percentiles of the distribution of income



Note: Impulse responses of the different income percentiles to a monetary policy shock using the proxy baseline SVAR model. The light-shaded areas are 68% bootstrap confidence bands.

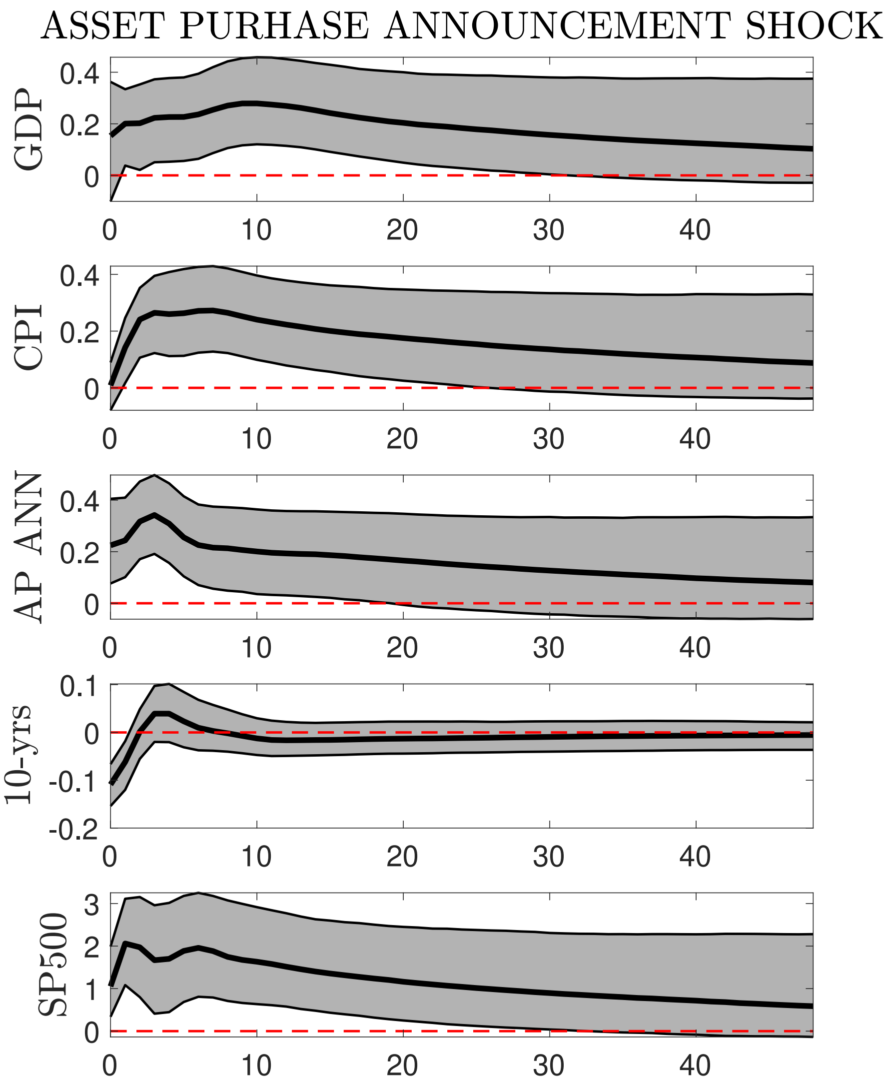
Figure B.4: Impulse responses from the baseline Proxy SVAR model - Measures of Financial Wealth



Note: Impulse responses of the measures of inequality and the different percentiles of the distribution of financial wealth to a monetary policy shock using the baseline proxy SVAR model. The light-shaded areas are 68% bootstrap confidence bands.

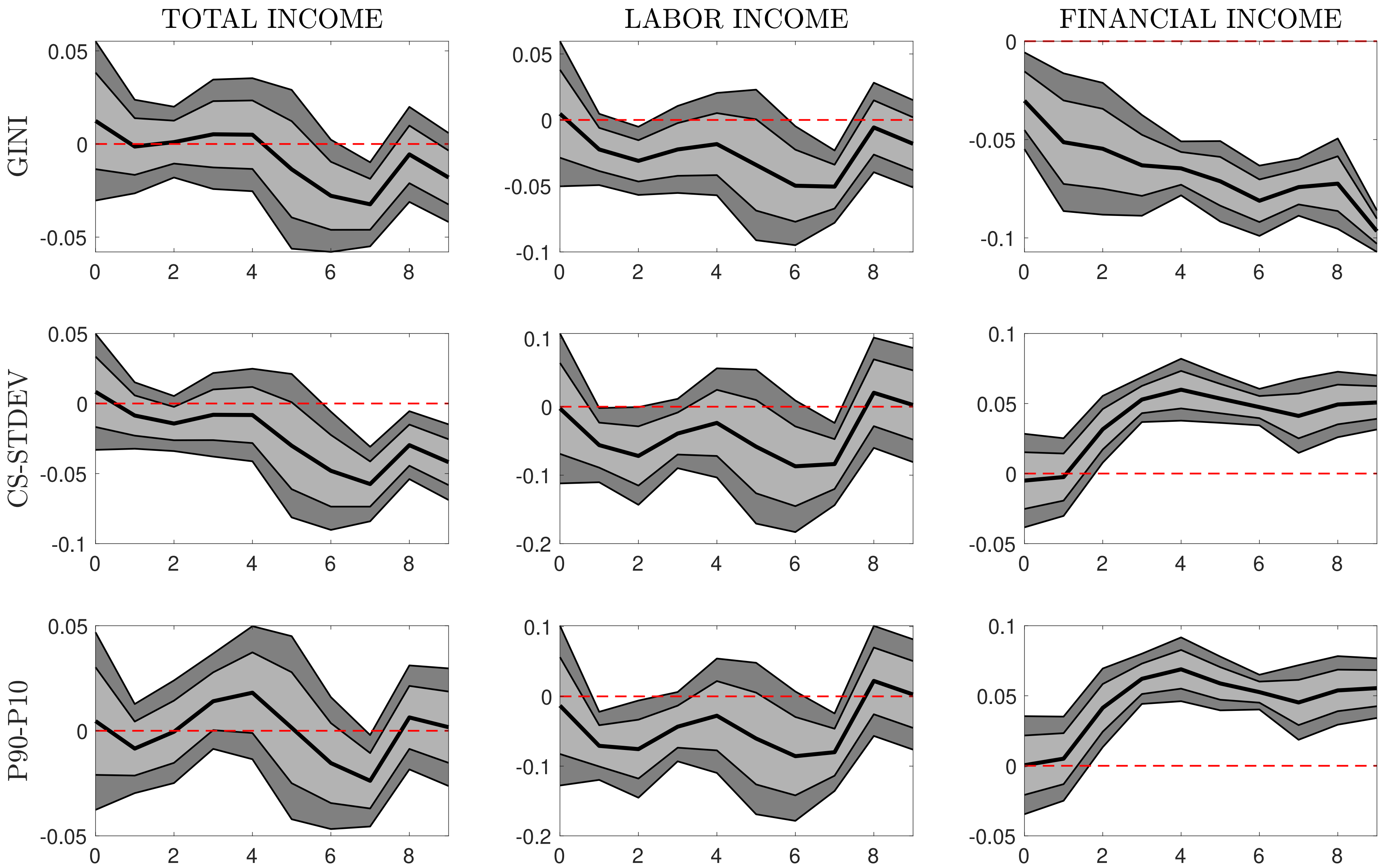
B.2 Unconventional monetary policy

Figure B.5: Impulse responses from the baseline SVAR model with sign restrictions



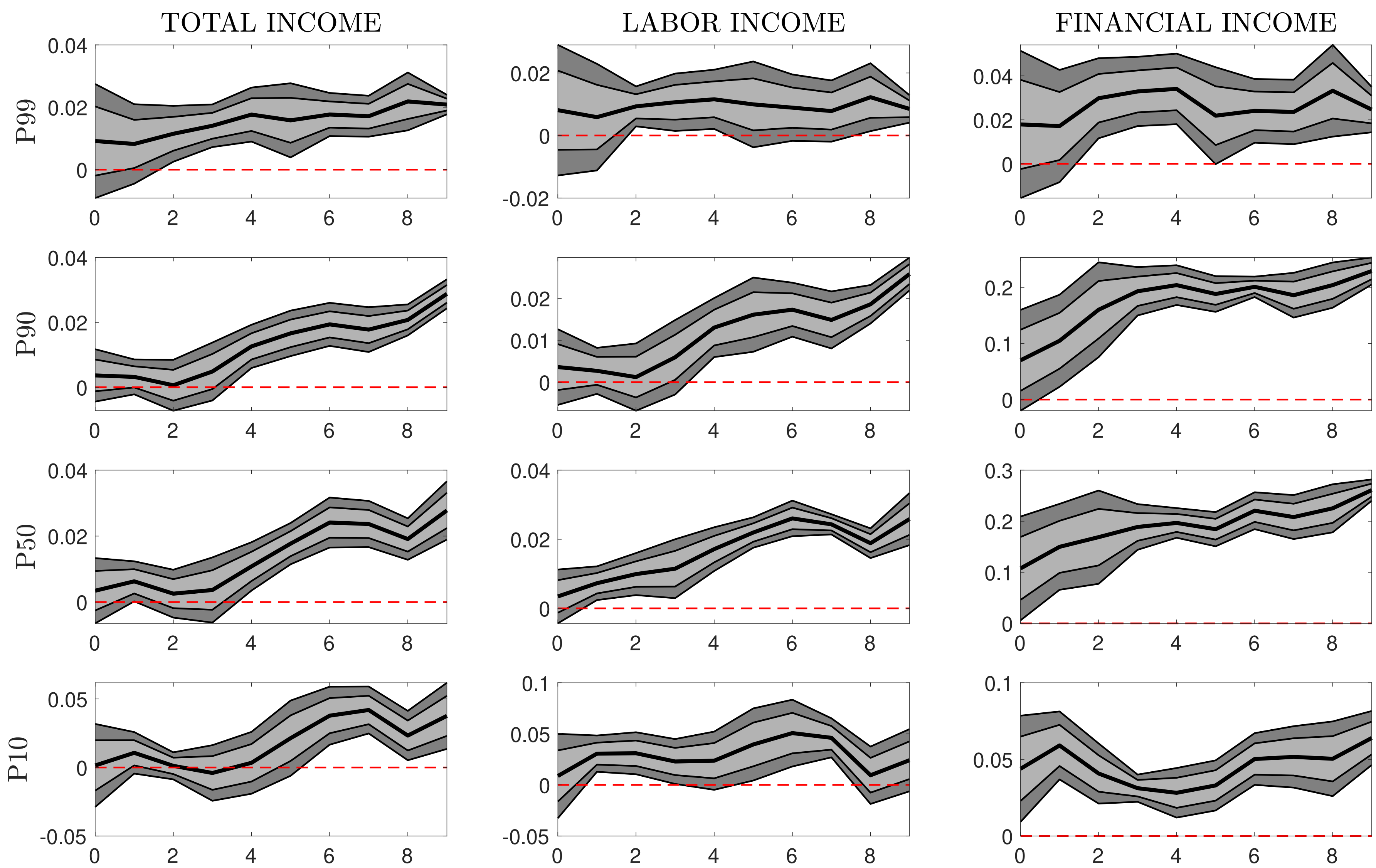
Note: Impulse responses of the different macroeconomic variables to an asset purchase announcement shock of one standard deviation. The light-shaded areas are 68% confidence bands. The solid line is the point-wise median.

Figure B.6: Impulse responses from the local projection - Income inequality measures



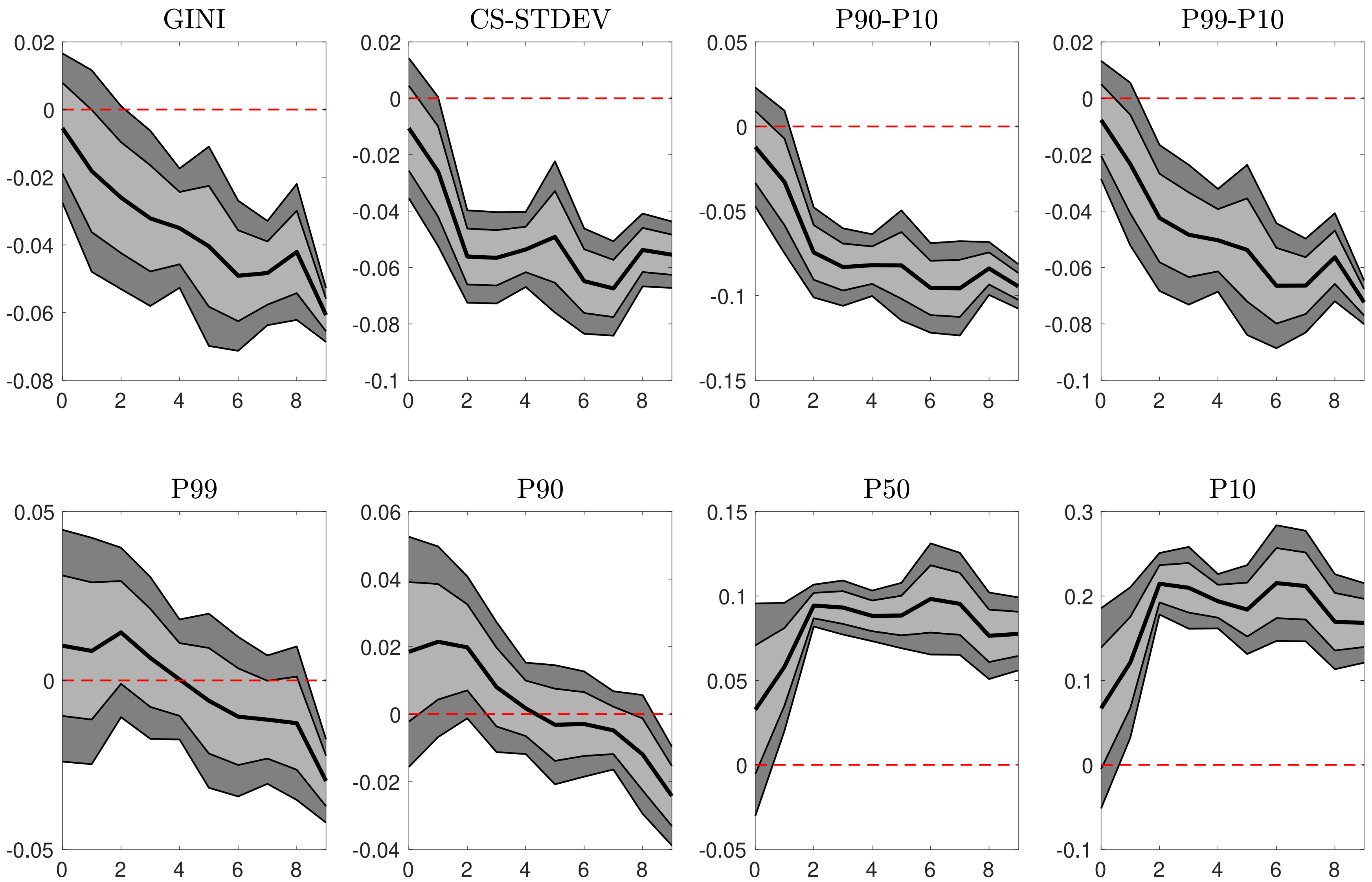
Note: Impulse responses to an asset purchase announcement shock of one standard deviation. The light-shaded and dark-shaded areas represent, respectively, 1 and 1.65 standard deviation confidence bands. The solid line is the OLS estimate.

Figure B.7: Impulse responses from the local projection - Percentiles of the distribution of income



Note: Impulse responses to an asset purchase announcement shock of one standard deviation. The light-shaded and dark-shaded areas represent, respectively, 1 and 1.65 standard deviation confidence bands. The solid line is the OLS estimate.

Figure B.8: Impulse responses from the local projection - Measures of Financial Wealth



Note: Impulse responses to an asset purchase announcement shock of one standard deviation. The light-shaded and dark-shaded areas represent, respectively, 1 and 1.65 standard deviation confidence bands. The solid line is the OLS estimate.

CHAPTER II

THE ASYMMETRIC UNEMPLOYMENT RESPONSE OF NATIVES AND FOREIGNERS TO MIGRATION SHOCKS*

Nicolò Maffei Faccioli[‡] and Eugenia Vella[†]

Abstract: This paper provides new evidence on the macroeconomic effects of net migration shocks in Germany using monthly data from 2006 to 2019 and a variety of identification strategies in a structural vector autoregression. Migration shocks are persistently expansionary, increasing industrial production, per capita net exports and tax revenue. In the labor market, they boost job openings and, after a year and a half, hourly wages in manufacturing. Unemployment falls for natives, driving a decline in total unemployment, while it rises for foreigners. Our analysis disentangles the effects of job-related migration from OECD countries and migration (including refugees) from less advanced economies. Using also quarterly data in a mixed-frequency SVAR, we shed light on the employment and participation responses for natives and foreigners, and also document that migration shocks increase per capita GDP, investment and hourly wages of the aggregate economy. Taken together, our results highlight a job-creation effect for natives and a job-competition effect for foreigners.

JEL classification codes: *C11, C32, E32, F22, F41.*
Keywords: *Migration, unemployment, job creation, job competition, mixed-frequency SVAR.*

Acknowledgements: Eugenia Vella and Nicolò Maffei Faccioli acknowledge financial support from the EU Horizon 2020 Marie Sklodowska-Curie Grant 798015 (EuroCrisisMove) and the La Caixa-Severo Ochoa International Doctoral Fellowship, respectively. Tommaso Bighelli provided research assistance in the preliminary phase of this project. For the mixed-frequency VAR, we have used the toolbox of Fabio Canova and Filippo Ferroni, which is available on their websites. We thank Francesco Furnaletto and Christoph Thoenissen for useful comments and suggestions, as well as participants in CEBRA'S 2020 virtual annual meeting.

[‡]Universitat Autònoma de Barcelona and Barcelona GSE. *e-mail:* nicolo.maffei@barcelonagse.eu.
[†]Corresponding author: University of Sheffield and Fundació MOVE (Markets, Organizations, and Votes in Economics). *e-mail:* e.vella@sheffield.ac.uk

1 INTRODUCTION

What is the macroeconomic impact of migration in the second-largest destination for migrants after the United States? In 2018, approximately 20.8 million people, or equivalently one in four, in Germany had a migrant background according to the microcensus results of the Federal Statistical Office (Destatis).[1] Immigration is a key determinant of changes in labor supply and, currently, the only source of population growth in the German economy, where unemployment has remained remarkably low in recent years.[2] This paper provides new macroeconometric evidence on the expansionary effects of net migration shocks in Germany and the asymmetric unemployment response between natives and foreigners to these shocks.

Understanding better the migration effects in the labor market and the macroeconomy is crucial for migration policy design and can also help to curb the rise in xenophobic movements. While a large literature has analyzed the impact of immigration on employment and wages using disaggregate data, the migration literature in the context of macroeconometric models is still limited due to a lack of data at high frequency over sufficiently long time. Interestingly, such data is available for Germany. Destatis has been collecting monthly data on the arrivals of foreigners (i.e., non-Germans) by country of origin on the basis of population registers at the municipal level since 2006. Registration is obligatory by the registration law of March 2002 ("Melderechtsrahmengesetz") and is necessary to obtain the income tax card required to sign any employment contract or to issue an invoice as self-employed, and also to rent an apartment.

Using monthly data on net migration flows for the period 2006:1-2019:10, we identify net migration shocks in a structural vector autoregression (SVAR) model using the Cholesky decomposition, traditional sign restrictions and mixed (sign and narrative) restrictions. Our analysis places special focus on the response of unemployment, which theoretically is ambiguous, depending on various channels, for instance how fast migrants enter the labor market and whether they do so as employed or as job seekers. In addition, if natives and immigrants are imperfect substitutes in production, increasing inflows exert stronger labor market competition on earlier immigrants than on natives.[3] Furthermore, there is potentially a job-creation effect stemming from the increase in both demand and

[1] A migrant background refers to having at least one parent who did not acquire German citizenship by birth. Naturalization rates have been low so that immigrants' offsprings often do not have the German citizenship.

[2] The strong decline in German unemployment from 2005 onwards and the exceptionally small increase during the Great Recession have been associated with the earlier labor market ("Hartz") reforms.

[3] Natives and immigrants are typically employed in different occupations, which makes them imperfect substitutes in production (Ottaviano and Peri (2012), Manacorda, Manning, and Wadsworth (2012), Llull (2018)). Usually, immigrants (natives) have a comparative advantage in manual-intensive (language-intensive) tasks.

labor supply, which can reduce unemployment.

We find that net migration shocks have expansionary effects in Germany, increasing industrial production, per capita net exports and tax revenue. The response of inflation appears statistically non-significant, but a subsample analysis reveals that this result comes from a combination of a positive response to job-related migration shocks from OECD countries and a negative response to migration shocks from less developed areas of the world, such as Africa. In the former case, a demand effect seems prevalent while in the latter case, where migration is predominantly low-skilled and often political in nature (including refugees), a supply effect prevails. Furthermore, once we include Syrian flows in our baseline net migration variable, we find that migration shocks are disinflationary, thus behaving like typical supply shocks.[4]

In the labor market, migration shocks boost job openings persistently and, after a year and a half, hourly wages in the manufacturing sector, while they reduce unemployment. The increase in vacancies highlights a job-creation effect and is in line with the inverse relation between vacancies and unemployment typically depicted by the Beveridge curve.[5] Interestingly, we uncover asymmetric unemployment responses between natives and foreigners: unemployment falls persistently for natives, driving the response of total unemployment, while it increases significantly after a year for foreigners. This asymmetrical response remains robust when we use for net migration subsamples for OECD and non-OECD countries and also when we examine data on the refugee wave from Syria, originally not included in our migration variable. Moreover, we find that the rise in foreigners' unemployment is reinforced in the case of predominantly low-skilled migration flows from less developed areas of the world, such as Syria or Africa. In the latter case, this increase becomes statistically significant immediately after the impact period. Importantly, the output and unemployment effects of migration shocks continue to hold if we combine our baseline sample with data on net flows from Syria.

The next step is to investigate more in depth the labor market impact of net migration shocks. Sign restrictions schemes in the literature typically restrict job-related immigration shocks to have on impact a positive effect on labor force participation and a negative effect on real hourly wages (see, e.g., Furlanetto and Robstad (2019)). Focusing on a wider notion of migration (including also non-OECD countries of origin), we find that net migration shocks decrease persistently the participation rate, while they increase hourly wages in manufacturing a year and a half after the shock. Immigration can boost productivity and wages over time when firms respond by expanding, investing, adjusting product specialization, adopting efficient technologies, and creating new businesses (Peri (2014)).

[4]For a theoretical analysis on the inflation response to immigration shocks, see, e.g., García and Guerra-Salas (2020).

[5]For the Beveridge curve in Germany, see, e.g., Figure 2 in Iftikhar and Zaharieva (2019).

While the positive response of wages is mainly driven by OECD migration shocks, the negative participation response is driven by flows from non-OECD countries. Generally, immigrants from Africa, Asia and South America, including asylum seekers, do not enter rapidly into the labor force (Furlanetto and Robstad (2019)).

To gain deeper insights, we investigate also quarterly data on participation and employment of natives and foreigners using a mixed-frequency SVAR model. The participation rate of natives increases significantly after approximately two years, while that of foreigners decreases roughly until then, driving the decrease in aggregate participation. Our results also show that the employment rate of natives (foreigners) increases significantly for approximately two years (one year). The positive responses of participation and employment rates for natives imply that the unemployment decrease for this group comes from the boost in employment and not because natives drop out of the labor market. Furthermore, the mixed-frequency SVAR exercise provides evidence that net migration shocks also increase real hourly wages of the aggregate economy as well as per capita GDP and investment, thus confirming the expansionary effects.

Taken together, our results highlight a job-creation effect of migration for natives and a job-competition effect on foreigners. On the one hand, jobs can be created directly by self-employed immigrants or entrepreneurs and indirectly by immigrant innovators (Constant (2014)). Migrants often complement and rarely substitute for natives, choosing locations where they fill shortages by accepting jobs that natives will not do. High-skilled immigrants boost technological adaptation and low-skilled immigrants foster occupational mobility, specialization, and human capital creation. In addition, immigrants raise demand, which increases job openings. On the other hand, the competition effect underscores the need for policymakers to address, through redistribution, challenges faced by foreigners while ensuring that the economy benefits from the expansionary effects of migration.

RELATED LITERATURE. Our paper contributes to the literature on the macroeconomic effects of migration.[6] A strand of this literature has performed steady-state analysis with search models, focusing on the U.S. economy. When natives and illegal immigrants serve as imperfect substitutes in production and compete for jobs in the same market, Liu (2010) shows that immigration increases the unemployment rate due to displacement (job competition) effects on natives (immigrants). If there are two labor markets, skilled natives are insulated from competition and can benefit from the rise in their marginal product of labor, which decreases their unemployment. In Chassamboulli and Palivos (2014), the job-creating response of firms leads to positive employment effects on natives. Under im-

[6] See Vella, Caballé, and Llull (2020) for a recent edited collection.

perfect substitutability between skilled and unskilled inputs, different search cost between natives and immigrants, cross-skill matching, and imperfect transferability of foreign human capital, Liu, Palivos, and Zhang (2017) find that the immigration influx of 2000-2009 reduced the unemployment rate for all workers. Under non-random hiring, Albert (2020) finds that the job-creation effect due to lower wages dominates the competition effect from undocumented immigration, decreasing the unemployment rate and raising wages for natives. Battisti et al. (2018) show that immigration has increased native welfare of both low and high-skilled workers in 20 OECD countries. For Germany, Iftikhar and Zaharieva (2019) find that the 25% immigration increase of 2012–2016 has a negative effect on the welfare of low skill workers in manufacturing whose unemployment rises due to competition, while all other workers gain. Our paper empirically validates with German data the job-competition effect for foreigners and the employment gains for natives.

Another body of literature has used SVAR models for empirical analysis (see, e.g., Furlanetto and Robstad (2019), Smith and Thoenissen (2019), Kiguchi and Mountford (2019)), d'Albis, Boubtane, and Coulibaly (2016), d'Albis, Boubtane, and Coulibaly (2019), and Schiman (2018)).[7] These studies typically highlight the expansionary effects of migration for the receiving economy, except for Smith and Thoenissen (2019) who find no statistically significant effect on per capita GDP and Kiguchi and Mountford (2019) who show a temporary reduction. Regarding the response of unemployment, Furlanetto and Robstad (2019) find it to be negative both for natives and foreigners in Norway, d'Albis, Boubtane, and Coulibaly (2016) show that family immigration reduces France's unemployment rate, and d'Albis, Boubtane, and Coulibaly (2019) find that the unemployment rate falls by 0.1 percentage points the year of the migration shock and for two years after the shock. On the other hand, Kiguchi and Mountford (2019) and Schiman (2018) show that unemployment increases after an immigration shock in the United States and Austria, respectively.[8] To the best of our knowledge, the only previous macroeconometric study examining the unemployment responses of natives and foreigners is Furlanetto and Robstad (2019) without finding job-competition effects. Providing evidence on the participation responses of natives and foreigners is another novel aspect of our paper.

Our aggregate time-series approach is also complementary to the large immigration literature with a microeconometric focus on the basis of disaggregate data, pioneered

[7]Business cycle analysis in migration models without labor frictions can be found in Mandelman and Zlate (2012), Smith and Thoenissen (2019), and Burriel, Fernández-Villaverde, and Rubio-Ramırez (2010). In the context of dynamic stochastic general equilibrium models with labor frictions, Lozej (2019) finds that immigration raises unemployment temporarily, but the effect may subsequently switch sign if wages fall sufficiently so that firms post vacancies. Kiguchi and Mountford (2019) stress that a temporary rise in unemployment may occur if immigration causes non-participants to enter the pool of job seekers.

[8]Smith and Thoenissen (2019) do not consider unemployment in their analysis.

by the seminal work of George Borjas (see, e.g., Borjas (2014)). Different approaches typically yield different results, but there is some consensus that some groups of native workers benefit through their relative complementarity with immigrants. Ottaviano and Peri (2012) and Manacorda, Manning, and Wadsworth (2012) provide evidence of the role of imperfect elasticity of substitution between natives and immigrants in the size of wage effects. Our paper is also related to studies exploring the differential impact of immigration on natives and foreigners specifically in Germany. Using administrative data and a labor-market equilibrium model, D'Amuri, Ottaviano, and Peri (2010) show that the immigration of the 1990s had little adverse effects on native wages and employment. Instead, it had adverse employment and wage effects on previous immigrants, driven by a higher degree of substitution between old and new immigrants and by wage rigidities. Felbermayr, Geis, and Kohler (2010) estimate a structural model of labor demand with annual survey data from the German Socio-Economic Panel. In a counterfactual exercise without restrictions for migration from new EU members before 2011, they find adverse wage and unemployment effects for incumbent foreigners, but positive effects for natives. Finally, using a difference-in-difference approach to assess the effect of a policy-induced migrant labor supply shock in 2016, Scharfbillig and Weißler (2019) find a negative employment effect only for other foreign residents. In a comprehensive framework allowing also to assess the expansionary effects of migration shocks, our paper provides new empirical evidence on the participation and (un)employment responses of natives and foreigners, while abstracting from wage effects for each group separately because of data limitations at high frequency.

