THE DECLINE OF THE LABOR SHARE: NEW EMPIRICAL EVIDENCE*

DRAGO BERGHOLT[§], FRANCESCO FURLANETTO[§] AND NICOLÒ MAFFEI FACCIOLI[‡]

June 14, 2018
First version, Very Preliminary

Abstract

We estimate a Structural Vector Autoregression (SVAR) model to evaluate the relative importance of at least four alternative factors in explaining the decline of the labor share in the US. Our model is driven by permanent shocks to wage mark-ups, price mark-ups, investment specific technology and offshoring. The SVAR is identified using robust restrictions derived in the context of a stylized macroeconomic model. We find that a decline in the bargaining power of workers, an increase in automation and an increase in profits are the most promising explanations for the decline of the labor share. A permanent reduction in the relative price of investment leads to a mild increase in the labor share. This suggests complementarity between labor and capital, thus, ruling out capital deepening as a main driver of the labor share decline.

1 INTRODUCTION

Since the beginning of the new century, the US labor's share of income has accelerated its decline, reaching its postwar lowest level in the aftermath of the Great Recession. Four main explanations have been proposed in the literature to rationalize the decline of the labor share which has occurred not only in the US but also at the global level within the large majority of countries and industries.

Karabarbounis and Neiman (2014) relate the decline of the labor share to investment-specific technological progress. Elsby, Hobijn, and Sahin (2013) argue that the behavior of wages and the capital-labor ratio are in contrast with the dynamics of the relative price of investment. Globalization, instead, and the process of off-shoring of intermediate goods production in developing countries in particular, is a more promising explanation for the decline in the labor share. Barkai (2018) documents that the decline in the labor

*The views expressed in this paper are those of the authors and do not necessarily reflect those of Norges Bank. The paper has benefited from discussions with Fabio Canova, Marc Giannoni, Jon Fernald, Valerie Ramey, Ivan Petrella, and Simcha Barkai.

§Norges Bank.

‡Universitat Autonoma de Barcelona, Barcelona GSE and Norges Bank.

share is accompanied by a decline in the capital share in the data. Moreover, he argues that the decline in the capital share is unlikely to be driven by unobserved capital. Importantly, he shows that a decline in competition is the only driving force able to generate simultaneously a decline in the labor and in the capital share in the context of the same theoretical model used by Karabarbounis and Neiman (2014). Finally, Blanchard and Giavazzi (2003) and Ciminelli, Duval, and Furceri (2017) among others show that a decline in the bargaining power of workers, proxied, respectively, by labour market deregulation and by major reforms to employment protection legislation (especially in the 1990s and in the 2000s for the latter) is responsible for a substantial fraction of the labor share decline.
While the strengths and weaknesses of these four explanations have been discussed in depth in the literature, an empirical analysis including all of them in the context of the same model is lacking. Our aim is to fill this gap by estimating a Structural Vector Autoregression (SVAR) whose identification is informed by a theoretical model that slightly modifies the framework proposed by Karabarbounis and Neiman (2014) (and used also by Barkai (2018)) to include in a stylized way the emergence of offshoring and variations in the bargaining power as possible drivers of the labor share decline. While the impact of the four permanent shocks on the sign of the labor share response depends largely on the chosen parameterization, the sign of other variables' responses to the four shocks are robust across different parameterizations. Following Canova and Paustian (2011), we use these "robust" restrictions derived from the theoretical model to separately identify the four shocks in the SVAR. Impact sign restrictions on the responses of real GDP, real wages, real profits and the ratio of price of intermediate imported goods over the price of investment goods are sufficient to set apart the four shocks. Notably, we make sure that the long-run properties of our theoretical model are satisfied by our empirical model by proposing, to the best of our knowledge, the first application for structural analysis of the priors for the long-run proposed recently by Giannone, Lenza, and Primiceri (2018).
We find that a reduction in the bargaining power of workers and a decline in the degree of competition in the economy generate a substantial decline of the labor share in our empirical model, in keeping with the analysis of Blanchard and Giavazzi (2003), Ciminelli et al. (2017) and Barkai (2018). In contrast, a decline in the relative price of investment, despite having substantial effects on GDP, leads to an increase in the labor share (and not to a decline) in the majority of draws consistent with our identification assumptions. This result can be used as (weak) evidence in favor of an elasticity of substitution between capital and labor lower than one. In fact, under that assumption, also the theoretical model generates an increase in the labor share in response to a decline in the relative price of investment. This is also confirmed by the response of the labor share to the wage bargaining shock, which is consistent with an elasticity smaller than one. The offshoring shock has also large effects on output and minor effects on the labor share.
From a quantitative point of view, the shocks originating in the labor market and in the rise of market power of firms are the main drivers of the labor share decline. Such an important role for the labor market shock calls for the use of additional data to be sure that we are not mixing it with other shocks, given that it might have multiple interpretations. In a first attempts in that direction, we use data on employment to disentangle a decline in the bargaining power of workers from an automation shock which share the features of the automation process described by Acemoglu and Restrepo (2016). We find that both shocks are important in the short-run, but automation eats most of the long-run response

of the wage bargaining shock.
The paper is organized as follows. Section 2 introduces the theoretical model and discusses the identification strategy. Section 3 presents the simulations of the models and the choice of robust sign restrictions. Section 4 describes the empirical methodology. Section 5 shows the results for the baseline version of our empirical model and for the case in which we disentangle the labor market shock from an automation shock. Finally, Section 6 concludes.