STRUCTURE. The rest of the paper is organized as follows. Section 2 lays out the data and econometric model, and Section 3 discusses our baseline empirical findings. Section 4 performs a subsample analysis for various geographical origins of migrants and also examines the impact of the recent refugee wave from Syria. Section 5 discusses the methodology and the results of a mixed-frequency SVAR. Finally, Section 6 concludes.

2 EMPIRICAL METHODOLOGY

In this section, we first describe the monthly data on net migration flows in Germany. Then, we present the details of the econometric model and the identification strategy.

2.1 MONTHLY DATA ON NET MIGRATION FLOWS

Since January 2006, Destatis has been collecting monthly data on the arrivals of foreigners by country of origin, defined as the country of last residence, on the basis of population registers at the municipal level. All geographical continents are covered (Europe, Asia, Australia and Oceania, America, and Africa). The exact list of countries is presented in the Appendix. The municipalities have a strong incentive to record new residents since their fiscal revenue depends on the number of registered, while they impose penalties on non-compliants with the mandatory registration. The difference between the numbers of arrivals and departures (de-registrations) produces the net migration figures, also available from Destatis.

Figure 1 shows the evolution of the net migration rate in Germany by various geographical origins over our sample period 2006:1-2019:10. The net migration rate is computed as the ratio of inflows minus outflows of non-Germans to the working-age population, multiplied by a thousand.[9] We observe a large increase during the period under study. Specifically, the total net migration rate (cyan line) rises from close to 0% in 2009 to 0.4% in 2014 and peaks at more than 1.8% with the refugee crisis in 2016. Notably, this significant increase is observed even if we exclude Syrian flows, which explain the bulk of the 2015-2016 spike (green line). We refer to this measure that is net of the Syrian flows as the baseline migration rate (blue line). Moreover, EU migration (orange line) is a key contributor to the rise in the net migration rate during the European sovereign debt crisis of 2009-2014. The surge is also certainly related to the Eastern enlargement of the EU.[10] Net migration flows from OECD countries are of smaller magnitude than those from the EU member states due to negative values mainly for Canada, the U.S., Australia, and Japan in various years. The net migration rate from Africa peaks in 2016 at around 0.1%. Finally, between 2016 and 2018 the total net migration rate fluctuates between 0.4% and 0.6% and after 2018 it tends to get stabilized close to 0.4%, which is higher than the level at the start of our sample.

We conduct below an in-depth empirical analysis to study the effects of the sizeable increase in net migration on the labor market and the macroeconomy in Germany. For the main analysis, we use the baseline migration rate (blue line), thus leaving Syrian flows out of our sample. We examine the effects of those flows in Section 4.

2.2 Econometric Model and Identification

We consider the following reduced-form VAR(p) model:

$$Y_t = C + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + u_t \qquad (1)$$

[9]The data is seasonally adjusted with JDemetra+ X13, consistently with Destatis.

[10]In 2011 free mobility started for the EU8 countries (Czech Republic, Estonia, Latvia, Lithuania, Hungary, Poland, Slovenia and Slovakia), which joined the EU in 2004, and in 2014 for Romania and Bulgaria, which joined in 2007.

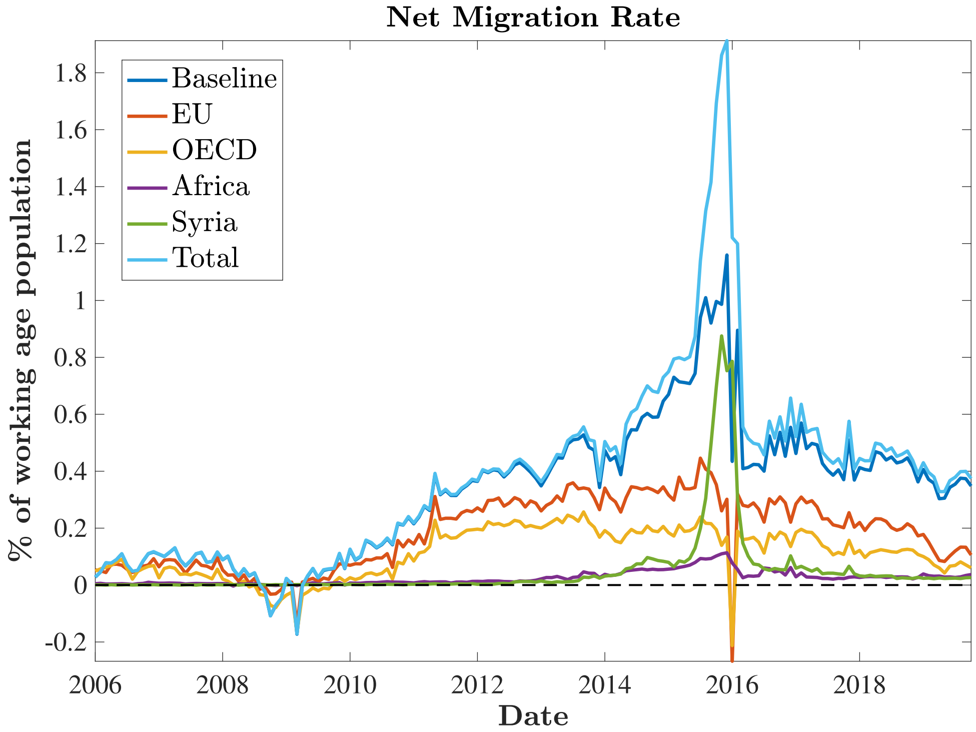


Figure 1: Net Migration Rate in Germany by Geographical Origin, 2006-2019

Note: The baseline migration rate excludes Syrian flows total net flows. EU migration refers to the EU-28 excluding Germany, thus covering 27 countries. From the group of OECD countries we exclude Chile, Colombia, and Mexico. Data comes from the Federal Statistical Office (Destatis).

where Y_t is a n x 1 vector containing n endogenous variables, C is a n x 1 vector of constants, $A_1, ..., A_p$ are n x n matrices of coefficients associated with the p lags of the dependent variable and $u_t \sim N(0_n, \Omega)$ is the reduced-form residual. In the baseline model, Y_t contains three variables in the following order: the net migration rate, the logarithm of the industrial production index, and the registered unemployment rate.[11] [12] To disentangle the unemployment responses of natives and foreigners, we also run the SVAR specification by replacing the registered unemployment rate with the unemployment of natives and foreigners. For these two variables, we take the ratios of the number of native and of foreigners who are registered unemployed to the economically active population (participants).

In further exercises, we add, one at a time, and order last in the system the following

[11]This is defined as the share of registered unemployed in the economically active population. The latter is computed as the sum of the number of residents in Germany who are in employment (from Destatis) and the number of registered unemployed (from the Federal Employment Agency - "Bundesagentur für Arbeit"). The industrial production index refers to the following sectors: mining and quarrying, manufacturing, energy and construction. Series in logs are multiplied by 100.

[12]We also consider quarterly data on per capita GDP in Section 5 where we use a mixed-frequency VAR model.

8 variables using data from Destatis: population[13], number of labor market participants, labor force participation rate, number of registered vacancies, real hourly wages in the manufacturing sector, real labor income tax revenue per capita, real tax revenues of the Federation per capita, real net exports per capita and the CPI index. We also perform an exercise in which we substitute the unemployment rate with the employment rate. All these variables enter in a logarithmic form except for the participation and employment rates.

The model is estimated using Bayesian methods with a flat prior such that the information in the likelihood is dominant. We use 3 lags of the dependent variable, which is the average of the AIC, BIC and HQC criteria. We also use alternative lags specifications as a robustness check in the Appendix (see Figure A.1). Let the mapping between reduced-form and structural disturbances be $u_t = S\epsilon_t$, where $\epsilon_t \sim N(0_n, I_n)$ is the n x 1 vector of unit variance structural disturbances. The model is structural in that the errors ϵ_t are mutually uncorrelated and have an economic interpretation. In the baseline specification, we define S as the Cholesky decomposition of Ω, thus as the unique lower triangular matrix such that $SS' = \Omega$, and give an economic interpretation to the first shock only (see, e.g., d'Albis, Boubtane, and Coulibaly (2016)).

We interpret the migration shock as the only one that has a contemporaneous effect on the net migration rate. Other shocks in the receiving country, which we we call "residual shocks" without giving a formal interpretation, such as business cycle or domestic labor supply shocks, affect net migration with a lag. While this assumption could be easily contested if we worked with annual or quarterly data, this is not the case with monthly data. The reason is simple: migration decisions motivated by positive conditions in the receiving country take some time to materialize in the statistics and, arguably, one month may be thought of as a lower-bound estimate of the period required. Let us provide an intuitive example. Suppose that someone decides to move to Germany because of current favorable economic developments there. It would certainly take some time before first acknowledging these developments, then taking the decision to move, start looking for a job and temporary accommodation, and finally registering with the authorities to be able to sign the employment contract and move to more permanent accommodation. It is difficult to argue that this process would take less than a month and will be even longer for those in need of a VISA. The reverse of this example can be applied to those leaving Germany.

As alternative identification strategy, we consider both traditional sign restrictions and more recently proposed mixed (sign and narrative) restrictions. A SVAR approach based on sign restrictions allows to disentangle the exogenous and the endogenous component

[13] This variable is interpolated from quarterly data.

of immigration in a system that fully takes into account feedback effects between different variables. Therefore, it addresses potential concerns about the response of immigration to the host country's economic conditions. Recently, Antolín-Díaz and Rubio-Ramírez (2018) developed a methodology which allows to impose that around selected historical events structural shocks and/or historical decompositions agree with some narrative information. For example, it is possible to impose that around a quarter (or several quarters) one specific shock has to be positive (or negative) or that this shock has to be the main (or the least important) driver of a variable or more (less) important than all the other shocks combined for a specific variable. The first is a restriction on the sign of the structural shocks. The second and the third are restrictions on the historical decompositions. Antolín-Díaz and Rubio-Ramírez (2018) focus on oil and monetary shocks and show that imposing only a few narrative sign restrictions may sharpen and even change the inference of a SVAR originally identified via traditional sign restrictions. Our model with net migration shocks constitutes an appropriate setup to incorporate narrative sign restrictions, previously employed in the study of immigration shocks by Furlanetto and Robstad (2019).

Table 1 reports our signs restrictions. We restrict the migration shock to have the highest positive impact response on net migration and a positive effect on industrial production. Alternatively, we make a joint use of the sign restriction that there is a positive effect on both industrial production and the net migration rate and of the narrative restriction that over the period 2014-2016 immigration is the biggest contributor to the net migration rate (see Figure 1). Notably, the response of unemployment after migration shocks, which is ambiguous as explained in the Introduction, is left unrestricted. The restrictions are imposed only on impact following Canova and Paustian (2011) and are implemented using the algorithm of Rubio-Ramírez, Waggoner, and Zha (2010).[14]

Table 1: Alternative identification strategy

	Sign Restrictions		
	Immigration Shock	Residual Shock 1	Residual Shock 2
Net migration rate	+	$\|b_{21}\| < b_{11}$	$\|b_{31}\| < b_{11}$
Industrial production	+	+	+
Unemployment rate	/	-	+

Note: b_{i1} denotes the impact response of variable i to a net migration shock.

[14]Results (available upon request) are very similar to the ones obtained with Cholesky if we impose instead that the migration shock explains the bulk of the variance decomposition of net migration in the first three months.

3 RESULTS

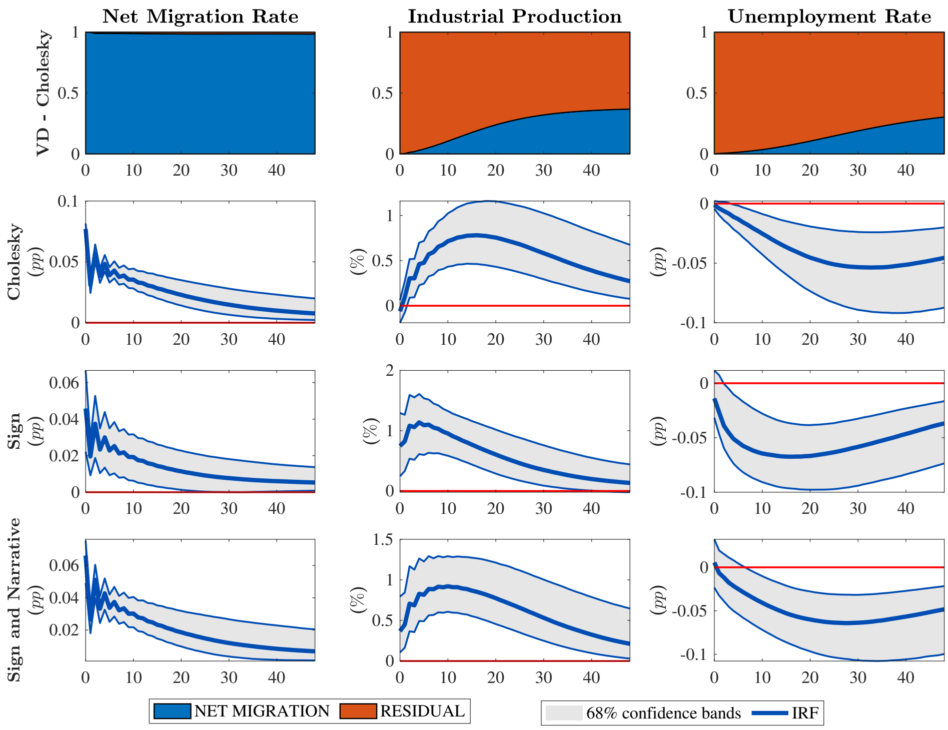
In this section, we present impulse response functions to one-standard-deviation net migration shocks. The continuous lines represent the posterior median at each horizon and the shaded areas indicate the 68th posterior probability region of the estimated impulse responses. The horizontal axis refers to time periods, measured by months.

3.1 THE EXPANSIONARY EFFECTS OF NET MIGRATION SHOCKS

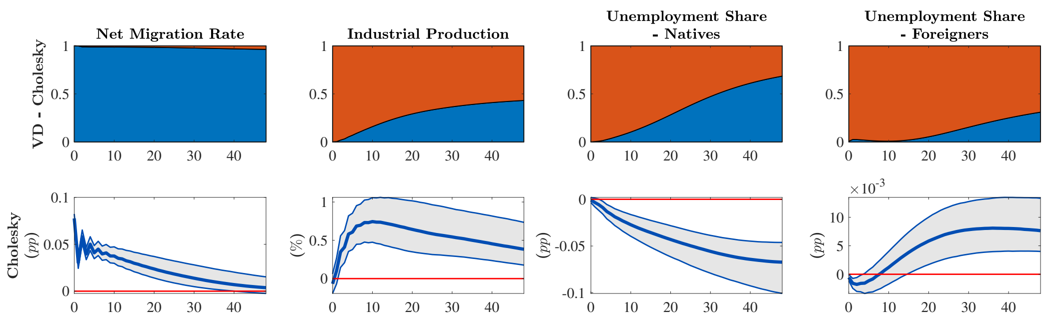
The second row of Figure 2a shows that a positive net migration shock increases persistently the net migration rate. The effects on the German economy are clearly expansionary as industrial production increases significantly after the first bimester and the total unemployment rate decreases significantly after the first quadrimester. These effects appear quantitatively important. The first row of Figure 2a shows that the migration shock explains entirely monthly fluctuations of the net migration rate. Regarding industrial production and unemployment, the migration shock explains around 40% and 25% respectively. Unsurprisingly, the other shocks in the system account for the bulk of fluctuations in industrial production and unemployment. These results are robust to different lag specifications, harmonized - instead of registered - unemployment, and different ordering of the net migration variable (see Figure A.1 in the Appendix).

In the third and fourth rows of Figure 2a we present impulse responses when the shocks are identified using sign restrictions. Specifically, the third row reports responses when we restrict the migration shock to be the one with the highest positive impact response of net migration and a positive effect on output. Recall that we leave the response of the unemployment rate to net migration shocks unrestricted. The fourth row reports responses when we make a joint use of the sign restriction of Table 1 and the narrative restriction that over the period 2014-2016 immigration is the biggest contributor to the net migration rate, as shown in Figure 1. In both cases, the persistent and expansionary effects of the net migration shock, namely the rise in industrial production and the fall in the unemployment rate, are confirmed. The responses exhibit very similar dynamics to the ones of the baseline framework, despite the use of a minimal amount of sign restrictions. We thus feel confident to use our baseline Cholesky approach to assess next the effects of net migration shocks on a variety of macroeconomic and labor market variables.

Figure 2: Impulse response functions to an one-standard-deviation net migration shock



(a) Different identification strategies in the SVAR with total unemployment



(b) Cholesky SVAR with unemployment of natives and foreigners

Note: The continuous lines represent the posterior median at each horizon and the shaded areas indicate the 68th posterior probability region of the estimated impulse responses. The horizontal axis refers to time periods, measured by months. VD denotes variance decomposition.

3.2 Asymmetric Unemployment Responses of Natives and Foreigners

Figure 2b shows the results when we augment our Cholesky SVAR with the unemployment shares of natives and foreigners in the labor force, which replace total unemployment in the baseline specification. The responses we obtain are asymmetric: the unemployment share of natives decreases significantly and persistently after the first two months, while the unemployment share of foreigners increases after slightly more than a year. In terms of magnitude, unemployment responds more strongly in the case of natives than of foreigners.

On the one hand, these results highlight competition effects from newly settled migrants on earlier migrants. This dynamic competition channel has been little analyzed until now as the literature has largely focused on the effects of immigration on natives. On the other hand, net migration shocks have largely beneficial effects in terms of unemployment for native workers, thus not confirming possible displacement effects.[15] As emphasized in the Introduction, migrants often complement and rarely substitute for native workers.

We check if these findings remain robust when we use subsamples on migration flows from OECD and nob-OECD countries and on the refugee wave from Syria in Section 4. We also investigate further the labor market responses of natives and foreigners in Section 5.

3.3 Other Key Macroeconomic and Labor Market Variables

Immigrant workers are over-represented in manufacturing and construction jobs in Germany (Iftikhar and Zaharieva (2019)), which is the largest manufacturing economy in Europe and one of the world's major manufacturing powerhouses.[16] In this section, we augment our baseline SVAR with real hourly wages in the manufacturing sector, for which monthly data is available, and also with other key variables listed in Section 2.2 (one at a time). The goal is to investigate the impact of net migration on labor supply, labor demand, hourly wages and inflation.

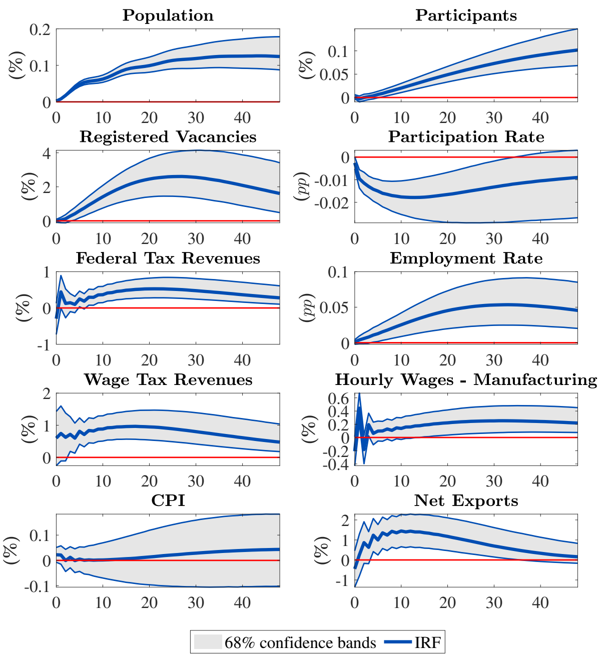
Figure 3a presents the impulse response functions, while variance decompositions are included in Figure 3b. The net migration shock increases persistently labor demand (vacancies), the employment rate, and also real hourly wages after around 18 months. The positive response of vacancies highlights a job-creation effect of migration and is in line with the inverse relation between vacancies and unemployment depicted by the Beveridge curve (see, e.g., Figure 2 in Iftikhar and Zaharieva (2019)). For inflation we do not find a statistically significant effect. Subsample analysis presented in the next section sheds light on this result.

Turning to labor supply effects, the shock leads to a protracted increase in the pool

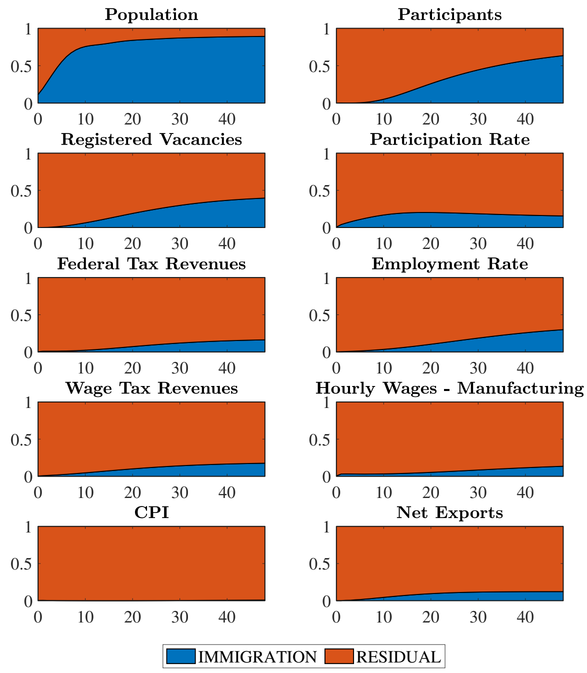
[15] Figure A.2a in the Appendix shows results when we break down the pool of unemployed into natives and foreigners. We observe a decline for natives and an increase for foreigners in line with Figure 2b. The total pool of unemployed decreases in line with Figure 2a.

[16] Germany has an exceptionally large employment share in manufacturing (around 25% in 2014).

Figure 3: Monthly baseline SVAR with additional variables



(a) Impulse response functions



(b) Variance decomposition

Note: The continuous lines represent the posterior median at each horizon and the shaded areas indicate the 68th posterior probability region of the estimated impulse responses. The horizontal axis refers to time periods, measured by months.

of labor force participants 5 months after the shock, which is outweighed though by a higher rise in population, resulting in a short-run decrease in the participation rate. This result is in contrast to the typical association of job-related migration shocks with an increase in participation (see, e.g., Furlanetto and Robstad (2019)). To understand better this result, we will conduct a subsample analysis in Section 4 and we will also investigate, through a mixed-frequency SVAR approach, quarterly data on participation of natives and foreigners in Section 5.

Given the positive impact on employment and wages, labor income tax revenue rises significantly a couple of months after the shock. The response of federal total tax revenue also appears positive 8 months after the shock. Regarding the impact on international trade, the net migration shock raises significantly net exports for more than 2 years (see Figure 3a). These findings further corroborate the expansionary macroeconomic effects of net migration in Germany.

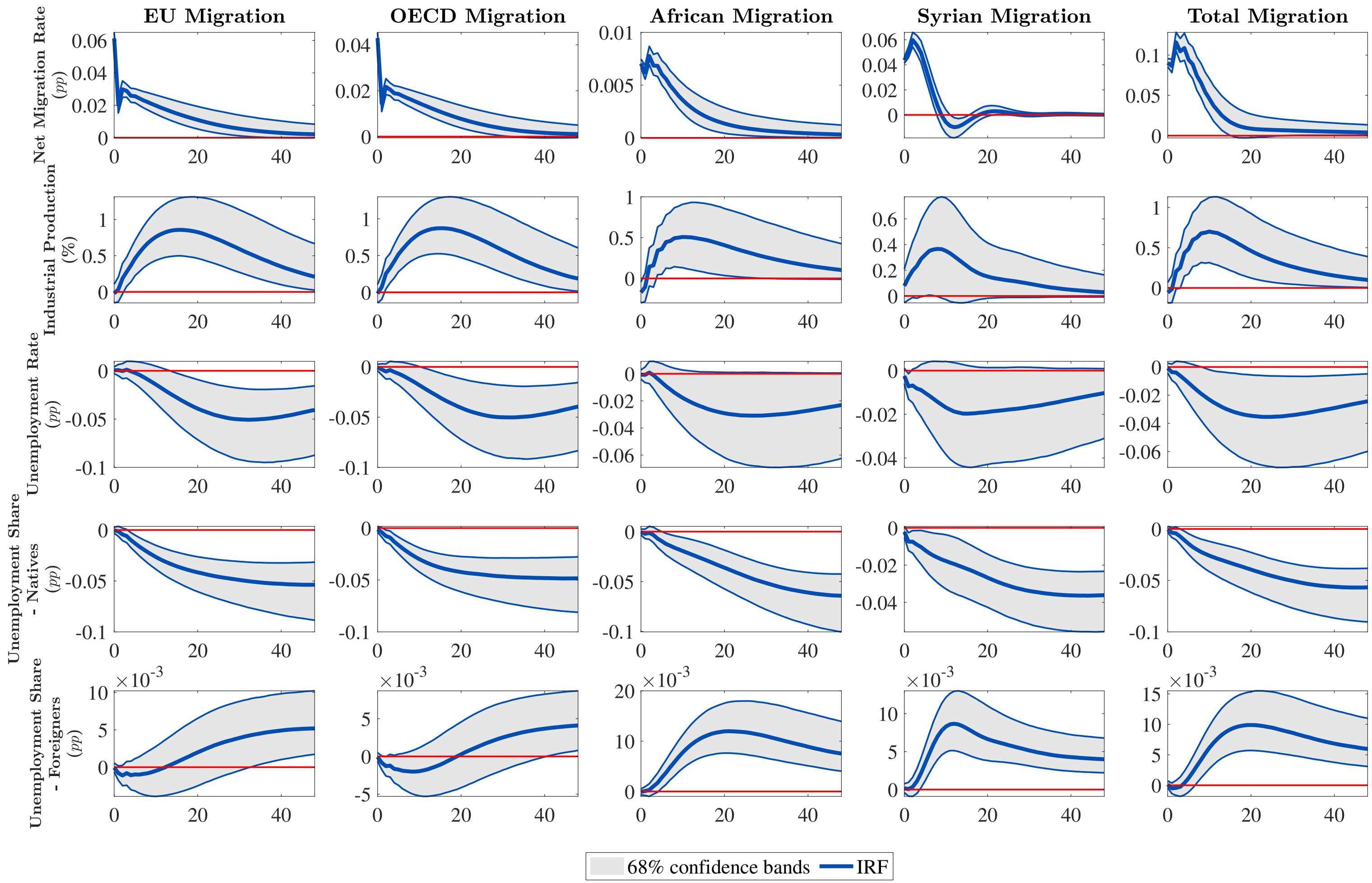
Finally, the variance decomposition in Figure 3b reveals that the net migration shock is the major driver of fluctuations in population over the entire horizon considered. This finding confirms that immigration is the main source of population growth in our sample. The effects are non-negligible for other variables, too. Net migration explains a large share of the variance of participants and vacancies, approximately 63% and 40% respectively, and a non-negligible share for the other variables, with the exception of the CPI index. Altogether, these findings stress the role of net migration in macroeconomic and labor market volatility.