2 THE THEORETICAL FRAMEWORK

In this section we summarize two theoretical models which motivate our identification strategy. The first model is purely neoclassical and allows us to derive long run restrictions for the empirical analysis. Importantly, changes in the labor share can occur due to (i) technical change (labor augmenting, capital augmenting or investment specific), (ii) market power distortions (in goods or labor markets), and (iii) globalization (thought of as offshoring to global value chains). The resulting framework captures, as a special case, the setup in e.g. Karabarbounis and Neiman (2014). The second model, which we refer to as New Keynesian, extends the first with various bells and whistles: nominal wage and price stickiness on the supply side, habit formation, investment adjustment costs, and variable capital utilization on the demand side. These frictions allow us complement the long run restrictions with various short run restrictions, given that they imply a fairly decent fit to the US business cycle. Importantly, both models have exactly the same long run properties. We describe the neoclassical model first, and then briefly summarize how bells and whistles are introduced into the system.

2.1 THE NEOCLASSICAL MODEL

The model economy is populated by a unit mass of firms and households. On the firm side, we distinguish between retailers, wholesalers, investment producers, and offshore producers.

2.1.1 RETAIL PRODUCER

A competitive retail goods producer combines wholesale goods according to the production technology

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

where $Z_{j,t}$ is output by wholesale firm j. Optimal demand for an individual good is $Z_{j,t} = P_{j,t}^{-\epsilon_{p,t}} Z_t$, where $P_{j,t}$ is the price of good j relative to the aggregate price index

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

Thus, the final good Z_t is chosen as the numeraire. It can be used for consumption or investment purposes. Market clearing dictates that

$$Z_t = C_t + X_t + M_t,$$

where C_t is consumption, X_t is investments, and M_t represents the goods shipped abroad to foreign value chains.

2.1.2 WHOLESALE PRODUCERS

Output of wholesale firm j is denoted by $Y_{j,t}$, where

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

and

$$F_{j,t} = \left[\alpha_l \left(A_{l,t}L_{j,t}\right)^{\frac{\gamma-1}{\gamma}} + (1-\alpha_l)\, O_{j,t}^{\frac{\gamma-1}{\gamma}}\right]^{\frac{\gamma}{\gamma-1}}.$$

Output requires three inputs: capital K_{t-1}, labor L_t, and imported offshore goods O_t. and $A_{l,t}$ and $A_{k,t}$ represent labor augmenting and capital augmenting technology shocks, respectively. The parameter η represents the elasticity of substitution between capital and the composite resource F, while γ determines the elasticity of substitution between labor and offshoring. Note that our production structure captures a few special cases: first, when $\eta \to \gamma$ we have symmetric factor technologies. Second, if $\alpha_l \to 1$, then we are back to the model analyzed by e.g. Barkai (2018) and Karabarbounis and Neiman (2014). Third, if also $\eta \to 1$, then we are back to the standard neoclassical growth model with Cobb Douglas production technology.

Firm j chooses factors of production and sets prices in order to maximize the profits. The maximization problem is constrained by a downward sloping demand curve as well as the market clearing condition $Z_{j,t} = Y_{j,t}$. We impose a symmetric equilibrium ($P_{j,t} = 1$, $Y_{j,t} = Y_t$, etc.) on the first order conditions and define the gross markup as $\mathcal{M}_{p,t} = MC_t^{-1}$ (price over nominal marginal costs). The representative firm's behavior can then be summarized by the production function as well as four optimality conditions (with respect to capital demand, labor demand, offshoring demand, and price setting):

$$\begin{aligned}
r_t^k \mathcal{M}_{p,t} &= (1-\alpha)\, A_{k,t}^{\frac{\eta-1}{\eta}} \left(\frac{Y_t}{K_{t-1}}\right)^{\frac{1}{\eta}} \\
W_t \mathcal{M}_{p,t} &= \alpha\alpha_l A_{l,t}^{\frac{\gamma-1}{\gamma}} \left(\frac{Y_t}{F_t}\right)^{\frac{1}{\eta}} \left(\frac{F_t}{L_t}\right)^{\frac{1}{\gamma}} \\
P_{O,t}\mathcal{M}_{p,t} &= \alpha(1-\alpha_l) \left(\frac{Y_t}{F_t}\right)^{\frac{1}{\eta}} \left(\frac{F_t}{O_t}\right)^{\frac{1}{\gamma}} \\
\mathcal{M}_{p,t} &= \frac{\epsilon_{p,t}}{\epsilon_{p,t}-1}
\end{aligned}$$

These equations imply that profits can be written as $\mathcal{D}_t = \frac{\mathcal{M}_{p,t}-1}{\mathcal{M}_{p,t}} Y_t$ or, alternatively, that firm revenues can be written as a time varying markup over producer costs:

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

We note, for completeness, that total value added and profits in the domestic economy are given by $GDP_t = Y_t - P_{O,t}O_t$ and $\mathcal{D}_t = (1 + s_{o,t}) \frac{\mathcal{M}_{p,t}-1}{\mathcal{M}_{p,t}} GDP_t$. $s_{o,t} = \frac{P_{O,t}O_t}{GDP_t}$ is the ratio of offshore inputs to value added. Also, $s_{l,t} + s_{k,t} + s_{d,t} = 1$, where $s_{l,t} = \frac{W_t L_t}{GDP_t}$, $s_{k,t} = \frac{r_t^k K_{t-1}}{GDP_t}$, and $s_{d,t} = (1 + s_{o,t}) \frac{\mathcal{M}_{p,t}-1}{\mathcal{M}_{p,t}}$. Importantly, the profit share responds to all shocks in the model as long as $s_{o,t} > 0$, and not only to the markup shock as in Barkai (2018).

2.1.3 INVESTMENT PRODUCER

A competitive investment good producer transforms the bundle X_t into final investments according to the production function $I_t = \Upsilon_t X_t$. The final good I_t is sold to households who accumulate capital. Profit maximization on behalf of the investment producer implies that

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

where $P_{I,t}$ is the price of investments in terms of consumption units.

2.1.4 OFFSHORING FIRM

We do this extension in the simplest way possible, and suppose that some of the intermediate goods are shipped abroad for assembly in a foreign value chain. In particular, the competitive offshoring transforms the quantity M_t into O_t using the technology $O_t = \Phi_t M_t$, where Φ_t is a productivity shock specific to the offshore sector. Profit maximization on behalf of the offshore firm implies that

$$P_{O,t} = \Phi_t^{-1},$$

where $P_{O,t}$ is the price of O_t in consumption units.

2.1.5 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. Optimal demand for each labor variety is $L_{n,t} = \left(\frac{W_{n,t}}{W_t} \right)^{-\epsilon_{w,t}} L_t$ and the aggregate wage index is

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

2.1.6 HOUSEHOLDS

Households derive utility from aggregate consumption and dis-utility from hours worked. The period utility of household n is equal to

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

where consumption is identical across households due to risk sharing. Household n maximizes $\mathbb{E}_t \sum_{s=t}^{\infty} \beta^{s-t} \mathcal{U}_s$ subject to an intertemporal budget constraint (which includes profits) and the law of motion for capital:

$$\begin{aligned} C_t + P_{I,t} I_t + B_t &= \mathcal{D}_t + W_{n,t} L_{n,t} + r_t^k K_{t-1} + (1 + r_{t-1}) B_{t-1} - Tax_t \\ K_t &= (1-\delta) K_{t-1} + I_t \end{aligned}$$

As before we impose a symmetric equilibrium ($W_{n,t} = W_t$, $L_{n,t} = L_t$, etc.) on the first order conditions and define the gross wage markup as $\mathcal{M}_{w,t} = \frac{W_t}{MRS_t}$ (wage over marginal rate of substitution between hours and consumption). The representative household's behavior can then be summarized by the budget constraint, the law of motion for capital, as well as six optimality conditions (with respect to consumption demand, bond holdings, labor supply, capital accumulation, investment demand, and wages):

$$\begin{aligned} \Lambda_t &= C_t^{-\sigma} \exp\left(-\Psi \frac{(1-\sigma) L_t^{1+\varphi}}{1+\varphi}\right) \\ \Lambda_t &= \beta \mathbb{E}_t \Lambda_{t+1} (1 + r_t) \\ W_t &= \mathcal{M}_{w,t} \Psi L_t^{\varphi} C_t \\ Q_t &= \beta \mathbb{E}_t \frac{\Lambda_{t+1}}{\Lambda_t} \left[r_{t+1}^k + Q_{t+1} (1-\delta) \right] \\ P_{I,t} &= Q_t \\ \mathcal{M}_{w,t} &= \frac{\epsilon_{w,t}}{\epsilon_{w,t} - 1} \end{aligned}$$

Fluctuations in $\mathcal{M}_{w,t}$ can be interpreted as changes in union power, labor supply, or other factors that influence the supply side of the labor market. This completes our description of the neoclassical growth model.

2.2 THE NEW KEYNESIAN EXTENSION

This section inorporates a few "bells and whistles" into the neoclassical model. For lack of a better word, we call this version New Keynesian. 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. The gains of having these frictions are twofold: first, the bells and whistles allow us to derive credible short run restrictions, without having to sacrifice any identification coming from the model's long run properties. Second, we can explore data on the nominal interest rate, as well as on nominal price or wage inflation. To this end we extend the model in the following way:

External habit formation: The period utility is changed to

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

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 $\xi_u \to \infty$).

2.3 Shock processes

Dynamics in both models are driven by six stochastic processes: $A_{l,t}$, $A_{k,t}$, $\mathcal{M}_{p,t}$, $\mathcal{M}_{w,t}$, Υ_t, and Φ_t. Their dynamics are as follows:

$$\frac{A_{l,t}}{A_{l,t-1}} = 1 + g_{l,t} = (1+g_l)\exp(z_{l,t}) \qquad \frac{A_{k,t}}{A_{k,t-1}} = 1 + g_{k,t} = (1+g_k)\exp(z_{k,t})$$

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

with forcing variables

$$\begin{aligned} z_{l,t} &= \rho_l z_{l,t-1} + \sigma_l \varepsilon_{l,t} & z_{k,t} &= \rho_k z_{k,t-1} + \sigma_k \varepsilon_{k,t} \\ z_{p,t} &= \rho_p z_{p,t-1} + \sigma_p \varepsilon_{p,t} & z_{w,t} &= \rho_w z_{w,t-1} + \sigma_w \varepsilon_{w,t} \\ z_{\Upsilon,t} &= \rho_\Upsilon z_{\Upsilon,t-1} + \sigma_\Upsilon \varepsilon_{\Upsilon,t} & z_{\Phi,t} &= \rho_\Phi z_{\Phi,t-1} + \sigma_\Phi \varepsilon_{\Phi,t} \end{aligned}$$

The innovations are normally distributed with zero mean and unit variance. A balanced growth path (BGP) is obtained if and only if the following conditions hold:

- $g_p = \sigma_p = g_w = \sigma_w = g_\Phi = \sigma_\Phi = 0$
- Either $\eta = 1$ or $g_k = \sigma_k = g_\Upsilon = \sigma_\Upsilon = 0$

3 MONTE CARLO SIMULATIONS

This section documents the distribution of impulse responses from a Monte Carlo exercise applied to both models. To this end we shut off all temporary innovations, and normalize all permanent shocks so that they eventually increase by 1 percent. We keep parameters that govern certain great ratios fixed (given that these ratios are known from data). This is done in order to ensure that all simulations start from the same point. We set $\beta = 0.99$, $\delta = 0.025$, $s_l = 0.59$, $s_k = 0.30$, $s_d = 0.11$, and $s_o = 0.25$. The remaining parameters are drawn from uniform distributions with support that spans common beliefs in the literature, see Table 1.