4 Geographical Origins of Migrants and the Refugee Wave

Empirical evidence suggests that the average education level of immigrants from developed and developing countries differs. In addition, so far we have not used data on the wave of predominantly low-skilled refugees from Syria, which increased immigration flows in Germany to about one million people in 2015 - 2016 (see also Figure 1). In this section, we study the effects of net migration shocks accounting for the geographical origin and the impact of refugee migration. To this end, we estimate the SVARs of Figure 2a (second row), Figure 2b and Figure 3 by changing the first variable to the net migration rate originating from the region of interest, namely EU countries, OECD countries, Africa, and Syria.[17] We also show findings when we add Syrian flows to the baseline

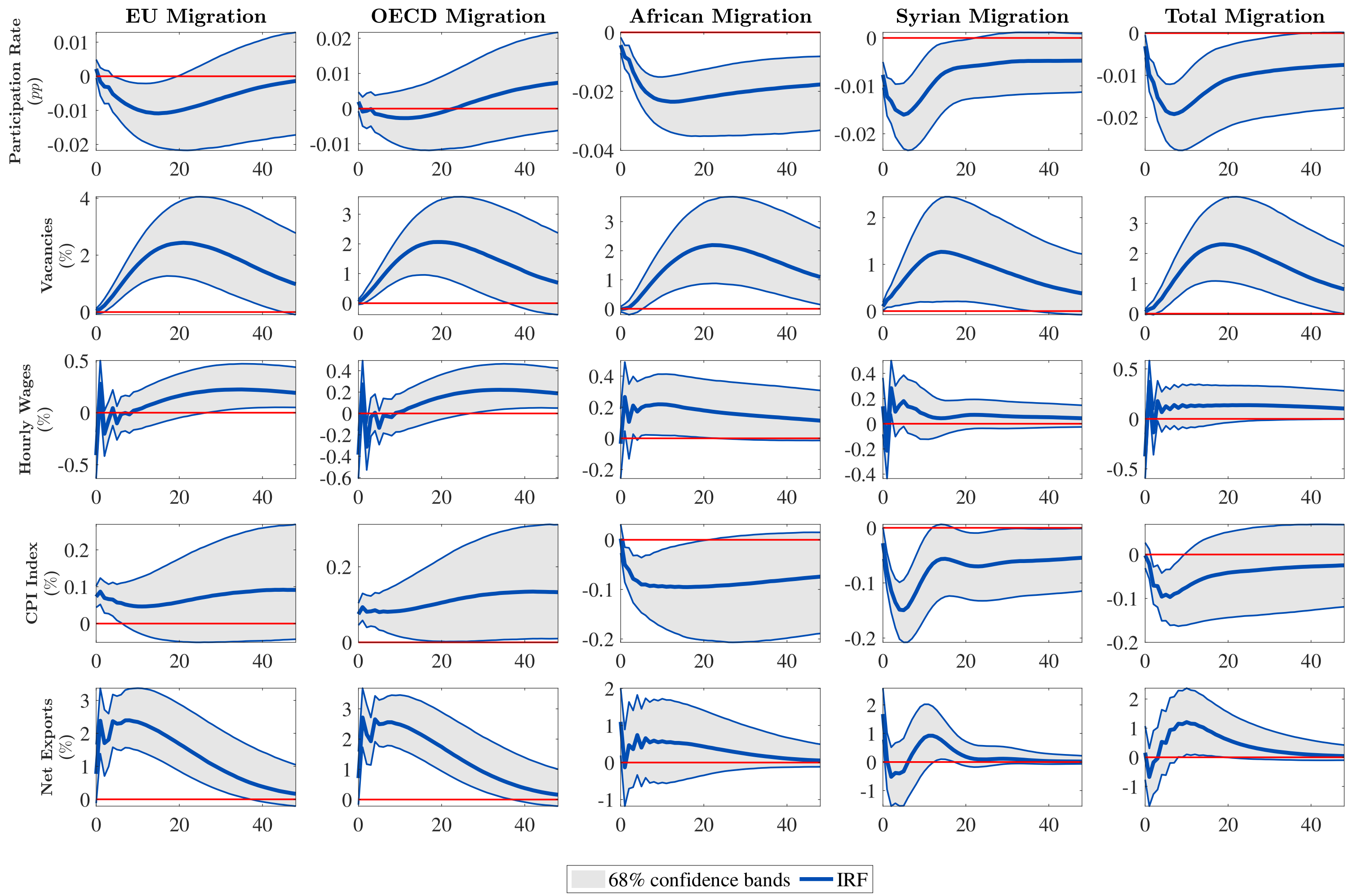
[17] Results for net migration flows from Asia are mainly driven by Syria and therefore look very similar. Net flows from South America were found to be little relevant for our analysis. Both sets of results are available upon request.

Figure 4: Subsample analysis and the refugee wave: Output and unemployment variables



Note: Total migration results from augmenting the net migration variable used in the baseline SVAR with net flows from Syria. The continuous lines represent the posterior median at each horizon and the shaded areas indicate the 68th posterior probability region of the estimated impulse responses. The horizontal axis refers to time periods, measured by months.

Figure 5: Subsample analysis and the refugee wave: Selected macroeconomic and labor market variables



Note: Total migration results from augmenting the net migration variable used in the baseline SVAR with net flows from Syria. The continuous lines represent the posterior median at each horizon and the shaded areas indicate the 68th posterior probability region of the estimated impulse responses. The horizontal axis refers to time periods, measured by months.

migration variable used until now.

Figure 4 shows responses for the net migration rate, industrial production, total unemployment rate, and the unemployment shares of natives and foreigners. Results remain qualitatively unchanged in all cases. Net migration shocks from EU and OECD countries have very similar effects (columns 1 and 2). In the case of net migration from Africa (column 3), the negative response of total unemployment is no longer statistically significant, while the increase of foreigners' unemployment becomes statistically significant shortly after the impact period and stronger in magnitude compared to both columns 1 and 2 and to Figure 2b. When our net migration variable includes only Syrian flows (column 4), the effects remain qualitatively the same, but loose statistical significance for industrial production and total unemployment, while they do maintain it for natives' and foreigners' unemployment. Importantly, all our results continue to hold and are statistically significant when we combine data on Syrian flows with our baseline data sample for net migration (column 5).

Figure 5 presents responses for the participation rate, vacancies, hourly wages in manufacturing, inflation, and net exports.[18] The negative response of the participation rate in Figure 3a is mainly driven by non-OECD economies (Africa or Syria), whereas the response is not statistically significant for net migration from OECD countries and is negative and significant only for nearly the first half of the time horizon for net migration from EU countries. As mentioned in Furlanetto and Robstad (2019), immigrants from Africa, Asia and South America are mostly those who do not enter rapidly into the labor force (as is the case for asylum seekers, for example). The positive response of vacancies is very robust in all cases shown in Figure 5. The positive response of hourly wages in the second half of the time horizon in Figure 3a is confirmed here for net migration from OECD (and EU) countries. The response of the CPI index is also insightful. Recall that the inflation response in Figure 3a was not statistically significant. The subsample analysis shows that this is explained by a positive response to OECD (and EU) migration shocks and a negative response to African and Syrian migration shocks. This finding suggests that for the former a demand effect is prevalent, while for the latter a supply effect. Once we include Syrian flows in our baseline net migration variable (column 5), we observe that migration shocks appear to decrease the CPI index, thus behaving like typical supply shocks. The response of net exports is also interesting. While net exports increase on impact and for a couple of months in the case of OECD (and EU) migration shocks, the response is substantially more muted if we examine African or Syrian migration.

Overall, the analysis in this section disentangles the effects of job-related migration

[18]Responses for population, participants, federal and wage tax revenues are included in Figure A.3 of the Appendix.

from OECD countries and predominantly low-skilled migration (including refugees) from less developed economies. This distinction matters qualitatively for the effects on inflation and quantitatively for participation, hourly wages, foreigners' unemployment and net exports.

5 Deeper Insights from a Mixed-Frequency SVAR

So far, we have shown that the participation rate falls after net migration shocks, but we have not examined the responses of natives and foreigners separately. Since data on participation (and employment) by nationality is available quarterly, in this section we proceed with a mixed-frequency SVAR. This approach allows us further to explore quarterly data on per capita consumption, investment, GDP, and real hourly wages for the aggregate economy.

5.1 The Model and Data

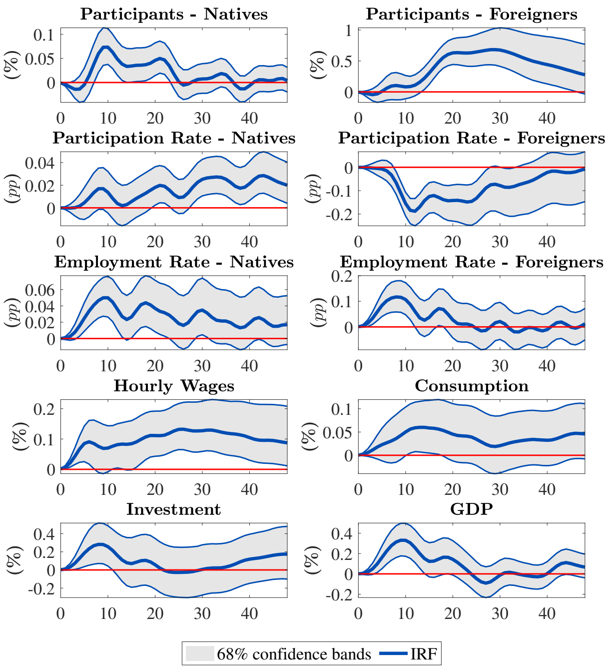
The main advantage of the mixed-frequency SVAR model is that we can assess the effects of net migration shocks on variables for which data is available at quarterly but not monthly frequency, while keeping our identifying restrictions unchanged. Estimation is carried along the lines of Schorfheide and Song (2015).

We complement the three variables of our baseline Cholesky SVAR (net migration rate, industrial production, unemployment rate) with (a) the participation rate and the logarithm of employed workers, as they convey relevant information to properly estimate the model, and (b) the following quarterly variables of interest (one at a time): participation rate of natives, participation rate of foreigners, the number of participants natives (in logs), the number of participants foreigners (in logs), the employment rate of natives, the employment rate of foreigners, real hourly wages for the total economy, per capita real consumption, per capita real investment, and per capita real GDP. We specify a flat rather than Minnesota prior in line with our monthly SVAR model and we include 6 lags of the dependent variable to ensure enough feedback.[19]

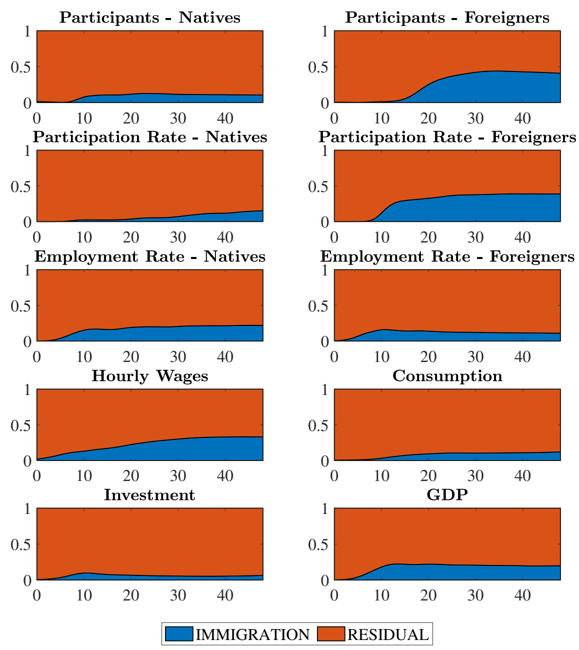
Data by nationality on the number of employed is available from the Eurostat's Labor Force Survey (LFS). By taking the sum of the numbers of employed and registered unemployed for natives and foreigners, we obtain a measure of native and non-native

[19] This might cause some impulse responses in 6a to gyrate. With a smaller number of lags, impulse responses are less robust especially if we exclude variables from the system. Results with 6 lags, instead, are robust even without participation and employed workers in the specification.

Figure 6: Mixed-Frequency SVAR



(a) Impulse response functions



(b) Variance decomposition

Note: The continuous lines represent the posterior median at each horizon and the shaded areas indicate the 68th posterior probability region of the estimated impulse responses. The horizontal axis refers to time periods, measured by months.

participants, respectively. We construct the employment rate by nationality by taking the corresponding ratio of the number of employed to the number of participants. Since population data distinguishing Germans and non-Germans is not available quarterly to compute participation rates by nationality, we rely again on the LFS data. Real hourly wages are defined as hourly gross wages and salaries from Destatis, deflated by the CPI index. The remaining macroeconomic variables (consumption, investment and GDP) are taken from the Destatis and FRED databases.

5.2 Participation and Employment Responses of Natives and Foreigners

Figure 6a shows impulse responses for the quarterly variables to a net migration shock. The participation rate of natives increases significantly after approximately two years, while that of foreigners decreases roughly until then, driving the decrease in the aggregate participation rate over the same period in Figure 3a. This result suggests that newly settled migrants enter the labor market only gradually, which can explain why it takes time for foreigners' unemployment to increase significantly in Figure 2b. As a result, the immediate rise in population outweighs the rise in total participants (Figure 3a).

Figure 6a also shows that the employment rate of natives (foreigners) increases significantly for approximately two years (one year). For natives, this increase coincides with an increase in participants. The participation and employment responses imply that the unemployment decrease for natives in Figure 2b is due to a boost in their employment following the net migration shock and not because natives respond by dropping out of the labor market. Possible displacement effects of net migration on natives are not found here.

For foreigners, the magnitude of the increase in the employment rate is roughly double compared to natives.[20] Over the same period, the number of foreigners' participants does not move, which matches well the initially insignificant response of foreigners' unemployment share in Figure 2b. The increase in the pool of foreigners participants a year after the shock is roughly ten times higher compared to natives, marking the gradual integration of newly settled immigrants in the labor market. As a result, foreigners' participation rate returns now to its pre-shock level after declining during the first year after the shock. This leads to stronger competition for jobs and higher unemployment among foreigners (see Figure 2b).

[20]Figure A.2a in the Appendix shows that the response of employed natives appears smaller in magnitude than that of foreigners, while Figure A.2b shows that the responses of participants and employment rate for foreigners and natives do not change if we use unemployment data from the LFS instead of the Federal Employment Agency.

In terms of variance decomposition, Figure 6b shows that net migration is an important driver of fluctuations in participation for foreigners, but less relevant for natives. The aggregate importance of net migration for participation is thus almost entirely driven by foreigners.

5.3 Responses for Aggregate Wages, Consumption, Investment and GDP

Figure 6a shows a significant and protracted increase in real hourly wages of the aggregate economy. Together with the positive response of hourly wages in the manufacturing sector in Figure 3a, our results indicate that, on average, net migration does not depress but, instead, boosts wages in Germany. The response of per capita investment is also statistically significant and positive for almost a year after the shock. A similar result, with higher statistical significance, is obtained for per capita GDP.[21] The response of per capita consumption is also positive but slightly significant (i.e., from month 10 to 17-18). The results from the variance decomposition in Figure 6b show that net migration shocks contribute to variability in hourly wages and per capita GDP, while they explain a small share of fluctuations in per capita consumption and investment. Overall, these results add to the positive responses of industrial production and per capita net exports and tax revenue from the monthly SVAR model, further confirming the expansionary effects of migration shocks in Germany.

6 Conclusion

The recession induced by the COVID-19 pandemic is expected to have long, deep, and pervasive consequences which can lead to an increase in migration flows in the most resilient economies. Considering also that migrants have been among the most vulnerable groups to infection, the sanitary and economic crisis could exacerbate xenophobic sentiments around the world. This paper contributes to a better understanding of the migration effects in the labor market and the macroeconomy, which is crucial both for migration policy design and for the effort required to curb the rise in xenophobic movements.

Using monthly and mixed-frequency SVAR models applied to data for Germany, we show that net migration shocks are expansionary, increasing persistently per capita output, investment, net exports and tax revenue. Shocks to net migration from OECD countries appear to be inflationary, reflecting demand forces, while the opposite is true for non-OECD migration shocks, for which supply-side effects are mostly prevalent.

[21] Figure A.4 in the Appendix plots the estimates of monthly GDP from the mixed-frequency SVAR.

In the labor market, migration shocks increase (over time) real hourly wages. The participation rate of foreigners decreases, driving a decrease in the aggregate participation rate, while that of natives increases. Migration shocks also boost new hirings and decrease the unemployment rate, in line with the inverse relation depicted by the Beveridge curve. Interestingly, we uncover asymmetric responses for the unemployment of natives and foreigners. Natives' unemployment decreases persistently, driving the response of total unemployment, while foreigners' unemployment increases. Newly settled migrants mainly from non-OECD countries enter the labor market at relatively slow pace. Our results highlight a job-creation effect of migration for natives and a job-competition effect for foreigners, which can be rationalized by imperfect substitutability of the two labor inputs in production. Job competition effects for foreigners are stronger in the case of predominantly low-skilled migration from non-OECD countries. With respect to wage effects for natives and foreigners, our study is unfortunately limited by the lack of data at high frequency. We leave theoretical investigations as a topic for future research.

References

Albert, Christoph (2020). "The Labor Market Impact of Immigration: Job Creation vs. Job Competition". In: *American Economic Journal: Macroeconomics* forthcoming.

Antolín-Díaz, Juan and Juan F Rubio-Ramírez (2018). "Narrative sign restrictions for SVARs". In: *American Economic Review* 108(10), pp. 2802–29.

Battisti, Michele et al. (2018). "Immigration, search frictions and redistribution: A quantitative welfare analysis". In: *Journal of the European Economic Association* 16, pp. 1137–1188.

Borjas, George J (2014). *Immigration Economics*. Harvard University Press.

Burriel, Pablo, Jesús Fernández-Villaverde, and Juan F Rubio-Ramırez (2010). "MEDEA: a DSGE model for the Spanish economy". In: *SERIEs* 1(1-2), pp. 175–243.

Canova, Fabio and Matthias Paustian (2011). "Business cycle measurement with some theory". In: *Journal of Monetary Economics* 58(4), pp. 345–361.

Chassamboulli, Andri and Theodore Palivos (2014). "A Search-equilibrium approach to the effects of immigration on labor market outcomes". In: *International Economic Review* 55, pp. 111–129.

Constant, Amelie F (2014). "Do migrants take the jobs of native workers?" In: *IZA World of Labor*.

d'Albis, Hippolyte, Ekrame Boubtane, and Dramane Coulibaly (2016). "Immigration policy and macroeconomic performance in France". In: *Annals of Economics and Statistics*(121/122), pp. 279–308.

d'Albis, Hippolyte, Ekrame Boubtane, and Dramane Coulibaly (2019). "Immigration and fiscal balance: Evidence from Western European countries". In: *Journal of Economic Dynamics and Control* 99, pp. 116–151.

D'Amuri, Francesco, Gianmarco IP Ottaviano, and Giovanni Peri (2010). "The labor market impact of immigration in Western Germany in the 1990s". In: *European Economic Review* 54(4), pp. 550–570.

Felbermayr, Gabriel, Wido Geis, and Wilhelm Kohler (2010). "Restrictive immigration policy in Germany: Pains and gains foregone?" In: *Review of World Economics* 146(1), pp. 1–21.

Furlanetto, Francesco and Ørjan Robstad (2019). "Immigration and the macroeconomy: Some new empirical evidence". In: *Review of Economic Dynamics* 34, pp. 1–19.

García, Benjamín and Juan Guerra-Salas (2020). *On the Response of Inflation and Monetary Policy to an Immigration Shock*. Tech. rep. Central Bank of Chile Working Paper No. 872.

Iftikhar, Zainab and Anna Zaharieva (2019). “General equilibrium effects of immigration in Germany: Search and matching approach”. In: *Review of Economic Dynamics* 31, pp. 245–276.

Kiguchi, Takehiro and Andrew Mountford (2019). “Immigration and Unemployment: A macroeconomic approach”. In: *Macroeconomic Dynamics* 23(4), pp. 1313–1339.

Liu, Xiangbo (2010). “On the macroeconomic and welfare effects of illegal immigration”. In: *Journal of Economic Dynamics and Control* 34, pp. 2547–2567.

Liu, Xiangbo, Theodore Palivos, and Xiaomeng Zhang (2017). “Immigration, skill heterogeneity, and qualification mismatch”. In: *Economic Inquiry* 55, pp. 1231–1264.

Llull, Joan (2018). “Immigration, wages, and education: A labour market equilibrium structural model”. In: *The Review of Economic Studies* 85(3), pp. 1852–1896.

Lozej, Matija (2019). “Economic migration and business cycles in a small open economy with matching frictions”. In: *Economic Modelling* 81, pp. 604–620.

Manacorda, Marco, Alan Manning, and Jonathan Wadsworth (2012). “The impact of immigration on the structure of wages: Theory and evidence from Britain”. In: *Journal of the European Economic Association* 10(1), pp. 120–151.

Mandelman, Federico S and Andrei Zlate (2012). “Immigration, remittances and business cycles”. In: *Journal of Monetary Economics* 59, pp. 196–213.

Ottaviano, Gianmarco IP and Giovanni Peri (2012). “Rethinking the effect of immigration on wages”. In: *Journal of the European Economic Association* 10(1), pp. 152–197.

Peri, Giovanni (2014). “Do immigrant workers depress the wages of native workers?” In: *IZA World of Labor*.

Rubio-Ramírez, Juan F, Daniel F Waggoner, and Tao Zha (2010). “Structural vector autoregressions: Theory of identification and algorithms for inference”. In: *The Review of Economic Studies* 77(2), pp. 665–696.

Scharfbillig, Mario and Marco Weißler (2019). “Heterogeneous displacement effects of migrant labor supply–quasi-experimental evidence from Germany”. In: *Available at SSRN 3429105*.

Schiman, Stefan (2018). *Labor supply shocks and the Beveridge curve: Empirical evidence from Austria*. Tech. rep. WIFO Working Papers.

Schorfheide, Frank and Dongho Song (2015). “Real-time forecasting with a Mixed-Frequency VAR”. In: *Journal of Business & Economic Statistics* 33(3), pp. 366–380.

Smith, Christie and Christoph Thoenissen (2019). “Skilled Migration and Business Cycle Dynamics”. In: *Journal of Economic Dynamics and Control*, p. 103781.

Vella, Eugenia, Jordi Caballé, and Joan Llull, eds. (2020). *Understanding Migration with Macroeconomics*. Palgrave Macmillan.

APPENDIX

A List of countries in the migration flows dataset of Destatis

OECD countries. Destatis provides data for European OECD countries as a whole. We also consider Australia, Canada, Israel, Japan, Korea Republic, New Zealand, and the United States. We thus do not include Chile, Colombia, Mexico.

EU countries (as of July 2013). Austria, Belgium, Bulgaria, Czech Republic, Croatia, Cyprus, Denmark, Estonia, Finland, France, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom.

Other European countries. Albania, Andorra, Belarus, Bosnia and Herzegovina, Iceland, Kosovo, Macedonia, Montenegro, Norway, Russian Federation, Serbia, Switzerland, Turkey, Ukraine, Rest of Europe.

Africa. Algeria, Angola, Cameroon, Cote d'lvoire, Egypt, Ethiopia, Ghana, Kenya, Central African Republic, Republic of Congo, Dem. Republic of Congo, Libya, Morocco, Namibia, Niger, Nigeria, Rwanda, Senegal, Somalia, South Africa, Tanzania, Tunisia, Uganda, Rest of Africa.

America. Argentina, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Honduras, Mexico, Nicaragua, Paraguay, Peru, United States, Uruguay, Venezuela, Rest of America.

Asia. Afghanistan, Arab Republic, Armenia, Azerbaijan, China, Georgia, India, Indonesia, Iran, Iraq, Israel, Japan, Jordan, Kazakhstan, Kuwait, Kyrgyzstan, the People's Republic of Korea, Democratic Republic of Korea, Lebanon, Pakistan, Philippines, Saudi Arabia, Singapore, Syria, Tajikistan, Thailand, United Arab Emirates, Uzbekistan, Vietnam, Yemen, Rest of Asia.

Australia and Oceania. Australia, New Zealand, Rest of Oceania.

B Bayesian estimation of the VAR model

Consider the reduced form VAR model presented in Section 2.2:

$$Y_t = C + \sum_{j=1}^{p} A_j Y_{t-j} + u_t$$

The process above can be stacked in a more compact form as follows:

$$\mathbf{Y} = \mathbf{X}B + \mathbf{U}$$

where:
1) $\mathbf{Y} = (Y_{p+1}, ..., Y_T)'$ is a $(T-p)$ x n matrix, with $Y_t = (Y_{1,t}, ..., Y_{n,t})'$.
2) $\mathbf{X} = (\mathbf{1}, \mathbf{Y}_{-1}, ..., \mathbf{Y}_{-p})$ is a $(T-p)$ x $(np+1)$ matrix, where $\mathbf{1}$ is a $(T-p)$ x 1 matrix of ones and $\mathbf{Y}_{-k} = (Y_{p+1-k}, ..., Y_{T-k})'$ is a $(T-p)$ x n matrix.
3) $\mathbf{U} = (u_{p+1}, ..., u_T)'$ is a $(T-p)$ x n matrix.
4) $B = (C, A_1, ..., A_p)'$ is a $(np+1)$ x n matrix of coefficients.
Vectorizing the equation above, we obtain:

$$\mathbf{y} = (I_n \otimes \mathbf{X})\beta + \mathbf{u}$$

where $\mathbf{y} = vec(\mathbf{Y})$, $\beta = vec(B)$, $\mathbf{u} = vec(\mathbf{U})$ and $\mathbf{u} \sim N(0, \Sigma \otimes I_{T-p})$.
Given the assumption of normality of the reduced-form errors, $u_t \sim N(0, \Sigma)$, we can express the likelihood of the sample, conditional on the parameters of the model and the set of regressors $\mathbf{X}$, as follows:

$$L(\mathbf{y}|\mathbf{X}, \beta, \Sigma) \propto |\Sigma \otimes I_{T-p}|^{-\frac{T-p}{2}} exp\left\{\frac{1}{2}(\mathbf{y} - I_n \otimes \mathbf{X}\beta)'(\Sigma \otimes I_{T-p})^{-1}(\mathbf{y} - I_n \otimes \mathbf{X}\beta)\right\}$$

Denote $\hat{\beta} = vec(\hat{B})$, where $\hat{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the OLS estimate, and let $S = (\mathbf{Y} - \mathbf{X}\hat{B})'(\mathbf{Y} - \mathbf{X}\hat{B})$ be the sum of squared errors. Then we can rewrite the likelihood as follows:

$$\begin{aligned} L(\mathbf{y}|\mathbf{X}, \beta, \Sigma) \propto & |\Sigma \otimes I_{T-p}|^{-\frac{T-p}{2}} exp\left\{\frac{1}{2}(\beta - \hat{\beta})'(\Sigma^{-1} \otimes \mathbf{X}'\mathbf{X})(\beta - \hat{\beta})\right\} \\ & exp\left\{-\frac{1}{2}tr(\Sigma^{-1}S)\right\} \end{aligned}$$

By choosing a non-informative (flat) prior for B and Σ that is proportional to $|\Sigma|^{-\frac{n+1}{2}}$, namely:

$$\begin{aligned} p(B|\Sigma) &\propto 1 \\ p(\Sigma) &\propto |\Sigma|^{-\frac{n+1}{2}} \end{aligned}$$

We can compute the posterior of the parameters given the data at hand using Bayes rule, as follows:

$$\begin{aligned} P(B,\Sigma|\mathbf{y},\mathbf{X}) &\propto L(\mathbf{y}|\mathbf{X},\beta,\Sigma)p(B|\Sigma)p(\Sigma) \\ &= |\Sigma|^{-\frac{T-p+n+1}{2}} exp\left\{\frac{1}{2}(\beta-\hat{\beta})'(\Sigma^{-1}\otimes\mathbf{X}'\mathbf{X})(\beta-\hat{\beta})\right\} exp\left\{-\frac{1}{2}tr(\Sigma^{-1}S)\right\} \end{aligned}$$

This posterior distribution is the product of a normal distribution for β conditional on Σ and an inverted Wishart distribution for Σ. Thus, we draw β conditional on Σ from:

$$\beta|\Sigma,\mathbf{y},\mathbf{X} \sim N(\hat{\beta},\Sigma\otimes(\mathbf{X}'\mathbf{X})^{-1})$$

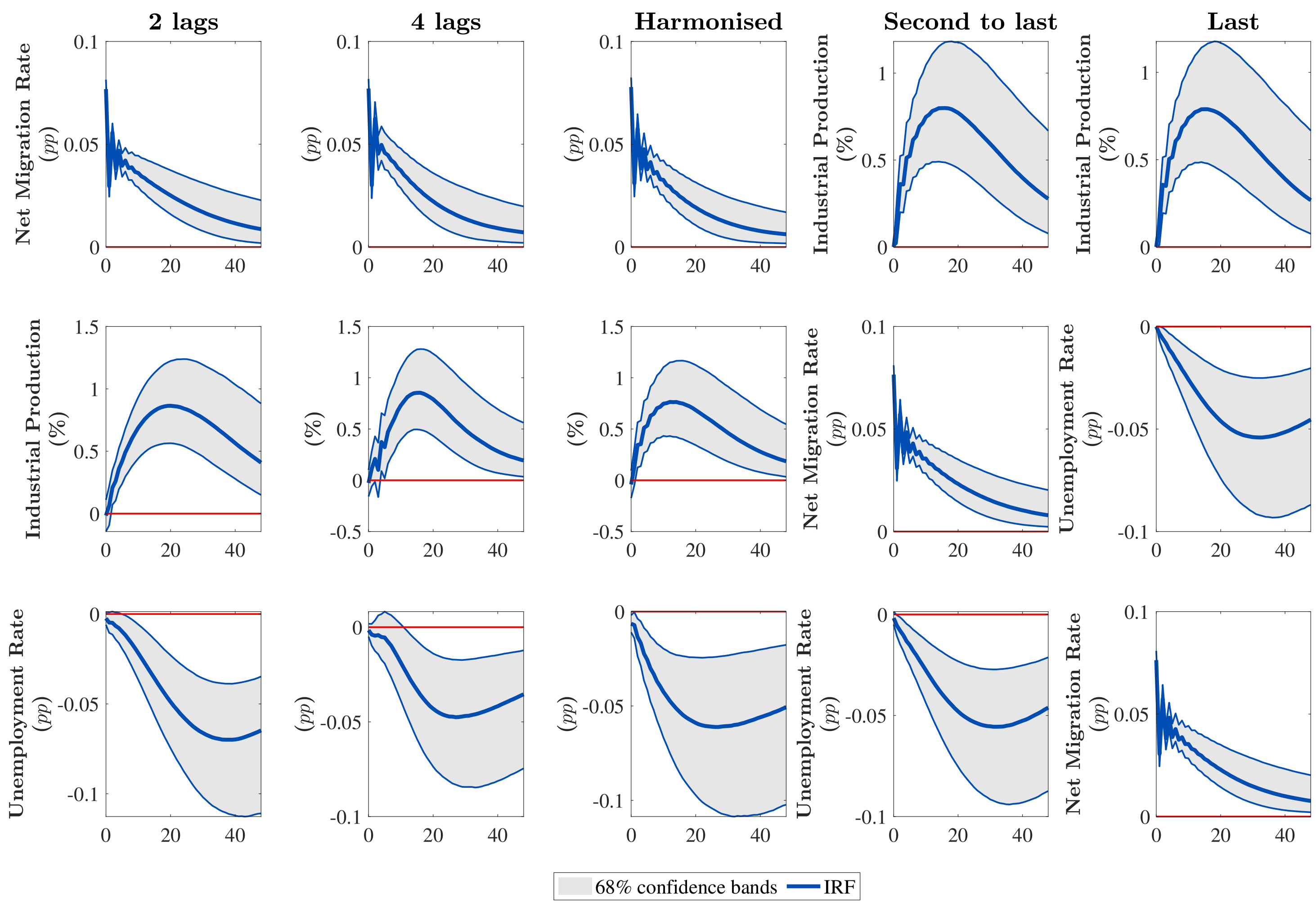
and Σ from:

$$\Sigma|\mathbf{y},\mathbf{X} \sim IW(S,v)$$

through Gibbs sampling, where $v = T - p - np - 1$.