Table 1: Bounds for the structural parameters

		Neoclassical		*New Keynesian*	
		LB	*UB*	*LB*	*UB*
"Deep" parameters					
σ	Inverse of intertemporal elasticity	1	5	1	5
φ	Inverse Frisch elasticity	1	5	1	5
η	Substitution between F and K	0.25	1.5	0.25	1.5
γ	Substitution between L and O	0.25	1.5	0.25	1.5
Dynamic parameters					
h	Habit formation	0	0	0	0.9
χ	Investment adjustment cost	0	0	0	10
ξ_u	Utilization cost	100	100	0.05	100
θ_p	Nominal price stickiness	0	0	0	0.9
θ_w	Nominal wage stickiness	0	0	0	0.9
γ_p	Degree of price indexation	-	-	0	0.75
γ_w	Degree of wage indexation	-	-	0	0.75
ρ_i	Interest rate inertia	-	-	0	0.9
ρ_π	Response to inflation	-	-	1.01	10
ρ_y	Response to output	-	-	0	1
Shocks' persistence					
ρ_l	Labor augmenting technology	0	0.75	0	0.75
ρ_k	Capital augmenting technology	0	0.75	0	0.75
ρ_p	Price markup	0	0.75	0	0.75
ρ_w	Wage markup	0	0.75	0	0.75
ρ_υ	Investment specific technology	0	0.75	0	0.75
ρ_ϕ	Offshoring technology	0	0.75	0	0.75

Note: LB $\rightarrow$ lower bound; UB $\rightarrow$ upper bound. The parameters θ_p and θ_w represent the probabilities of being stuck with old prices and wages in the Calvo model. They do not appear in our model because we use Rotemberg pricing. However, we exploit the first order equivalence between Calvo and Rotemberg pricing in order to back out ξ_p and ξ_w, given θ_p and θ_w. Indexation parameters and Taylor rule coefficients are set to arbitrary numbers in the neoclassical model (where they do not have real effects).

The results of the simulation are represented in Appendix A for the Neoclassical model and in Appendix B for the New Keynesian one. Given these, we construct, following Canova and Paustian (2011), "robust" sign restrictions that are used to identify the VAR model presented in the next section. Figure 9 and 10 show the impulse responses of, respectively, a positive permanent price markup shock ($\mu_p \uparrow$) and a positive wage markup shock ($\mu_w \uparrow$) . The distinctive feature of the former shock is the negative co-movement between GDP and profits. The latter shock, instead, exhibits a negative co-movement between GDP and wages. These are features, within the model, present only in these two shocks and this is how we disentangle different shocks in the VAR with sign restrictions. Figures 11 and 12 present the impulse responses of an investment specific shock and of an offshoring shock. What distinguish these two shocks are the responses of P_i and P_o, namely investment specific technology shocks are the only ones affecting P_i and offshoring shocks are the only ones affecting P_o. Thus, the ratio $\frac{P_i}{P_o}$ decreases after a positive investment specific techology shock, but increases subsequent to an offshoring shock.

Finally, we include a demand shock to characterize business cycle fluctuations. Demand shocks are characterized by a positive co-movement between GDP and inflation, feature not present for the remaining shocks, which experience, instead, a decrease in inflation. The sign restrictions are summarized in Table 2.

Sign Restrictions					
Variables	$\mu_w \downarrow$	$\mu_p \downarrow$	$\upsilon \uparrow$	$\phi \uparrow$	$Demand$
Real GDP	+	+	+	+	+
Real Wages	-	+	+	+	NA
Real Profits	NA	-	+	+	NA
Inflation	-	-	-	-	+
$\frac{P_i}{P_o}$	NA	NA	-	+	NA
Labor Share	NA	NA	NA	NA	NA

4 Econometric Methodology:

Consider the following reduced form VAR model:

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

where $u_t \sim N(0, \Sigma)$ is the reduced form residual, Y_t is a nx1 vector containing all the n endogenous variables, $A_1, ..., A_p$ are nxn matrices of coefficients associated to the p lags of the dependent variable and C is a nx1 vector of constants.
The model in (1) can be rewritten in terms of levels and differences, as follows:

$$\Delta Y_t = C + \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + u_t \tag{2}$$

where $\Pi = (A_1 + \cdots + A_p) - I_n$ and $\Gamma_j = -(A_{j+1} + \cdots + A_j)$, with $j = 1, ..., p-1$. Giannone, Lenza and Primiceri (2018) propose a conjugate prior for Π which disciplines the long-run behavior of the VAR, based on economic theory. Let H be an invertible nxn matrix and rewrite (2) in the following form:

$$\Delta Y_t = C + \Lambda \tilde{Y}_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + u_t \tag{3}$$

where $\tilde{Y}_{t-1}$ is a vector of independent linear combinations of the endogenous variables in Y_t and $\Lambda = \Pi H^{-1}$ is a nxn matrix of coefficients representing the effects of these linear combinations on ΔY_t. In this framework, the elicitation of a prior for the long-run behaviour of the VAR reduces to the choice of a prior for Λ, conditional on the choice of a matrix H. Following Giannone, Lenza and Primiceri (2018), we specify the prior on the matrix of loadings Λ as follows:

$$\Lambda_{.i} | H_{i.} \sim N(0, \tilde{\phi}_i(H_{i.})\Sigma) \tag{4}$$

where $\Lambda_{.i}$ refers to the i-th column of Λ, $H_{i.}$ refers to the i-th row of the matrix H and $\tilde{\phi}_i(H_{i.})$ is a scalar hyperparameter defined as follows:

$$\tilde{\phi}_i(H_{i.}) = \frac{\phi_i^2}{(H_{i.}\bar{Y}_0)} \tag{5}$$

where $\bar{Y}_0$ is a nx1 vector containing the average of the p initially discarded observations. Following Giannone, Lenza and Primiceri (2015), we use a hierarchical interpretation of the model and set $\tilde{\phi}_i(H_{i.})$ based on its posterior distribution, which combines the marginal likelihood and an hyperprior on ϕ_i centered around one with standard deviation equal to one. Thus, a fully Bayesian inference is conducted on the tightness of the prior, based on the hierarchical interpretation of the model. This methodology has a number of advantages: it improves significantly the out-of-sample forecasting accuracy of the VAR, especially with a large system of variables, delivers precise estimates of the impulse responses to structural shocks and reduces the burden of subjective choices in the setting of the priors.

Being conjugate, this prior can be implemented in a VAR in levels as (1) using Theil Mixed estimation, which consists of adding a set of n artifical dummies to the original sample and then conducting inference with the augmented sample. We construct the following set of dummy observations:

$$Y_i^+ = \frac{H_{i.}\bar{Y}_0}{\phi_i}(H^{-1})_{.i} \quad , \quad i = 1, ..., n$$

Notice that the process represented in (1) can be stacked in a more compact form as follows:

$$\mathbf{Y} = \mathbf{X}B + \mathbf{U} \tag{6}$$

where:

1) $\mathbf{Y} = (Y_{p+1}, ..., Y_T)$ is a $(T-p)$ x n matrix, with $Y_t = (Y_{1t} ..., Y_{nt})$.

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, for $k = 1, .., p$.

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.

Then, the artificial dummy observations are added on top of the data matrices $\mathbf{Y}$ and $\mathbf{X}$, as follows:

$$\mathbf{Y}^* = [\mathbf{Y}^+, \mathbf{Y}]'$$
$$\mathbf{X}^* = [\mathbf{X}^+, \mathbf{X}]'$$

where $\mathbf{Y}^+ = (Y_1^+, ..., Y_n^+)$ is a nxn matrix and $\mathbf{X}^+ = (\mathbf{0}, \mathbf{Y}^+, ..., \mathbf{Y}^+)$ is a nx$(np+1)$ matrix. The process in (6) can be rewritten as follows:

$$\mathbf{Y}^* = \mathbf{X}^* B + \mathbf{U}^* \tag{7}$$

Vectorizing (7), we obtain:

$$\mathbf{y}^* = (I_n \otimes \mathbf{X}^*)\beta + \mathbf{u}^* \tag{8}$$

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

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

where $\hat{\beta} = vec(\hat{B})$, $\hat{B} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Y}^*$ and $S = (\mathbf{Y}^* - \mathbf{X}^*\hat{B})'(\mathbf{Y}^* - \mathbf{X}^*\hat{B})$ is the sum of squared errors.
In their baseline, Giannone, Lenza and Primiceri (2018) use the prior for the long-run in combination with a Minnesota prior. We follow the same approach in what follows. Given a choice of Ψ, d and b, the Minnesota prior takes the following form:

$$\Sigma \sim IW(\Psi, d)$$
$$\beta|\Sigma \sim N(b, \Sigma \otimes \Omega)$$

and leads, in combination with the likelihood presented above, to the following posterior distributions for β and Σ:

$$\Sigma|\mathbf{y}^* \sim IW(\Psi + S^* + (\hat{B}^* - \mathbf{b})'\Omega^{-1}(\hat{B}^* - \mathbf{b}), T - p + d + n + 1)$$
$$\beta|\Sigma, \mathbf{y}^* \sim N(\hat{\beta}^*, \Sigma \otimes (\mathbf{X}^{*\prime}\mathbf{X}^* + \Omega^{-1})^{-1})$$

where $S^* = \hat{\mathbf{u}}^{*'}\hat{\mathbf{u}}^*$, $\hat{\mathbf{u}}^{*'} = \mathbf{Y}^* - \mathbf{X}^*\hat{B}^*$, $\hat{B}^* = (\mathbf{X}^{*\prime}\mathbf{X}^* + \Omega^{-1})^{-1}(\mathbf{X}^{*\prime}\mathbf{Y}^* + \Omega^{-1}\mathbf{b})$, $\hat{\beta}^* = vec(\hat{B}^*)$ and $b = vec(\mathbf{b})$. Given the Gaussian-inverse Wishart form, draws of the reduced form parameters from the posterior distribution are obtained using the Gibbs sampler.
In order to map the economically meaningful structural shocks from the reduced form estimated shocks, we need to impose restrictions on the variance covariance matrix previously estimated. In particular, let $u_t = A\epsilon_t$, where $\epsilon_t \sim N(0, I_n)$ and A is such that $AA' = \Sigma$. In what follows, we assume that A is a Cholesky decomposition of Σ. In order to identify all the shocks in the system, we need additional $\frac{n(n-1)}{2}$ conditions. The additional (robust) sign restrictions, derived from the previous section, are imposed using the QR decomposition algorithm proposed by Rubio-Ramirez, Waggoner and Zha (2010), as follows:

1. Make a draw from a $MN(0_n, I_n)$ and perform a QR decomposition of the matrix with the diagonal of R normalized to be positive, where $QQ' = I_n$.