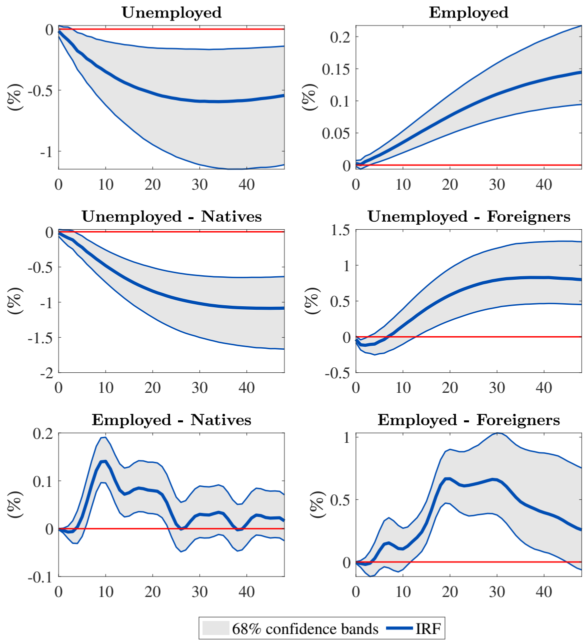
C Robustness checks and additional results

Figure A.1: Impulse response functions to an one-standard-deviation net migration shock: Robustness

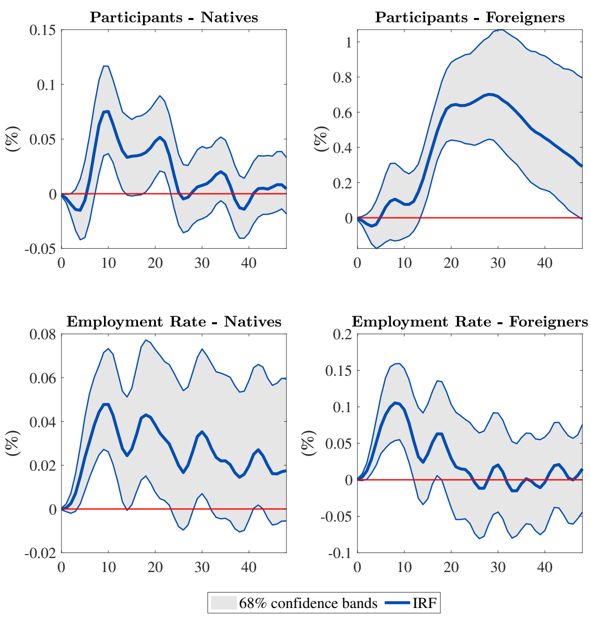


Note: In column 3, we replace the registered unemployment rate with harmonized unemployment. In columns 4 and 5, the net migration rate is ordered in the SVAR second to last and last, respectively. The continuous lines represent the posterior median at each horizon and the shaded areas indicate the 68th posterior probability region of the estimated impulse responses. The horizontal axis refers to time periods, measured by months.

Figure A.2: More impulse response functions to an one-standard-deviation net migration shock



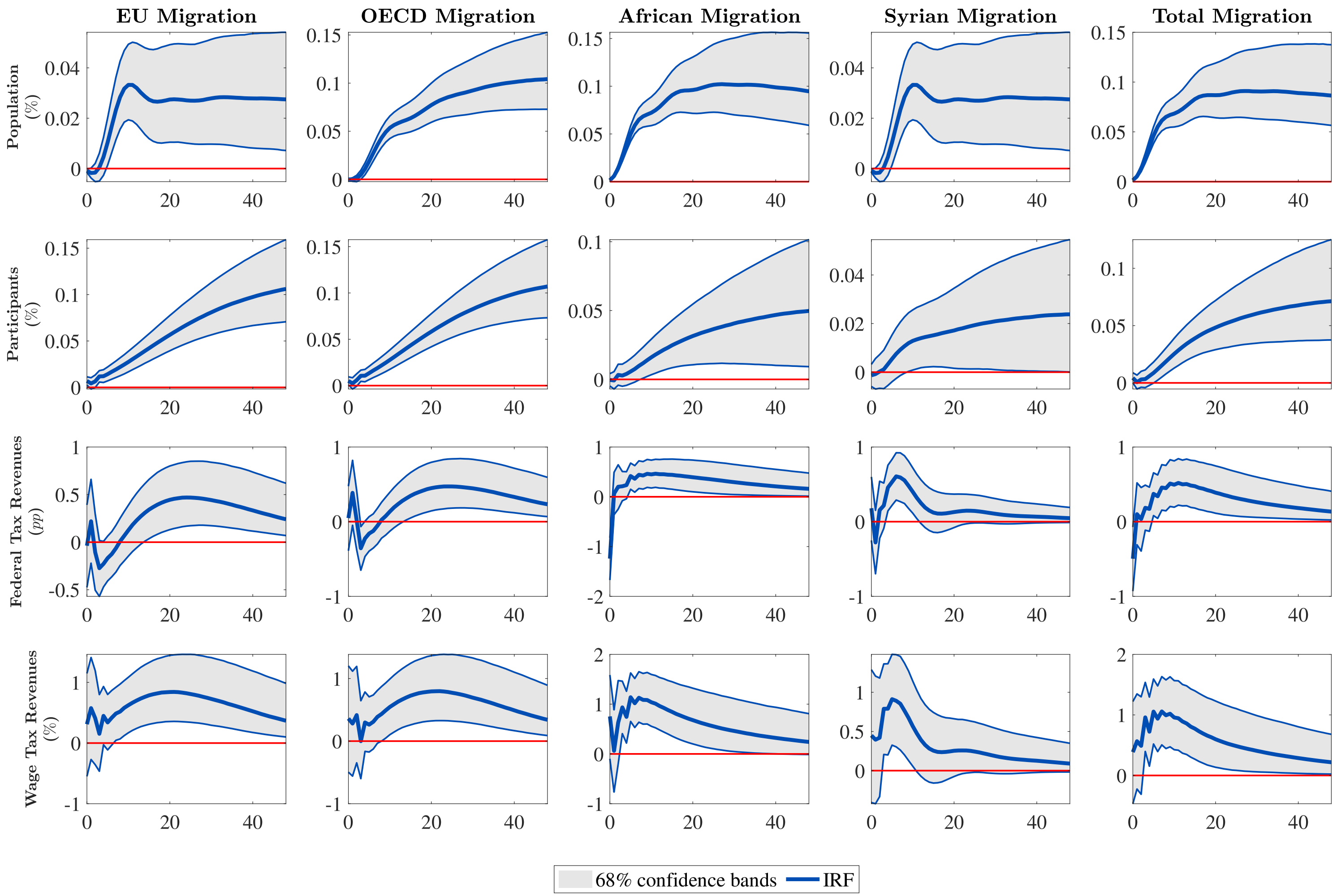
(a) Additional variables in the monthly SVAR



(b) LFS data in the Mixed-Frequency SVAR

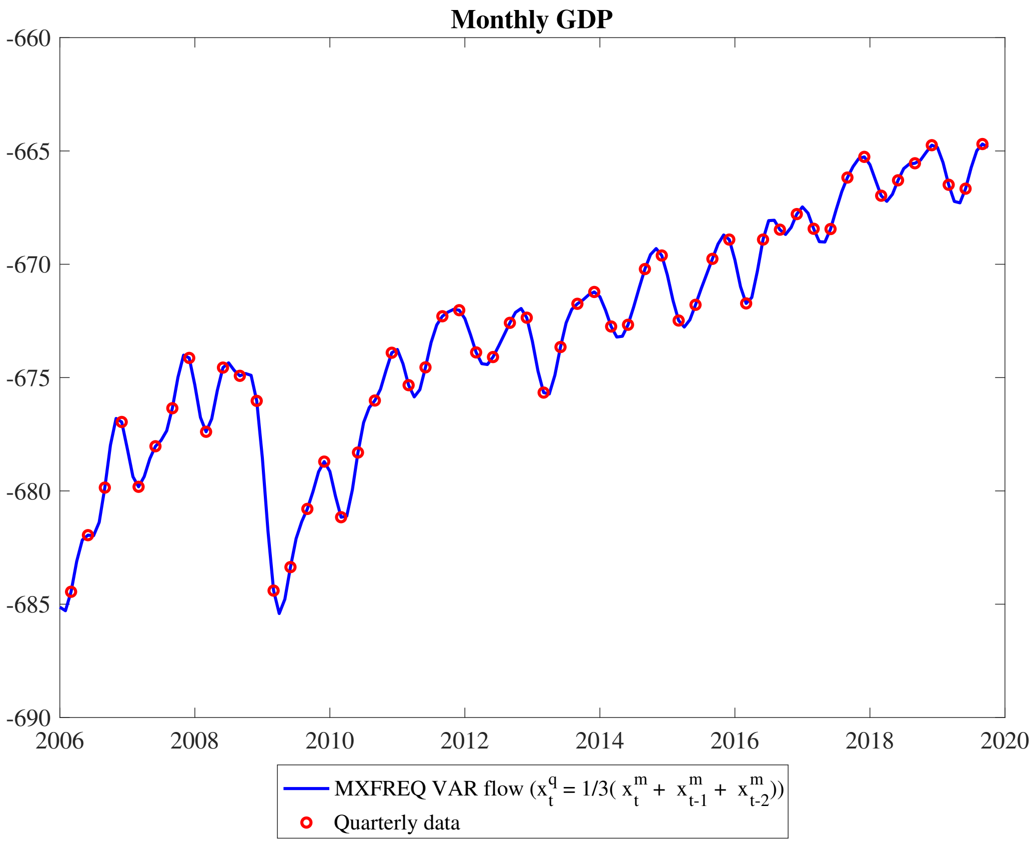
Note: The continuous lines represent the posterior median at each horizon and the shaded areas indicate the 68th posterior probability region of the estimated impulse responses. The horizontal axis refers to time periods, measured by months. In the bottom panel, variables are constructed as explained in Section 4.1, using both employment and unemployment data from the Eurostat Labor Force Survey.

Figure A.3: Subsample analysis and the refugees wave: Additional variables



Note: Total migration results from augmenting the net migration variable used in the baseline SVAR with net flows from Syria. The continuous lines represent the posterior median at each horizon and the shaded areas indicate the 68th posterior probability region of the estimated impulse responses. The horizontal axis refers to time periods, measured by months.

Figure A.4: Estimated monthly GDP (in logs)



CHAPTER III

THE DECLINE OF THE LABOR SHARE: NEW EMPIRICAL EVIDENCE*

DRAGO BERGHOLT§, FRANCESCO FURLANETTO† AND NICOLÒ MAFFEI FACCIOLI‡

October, 2019

Abstract: We estimate a structural vector autoregressive model in order to quantify four main explanations for the decline of the US labor income share: (i) rising market power of firms, (ii) falling market power of workers, (iii) higher investment-specific technology growth, and (iv) the widespread emergence of automation or robotization in production processes. Identification is achieved with theory robust sign restrictions imposed at medium-run horizons. The restrictions are derived from a stylized macroeconomic model of structural change. Across specifications we find that automation is the main driver of the long-run labor share. Firms' rising markups can, however, account for a significant part of the accelerating labor share decline observed in the last 20 years. Our results also point to complementarity between labor and capital, thus ruling out capital deepening as a major force behind declining labor shares. If anything, investment-specific technology growth has contributed to higher labor income shares in our sample.

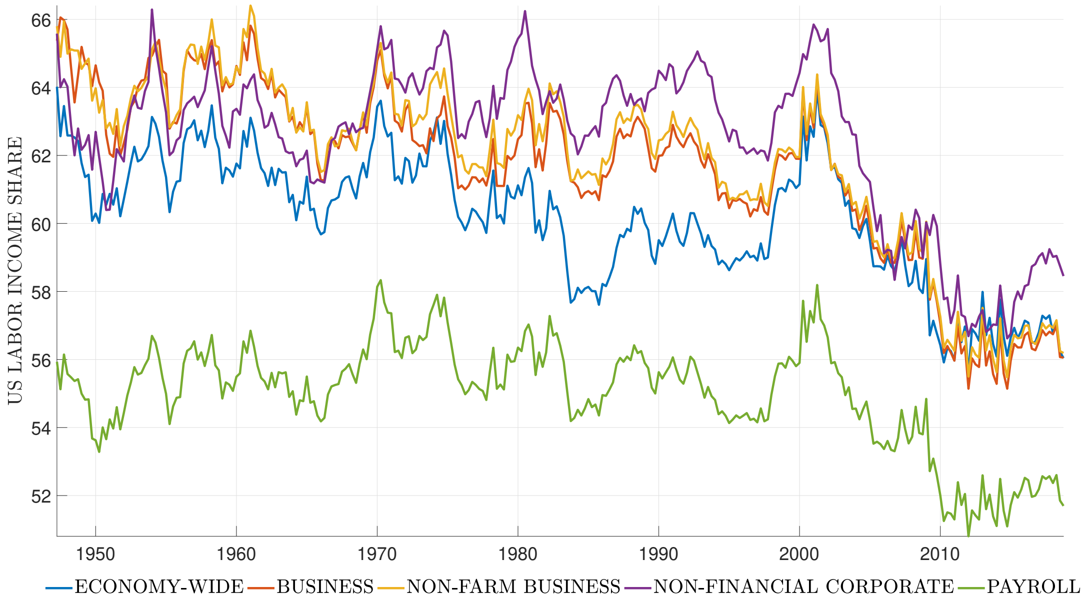
Keywords: *Labor income share, secular trends, technological progress, market power.*
JEL Classification: *E2, D2, D4, J3, L1.*

1 INTRODUCTION

Labor's share of national income has fallen in many countries in the last decades. In the US, the labor income share has accelerated its decline since the beginning of the

*This working paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank. We are grateful to Martin Beraja who kindly provided a detailed, written review of this paper. We have also benefited from discussions with Knut Are Aastveit, Alice Albonico, Alberto Alesina, Jonas Arias, Paul Beaudry, Hilde Bjørnland, Fabio Canova, Ferre De Graeve, Davide Debortoli, Herman van Dijk, Thorsten Drautzburg, Romain Duval, Jan Eeckhout, Gauti Eggertsson, Martin Eichenbaum, Mike Elsby, John Fernald, Luca Fornaro, Luca Gambetti, Marc Giannoni, Òscar Jordà, Jesper Lindé, Albert Marcet, Joseba Martinez, Karel Mertens, Magne Mogstad, Giuseppe Moscarini, Evi Pappa, Ivan Petrella, Sophie Piton, Giorgio Primiceri, Valerie Ramey, Jake Robbins, Juan Rubio-Ramírez, Aysegul Sahin, Raül Santaeulàlia-Llopis, Jae Sim, Patrizio Tirelli, Mathias Trabandt, Gianluca Violante, Christian Wolf and Maik Wolters. Finally, we would like to thank participants at various conferences and research seminars for their useful comments.
§Norges Bank. P.O. Box 1179 Sentrum, 0107 Oslo, Norway. E-mail: drago.bergholt@norges-bank.no.
†Norges Bank. P.O. Box 1179 Sentrum, 0107 Oslo, Norway. Corresponding author. E-mail: francesco.furlanetto@norges-bank.no.
‡Universitat Autònoma de Barcelona and Barcelona GSE. E-mail: nicolo.maffei@barcelonagse.eu. The author gratefully acknowledges financial support from the "La Caixa-Severo Ochoa International Doctoral Fellowship".

Figure 1: The US labor income share over time



Source: Bureau of Labor Statistics and authors' calculations.

new century, reaching its postwar lowest level in the aftermath of the Great Recession (Elsby, Hobijn, and Sahin, 2013). Figure 1 documents the evolution of five alternative measures of the US labor income share (subsection 5.2 provides a detailed description of the measures). While each series implies a somewhat different trend, they have all gone through a clear fall in the last 20 years. In addition, Gutierrez and Piton (2019) show that, while the evidence for a global decline of the labor share across major economies is weaker, the recent decline in the US is undisputed, with several potential implications for policy and welfare. Yet, a consensus view regarding the main structural forces at play is still lacking. The aim of this paper, therefore, is to empirically evaluate and quantify some of the main explanations for observed labor share trends in the US economy. We do this using a combination of economic theory and time series techniques applied to US macroeconomic data.

We consider four explanations with rather broad appeal in the literature: first, a number of recent studies have argued that rising market power among firms has crowded out labor's share of income (Barkai, 2018; De Loecker and Eeckhout, 2017; Eggertsson, Robbins, and Wold, 2018; Gutierrez and Philippon, 2017). These studies find evidence of declining competition and increasing market concentration. The claim is that trends in firms' market power has spurred profit growth at the expense of labor income. A second take on the labor share decline concerns technological progress in the form of automation or robotization (Acemoglu and Restrepo, 2019, 2018; Autor and Salomons, 2018; Leduc and Liu, 2019; Martinez, 2018). Acemoglu and Restrepo (2019), for example, argue that many tasks previously done by workers are currently being automated on a relatively large scale. They find that automation leads to lower employment and stagnant wages, thus lowering the labor share of income. A third group of arguments focuses on labor market institutions such as unions and minimum wages (Piketty, 2014). Along these lines, Blanchard and Giavazzi (2003) and Ciminelli, Duval, and Furceri (2018) find that a decline in the bargaining power of workers, proxied, respectively, by labor market deregulation and by major reforms in employment protection legislation, may be responsible for substantial movements in the labor share. Finally, the fourth explanation we consider puts forward a major role for capital-biased technology growth. Karabarbounis and Neiman (2014) in particular use the relative price of investment as a proxy for investment-specific technological progress, and find that capital deepening measured in this way may account for declining labor shares in a number of countries including the US. Importantly, cheaper capital should imply lower labor income shares only if labor and capital are net substitutes, which is exactly what Karabarbounis and Neiman (2014) find in their data.

While a large literature has discussed each of these four explanations in isolation, an empirical analysis including all of them in the context of the same model is lacking. Our aim is to fill this gap. To this end we estimate a structural vector autoregression (SVAR) with permanent shocks. These shocks are interpreted as candidate explanations for low frequency changes in the labor share. We identify them using theory robust sign restrictions á la Canova and Paustian (2011), imposed on impulse response functions at medium-run horizons. "Theory robust" in this context means that the restrictions hold across a broad set of parameterizations in a benchmark, macroeconomic model. Our approach involves two steps: first, we set up a fairly stylized, yet flexible model of structural change. It incorporates the four candidate explanations of interest and nests, as special cases, several of the models used to study declining labor shares (including those used by Karabarbounis and Neiman (2014) and Barkai (2018)). We then consider the macroeconomic implications of each candidate explanation under a broad set of model parameterizations. In particular, we show that they can be separately identified by a combination of medium-run sign restrictions that are mutually exclusive and jointly exhaustive. Second, this set of restrictions is used to identify the structural shocks in the empirical model. As a byproduct, we can also obtain indirect empirical evidence on the elasticity of substitution between labor and capital—arguably a key parameter for labor share dynamics. Importantly, we show that our identification scheme holds for about any value of this parameter, and the estimated impulse responses from the SVAR can be used to infer whether the capital-labor substitution elasticity is bigger or smaller than one.

The econometric approach used in this paper differs fundamentally from typical approaches in the existing literature on labor shares: while most studies draw inference based on cross-sectional variation in microeconomic data (at the firm or sectoral level), we instead exploit the macroeconomic time series implications of permanent, but aggregate shocks. Moreover, we use the SVAR framework to study medium-run trends rather than short-run fluctuations, as normally done in the business cycle literature. To the best of our knowledge, this is the first paper using sign restrictions to identify several permanent shocks. Finally, we stress that our estimation approach addresses a well-known issue in the literature on factor substitution and biased technical change: Diamond, McFadden, and Rodriguez (1978) amongst others argue that factor elasticities and technology cannot be jointly identified in a theoretical model like ours (see León-Ledesma, McAdam, and Willman (2010) for further discussion of this so-called "impossibility theorem"). We confirm that this is likely to be the case if model equations are estimated directly, but that the sign restriction approach used here can get around the issue.

The empirical model is estimated on data covering the period 1983Q1-2018Q3. With the estimated model at hand, we set out to shed light on the observed labor share decline in the US economy. Our main results can be summarized as follows: first, we find that the labor income share falls permanently after a rise in automation or a rise in firms' market power, but increases permanently in response to higher investment-specific technology growth. The labor share response to a decline in workers' market power is negative in the short run, but unclear and not significantly different from zero in the long run. Importantly, although we cannot pinpoint the exact value of the substitution elasticity between labor and capital, the latter two findings are only consistent with net complementarity. Our second result concerns the main drivers of the labor share. We find that automation accounts for the bulk of labor share fluctuations in our sample. The second most important factor is firms' market power, at least in the medium to long run. Labor markups have some explanatory power in the very short run while investment-specific technology only plays a minor role. Our third result sheds light on the causes of the accelerating labor share decline observed in the last 20 years. A historical decomposition reveals that this decline is driven both by automation and firms' rising market power, with the latter becoming increasingly important after the Great Recession. Turning to investment-specific or capital-biased technology, we find that this kind of shock, if anything, has led to an increase in the labor share throughout the 2000s.

Our empirical findings help to assess various the explanations of declining labor shares. They are well in line with the view that tasks previously done by human workers have been taken over by robots on a significant scale in recent years (Acemoglu and Restrepo, 2019; Autor and Salomons, 2018; Leduc and Liu, 2019; Martinez, 2018), but also with stories about increased market concentration (Barkai, 2018; De Loecker and Eeckhout, 2017; Eggertsson et al., 2018; Gutierrez and Philippon, 2017). Moreover, while our results confirm that capital deepening in the form of investment-specific technology growth has taken place in the last decades (Karabarbounis and Neiman, 2014), we find that net complementarity between labor and capital has led to a crowding-in of labor rather than the opposite.

Interestingly, we reach our conclusions using a fundamentally different approach than the aforementioned studies. Moreover, our results do not seem to suffer from the timing issue put forward by Elsby et al. (2013), who note that some of the explanations for falling labor shares rely on trends that started decades before the labor share decline. All in all, the important role for automation emerging from the SVAR can be explained on intuitive grounds. A positive automation shock increases output in the medium run and lowers wages and total hours, in keeping with the effects discussed in Acemoglu and Restrepo (2019). With the labor share defined as total labor income over output, we emphasize that the response to automation of each of these variables favors a decline of the labor share. Put simply, the numerator of the labor share decreases, while the denominator increases. No other shock generates such a negative co-movement between wages and labor productivity. Note, however, that the automation shock is redistributive in nature and does not have important aggregate effects on output. This is hardly surprising since countercyclical wages and hours are not a characteristic of economic fluctuations.

The literature on falling labor income shares has exploded in recent years and several explanations have been proposed in addition to those included in our baseline model: Rognlie (2015) focuses on developments in the housing sector and finds that more expensive residential investment and increased land scarcity have led to higher (housing) capital shares at the expense of labor income. Giannoni and Mertens (2019) document how outsourcing—firms' contracting of labor-intensive activities to external companies—can explain why labor shares are falling within many industries, although the aggregate effects are milder. Glover and Short (2017) formalize a hypothesis linking an aging workforce to the declining labor share. Kaymak and Schott (2018) focus on the manufacturing sector and emphasize the role played by corporate tax cuts. Finally, Elsby et al. (2013) argue that globalization, and the process of off-shoring of intermediate goods production to developing countries in particular, is a promising candidate for the decline in the labor share. In later sections, we discuss how our identification approach and main results might be interpreted in light of some alternative explanations.

An important strand of the literature has focused on issues related to the measurement of labor income. The seminal paper by Elsby et al. (2013), for example, discusses how mis-measurement of income earned by the self-employed may exaggerate the recent decline in the labor income share. More recently, Koh, Santaeulalia-Llopis, and Zheng (2018) argue that the long-run post-war trend in labor income may be driven by the capitalization of intellectual property products (IPP). In fact, the Bureau of Economic Analysis (BEA) has revised the treatment of IPP from an accounting perspective by attributing the entire rents from IPP investment to capital income. This choice affects the long-run trend in the labor share series but not its steep decline in recent years. Finally, disentangling the capital share of income and the profit share of income has proven to be challenging from an empirical point of view. Barkai (2018) argues that pure profits have increased substantially in recent years, while the capital share has decreased. Karabarbounis and Neiman (2018) claim that the residual payments (referred to as "factorless income") obtained after measuring the labor share and the capital share cannot be interpreted as pure profits and may reflect measurement error in the capital stock or in the rental rate of capital. In order to reduce the unavoidable issues related to the measurement of the labor share and the profit share, we will conduct an extensive sensitivity analysis using alternative measures for those variables.

The rest of the paper is organized as follows: Section 2 describes a theoretical model of structural change. Section 3 derives the set of theory robust sign restrictions, lays out the econometric methodology and discusses identification. Section 4 documents our main empirical results. Section 5 provides a battery of robustness tests and extensions. Finally, Section 6 concludes.

2 Theoretical framework

Our baseline, theoretical framework is the standard neoclassical growth model, but we add a few, simple extensions that allow us to consider trends in the labor share. Importantly, in our setup the labor share can change due to (i) investment-specific technical change, (ii) automation of labor-intensive production tasks, (iii) distortions in labor markets, and (iv) changes in the market power of firms. The resulting framework is, with minor deviations, similar to those used by Karabarbounis and Neiman (2014), Barkai (2018) and Caballero, Farhi, and Gourinchas (2017), amongst others.

The model economy is populated by a unit mass of firms and households. For convenience we also distinguish between retailers, investment producers, and conventional (wholesale) firms. In the labor market we make a distinction between individual workers and a labor union that rents workers' services in order to provide labor to firms.

2.1 RETAILERS

A competitive retailer combines individual goods in order to produce an aggregate, final good. The aggregation technology is standard:

$$Y_t = \left(\int_0^1 Y_{j,t}^{\frac{\epsilon_{p,t}-1}{\epsilon_{p,t}}} dj \right)^{\frac{\epsilon_{p,t}}{\epsilon_{p,t}-1}}$$

$Y_{j,t}$ is output by firm j and $\epsilon_{p,t}$ is a time varying elasticity of substitution between inputs. The retailer chooses inputs in order to maximize profits. Optimal demand towards firm j's output follows:

$$Y_{j,t} = P_{j,t}^{-\epsilon_{p,t}} Y_t$$

$P_{j,t}$ is the price of good j relative to the aggregate price index specified below. This downward sloping demand function equips firms with market power and allows them to charge a markup over marginal costs when they set their own prices. The optimal price index is given by

$$1 = \left(\int_0^1 P_{j,t}^{1-\epsilon_{p,t}} dj \right)^{\frac{1}{1-\epsilon_{p,t}}}.$$

Thus, we choose the final good Y_t as the numeraire. It can be used for consumption or investment purposes. Market clearing dictates that

$$Y_t = C_t + X_t, \tag{1}$$

where C_t denotes consumption and X_t represents raw investments.