2. Compute $IRF_j = C_jAQ'$, where C_j are the reduced form impulse responses, for $j = 0, ..., J$. If the set of IRFs satisfy the sign restriction, store them. If not, discard them.

3. Repeat 1. and 2. until you stored 1000 impulse responses.

5 Empirical Results

In this section, we present the results derived from our baseline VAR model, which is estimated on quarterly data in levels from 1983Q1 to 2017Q4 for the US. The set of endogenous variables Y_t contains six variables, in the following order: real GDP per capita, real wages, real corporate profits after tax per capita, inflation, the ratio between the price of investment and the price of offshoring and the BLS measure of the labor share. All the variables are expressed in natural logarithms multiplied by 100. The baseline model is estimated using 4 lags and implementing the restrictions of Table 2 on impact. The matrix H presented in the previous section suggests a prior on the long-run dynamics of the system and is chosen based on economic theory. We follow the baseline of Giannone et al. (2018) and use the prior that output, wages and profits are likely cointegrated, whereas the ratio $\frac{P_i}{P_o}$, inflation and the labor share are likely not. In other terms, the prior is centered on a balanced growth path. This is effectively performed by defining H as follows:

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus, $\tilde{Y}_t = HY_t$. The construction of this matrix is representative of a large set of DSGE models a la Smets and Wouters (2007), but it's not fully robust within our framework in which we don't necessarily have a balanced growth path, which is captured as a special case. The fact that this is not a robust feature in all the parametrizations of our model is not an issue, as we don't impose the prior dogmatically.

This fully probabilistic approach is particularly convenient because it doesn't require to take a stand on the cointegration relations and common trends of the variables in the system, but it only puts forward their possible existence. Additionally, it is aimed at solving a pathological issue in flat-prior VARs, namely the fact that initial conditions explain an implausibly large share of the low-frequency variation of the data, generating poor out-of-sample forecasts and inaccurate impulse responses, especially in the long-run. This is particularly important in this context, in which we are after long-run effects and permanent shocks. As underlined by Giannone, Lenza and Primiceri (2018), a simple Minnesota prior alone, for instance, doesn't solve the overfitting problem. This issue was addressed in early work by the introduction of the sum-of-coefficients prior by Doan, Litterman, and Sims (1984) and, in the context of Bayesian VARs, by Sims and Zha (1998). This last approach incorporates the idea of a prior which expresses disbelief in the excessive predictive power of initial conditions and consists in the specification of H as an identity matrix. This framework, however, doesn't allow for a distinction in the tightness of the prior for likely non-stationary and likely stationary variables. With the introduction of the prior for the long-run, we are shrinking VAR coefficients such that little predictive power is given to initial conditions and especially for non-stationary variables. In view of the foregoing, we are motivated in using the priors for the long run in order to explain the low-frequency behavior of the variables in the system, and particularly of the labor share.

5.1 The Baseline VAR model

Figure 13 presents the impulse responses derived from our model. The x axis represents the horizon considered, $h = 0, 1, ..., 20$ (quarters), and the y axis the response. Each column represents a particular shock and the ordering is the one presented in Table 2.

The first column shows the effect of a negative wage markup shock, which can be interpreted, for instance, as a decrease in the bargaining power of workers. In line with the theoretical model, real GDP per capita increases and real wages decrease significantly and persistently. Without restricting the response of real profits, we obtain, consistent with the theoretical framework, a persistent increase in the majority of draws and $\frac{P_i}{P_0}$ doesn't experience a significant change. The interesting feature is the response of the labor share, which decreases significantly and persistently over the horizon considered and it is consistent with complementarity between labor and capital.

The second column depicts the responses to a negative markup shock, namely a decrease in the market power of firms. In line with the theoretical model, a decrease in markups leads to a persistent increase in real GDP and wages, to a strong decrease in profits and no appreciable change in the ratio $\frac{P_i}{P_0}$. The labor share increases persistently and significantly.

The third column presents the effects of an investment specific technology shock. GDP, wages and profits increase, whereas the ratio $\frac{P_i}{P_0}$ decreases persistently. interestingly, the labor share doesn't experience a significant change and, if anything, it goes in the direction of an increase for the majority of draws, giving some weak evidence to an elasticity of substitution between capital and labor being smaller than one.

The fourth column shows the responses to an offshoring shock. GDP, wages, profits and $\frac{P_i}{P_0}$ increase persistently, especially the former two. Labor share, instead, doesn't experience a statistically significant change and, if anything, shows an increasing pattern, again consistent with complementarity between capital and labor.

The responses to a demand shock are represented in the last column. Both GDP and inflation increase and decay to zero in the long-run, whereas the other variables don't experience a significant change.

Likewise the theoretical model, $\frac{P_i}{P_0}$ moves basically only in response to investment specific and offshoring shocks and, without having to impose any restriction, the responses of the labor share to different shocks seem to all support complementarity between labor and capital, ruling out capital deepening as a possible explanation of the decline of the labor share.

Figure 14 presents the variance decomposition of our model. Therefore, the y axis now represents the share of the variance of a given variable attributable to each shock. In line with the impulse response analysis presented above, wage markup, investment specific, offshoring and demand shocks explain largely variations in GDP, whereas markup shocks play a smaller role. The picture is slightly different for wages, where markup shocks play a bigger role to the detriment of investment specific and demand ones. In line with the theoretical framework, the variation in profits is mainly explained by markup shocks and variation in $\frac{P_i}{P_0}$ mostly by investment specific and offshoring shocks, with a small bite for demand at short horizons. Finally, labor share seems to be driven by two different set of stories, both at the short and long-run. The former is the one proposed by Barkai (2018), which puts forward a key role of the increase in markups in the decline of the labor share, which, at first glance, seems to explain approximately 42% of the variation of the labor share. The latter is a combination of stories: wage bargaining, automation, demographics. It explains 36% of the variation in the labor share. This last result motivates us to dig deeper in this story and to try to further disentangle this shock.