2.2 INVESTMENT PRODUCERS

Following Fisher (2006), we suppose that a competitive investment goods producer transforms raw investments X_t into final investment goods. The production technology for this activity is given by

$$I_t = \Upsilon_t X_t. \tag{2}$$

Changes in Υ_t represent investment-specific technological progress. The final good I_t is sold to households, who accumulate capital. We denote by $P_{I,t}$ the unit price of final investments relative to final consumption. Profit maximization on behalf of the investment producer leads to the optimality condition

$$P_{I,t} = \Upsilon_t^{-1}, \tag{3}$$

which in turn implies the zero profit condition $P_{I,t} I_t = X_t$. Karabarbounis and Neiman (2014) find that falling investment prices can explain a major share of the observed labor share decline in many countries, including the US.

2.3 LABOR UNION

A competitive labor union combines hours from individual workers using the technology

$$L_t = \left(\int_0^1 L_{n,t}^{\frac{\epsilon_{w,t}-1}{\epsilon_{w,t}}} dn \right)^{\frac{\epsilon_{w,t}}{\epsilon_{w,t}-1}},$$

where $L_{n,t}$ is hours supplied by worker n. $\epsilon_{w,t}$ is a time varying elasticity of substitution between labor varieties. Optimal demand for worker n's services follows:

$$L_{n,t} = \left(\frac{W_{n,t}}{W_t} \right)^{-\epsilon_{w,t}} L_t$$

$W_{n,t}$ is the unit cost of worker n while W_t is the optimal, aggregate wage index:

$$W_t = \left(\int_0^1 W_{n,t}^{1-\epsilon_{w,t}} dn \right)^{\frac{1}{1-\epsilon_{w,t}}}$$

2.4 HOUSEHOLDS

There is a unit mass of optimizing households in the economy. Household $n \in [0,1]$ derives utility from consumption and dis-utility from work activities. The period utility is equal to:

$$\mathcal{U}_{n,t} = \frac{C_{n,t}^{1-\sigma}}{1-\sigma} \exp \left(-\Psi \frac{(1-\sigma) L_{n,t}^{1+\varphi}}{1+\varphi} \right).$$

These preferences allow for a balanced growth path when the intertemporal substitution elasticity differs from one, as shown by King, Plosser, and Rebelo (1988). Household n maximizes $\mathbb{E}_t \sum_{s=t}^{\infty} \beta^{s-t} \mathcal{U}_{n,s}$, where β is a time discount factor. Maximization is subject to two constraints. The first is an intertemporal budget constraint:

$$C_{n,t} + P_{I,t} I_{n,t} + B_{n,t} \leq W_{n,t} L_{n,t} + r_t^k K_{n,t-1} + \mathcal{D}_{n,t} + (1 + r_{t-1}) B_{n,t-1} - T_{n,t}$$

Labor income, capital income and profit income are denoted by $W_{n,t}L_{n,t}$, $r_t^k K_{n,t-1}$, and $\mathcal{D}_{n,t}$ respectively. r_t^k is the competitive rental price on the current capital stock in place, $K_{n,t-1}$. $B_{n,t}$ represents the amount of one-period bonds purchased in period t with return r_t. Finally, T_t is a lump-sum tax levied by the government. The second constraint is the law of motion for capital:

$$K_{n,t} \leq (1-\delta) K_{n,t-1} + I_{n,t}$$

where δ is the capital depreciation rate. We assume perfect risk-sharing across households. This allows us to consider a symmetric equilibrium ($W_{n,t} = W_t$, $L_{n,t} = L_t$, etc.) with a representative household. The representative household's behavior can be summarized by the budget constraint, the law of motion for capital, as well as five optimality conditions. We define the gross wage markup as $\mathcal{M}_{w,t} = \frac{W_t}{MRS_t}$, where MRS_t is the marginal rate of substitution between labor and consumption. Optimality conditions are stated below:

$$\Lambda_t = C_t^{-\sigma} \exp\left(-\Psi \frac{(1-\sigma) L_t^{1+\varphi}}{1+\varphi}\right) \tag{4}$$

$$\Lambda_t = \beta \mathbb{E}_t \Lambda_{t+1} (1 + r_t) \tag{5}$$

$$W_t = \mathcal{M}_{w,t} \Psi L_t^{\varphi} C_t \tag{6}$$

$$P_{I,t} = \beta \mathbb{E}_t \frac{\Lambda_{t+1}}{\Lambda_t} \left[r_{t+1}^k + P_{I,t+1} (1-\delta)\right] \tag{7}$$

$$\mathcal{M}_{w,t} = \frac{\epsilon_{w,t}}{\epsilon_{w,t} - 1} \tag{8}$$

The evolution of $\mathcal{M}_{w,t}$ is exogenous from the household's point of view. It can be triggered by changes in union power, but also by leisure preferences, demographics, or other factors that influence the supply side of the labor market. Drautzburg, Fernández-Villaverde, and Guerrón-Quintana (2017), for example, provide narrative evidence of the macroeconomic importance of workers' bargaining power. We do not take a stand on the particular drivers of $\mathcal{M}_{w,t}$, but simply refer to them as wage or labor markup shocks.

2.5 Monopolistic firms

There is a unit measure of monopolistically competitive firms in the economy. Their output is produced with labor and capital. Firm $j \in [0,1]$ sets its own price in order to maximize profits $\mathcal{D}_{j,t}$:

$$\mathcal{D}_{j,t} = P_{j,t} Y_{j,t} - W_t L_{j,t} - r_t^k K_{j,t-1}$$

Profit maximization is subject to the downward sloping demand from retailers, as well as a production technology featuring constant elasticity of substitution:

$$Y_{j,t} = \left[\alpha_{l,t} \left(A_{l,t} L_{j,t}\right)^{\frac{\eta-1}{\eta}} + \alpha_{k,t} \left(A_{k,t} K_{j,t-1}\right)^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}$$

η represents the elasticity of substitution between capital and labor. This production function includes three distinct technological processes: $A_{l,t}$ and $A_{k,t}$, respectively, represent the conventional labor-augmenting and capital-augmenting technology innovations. $\alpha_{k,t}$, in contrast, is interpreted as an automation shock that makes output more capital intensive at the expense of labor. Its microeconomic foundation is derived by Acemoglu and Restrepo (2018) and the references therein. They consider a framework where a continuum of tasks is produced within a production unit such as a firm. Some tasks require labor, but for others labor and capital are perfect substitutes. Automation in this context is interpreted as a shift in the share of tasks that can be produced with capital. Acemoglu and Restrepo (2018) show how one can aggregate the tasks in order to establish a production function like ours, with time-varying weights $\alpha_{l,t}$ and $\alpha_{k,t}$. Importantly, $\alpha_{l,t}$ and $\alpha_{k,t}$ are decreasing and increasing in the degree of automation, respectively. We follow Caballero et al. (2017) by restricting attention to a baseline case where automation implies that $\alpha_{l,t} = \bar{\alpha} - \alpha_{k,t}$. As before, we consider a representative firm in the symmetric equilibrium ($P_{j,t} = 1$, $Y_{j,t} = Y_t$, etc.) and define the firm's gross markup as $\mathcal{M}_{p,t} = MC_t^{-1}$

(price over nominal marginal costs). Firm behavior can then be summarized by the production function as well as the following optimality conditions:

$$r_t^k \mathcal{M}_{p,t} = \alpha_{k,t} A_{k,t}^{\frac{\eta-1}{\eta}} \left(\frac{Y_t}{K_{t-1}} \right)^{\frac{1}{\eta}} \tag{9}$$

$$W_t \mathcal{M}_{p,t} = \alpha_{l,t} A_{l,t}^{\frac{\eta-1}{\eta}} \left(\frac{Y_t}{L_t} \right)^{\frac{1}{\eta}} \tag{10}$$

$$\mathcal{M}_{p,t} = \frac{\epsilon_{p,t}}{\epsilon_{p,t} - 1} \tag{11}$$

The last equation defines the optimal, time-varying markup from firms' point of view. Firm revenues follow:

$$Y_t = \mathcal{M}_{p,t} \left(W_t L_t + r_t^k K_{t-1} \right)$$

Movements in $\mathcal{M}_{p,t}$ can be caused by changes in market concentration, segmentation, product specialization, or other factors that affect the degree of competition between firms (Barkai, 2018). We do not take a stand on the particular drivers of $\mathcal{M}_{p,t}$, but simply refer to them as price or firm markup shocks.

2.6 AGGREGATION AND INCOME ACCOUNTING

Market clearing in labor and capital markets dictate that:

$$L_t = \int_0^1 L_{j,t} dj \qquad K_{t-1} = \int_0^1 K_{j,t-1} dj \qquad \mathcal{D}_t = \int_0^1 \mathcal{D}_{j,t} dj$$

We suppose that bonds are in zero net supply and sum up over all households' budget constraints in order to express aggregate income:

$$\begin{aligned} Y_t &= C_t + P_{I,t} I_t \\ &= W_t L_t + r_t^k K_{t-1} + \mathcal{D}_t \end{aligned}$$

Income shares in our simple model are defined accordingly:

$$s_{l,t} = \frac{W_t L_t}{Y_t} \qquad s_{k,t} = \frac{r_t^k K_{t-1}}{Y_t} \qquad s_{d,t} = \frac{\mathcal{D}_t}{Y_t}$$

Moreover, $s_{l,t} + s_{k,t} + s_{d,t} = 1$. At this point it is useful to evaluate how the labor income share in our simple model reacts to structural shocks at low frequencies. To this end we define a long-run equilibrium as the non-stochastic equilibrium outcome once all shock dynamics have settled down. In the appendix we show that:

$$\bar{s}_{l,t} = \frac{1}{\bar{\mathcal{M}}_{p,t}} \left[1 - \bar{\alpha}_{k,t}^{\eta} \left(\frac{\beta^{-1} - (1-\delta)}{\bar{\Upsilon}_t \bar{A}_{k,t}} \bar{\mathcal{M}}_{p,t} \right)^{1-\eta} \right],$$

where long-run equilibrium variables are denoted by a bar. A few remarks are in place: first, the long-run labor share is not affected by labor-augmenting technology or markups in the labor market. Thus, only short- to medium-run fluctuations in the labor share can be accounted for by these shocks according to our model. Second, higher firm markups or more automation both imply a decline in the long-run labor share. This is true regardless of the degree of substitutability between capital and labor. Third, the long-run effects of investment-specific and capital-augmenting technology shocks on the labor share are observationally equivalent. For this reason it is sufficient to consider only one of the two shocks, as long as the focus is on low frequency dynamics. Finally, whether or not a rise in $\bar{\Upsilon}_t$ (or $\bar{A}_{k,t}$) reduces labor's share of income depends crucially on η: the labor share unambiguously falls if $\eta > 1$, and unambiguously rises if $\eta < 1$. The knife-edge case with Cobb-Douglas production ($\eta = 1$) implies no change in the long-run labor share in response to factor-augmenting shocks. We further describe the identification challenge associated with η and how we address it in Section 3.

2.7 Shock processes

Given the preceding discussion, we restrict attention to four stochastic shock processes: exogenous innovations to firms' price markup $\mathcal{M}_{p,t}$, to labor's wage markup $\mathcal{M}_{w,t}$, to investment-specific technology Υ_t, and to the automation parameter $\alpha_{k,t}$. The processes are assumed to follow a random walk:

$$\frac{\mathcal{M}_{p,t}}{\mathcal{M}_{p,t-1}} = 1 + g_{p,t} = (1 + g_p)\exp(z_{p,t}) \qquad \frac{\mathcal{M}_{w,t}}{\mathcal{M}_{w,t-1}} = 1 + g_{w,t} = (1 + g_w)\exp(z_{w,t})$$
$$\frac{\Upsilon_t}{\Upsilon_{t-1}} = 1 + g_{\Upsilon,t} = (1 + g_\Upsilon)\exp(z_{\Upsilon,t}) \qquad \frac{\alpha_{k,t}}{\alpha_{k,t-1}} = 1 + g_{\alpha_k,t} = (1 + g_{\alpha_k})\exp(z_{\alpha_k,t})$$

The innovations themselves are autoregressive processes:

$$z_{p,t} = \rho_p z_{p,t-1} + \sigma_p \varepsilon_{p,t} \qquad z_{w,t} = \rho_w z_{w,t-1} + \sigma_w \varepsilon_{w,t}$$
$$z_{\Upsilon,t} = \rho_\Upsilon z_{\Upsilon,t-1} + \sigma_\Upsilon \varepsilon_{\Upsilon,t} \qquad z_{\alpha_k,t} = \rho_{\alpha_k} z_{\alpha_k,t-1} + \sigma_{\alpha_k} \varepsilon_{\alpha_k,t}$$

It is assumed that $\varepsilon_{p,t}$, $\varepsilon_{w,t}$, $\varepsilon_{v,t}$ and $\varepsilon_{\alpha_k,t}$ are independently drawn from a normal distribution with mean zero and unit variance. We stress that the shock processes specified here in general imply separate stochastic trends for all variables of interest in the model. A common stochastic trend is obtained only in a particular special case: if the automation shock as well as both markup shocks are absent (or if all three shocks are temporary), and at the same time $\eta = 1$, then one is back to the standard, neoclassical growth model with constant long-run income shares.

3 Empirical strategy

We have already seen how the substitution elasticity η determines the response of labor's income share to factor-augmenting technical change. An observed fall in the labor share, for example, can be attributed to the combination of rising investment-specific technology (Υ_t) and net substitutability between capital and labor ($\eta > 1$), but equally well to declining investment technology and net complementarity. Herein lies a potentially serious identification problem, as neither Υ_t nor η are observed. This issue is well-known in the literature, and has led researchers to suggest that one cannot simultaneously identify the capital-labor elasticity and biased technical change. Diamond et al. (1978), for example, derive non-identification from the sign patterns produced by a fairly general class of neoclassical production functions.[1] More recently, León-Ledesma et al. (2010) claim that the "impossibility theorem" developed by Diamond et al. (1978) represents the received wisdom in the literature.

Faced with this challenge, applied researchers have typically opted for one of two strategies: the first is to obtain a direct measure, or at least a proxy, of technical change. This is the route taken by Karabarbounis and Neiman (2014), who use relative investment prices to measure investment technology, as well as by Acemoglu and Restrepo (2019), who proxy automation by the number of robots per worker. One can then regress the labor share on the obtained measure. Typically, this single-equation approach derives inference from cross-sectional variation in microeconomic data at the firm or sectoral level. The second strategy is to directly estimate the full theoretical model or at least parts of it, either by maximum likelihood (with or without priors) or by moment matching. The idea, then, is to achieve identification from the cross-equation restrictions embedded in the system of model equations. Equipped with the estimated model, one can use the Kalman filter to estimate unobserved drivers of the labor share. While both of these approaches bear some merits, there are important reasons why we prefer a fundamentally different identification strategy: first and foremost, since our goal is to quantify the relative importance of four different labor share drivers—all unobservable—it is not sufficient to exploit proxy variables for only one or two shocks in a single equation setting. Rather, we need an identification strategy which allows us to identify and quantify all four shocks simultaneously within the same system. Second, as we show later in this section, η has very little influence on macroeconomic variables in the model other than the labor share. System estimation where the model is fitted quantitatively to empirical moments is, therefore, subject to a problem of weak identification unless additional assumptions are made.[2]

These concerns call for an alternative strategy, still heavily guided by economic theory, but in which theoretically consistent sign patterns of impulse response functions are exploited as a means to sidestep the identification issue. This is exactly the approach we take here. The identification problem is addressed using a two-step procedure: first we conduct a careful analysis of the theoretical model in order to arrive at identification restrictions—in terms of signs—which are robust to a broad range of values for the model's parameters (including η). Importantly, these theory robust restrictions do not involve the labor share itself. Second, we impose the derived sign restrictions on a flexible time series model in order to estimate the evolution of shocks and their effects on the labor share. Of course, one could go ahead and estimate an empirical time series model as we do, and then impose sign restrictions on the labor share itself. However, any sign restriction imposed on the labor share (in response to shocks that move relative factor prices, such as investment-specific technology) would implicitly assume either net complementarity or net substitutability between labor and capital. We nevertheless do this exercise as a sensitivity check and show that the main results are robust to additional elasticity

[1] Among others, the authors show that the data generated by one particular production function can be perfectly replicated by another production function exhibiting different elasticities and different technical bias. See also the discussions by Kumar and Gapinski (1974) and Thursby (1980).

[2] See León-Ledesma and Satchi (2018) for an application. When estimating their model, the authors impose restrictions on the elasticity of substitution between labor and capital.

Table 1: Parameter bounds for the Monte Carlo exercises

		Benchmark model		
		M	*LB*	*UB*
Initial income shares				
s_l	Labor income share	0.6	0.5	0.7
s_k	Capital income share	0.3	0.225	0.375
s_d	Profit income share	0.1	0.075	0.125
"Deep" parameters				
σ	Inverse of intertemporal elasticity	3	1	5
φ	Inverse Frisch elasticity	3	1	5
η	Substitution between labor and capital	1	0.5	1.5
Shocks' persistence				
ρ_p	Firms' markup growth	0.25	0	0.5
ρ_w	Labor's markup growth	0.25	0	0.5
ρ_υ	Investment specific technology growth	0.25	0	0.5
ρ_{α_k}	Automation growth	0.25	0	0.5

Note: Bounds for the uniform distributions. Notation: M $\rightarrow$ median; LB $\rightarrow$ lower bound; UB $\rightarrow$ upper bound. The parameters σ_p, σ_w, σ_υ, and σ_{α_k} are normalized so that impulse responses are computed conditional on a long-run change in $\mathcal{M}_{p,t}$, $\mathcal{M}_{w,t}$, Υ_t, and $\alpha_{k,t}$ of 1 percent.

restrictions. The rest of this section lays out the details of our empirical strategy.

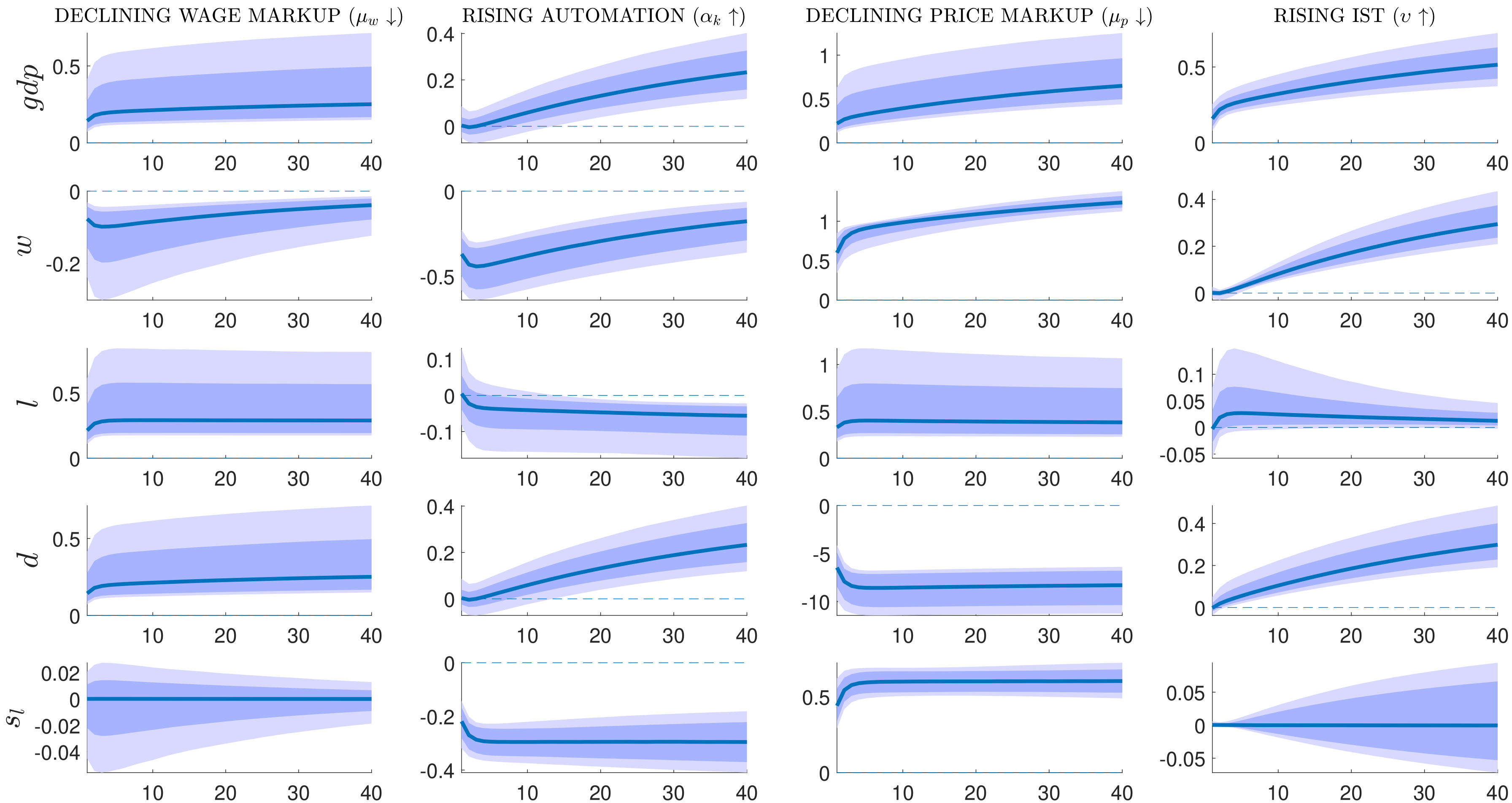
3.1 Step one – Theory robust sign restrictions

The objective in step one is to establish a set of theory robust restrictions that we can use to separately identify the potential structural forces at play in the empirical model. The exercise follows along the lines of Canova and Paustian (2011) and involves the following stages: first, we make one independent draw from a uniform distribution specific to each of the model's structural parameters, and gather the resulting parameter values in a vector Θ. Second, we solve the model conditional on Θ. Third, we compute and save the impulse responses implied by the model solution. Stages 1-3 are repeated 10,000 times.[3] This exercise leaves us with a distribution of impulse responses that can be used to establish combinations of sign restrictions unique to each shock under consideration.

Further details about the inferred identification scheme are laid out below, but first we make a few comments regarding the numerical approximations involved. We use perturbation methods to solve the model, which means that we must choose an initial point to start simulations from. Two issues arise here: first, the elasticity of labor's income share to various shocks depends on the initial income shares when those shocks are realized, and the model is consistent with a continuum of distinct, initial income shares. Second, $\alpha_{l,t}$ and $\alpha_{k,t}$ are not dimension-free, regardless of which starting point we consider (see Cantore, León-Ledesma, McAdam, and Willman (2014) for discussion of the latter issue). Therefore, for every simulation we draw initial income shares and add them to the

[3]Parameter combinations that violate saddle path stability are discarded.

Figure 2: Simulation results from the baseline theoretical model



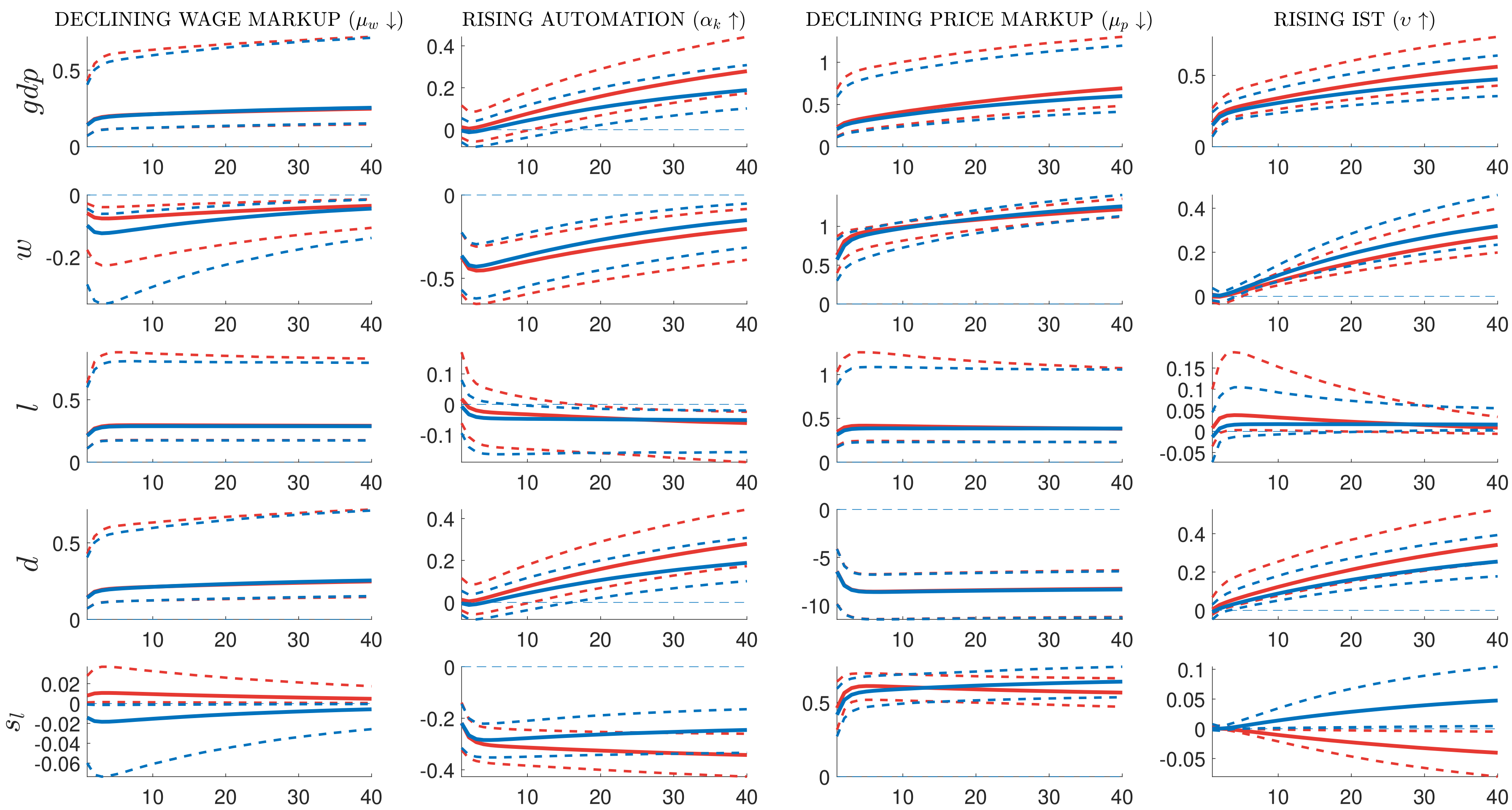
Note: Median (solid line), 90%, and 68% credible bands based on 10,000 draws. Income shares are expressed in percentage point deviations from initial values. Remaining variables are expressed in percentage deviations.

parameter vector Θ. In turn, the model is re-parameterized conditional on realized values for initial income shares. The re-parametrization follows along the lines of Cantore and Levine (2012). Initial equilibrium values of certain great ratios are fixed by setting $\beta = 0.99$ and $\delta = 0.025$. Without loss of generality we also start the simulations at $\bar{A}_{l,t} = \bar{\Upsilon}_t = \bar{L}_t = 1$. Finally, the volatility parameters σ_p, σ_w, σ_v, and σ_{α_k} are normalized so that impulse responses are computed conditional on a long-run change in $\mathcal{M}_{p,t}$, $\mathcal{M}_{w,t}$, Υ_t, and $\alpha_{k,t}$ of 1 percent. Remaining variables follow endogenously. Table 1 reports chosen bounds for the uniform distributions of parameters and initial income shares. We choose relatively wide bands for the latter, so that the initial labor income share can take all values observed in the post-war US economy (see Figure 1). Moreover, also the parameter bounds span commonly used values in the literature. The elasticity of substitution between labor and capital, for example, is centered around unity with support between 0.5 and 1.5. Applied work commonly assumes $\eta = 1$ (Cobb-Douglas production), although many empirical estimates are somewhat smaller (León-Ledesma et al., 2010). Karabarbounis and Neiman (2014), in contrast, find numbers around 1.2 or even higher.

Figure 2 summarizes the distribution of impulse responses derived from the Monte Carlo exercise. In the figure, we have normalized the two markup shocks so that the long-run effect on output is positive. Thus, all shocks considered here will eventually cause a rise in output. Our first part of the identification scheme comes from the observation that wages inevitably decline following labor markup and automation shocks, but rise in response to firm markup and investment-specific technology shocks. As such, we will attribute un-forecastable, negative co-movement between GDP and wages to labor markups or automation. We further disentangle these two by exploiting their contrasting implications for hours worked: a decline in the wage markup implies more competition among workers and is, therefore, a positive supply shock in the labor market. Working hours rise as a result. Automation, in contrast, reduces the need for firms to hire workers. As such, automation is a negative labor demand shock. Note that an increase in automation leads to a decline in wages and hours: these are precisely the macroeconomic effects of automation documented both theoretically and empirically by Acemoglu and Restrepo (2019, 2018). We remark that an increase in automation has small aggregate effects on output (in some cases even negative in the short run) but a strong re-distributive effects with large displacement of labor in favor of capital and profits. In order to distinguish between innovations to firms' markup and investment-specific technology, we note that the former leads to a decline in profits, while profits rise in response to an increase in investment-specific technology. The intuition is simple: stronger competition between firms implies lower margins and, therefore, lower profits. Higher investment productivity, on the other hand, leads to an abundance of capital and higher output. This results in more profits, even though profit margins might be unchanged. For completeness, Figure 2 also reports the impulse responses of labor income shares. Consistent with the earlier discussion, wage markup and investment technology shocks can raise or lower the labor share, depending on whether η is higher or lower than one. The median response to both shocks is exactly zero, as the distribution of η is centered around unity.

Figure 3 documents the impulse responses when we redo the simulation exercise but restrict the distribution of η. In the first case, values of η are drawn from a uniform distribution with support $[1, 1.5]$. In the second case, we instead consider values in the range $[0.5, 1]$. The remaining parameter distributions are as before. As seen from the

Figure 3: Simulation results from the baseline model: $\eta < 1$ vs. $\eta > 1$



Note: Comparison of Monte Carlo results based on 10,000 draws in i) the model with $\eta < 1$ (blue) and ii) the model with $\eta > 1$ (red). Median (solid line) and 90% credible bands (dotted lines). Income shares are expressed in percentage point deviations from initial values. Remaining variables are expressed in percentage deviations.

Table 2: Baseline sign restrictions

	Labor's $\mathcal{M}_w \downarrow$	Automation $\alpha_k \uparrow$	Firms' $\mathcal{M}_p \downarrow$	IST $\Upsilon \uparrow$
GDP	+	+	+	+
Wages	-	-	+	+
Hours	+	-	/	/
Profits	/	/	-	+

Note: Sign restrictions on impulse responses in the empirical models. The restrictions are imposed at quarter 16 in the baseline specification.

figure, an increase in investment-specific technology, for example, unambiguously lowers the labor share if $\eta > 1$, while the labor share increases if $\eta < 1$. More intriguingly, the labor share is the only variable that depends *quantitatively* on the parametrization of η. For all other variables, the impulse responses are very similar. This implies an important insight—if we were to estimate the model and its parameters directly, then it would be difficult to obtain a sharp identification of the shocks driving labor income shares from the quantitative responses of GDP, wages, and so on. If anything, the results in Figure 3 suggest that the "impossibility theorem" by Diamond et al. (1978) applies also in a context where a system of model equations is fitted quantitatively to empirical moments. For this reason we choose to infer whether or not η is larger than one indirectly.

A potential issue with the analysis so far concerns the measurement of profit income, which in data might be distorted by the inclusion of some unobserved, intangible capital (Karabarbounis and Neiman, 2018). However, as shown in Appendix A.3.1, our sign restrictions hold even if one takes the extreme view that *all capital income* is counted as profits in data. As an additional robustness test, we also analyze the role of real and nominal frictions, and find that impulse response signs are unaffected by these from quarter 16 and onwards (see Appendix A.3.2). Our restrictions are satisfied in the medium run also in a version of the model with sticky investment prices (results are available upon request), although the *impact* responses might differ (see Basu, Fernald, and Liu (2012) for further discussion). Next, we lay out the details of the second step in our empirical strategy.

3.2 Step two – Empirical specification

The simulation results just described allow us to construct theory robust sign restrictions which separately identify all four shocks under consideration. The sign restrictions used in our baseline SVAR model are derived from Figure 2 (or Figure 3) and summarized in Table 2. Combined, they account for all variation in data. Note however, that the signs need not hold in the short run. Rather, we use them as medium- to long-run restrictions in the empirical analysis. Our baseline identification scheme is one where the signs are imposed 16 quarters after shocks are realized, although alternative frequencies are explored in the robustness section. The focus on permanent shocks and the use of medium-run restrictions set us apart from the standard use of SVARs to study business cycle fluctuations as in Furlanetto, Ravazzolo, and Sarferaz (2019) among others.

For the empirical analysis we consider the following reduced form VAR model:

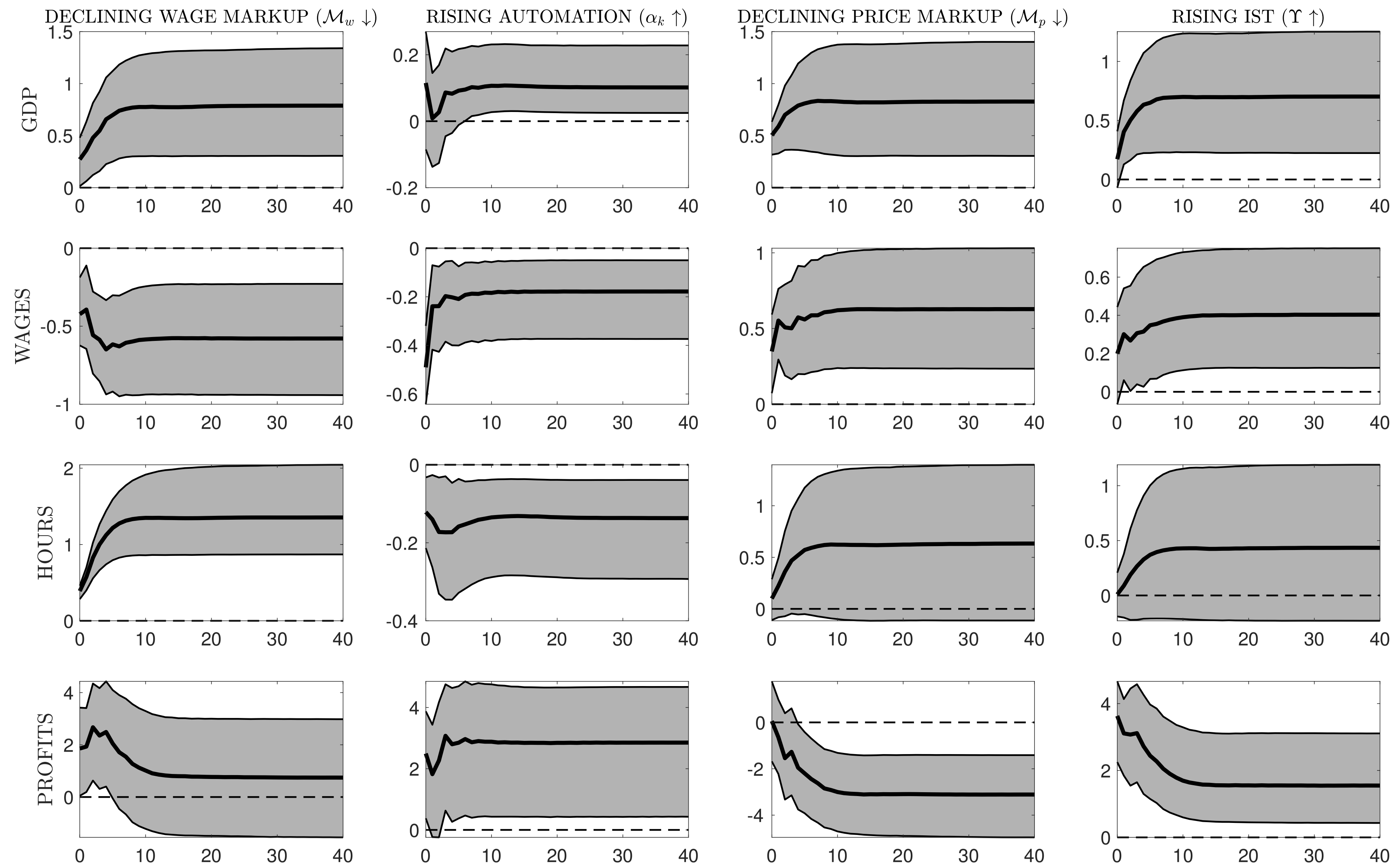
$$Y_t = C + \sum_{j=1}^{p} A_j Y_{t-j} + u_t \tag{12}$$

where Y_t is a nx1 vector containing all the endogenous variables, C is a nx1 vector of constants, $A_1, ..., A_p$ are the nxn matrices of coefficients associated with the p lags of the dependent variable and $u_t \sim N(0, \Sigma)$ is the nx1 vector of reduced form residuals. We estimate the VAR model using Bayesian methods and the variables in first differences. This specification of the empirical model is motivated by our theoretical framework where all variables follow separate stochastic trends conditional on the four shocks under consideration (wage markup, price markup, automation and investment-specific technology). Thus, we consider an empirical framework with *permanent shocks*. We specify flat priors for the reduced form parameters so that the posterior distribution has the usual Normal-Inverse-Wishart form and the information in the likelihood is dominant. In order to map the economically meaningful structural shocks from the estimated residuals, we need to impose restrictions on the variance-covariance matrix previously estimated. In particular, let $u_t = A\epsilon_t$, where $\epsilon_t \sim N(0, I_n)$ is the nx1 vector of structural disturbances with unit variance. A is a non-singular parameter matrix such that $AA' = \Sigma$. In order to identify all the shocks in the system, we need at least $\frac{n(n-1)}{2}$ additional restrictions. The sign restrictions summarized in Table 2, which are mutually exclusive and jointly exhaustive, are sufficient to set apart our four structural shocks of interest. The signs are imposed using the QR decomposition algorithm proposed by Rubio-Ramírez, Waggoner, and Zha (2010).[4]

Our dataset is quarterly and spans the period 1983Q1-2018Q3. Consistent with the identification scheme summarized in Table 2, the set of endogenous variables Y_t includes four variables for the US economy: real GDP per capita, real hourly wages, hours worked per capita, and real per capita corporate profits after tax with inventory valuation and capital consumption adjustments. The first three variables are taken for the nonfarm business sector so that their combination results in BLS's headline measure of the labor share. The latter variable is taken from the BEA and has been used by De Loecker and Eeckhout (2017) to externally validate their measure of profits, although they focus on the non-financial corporate sector. We take the log of all variables and then the first difference. The resulting series are multiplied by 100. The baseline model is estimated using 4 lags. As mentioned in the previous section, we impose our sign restrictions after 16 quarters, since at that horizon they are satisfied for nearly all parameterizations in our theoretical model (cf. in particular the response of output to an automation shock and the response of hours to an investment-specific change). Nonetheless, we checked the robustness of our main results by changing the horizon at which the medium-run restrictions are imposed and the number of lags we include in the system (see Section 5). The impulse responses of the labor share are then backed out from the impulse responses of real GDP, real wages and hours worked. Specifically, as the variables in the system are in natural logarithms, the impulse responses of the labor share can be simply computed as a linear combination

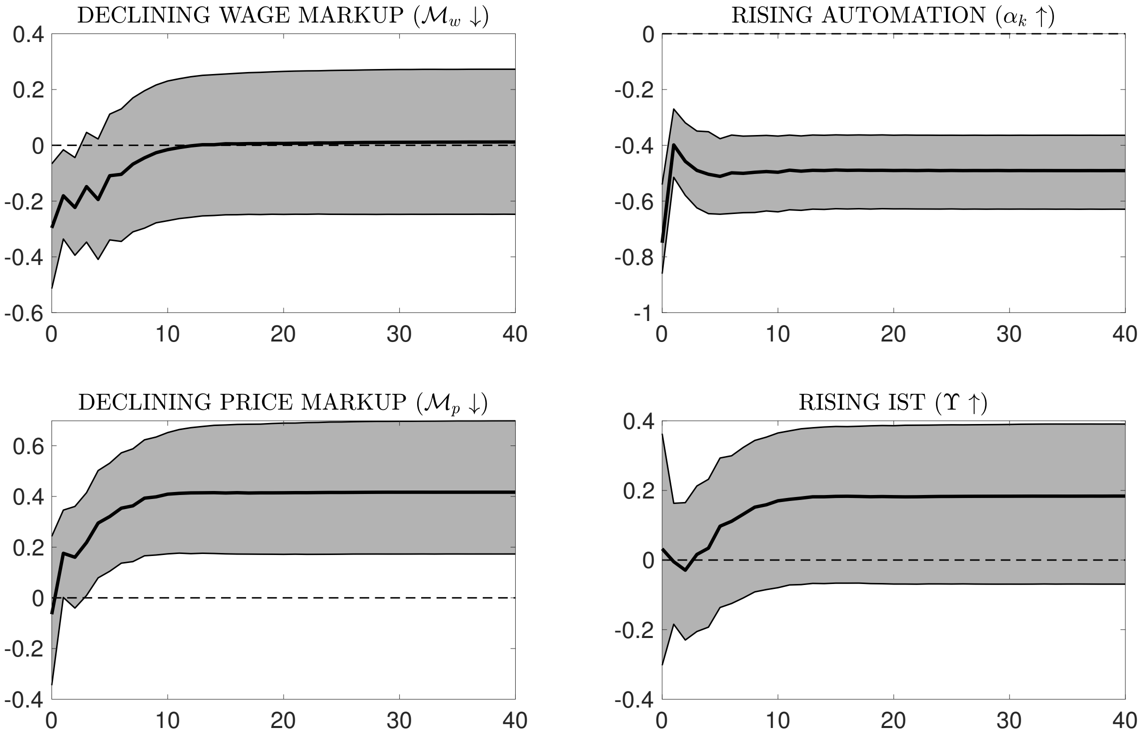
[4] Additional details on the Bayesian estimation of the reduced form VAR model and on the QR algorithm are provided in Appendix B.

Figure 4: Empirical impulse responses from the baseline VAR model



Note: Posterior distributions of cumulated impulse responses to an estimated shock of one standard deviation using the baseline identifying restrictions. Median (solid line) and 68% probability density intervals (shaded area) based on 10,000 draws. The median and the percentiles are defined at each point in time.

Figure 5: Implied labor share responses to structural change



Note: Posterior distributions of cumulated impulse responses of the labor income share to an estimated shock of one standard deviation using the baseline identifying restrictions. Median (solid line) and 68% probability density intervals (shaded area) based on 10,000 draws. The median and the percentiles are defined at each point in time.

of the impulses of its components:

$$IRF_{LS,j} = IRF_{wages,j} + IRF_{hours,j} - IRF_{GDP,j} \quad \text{for } j = 0, \ldots, J$$

The same approach is used when we compute variance decompositions as well as the historical decomposition of the labor share data.

4 Results

This section documents our main empirical results, obtained from the estimated SVAR model.

4.1 Labor share responses to structural change

We first use the estimated model to ask the following question: how does the labor income share respond to permanent changes in wage markups, automation, price markups and investment-specific technology? Empirical cumulated impulse responses for the four variables included in the SVAR are reported in Figure 4. The implied labor share responses are documented in Figure 5. In both figures the horizontal axis measures time in quarters from impact to 40 quarters after innovations have occurred. The vertical axis represents the responses in percent.

We start by considering a negative wage markup shock. This shock can be interpreted, for example, as a decrease in the bargaining power of workers. It leads to higher GDP and hours, while wages drop. Also, without any restrictions on profits we obtain a persistent rise in the majority of draws. This is consistent with the theoretical framework. The more interesting feature is the labor share response. The median response decreases significantly in the short run but then goes back towards zero. Recall that the theoretical model implies a zero long-run effect on the labor share of shocks to the wage markup. Intriguingly, a short-run decline in the labor share after falling wage markups is only consistent with complementarity between labor and capital. This is our first piece of indicative evidence about the likely size of η.

Next we consider the responses to a positive automation shock, identified by a rise in GDP at quarter 16, combined with negative wage and hours responses in that period. While the long-run dynamics of these variables are in line with the identification scheme, the very short-run effect on GDP is ambiguous. This is consistent with findings by Acemoglu and Restrepo (2018), who argue that automation might reduce economic activity in a transition period as firms and workers prepare for more automated production technologies. Without restricting profits, we also obtain a positive response as in the theoretical framework. The labor share, in contrast, decreases substantially and on a permanent basis.

The macroeconomic responses to an expansionary price markup shock are reported in the third column in Figure 4. This shock is assumed to raise output and wages, while at the same time lowering profits at quarter 16. We note that hours, which are left unrestricted, increase for the bulk of draws. More importantly, the labor income share that is plotted in Figure 5 rises unambiguously as in the theoretical model, at least when we consider responses beyond the very short run. At lower frequencies the median labor share response is sizeable.

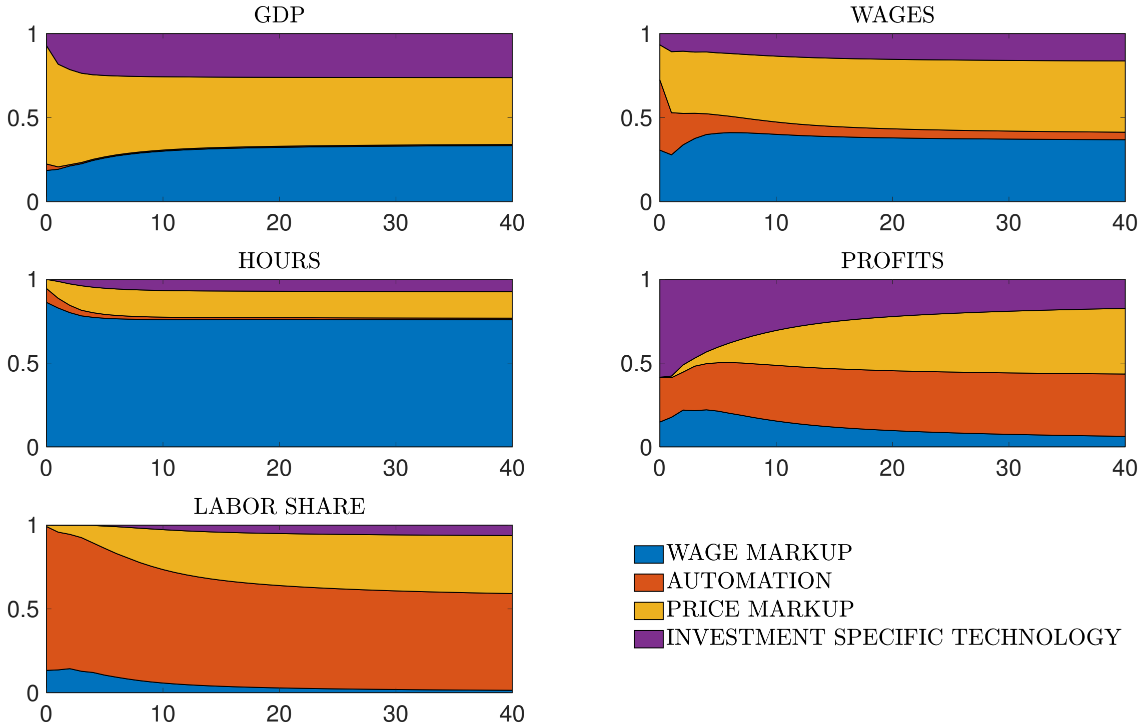
Finally, the last column in Figure 4 documents how an investment-specific technology shock affects the observables in our model. GDP, wages and profits increase by assumption (at quarter 16), but hours tend to rise too. More interestingly, after a few quarters the labor share responds positively in the vast majority of draws. This is shown in Figure 5. Thus, the VAR is informative about the sign of the labor share response despite not imposing any restriction on this variable. From a theoretical point of view, the investment shock implies rising productivity of capital relative to labor. A positive labor share response in our empirical model is, therefore, consistent with an elasticity of substitution between labor and capital smaller than one. This is our second piece of evidence in favor of net capital-labor complementarity.

4.2 What are the main drivers of the labor share?

Next we ask the model to quantify the relative importance of the four, structural shocks under consideration. To this end we compute the share of the variance of a given variable attributable to each shock in the system. This is done at different frequencies from impact to 40 quarters ahead. Figure 6 shows the results.

Importantly, we find that at least half of the variation in the labor income share is due to automation. The role of automation is even more prominent in the short run, where it

Figure 6: Variance decompositions at different frequencies

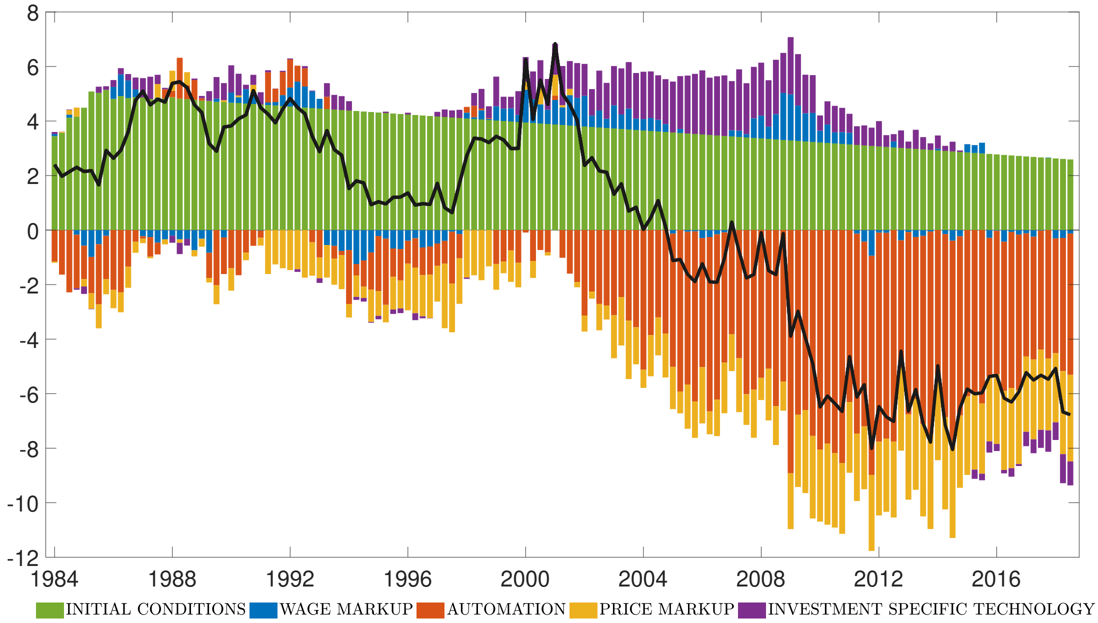


Note: The colored areas represent the point-wise median contributions of each identified shock to the forecast error variance of each variable (in levels) at horizons $j = 0, 1, \ldots, 40$ using the baseline identifying restrictions.

accounts for over 80% of fluctuations. At longer horizons, the remaining fraction of labor share fluctuations is mostly attributable to price markup shocks, while investment-specific technology only plays a very minor role. Wage markups have some explanatory power in the short run but their importance becomes negligible at longer horizons. This latter results is consistent with our theoretical model where wage markups are irrelevant for the labor share in the long run.

All in all, automation and firms' markups are dominant drivers of the US labor income share, while investment or capital biased technology are not. Such an important role for automation emerging from the SVAR can be explained on intuitive grounds. A positive automation shock increases output in the medium run and lowers wages and total hours, in keeping with the effects discussed in Acemoglu and Restrepo (2019). With the labor share defined as total labor income over output, we remark that the response of each variable favors a decline of the labor share. Put simply, the numerator of the labor share decreases, while the denominator increases. No other shock generates such a negative co-movement: investment-specific technology shocks increase wages, a decline in the bargaining power of workers increases total hours worked and, finally, an increase in markups generates a decline in output. Each of these effects in isolation pushes for an increase of the labor share. Of course, the response of the other variables may overturn the sign of the labor share response (as is the case for price markup and wage markup shocks). Nevertheless, only the automation shock generates the right movements in all components of the labor share, so that its decline becomes quantitatively important.

Figure 7: A historical decomposition of the labor share



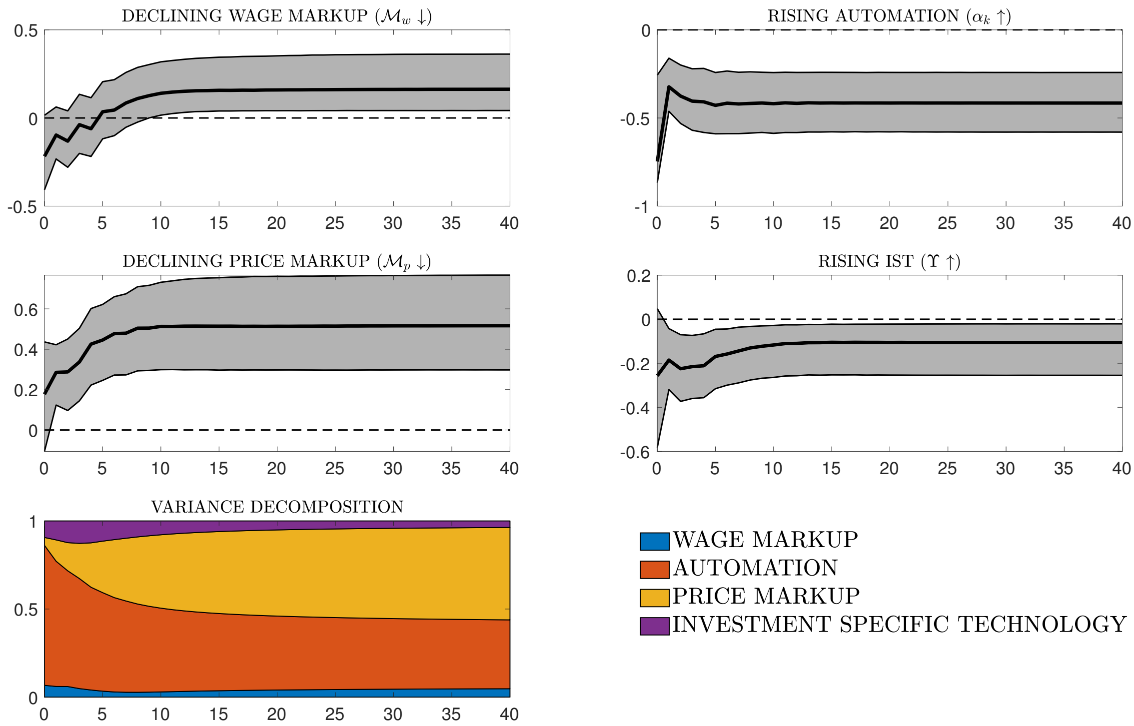
Note: The colored bars present the evolution of the labor share, in deviations from its mean, attributable to the deterministic components (initial conditions) of the VAR and to each structural shock, based on the median-target model of Fry and Pagan (2011).

Next we discuss the role of the four identified shocks for remaining variables in the system. A couple of remarks are warranted: first, investment-specific technology as well as wage and price markup shocks explain the bulk of variations in GDP and wages. Fluctuations in hours, in contrast, are mainly driven by wage markups and to some extent by price markups. At first glance it might seem surprising that automation, while being important for the labor share, plays such a minor role for the macroeconomy. Note however, that automation shocks are re-distributive by nature: they shift the composition of factor use but do not necessarily lead to large changes in aggregate activity (see Acemoglu and Restrepo (2018) for further discussion). In addition, they generate countercyclical wages and hours that are not standard features of economic fluctuations, despite being very useful to generate the labor share decline. Turning to profits, they are well explained by investment-specific technology in the short to medium run, while price markups and automation have significant explanatory power in the long run. Our results are broadly in line with common findings in the literature on estimated macroeconomic models where these shocks are quantified. Note, however, our departure from that literature by focusing on permanent rather than temporary shocks.

4.3 What caused the observed labor share decline?

Our final result concerns the relative importance of different explanations for the labor share decline observed in data. To this end we carry out a historical decomposition of the labor share. Figure 7 displays the labor share decomposition in deviations from its mean. A brief remark about the deterministic component (initial conditions) is warranted. This component can be interpreted as our model-based forecast of the labor share in the very

Figure 8: A thought experiment



Note: Posterior distributions of cumulated impulse responses of the labor income share to an estimated shock of one standard deviation using the baseline identifying restrictions and imposing net substitution between capital and labor. Median (solid line) and 68% probability density intervals (shaded area) based on 10,000 draws. The median and the percentiles are defined at each point in time. The colored areas represent the point-wise median contributions of each identified shock to the forecast error variance of the labor share at horizons $j = 0, 1, \ldots, 40$.

beginning of the sample, given the estimated VAR coefficients and the initially discarded observations. That forecast entails an evolution of the labor share broadly in line with its initial observations. That is, the deterministic component does not play a big role in explaining the secular labor share decline (see Giannone, Lenza, and Primiceri (2018) for a discussion on the pathological behavior of initial conditions in VAR models).

Turning to the structural shocks of interest, it is clear that, according to our model, automation and firms' rising market power are key for understanding the post-2000 labor share evolution. Automation has become an increasingly important factor since the early 2000s, while the role of price mark-up shocks is particularly important after 2009, in keeping with the decomposition presented in Giannoni and Mertens (2019). These results corroborate well with the view put forward by Elsby et al. (2013): reasonable explanations for the labor share decline should be consistent with the timing of this decline. Automation and the large increase in profits are recent phenomenons whose timing correlates well with the sharp decline of the labor share. In contrast, investment-specific technological progress and the decline in unionization, which could proxy a decline in wage mark-ups, started long before the beginning of the new century.