Figure 15 displays the historical decomposition of the labor share in deviations from its mean. A couple of facts stand out. In line with the motivation for the usage of the priors for the long-run, the deterministic components (initial conditions) don't play a big role in explaining the low frequency variation of the labor share. Before the 2000s, changes in the labor share were mostly driven by negative wage markup shocks, especially between 1992 and 1999. After the 2000s, it seems that the strong decrease in the labor share is to large extent a story of rising market power of firms, even though the decline in wage markups still plays a considering role. Consistently with the responses presented in Figure 13, both offshoring and investment specific technology shocks, if anything, pushed the labor share upwards, especially after the 2000s.

5.2 Introducing an "automation" shock

In order to disentangle the wage markup shock, we can think about two different kind of shocks: a wage bargaining shock, which increases GDP, decreases wages and increases employment, and an automation shock, which leads to an increase in GDP, decrease in wages but also a decrease in employment. This is implemented in our VAR framework by augmenting the system of endogenous variables, including hours per capita ordered second to last in the system of variables presented above.
Figure 16 presents the impulse responses of the variables in the system to the five different shocks. The labor share decline in response to the wage bargaining shock seems to be less persistent in the case in which the automation shock is included, whereas all the other shocks show the same behavior of Figure 13. The automation shock has a strong and persistent negative effect on the labor share. Employment increases persistently after a wage bargaining shock and, even without imposing the restriction, in case of an investment specific and offshoring shock. It decreases on impact in response to an automation shock but subsequently becomes insignificant.
These results are reflected in the variance decomposition analysis, presented in Figure 17. Clearly, the automation shock seems to take away most of the importance of the wage bargaining shock both at a high and low frequency, giving first evidence of the importance of this channel. Despite this, the wage bargaining shock seem to be still relevant at a short horizon.

6 Conclusions

This paper sheds new light on the factors driving the decline of the US labor share in the last decades. Estimating a Structural VAR model with robust sign restrictions derived from a stylized DSGE model, we document that the decline of the labor share is, to large extent, explained by a decrease in the bargaining power of workers, especially in the 1990s, and by an increase in the profit share of firms, especially after the 2000s. Additionally, the labor share reacts positively, in the vast majority of draws consistent with our sign restrictions, to a positive investment specific technology shock and to an increase in offshoring of the labor-intensive component of the supply chain. Altogether, these results provide evidence of an elasticity of substitution between capital and labor smaller than one, ruling out capital deepening as a potential explanation for the decline in the labor share. In a first attempt to disentangle the shock originating in the labor market into a wage bargaining and an automation shock, we observe that both a decrease in wage bargaining and an increase in automation lead to a decrease in the labor share. Both shocks appear to be relevant in the short-run, but automation takes over in the long-run.

References

Acemoglu, D. and P. Restrepo (2016, May). The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares and Employment. NBER Working Papers 22252, National Bureau of Economic Research, Inc.

Barkai, S. (2018). Declining labor and capital shares. *Manuscript*.

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.

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

Ciminelli, G., R. Duval, and D. Furceri (2017). Employment protection deregulation and labor shares in advanced economies. *IMF Working Paper*.

Doan, T., R. Litterman, and C. Sims (1984, 01). Forecasting and conditional projection using realistic prior distributions. *3*, 1–100.

Domenico Giannione, M. L. and G. Primiceri (2015). Prior selection for vector autoregressions. *The Review of Economic Studies, 82(4), October 2015, pp. 1342-1345.*.

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.

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

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

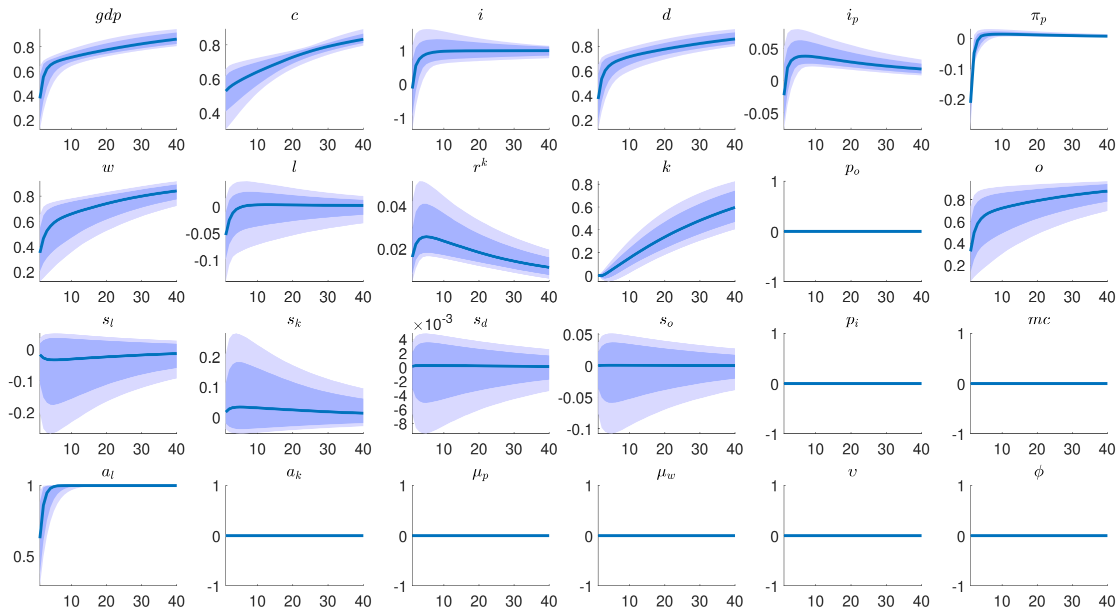
Rubio-Ramrez, J. F., D. F. Waggoner, and T. Zha (2010). Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies 77*(2), 665–696.