Finally, we find that investment-specific shocks, if anything, have led to a mild increase in the labor share. This is particularly true during most of the 2000s. Again, the conditional labor share increase following rising investment-specific technology suggests net complementarity between labor and capital. Karabarbounis and Neiman (2014), in contrast, argue that labor and capital may be net substitutes. Given the debate on the degree of labor-capital substitutability, we now consider the following thought experiment: for the sake of argument, we suppose for a second that capital and labor are net substitutes and we restrict the labor share accordingly in the SVAR. We thus impose that the labor share is procyclical in response to price and wage mark-up shocks and countercyclical in response to automation and investment-specific shocks (at a horizon of 16 quarters ahead). The impulse responses presented in Figure 8 reflect the identification assumptions and are thus not particularly informative. However, the variance decomposition obtained from this non-agnostic exercise is far more interesting. In fact, although we impose that positive investment-specific shocks must lower the labor share, these shocks turn out to be quantitatively unimportant. We see in Figure 8 that automation and price mark-up shocks are still the dominant drivers of the labor share decline, as in our baseline model. This exercise constitutes, we believe, an important validation for the main results of our paper.

5 Robustness and extensions

In this section we check the robustness of our baseline results to a battery of sensitivity checks and perform a number of extensions to the baseline specification. We include all the figures related to this section in Appendix C.

5.1 Alternative horizons, lag specifications, samples and priors

The baseline SVAR model presented in the previous section is estimated using 4 lags, imposing the sign restrictions of Table 2 at a horizon of 16 quarters ahead, using the variables in differences with a flat prior and on a quarterly sample that spans 1983Q1-2018Q3. We check the robustness of our results to changes in all of these specifications. For the sake of exposition, we present only the variance decompositions of the labor share corresponding to the different sensitivity checks, but the complete set of results is available upon request. The first two rows of Figure C.1 present the variance decompositions of the labor share using, respectively, different horizons and lag specifications. Changing the horizon at which the sign restrictions are imposed does not seem to affect the results presented in the previous section. The same is true if we use a different lag specification, although the role of price markups in explaining labor share fluctuations becomes slightly higher at long-run horizons when we include more feedback in the system. In the first two panels of the third row, we first expand the sample to go back to 1948Q1 and then restrict it from 1990Q1 onwards. Interestingly, price markups seem to have significantly less explanatory power in the first decades after the second World War. This evidence supports the view that firms' market power started to rise in the beginning of the 1980s and then accelerated in the 1990s and 2000s. The third panel, instead, presents the variance decomposition of the labor share using annual data in the estimation of the baseline model. The three panels of the fourth row refer to three different exercises. In the previous two, we use two different prior specifications: we estimate the VAR in levels using the dummy observation prior proposed by Sims and Zha (1998) and the priors for the long run (PLR) of Giannone et al. (2018), which resemble our baseline specification in differences when infinitely tight. Differently from our baseline empirical framework, in these cases shocks do not necessarily have permanent effects. In the latter, we consider the median-target impulse responses proposed by Fry and Pagan (2011). Overall, the results are in line with our baseline, although price markups seem slightly less relevant when we use the VAR in levels with the sum of coefficients prior or PLR.

5.2 Alternative measures of the labor share

In this subsection we present the robustness of our baseline results to different measures of the labor share. A first step in this direction is to focus on the business sector, which includes the farm sector and therefore covers a larger share of the economy compared with our baseline specification. We adjust the other variables in the system (real GDP, real hourly wages and hours) accordingly.

As discussed in depth in the literature, BLS's headline measure of the labor share suffers from measurement issues related to the treatment of self-employment (proprietors') income. This income component is ambiguous in that it reflects returns on both work effort and investment, and thus there is no straightforward way to isolate which part accrues to labor and which to capital. The way that the BLS tackles this problem is by assuming that the self-employed pay themselves the average hourly wage of workers on payroll. Therefore, the part of self-employment income accruing to labor is obtained by multiplying the average hourly wage by the hours worked by the self-employed. The remaining part of proprietors' income accrues to capital. The headline measure of the labor share is then constructed as follows:

$$s_{l,t} = \frac{W_t L_t}{Y_t} = \frac{W_t^p L_t^p}{Y_t} + \frac{W_t^s L_t^s}{Y_t}$$

$W_t = W_t^p = W_t^s$ denotes the average hourly wage of workers on payroll, L_t^p is hours of payroll workers and L_t^s is hours of self-employed. As noted by Elsby et al. (2013), this assumption results in an overstatement of the overall decline of the labor share and implies a negative capital share in the proprietors' sector in the 1980s, which is clearly an implausible outcome and casts doubts on the reliability of this measure.

One way to address the measurement issues related to the self-employed is to focus on the payroll share, $s_{l,t}^p = \frac{W_t^p L_t^p}{Y_t}$, thus taking proprietors' income out of the picture and considering a measure which unambiguously reflects payments to labor only. This can be conveniently implemented in our baseline empirical framework by simply substituting, in our set of endogenous variables, hours in the nonfarm business sector L_t with hours of payroll workers in the nonfarm business sector L_t^p, which are available on the website of the BLS. Then the responses of the payroll share to the different shocks for the nonfarm business sector are constructed exactly as in the previous section.

Another alternative to bypass this problem is to consider the labor share for the nonfinancial corporate sector, as proposed by Karabarbounis and Neiman (2014). This measure does not suffer from the issues related to proprietors' income as corporations must declare payrolls and profits separately for fiscal purposes. Thus, there is no ambiguity in the way income is treated and how it accrues to labor or capital. However, this comes at the cost of focusing on a smaller share of the US economy. In our empirical framework, we can restrict our interest to real GDP per capita, real wages, hours per capita and real profits per capita for the nonfinancial corporate sector. The labor share for the nonfinancial corporate sector is then obtained by combining the first three variables, as shown in the previous section.

Finally, we can also consider the economy-wide measure proposed by Kravis (1959) and used by Gomme and Rupert (2004). The assumption underlying this measure is that the self-employment labor share is the same as that of the overall economy. Specifically, the economy-wide measure is defined as follows:

$$s_{l,t}^{ew} = \frac{W_t^p L_t^p}{Y_t - Y_t^s}$$

where Y_t^s is proprietors' income without consumption allowances and inventory valuation adjustment. This is effectively performed in our baseline empirical framework by including in the system $Y_t - Y_t^s$ instead of GDP and hours of payroll workers instead of L_t.

We present the results of different measures of the labor share (business sector, payroll, non-financial corporate and economy-wide) using our baseline restrictions of Table 2 in Figure C.2. Our baseline results are robust across different definitions of the labor share. In particular, the responses of the labor share to investment-specific technology shocks are positive for the bulk of the draws and negative on impact in response to wage markups regardless of the measure used, confirming the evidence in favor of capital-labor complementarity. Price markups become slightly more important for low frequency fluctuations in the labor share when we restrict our analysis to the non-financial corporate sector. This result is not too surprising given a relatively large increase in the profits-to-GDP ratio for the non-financial corporate sector compared with the overall economy.

5.3 Additional restrictions

In our baseline specification of Table 2, we have shown a minimum set of identifying restrictions that are sufficient to set apart the four economically meaningful structural shocks under consideration. Our theoretical model, however, gives us additional sign restrictions that we could potentially exploit to sharpen the identification of the structural shocks. Figure C.3 shows the impulse responses and variance decompositions of BLS's headline measure of the labor share to the different structural shocks adding additional restrictions to the minimal set of Table 2. In the first row, we add the restriction that hours increase at horizon 16 in response to a negative price markup shock. In the second, we impose the restriction that hours increase at horizon 16 in response to a positive investment-specific technology shock. In the third, we restrict all the variables in the system: hours increase in response to both price markup and investment-specific technology shocks at horizon 16, and profits increase in response to wage markups and automation at horizon 16. Our baseline results are largely confirmed by adding additional restrictions. If anything, the evidence in favor of capital-labor complementarity becomes stronger once we restrict hours to increase in response to investment technology.

5.4 MORE VARIABLES

To externally validate our empirical methodology and, generally, the identification of the different shocks in the system, we can include additional variables in the VAR and analyze the behavior of these in response to the different shocks in the system. In what follows, we add, one by one, the following *unrestricted* variables to the system: relative price of investment, real investment per capita, routine and non-routine employment. We then estimate the augmented SVAR with the same baseline restrictions exposed in Table 2. The results are reported in Figure C.4. The first row presents the impulse responses and the variance decomposition of the relative price of investment. In line with the theoretical model, movements in the price of investment are explained mainly by investment-specific technology shocks, which lead to a permanent decrease in this variable. This is reassuring because we would expect the investment-specific shock to lead to a long-run decrease in the relative price of investment, if correctly identified. The second row presents the results for real investment. Investment increases significantly and permanently in response to wage markup, price markup and investment-specific technology shocks. Two results stand out: first, in line with the work of Gutierrez and Philippon (2017), price markups are important drivers of investment, and thus might help to explain its recent slowdown. Second, automation does not seem to affect this variable at all, in line with its small effect on aggregate variables observed in Section 4. The third and fourth rows show the responses of routine and non-routine per capita employment to the different structural disturbances using the series by Zhang (2019), which in turn updates the data constructed by Jaimovich and Siu (2019). Interestingly, while wage markups are important drivers of both routine and non-routine employment, automation has a much stronger negative effect on routine employment at short horizons. Moreover, price markups are also more important for routine than non-routine employment. When compared to the variance decomposition of hours in Figure 6, non-routine employment has an extremely similar behavior to hours, whereas routine employment features a much larger role, on impact, for automation and, over the entire horizon, for price markups.

5.5 TAKING STOCK—HOW TO INTERPRET OUR IDENTIFIED SHOCKS

In this section we further discuss the interpretation of the two main drivers of the labor share according to our empirical model, i.e. automation and rising price markups. When interpreted literally, the price markup shock captures a decline in domestic competition in US industries—for example driven by increasing entry costs, lax antitrust enforcement or lobbying.[5] This interpretation is perhaps the most widespread and consistent with the evidence presented by Barkai (2018), De Loecker and Eeckhout (2017), Eggertsson et al. (2018), Gutierrez and Philippon (2017), and Grullon, Larkin, and Michaely (2019). All these papers argue that the observed increase in market concentration is associated with higher firm profitability, mainly because of rising profit margins rather than improved technological efficiency.

However, a more benign interpretation of the increase in market concentration points to technical change, perhaps in combination with barriers to entry. Autor, Dorn, Katz, Pat-

[5] Similar stories would also fit well, such as a rise in firms' monopsony power in labor markets, as studied by Alpanda (2019).

terson, and Van Reenen (2017) find that the increase in concentration in several industries reflects changes in the economic environment, favoring a few very productive "superstar" firms. Kehrig and Vincent (2018) obtain similar results but with a focus on the manufacturing sector. These results may, at least in part, reflect better search technologies which favor certain firms or industries. For example, the rise in concentration could come about due to the emergence of platform competition or advances in information technology, as argued by Aghion, Bergeaud, Li, Klenow, and Boppart (2019). It is not clear where such a technology driven rise in market concentration and profit income would show up in our framework. This depends to a large extent on the associated productivity effects: on the one hand, a rise in market concentration is likely to hamper economic activity and thereby reduce aggregate output, consistent with our identified markup shock. But simultaneous productivity improvements (e.g. due to the scale effects of advances in platform technologies), on the other hand, are likely to work the opposite way. Indeed, aggregate output may even rise if the productivity effects are sufficiently powerful, implying a positive conditional correlation between output and profits (but still with contractionary effects on the labor market). If the productivity effect dominates, then a rise in profit revenues will necessarily be interpreted as an automation shock in our model.

Finally, it is important to stress that this paper is rather silent regarding the role of globalization. An automation shock, for example, could also capture the effects of globalization, import competition or offshoring out of the US economy, as originally proposed by Elsby et al. (2013). These shocks may indeed imply an increase in output and a decline in wages and hours worked, although there is no agreement in the literature on the magnitude of such displacement effects of globalization on the US labor market. We remark, however, that recent analysis of the labor share at the industry or establishment level downplay the importance of open economy factors for labor share dynamics (Giannoni and Mertens, 2019; Autor et al., 2017). Nevertheless, we acknowledge that disentangling the automation narrative from globalization narratives is an important topic for future research.

6 Conclusions

The labor share of national income has fallen in many countries in the last decades. In the US, the labor income share has accelerated its decline since the beginning of the new century, reaching its postwar lowest level in the aftermath of the Great Recession. While this observation has led to substantial interest, both among policy makers and in the popular press, a consensus view regarding the structural forces at play is still lacking. In this paper, we quantify and interpret four main explanations for the secular decline of the US labor income share. To this end we estimate a time series model with permanent shocks, identified with theory robust sign restrictions. To the best of our knowledge, this is the first paper to quantify the relative importance of these forces within a unified framework. Moreover, our econometric approach to achieve identification differs fundamentally from previous literature: while most studies draw inference based on cross-sectional variation in microeconomic data (at the firm or sectoral level), we instead exploit the time series properties of macroeconomic data, thus providing a potentially useful complement to the existing literature.

Our main empirical results can be summarized as follows: first, in the postwar US economy, automation and firms' rising market power unambiguously lower the labor income share, while capital deepening in the form of higher investment-specific technology growth tends to raise it. The latter result suggests that labor and capital are net complements as factors of production. Second, the estimated model assigns a major role for automation and firm markups as drivers of labor's income share, especially at lower frequencies. The labor share implications of shocks to labor's markup and investment efficiency, in contrast, are not supported by aggregate time series data on labor income. Third, we decompose the historical evolution of the US labor income share and find that most of the pre-crisis decline can be attributed to automation, while firms' rising market power has been the main source of lower labor shares since the Great Recession. Interestingly, our historical decomposition suggests that investment-specific technology has tended to raise the US labor income share, at least in the 2000s. This latter finding is consistent with the view that investment-specific technology growth has indeed taken place in recent decades, but that this has stimulated labor's income share because of complementarity between labor and capital. Finally, we document that our empirical results are robust to a large battery of robustness tests, including various identification assumptions and the use of different measures of labor's income share.

While this paper offers a benchmark account of the evolution of labor shares in the post-war US economy, we ignore distributional aspects such as those studied by Moll, Rachel, and Restrepo (2019). Extending our setup to study questions related to the cross-sectional effects of automation and market power will likely be an important topic for our future research.

REFERENCES

Acemoglu, D. and P. Restrepo (2018). The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares and Employment. *American Economic Review 108*(6), 1488–1542.

Acemoglu, D. and P. Restrepo (2019). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy* (Forthcoming).

Aghion, P., A. Bergeaud, H. Li, P. Klenow, and T. Boppart (2019). A Theory of Falling Growth and Rising Rents. Working Paper Series 2019-11, Federal Reserve Bank of San Francisco.

Alpanda, S. (2019). Business Cycles with Oligopsonistic Competition in Labor Markets. Technical report, University of Central Florida.

Autor, D., D. Dorn, L. Katz, C. Patterson, and J. Van Reenen (2017). The Fall of the Labor Share and the Rise of Superstar Firms. CEPR Discussion Papers 12041, C.E.P.R. Discussion Papers.

Autor, D. and A. Salomons (2018). Is automation labor displacing? Productivity growth, employment, and the labor share. *Brookings Papers on Economic Activity 49*(1 (Spring), 1–87.

Barkai, S. (2018). Declining Labor and Capital Shares. Technical report, London Business School.

Basu, S., J. Fernald, and Z. Liu (2012). Technology Shocks in a Two-Sector DSGE Model. 2012 Meeting Papers 1017, Society for Economic Dynamics.

Blanchard, O. and F. Giavazzi (2003). Macroeconomic Effects of Regulation and Deregulation in Goods and Labor Markets. *The Quarterly Journal of Economics 118*(3), 879–907.

Caballero, R. J., E. Farhi, and P.-O. Gourinchas (2017). Rents, Technical Change, and Risk Premia Accounting for Secular Trends in Interest Rates, Returns on Capital, Earning Yields, and Factor Shares. *American Economic Review 107*(5), 614–620.

Canova, F. and M. Paustian (2011). Business cycle measurement with some theory. *Journal of Monetary Economics 58*(4), 345–361.

Cantore, C., M. León-Ledesma, P. McAdam, and A. Willman (2014). Shocking Stuff: Technology, Hours, And Factor Substitution. *Journal of the European Economic Association 12*(1), 108–128.

Cantore, C. and P. Levine (2012). Getting normalization right: Dealing with dimensional constants in macroeconomics. *Journal of Economic Dynamics and Control 36*(12), 1931–1949.

Ciminelli, G., R. A. Duval, and D. Furceri (2018). Employment Protection Deregulation and Labor Shares in Advanced Economies. IMF Working Papers 18/186, International Monetary Fund.

De Loecker, J. and J. Eeckhout (2017). The Rise of Market Power and the Macroeconomic Implications. NBER Working Papers 23687, National Bureau of Economic Research, Inc.

Diamond, P., D. McFadden, and M. Rodriguez (1978). Measurement of the Elasticity of

Factor Substitution and Bias of Technical Change. In M. Fuss and D. McFadden (Eds.), *Production Economics: A Dual Approach to Theory and Applications*, Volume 2 of *Histoy of Economic Thought Chapters*, Chapter 5. McMaster University Archive for the History of Economic Thought.

Drautzburg, T., J. Fernández-Villaverde, and P. Guerrón-Quintana (2017). Political Distribution Risk and Aggregate Fluctuations. NBER Working Papers 23647, National Bureau of Economic Research, Inc.

Eggertsson, G. B., J. A. Robbins, and E. G. Wold (2018). Kaldor and Pikettys Facts: The Rise of Monopoly Power in the United States. NBER Working Papers 24287, National Bureau of Economic Research, Inc.

Elsby, M., B. Hobijn, and A. Sahin (2013). The Decline of the U.S. Labor Share. *Brookings Papers on Economic Activity 44*(2 (Fall)), 1–63.

Fisher, J. D. M. (2006). The Dynamic Effects of Neutral and Investment-Specific Technology Shocks. *Journal of Political Economy 114*(3), 413–451.

Fry, R. and A. Pagan (2011). Sign Restrictions in Structural Vector Autoregressions: A Critical Review. *Journal of Economic Literature 49*(4), 938–960.

Furlanetto, F., F. Ravazzolo, and S. Sarferaz (2019). Identification of Financial Factors in Economic Fluctuations. *The Economic Journal 129*(617), 311–337.

Giannone, D., M. Lenza, and G. E. Primiceri (2018). Priors for the long run. Working Paper Series 2132, European Central Bank.

Giannoni, M. and K. Mertens (2019). Outsourcing, Markups and the Labor Share. Manuscript.

Glover, A. and J. Short (2017). Demographic Origins of the Decline in Labors Share. Manuscript.

Gomme, P. and P. Rupert (2004). Measuring labor's share of income. *Federal Reserve Bank of Cleveland Policy Discussion Papers*.

Grullon, G., Y. Larkin, and R. Michaely (2019). Are US Industries Becoming More Concentrated? *Review of Finance 23*(4), 697–743.

Gutierrez, G. and T. Philippon (2017). Declining Competition and Investment in the U.S. NBER Working Papers 23583, National Bureau of Economic Research, Inc.

Gutierrez, G. and S. Piton (2019). Revisiting the Global Decline of the (Non-Housing) Labor Share. Technical Report 811, Bank of England.

Jaimovich, N. and H. E. Siu (2019). Job Polarization and Jobless Recoveries. *Review of Economics and Statistics* (Forthcoming).

Karabarbounis, L. and B. Neiman (2014). The Global Decline of the Labor Share. *The Quarterly Journal of Economics 129*(1), 61–103.

Karabarbounis, L. and B. Neiman (2018). Accounting for Factorless Income. NBER Working Papers 24404, National Bureau of Economic Research, Inc.

Kaymak, B. and I. Schott (2018). Corporate Tax Cuts and the Decline of the Labor Share. Manuscript.

Kehrig, M. and N. Vincent (2018). The Micro-Level Anatomy of the Labor Share Decline. NBER Working Papers 25275, National Bureau of Economic Research, Inc.

King, R. G., C. I. Plosser, and S. T. Rebelo (1988). Production, growth and business cycles: I. the basic neoclassical model. *Journal of Monetary Economics 21*(2–3), 195 – 232.

Koh, D., R. Santaeulalia-Llopis, and Y. Zheng (2018). Labor Share Decline and Intellectual Property Products Capital. Working Papers 873, Queen Mary University of London, School of Economics and Finance.

Kravis, I. B. (1959). Relative income shares in fact and theory. *The American Economic Review 49*(5), 917–949.

Kumar, T. K. and J. H. Gapinski (1974). Nonlinear Estimation of the CES Production Parameters: A Monte Carlo Study. *The Review of Economics and Statistics 56*(4), 563–567.

Leduc, S. and Z. Liu (2019). Robots or Workers: A Macro Analysis of Automation and Labor Markets. *Federal reserve Bank of San Francisco Working Paper Series* (17).

León-Ledesma, M. A., P. McAdam, and A. Willman (2010). Identifying the Elasticity of Substitution with Biased Technical Change. *American Economic Review 100*(4), 1330–1357.

León-Ledesma, M. A. and M. Satchi (2018). Appropriate Technology and Balanced Growth. *The Review of Economic Studies 86*(2), 807–835.

Martinez, J. (2018). Automation, Growth and Factor Shares. Manuscript.

Moll, B., L. Rachel, and P. Restrepo (2019). Uneven Growth: Automation's impact on Income and Wealth Inequality. Manuscript.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

Rognlie, M. (2015). Deciphering the Fall and Rise in the Net Capital Share: Accumulation or Scarcity? *Brookings Papers on Economic Activity 46*(1 (Spring), 1–69.

Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010). Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies 77*(2), 665–696.

Sims, C. A. and T. Zha (1998). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review 39*(4), 949–968.

Thursby, J. (1980). Alternative CES Estimation Techniques. *The Review of Economics and Statistics 62*(2), 295–299.

Zhang, W. (2019). Deciphering the causes for the post-1990 slow output recoveries. *Economics Letters 176*(C), 28–34.

APPENDIX

A ADDITIONAL DETAILS ON THE THEORETICAL MODEL

A.1 THE INITIAL STEADY STATE USED FOR SIMULATIONS

The steady state in the baseline, theoretical model follows recursively given an initialization of s_l, s_k, s_d, A_l, Υ and L:

$$
\begin{aligned}
r &= \beta^{-1} - 1 \\
P_I &= \Upsilon^{-1} \\
r^k &= P_I\left(r + \delta\right) \\
\alpha_l &= \frac{s_l}{1 - s_d} \\
\alpha_k &= \frac{s_k}{1 - s_d} \\
\mathcal{M}_p &= \frac{1}{1 - s_d} \\
A_k &= r^k s_k^{\frac{1}{\eta-1}} \left(\frac{\mathcal{M}_p}{\alpha_k}\right)^{\frac{\eta}{\eta-1}} \\
Y &= \left(\frac{\alpha_l}{\mathcal{M}_p s_l}\right)^{\frac{\eta}{\eta-1}} A_l L \\
W &= s_l \frac{Y}{L} \\
K &= \frac{s_k Y}{r^k} \\
I &= \delta K \\
X &= P_I I \\
C &= Y - X \\
\mathcal{D} &= s_d Y \\
\epsilon_p &= \frac{\mathcal{M}_p}{\mathcal{M}_p - 1} \\
\mathcal{M}_w &= \mathcal{M}_p \\
\epsilon_w &= \frac{\mathcal{M}_w}{\mathcal{M}_w - 1} \\
\Psi &= \frac{W}{\mathcal{M}_w L^{\varphi} C} \\
\Lambda &= C^{-\sigma} \exp\left(-\Psi \frac{\left(1 - \sigma\right) L^{1+\varphi}}{1 + \varphi}\right)
\end{aligned}
$$

The steady state of the New Keynesian model (see Appendix A.3.2) is identical, except that we also have to solve for nominal variables: given a choice of gross inflation Π (we set $\Pi = 1$), we have $i_p = \frac{\Pi}{\beta} - 1$ and $\Pi_w = \Pi$.

A.2 LONG-RUN INCOME SHARES

Given the definition of a long-run equilibrium in the main text, we set out to derive expressions of long-run income shares. We start with the profit income share. It follows from the definition of profits and optimal price-setting behavior:

$$\bar{s}_{d,t} = \frac{\bar{\mathcal{D}}_t}{\bar{Y}_t} = 1 - \frac{1}{\bar{\mathcal{M}}_{p,t}}$$

Thus, the long-run profit income share depends only on firms' markup, which is assumed exogenous in the baseline model. In order to derive the long-run capital income share we note that

$$\bar{r}^k_t = \bar{\Upsilon}_t^{-1}\left[\beta^{-1} - (1-\delta)\right]$$

The expression for firms' optimal capital demand can then be used to arrive at the following long-run capital share:

$$\bar{s}_{k,t} = \frac{\bar{r}^k_t \bar{K}_{t-1}}{\bar{Y}_t} = \left(\frac{\bar{\alpha}_{k,t}}{\bar{\mathcal{M}}_{p,t}}\right)^{\eta}\left(\frac{\beta^{-1} - (1-\delta)}{\bar{\Upsilon}_t \bar{A}_{k,t}}\right)^{1-\eta}$$

This expression shows that automation (firms' markup) raises (lowers) the capital income share. The effects of investment-specific or capital-biased technologies depend qualitatively on whether or not η is higher than one. Labor-augmenting technology and labor markups have no long-run effects on the capital share. Finally, the labor income share is found by substituting the two expressions derived above into the identity $\bar{s}_{l,t} + \bar{s}_{k,t} + \bar{s}_{d,t} = 1$:

$$\bar{s}_{l,t} = \frac{\bar{W}_t \bar{L}_t}{\bar{Y}_t} = \frac{1}{\bar{\mathcal{M}}_{p,t}}\left[1 - \bar{\alpha}^{\eta}_{k,t}\left(\frac{\beta^{-1} - (1-\delta)}{\bar{\Upsilon}_t \bar{A}_{k,t}}\bar{\mathcal{M}}_{p,t}\right)^{1-\eta}\right]$$

This is the equation used in the main text.

A.3 ALTERNATIVE THEORETICAL ASSUMPTIONS

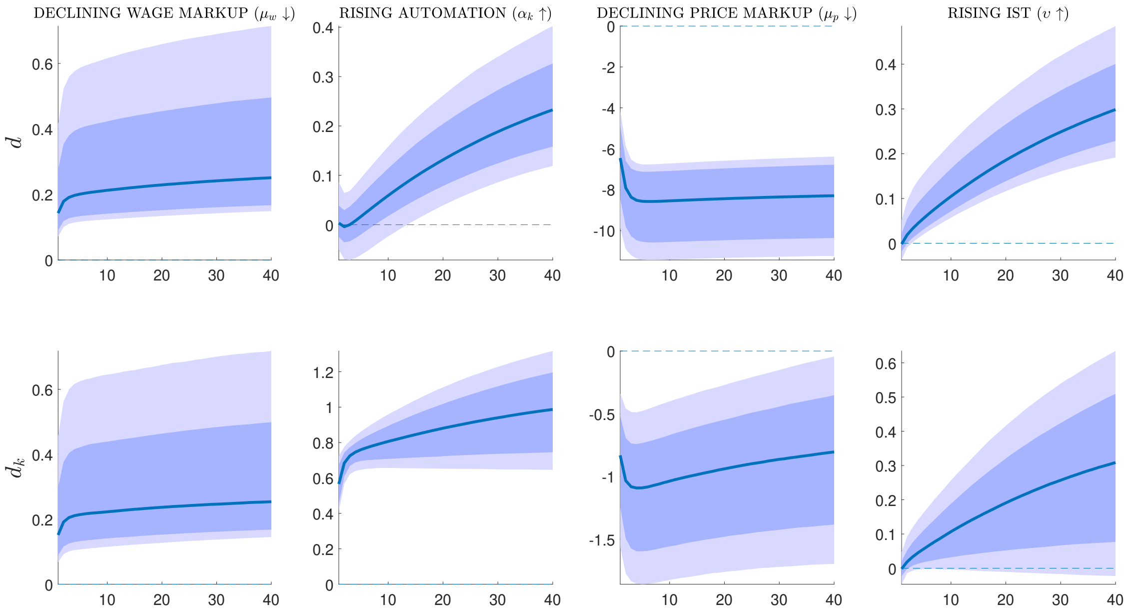
This section documents how robust our identifying sign restrictions are to (i) mis-measurement of profit income, and (ii) the inclusion of various real and nominal frictions.

A.3.1 MEASUREMENT OF PROFITS

A potential issue with the analysis in Section 3.1 in the main text concerns our measurement of profit income. The profit variable displayed in Figure 2 and Figure 3 is model-consistent and interpreted as sales net of factor payments. However, empirical measurements of profits might be distorted by the inclusion of some unobserved, intangible capital income (Karabarbounis and Neiman, 2018). Therefore, as a robustness check we now take the extreme view that *all capital income* is counted as profits in data, and simply refer to profit revenues $D_{k,t}$ as non-labor income:

$$D_{k,t} = D_t + r^k_t K_{t-1} = Y_t - W_t L_t$$

Figure A.1: Monte Carlo results: pure profits vs. non-labor income in the baseline model



Note: Median (solid line), 90%, and 68% credible bands based on 10000 draws. Pure profits (d) and non-labor income (d_k) are both expressed in percentage deviations from initial values.

Given this new measure, we re-evaluate the model's implied sign restrictions. Figure A.1 compares impulse responses of pure profits, D_t, with those of non-labor income $D_{k,t}$. The medium- to long-run signs of either variable are largely identical for all shocks. Our only disclaimer in this regard is that, conditional on investment-specific technology shocks, about 6% of the models imply a decline in $D_{k,t}$ at horizons relevant for our sign restrictions.

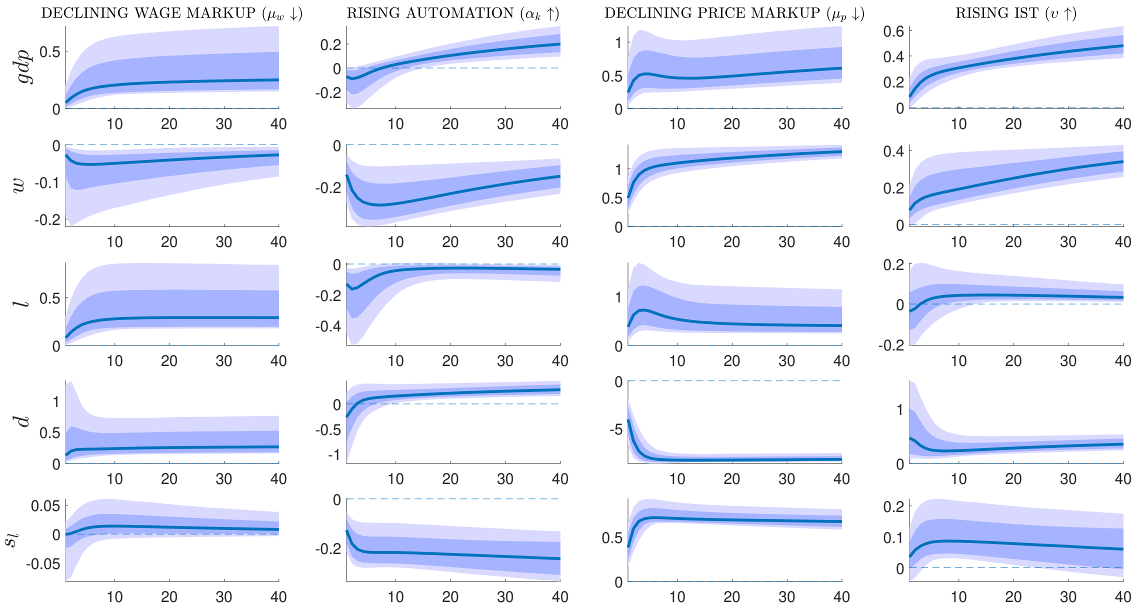
A.3.2 Real and nominal frictions

The theoretical model presented in the main text abstracts from a number of commonly used real and nominal frictions. One potential concern, therefore, is that our sign restrictions might be violated at certain frequencies if these frictions are included. This section incorporates a few "bells and whistles" into the baseline, theoretical model. We add (i) habit formation in consumption, (ii) adjustment costs in investments, (iii) variable capital utilization, (iv) nominal price stickiness, and (v) nominal wage stickiness. We also allow for partial indexation to past inflation in price and wage setting. Finally, we specify (vi) a Taylor-type rule for monetary policy. While the two models share identical long-run properties, the extended version implies different dynamics in the short to medium run. A brief summary of the additions to our baseline model follows:

External habit formation: The period utility is changed to

$$\mathcal{U}_t = \frac{(C_t - hC_{t-1})^{1-\sigma}}{1-\sigma} \exp\left(-\Psi \frac{(1-\sigma)\, L_t^{1+\varphi}}{1+\varphi}\right).$$

Figure A.2: Monte Carlo results: New Keynesian model with additional bells and whistles



Note: Median (solid line), 90%, and 68% credible bands based on 10000 draws. Income shares are expressed in percentage point deviations from initial values. Remaining variables are expressed in percentage deviations.

Investment adjustment costs: We assume a convex investment adjustment cost, so that

$$K_t = (1-\delta)\,K_{t-1} + \left[1 - \frac{\chi}{2}\left(\frac{I_t}{I_{t-1}} - 1\right)^2\right] I_t.$$

Variable capital utilization: Wholesale firms rent effective capital services $\bar{K}_t = U_t K_{t-1}$, where U_t is the utilization rate of capital. Higher utilization comes at a cost $AC_{u,t}$ paid by households who own the capital, where

$$AC_{u,t} = \xi_u'\,(U_t - 1) + \frac{\xi_u \xi_u'}{2}\,(U_t - 1)^2\,.$$

Nominal price stickiness: We incorporate price stickiness á la Rotemberg (1982). Nominal price adjustments are costly for wholesale firms. We also allow for partial indexation to past inflation and specify the cost function as

$$AC_{p,t} = \frac{\xi_p}{2}\left(\frac{\Pi_{jp,t}}{\Pi_{p,t-1}^{\gamma_p}\Pi_p^{1-\gamma_p}} - 1\right)^2 Y_t.$$

Nominal wage stickiness: Wage stickiness á la Rotemberg (1982) is the final extension. Nominal wage adjustments come at a cost paid by households:

$$AC_{w,t} = \frac{\xi_w}{2}\left(\frac{\Pi_{nw,t}}{\Pi_{p,t-1}^{\gamma_w}\Pi_p^{1-\gamma_w}} - 1\right)^2 L_t.$$

Monetary policy: Nominal rigidities imply the need to specify a nominal anchor. To this end we assume a Taylor type rule for the policy rate $i_{p,t}$:

$$1 + i_{p,t} = (1 + i_{p,t-1})^{\rho_i} \left[(1 + i_p) \left(\frac{\Pi_{p,t}}{\Pi_p} \right)^{\rho_\pi} \left(\frac{GDP_t}{GDP_{t-1}} \right)^{\rho_y} \right]^{1-\rho_i}$$

The Fisher equation $(1 + i_{p,t}) = (1 + r_t)\, \Pi_{t+1}$ links nominal to real outcomes. We also note that wage adjustment costs enter $s_{l,t}$, utilization adjustment costs enter $s_{k,t}$, while price adjustment costs enter $s_{d,t}$. However, these shares still sum to one, and the long run properties of the model are unaffected. Finally, we note that the New Keynesian model captures the neoclassical setup as a special case ($h = \chi = \xi_p = \xi_w = 0$ and $\xi_u \to \infty$).

Figure A.2 documents the distributions of theoretical impulse responses when we also draw parameters from the extended model. Importantly, the impulse responses are qualitatively similar across models even after a few periods, and the signs are identical from quarter 16 and onwards. We conclude, therefore, that the sign restrictions used in the main text are robust to the inclusion of real and nominal frictions.

B Bayesian Estimation of the VAR model

Consider the reduced form VAR model presented in Section 3.2:

$$Y_t = C + \sum_{j=1}^{p} A_j Y_{t-j} + u_t$$

The process above can be stacked in a more compact form as follows:

$$\mathbf{Y} = \mathbf{X}B + \mathbf{U}$$

where:
1) $\mathbf{Y} = (Y_{p+1}, ..., Y_T)'$ is a $(T - p)$ x n matrix, with $Y_t = (Y_{1,t}, ..., Y_{n,t})'$.
2) $\mathbf{X} = (\mathbf{1}, \mathbf{Y}_{-1}, ..., \mathbf{Y}_{-p})$ is a $(T - p)$ x $(np + 1)$ matrix, where $\mathbf{1}$ is a $(T - p)$ x 1 matrix of ones and $\mathbf{Y}_{-k} = (Y_{p+1-k}, ..., Y_{T-k})'$ is a $(T - p)$ x n matrix.
3) $\mathbf{U} = (u_{p+1}, ..., u_T)'$ is a $(T - p)$ x n matrix.
4) $B = (C, A_1, ..., A_p)'$ is a $(np + 1)$ x n matrix of coefficients.
Vectorizing the equation above, we obtain:

$$\mathbf{y} = (I_n \otimes \mathbf{X})\beta + \mathbf{u}$$

where $\mathbf{y} = vec(\mathbf{Y})$, $\beta = vec(B)$, $\mathbf{u} = vec(\mathbf{U})$ and $\mathbf{u} \sim N(0, \Sigma \otimes I_{T-p})$.
Given the assumption of normality of the reduced-form errors, $u_t \sim N(0, \Sigma)$, we can express the likelihood of the sample, conditional on the parameters of the model and the set of regressors $\mathbf{X}$, as follows:

$$L(\mathbf{y}|\mathbf{X}, \beta, \Sigma) \propto |\Sigma \otimes I_{T-p}|^{-\frac{T-p}{2}} exp\left\{ \frac{1}{2}(\mathbf{y} - I_n \otimes \mathbf{X}\beta)'(\Sigma \otimes I_{T-p})^{-1}(\mathbf{y} - I_n \otimes \mathbf{X}\beta) \right\}$$

Denote $\hat{\beta} = vec(\hat{B})$, where $\hat{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the OLS estimate, and let $S = (\mathbf{Y} -$

$\mathbf{X}\hat{B})'(\mathbf{Y} - \mathbf{X}\hat{B})$ be the sum of squared errors. Then we can rewrite the likelihood as follows:

$$L(\mathbf{y}|\mathbf{X}, \beta, \Sigma) \propto |\Sigma \otimes I_{T-p}|^{-\frac{T-p}{2}} exp\left\{\frac{1}{2}(\beta - \hat{\beta})'(\Sigma^{-1} \otimes \mathbf{X}'\mathbf{X})(\beta - \hat{\beta})\right\}$$
$$exp\left\{-\frac{1}{2}tr(\Sigma^{-1}S)\right\}$$

By choosing a non-informative (flat) prior for B and Σ that is proportional to $|\Sigma|^{-\frac{n+1}{2}}$, namely:

$$p(B|\Sigma) \propto 1$$
$$p(\Sigma) \propto |\Sigma|^{-\frac{n+1}{2}}$$

We can compute the posterior of the parameters given the data at hand using Bayes rule, as follows:

$$P(B, \Sigma|\mathbf{y}, \mathbf{X}) \propto L(\mathbf{y}|\mathbf{X}, \beta, \Sigma)p(B|\Sigma)p(\Sigma)$$
$$= |\Sigma|^{-\frac{T-p+n+1}{2}} exp\left\{\frac{1}{2}(\beta - \hat{\beta})'(\Sigma^{-1} \otimes \mathbf{X}'\mathbf{X})(\beta - \hat{\beta})\right\} exp\left\{-\frac{1}{2}tr(\Sigma^{-1}S)\right\}$$

This posterior distribution is the product of a normal distribution for β conditional on Σ and an inverted Wishart distribution for Σ. Thus, we draw β conditional on Σ from:

$$\beta|\Sigma, \mathbf{y}, \mathbf{X} \sim N(\hat{\beta}, \Sigma \otimes (\mathbf{X}'\mathbf{X})^{-1})$$

and Σ from:

$$\Sigma|\mathbf{y}, \mathbf{X} \sim IW(S, v)$$

through Gibbs sampling, where $v = T - p - np - 1$.

The identification procedure described in subsection 3.2 is performed using the algorithm of Rubio-Ramírez et al. (2010), which consists of the following steps, for each given draw from the posterior of the reduced-form parameters:

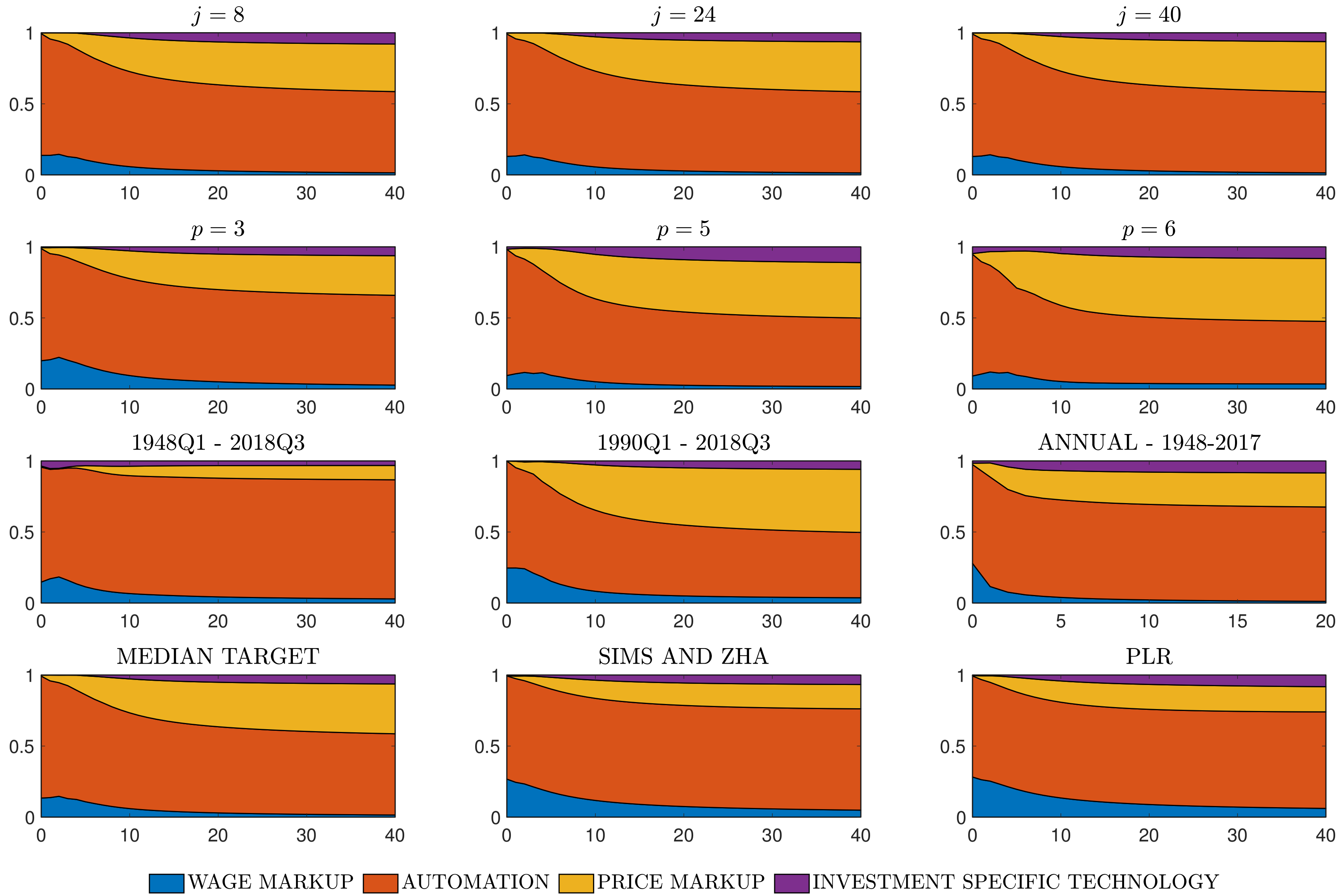
1. Draw a nxn matrix W from $N(0_n, I_n)$ and perform a QR decomposition of W, with the diagonal of R normalized to be positive and $QQ' = I_n$.

2. Let S be the lower triangular Cholesky decomposition of Σ and define $A = SQ'$. Compute the candidate impulse responses as $IRF_j = C_jA$, where C_j are the reduced form impulse responses from the Wold representation, for $j = 0, ..., J$. If the set of impulse responses satisfies all the sign restrictions, store them. If not, discard them and go back to the first step.

3. Repeat steps 1 and 2 until M impulse responses that satisfy the sign restrictions are obtained. The resulting set A, together with the reduced-form estimates, characterizes the set of structural VAR models that satisfy the sign restrictions.

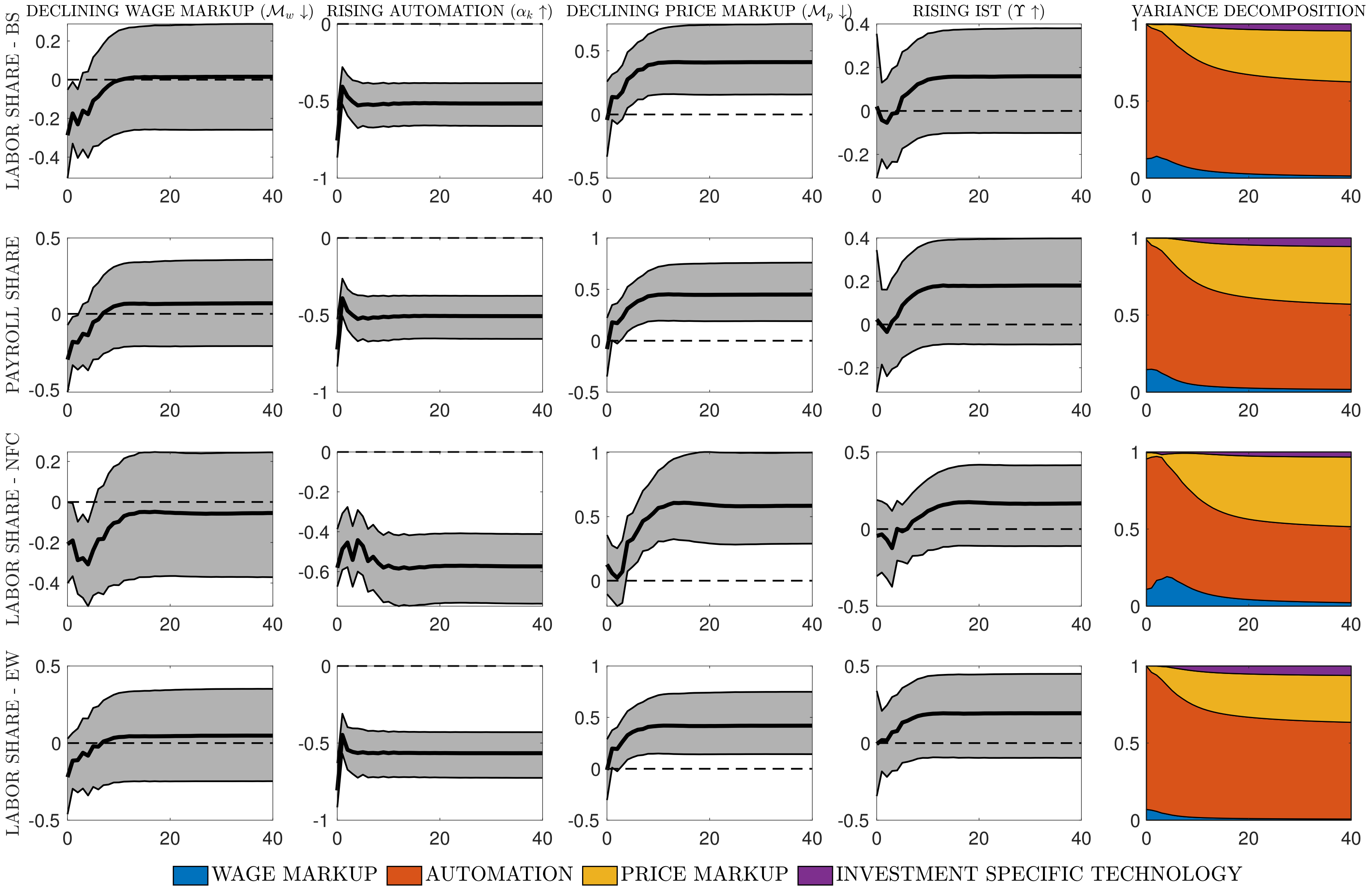
C ADDITIONAL RESULTS

Figure C.1: Empirical impulse responses from the baseline VAR model - Sensitivity checks



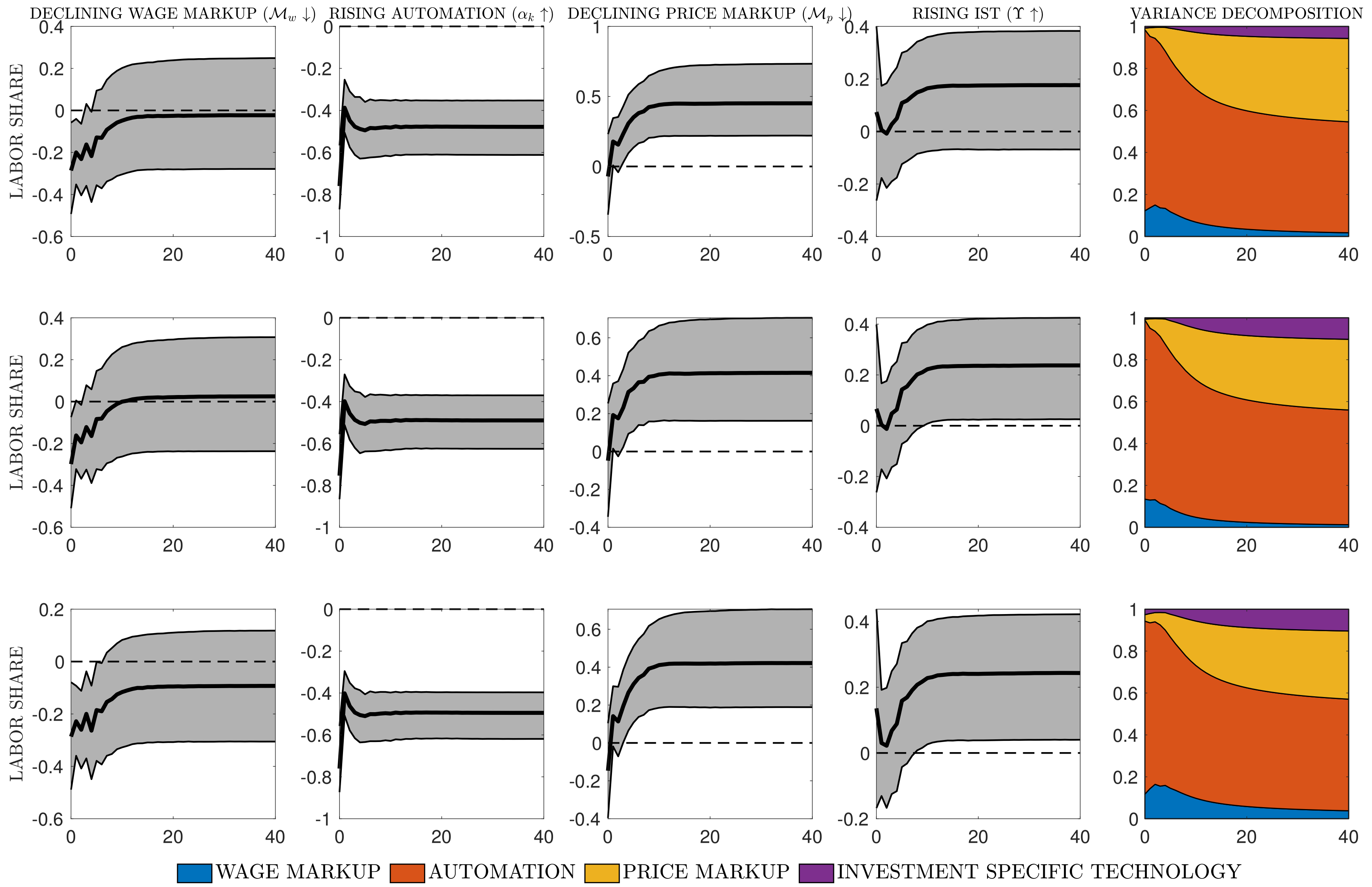
Note: The colored areas represent the point-wise median contributions of each identified shock to the forecast error variance of the labor share (in levels) at horizons $j = 0, 1, \dots, 40$ using the baseline identifying restrictions.

Figure C.2: Empirical impulse responses from the baseline VAR model - Alternative measures



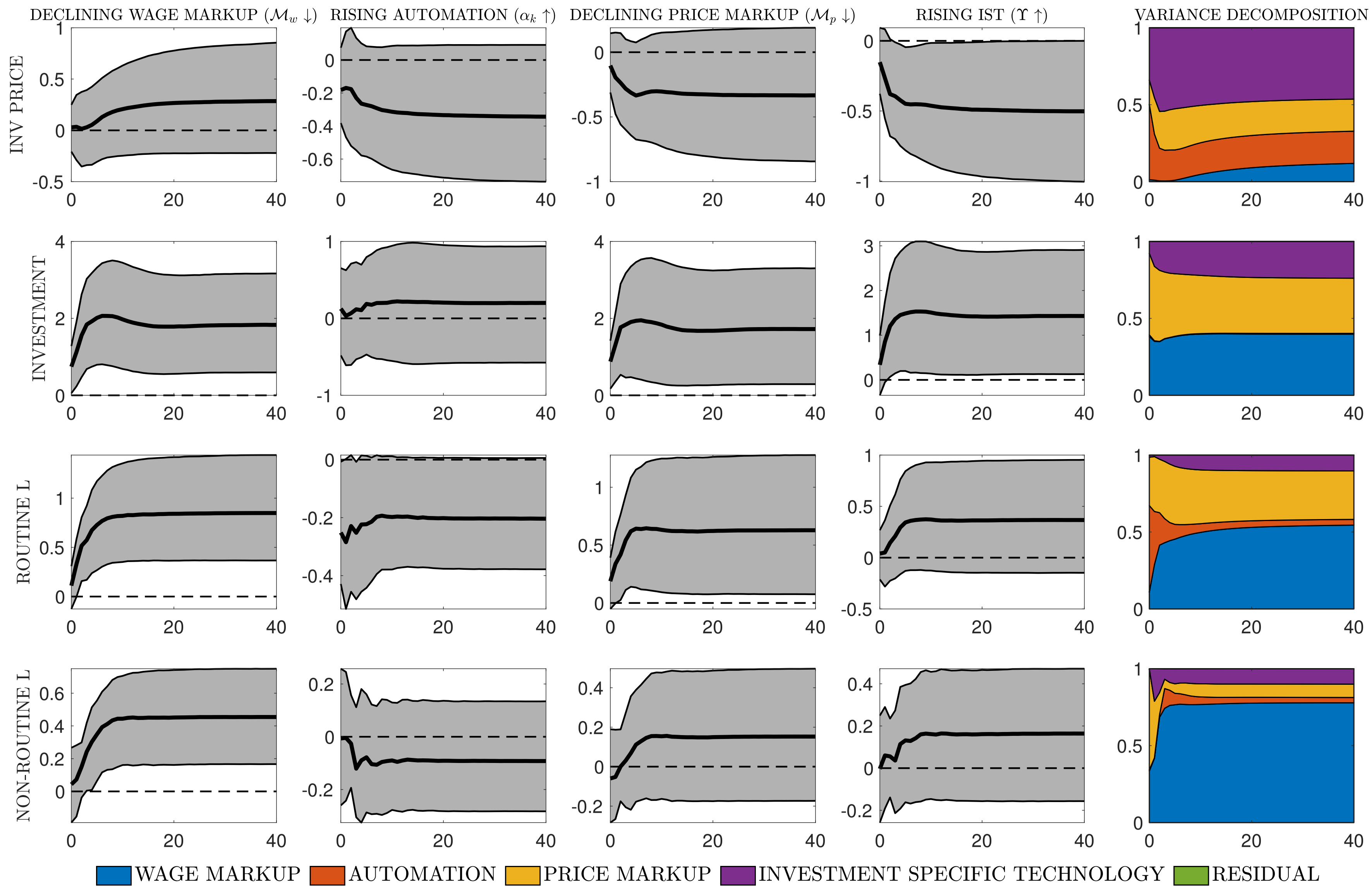
Note: Posterior distributions of cumulated impulse responses to an estimated shock of one standard deviation using the baseline identifying restrictions. Median (solid line) and 68% probability density intervals (shaded area) based on 10,000 draws. The median and the percentiles are defined at each point in time. The colored areas represent the point-wise median contributions of each identified shock to the forecast error variance of the labor share (in levels) at horizons $j = 0, 1, \ldots, 40$ using the baseline identifying restrictions.

Figure C.3: Empirical impulse responses from the baseline VAR model - Alternative identification



Note: Posterior distributions of cumulated impulse responses to an estimated shock of one standard deviation using the baseline identifying restrictions. Median (solid line) and 68% probability density intervals (shaded area) based on 10,000 draws. The median and the percentiles are defined at each point in time. The colored areas represent the point-wise median contributions of each identified shock to the forecast error variance of the labor share (in levels) at horizons $j = 0, 1, \ldots, 40$ using the alternative identifying restrictions.

Figure C.4: Empirical impulse responses from the baseline VAR model - More variables



Note: Posterior distributions of cumulated impulse responses to an estimated shock of one standard deviation using the baseline identifying restrictions. Median (solid line) and 68% probability density intervals (shaded area) based on 10,000 draws. The median and the percentiles are defined at each point in time. The colored areas represent the point-wise median contributions of each identified shock to the forecast error variance of each variable (in levels) at horizons $j = 0, 1, \ldots, 40$ using the alternative identifying restrictions.