Sims, C. A. and T. Zha (1998). Bayesian methods for dynamic multivariate models. *International Economic Review 39*(4), 949–968.

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

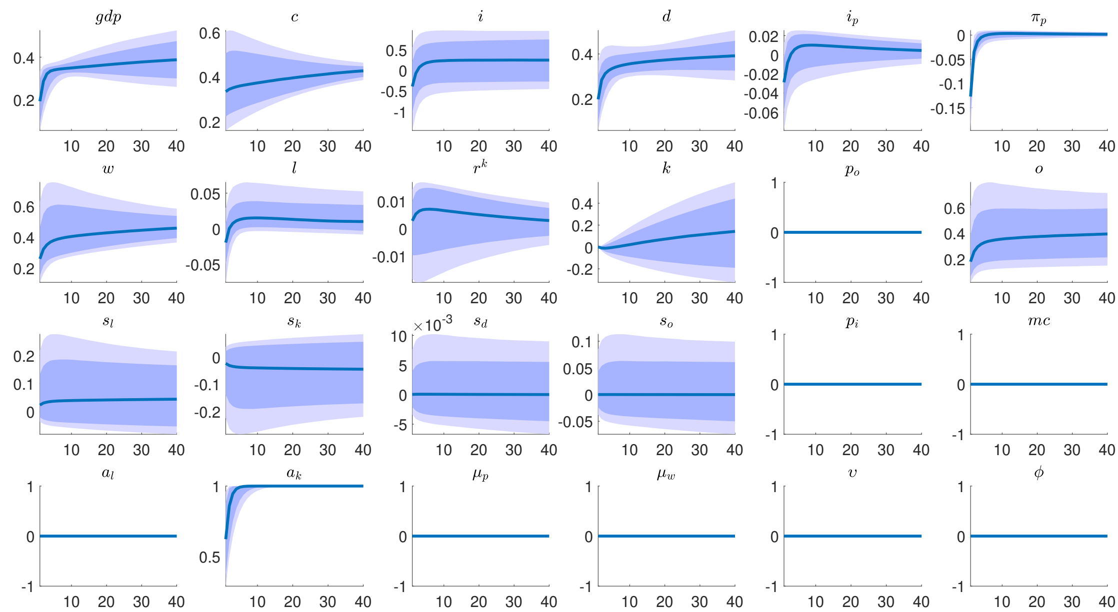
A IMPULSE RESPONSES FROM THE NEOCLASSICAL MODEL

Figure 1: Permanent labor augmenting productivity shock



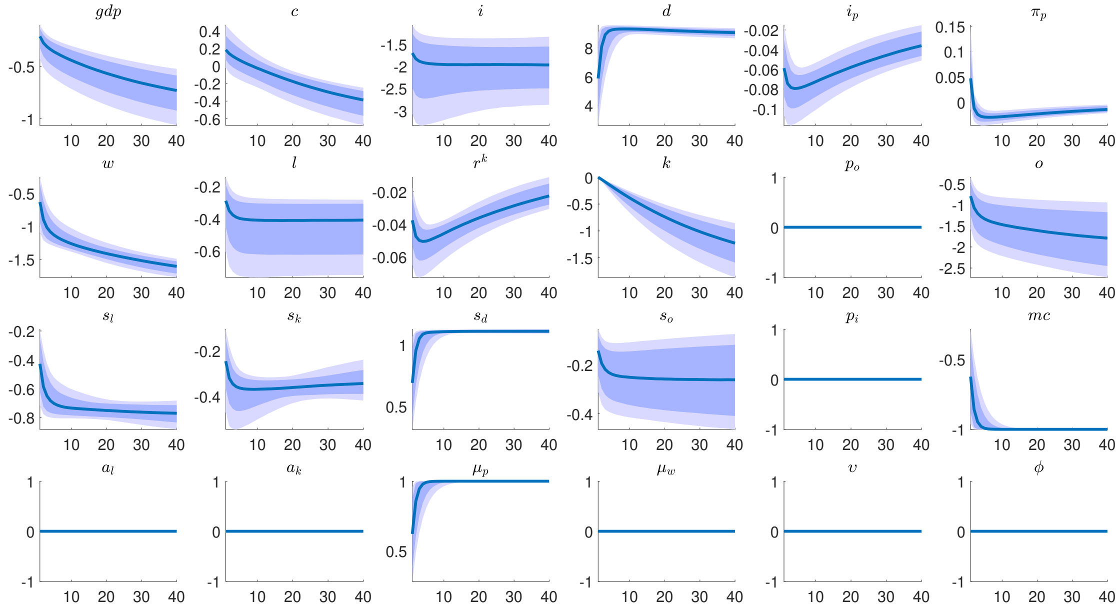
Note: Median (solid line), 90%, and 68% credible bands based on 10000 draws. Income shares and interest rates are expressed in percentage point deviations from initial values. Remaining variables are in percentage deviations. The shocks are shown in the last row.

Figure 2: Permanent capital augmenting productivity shock



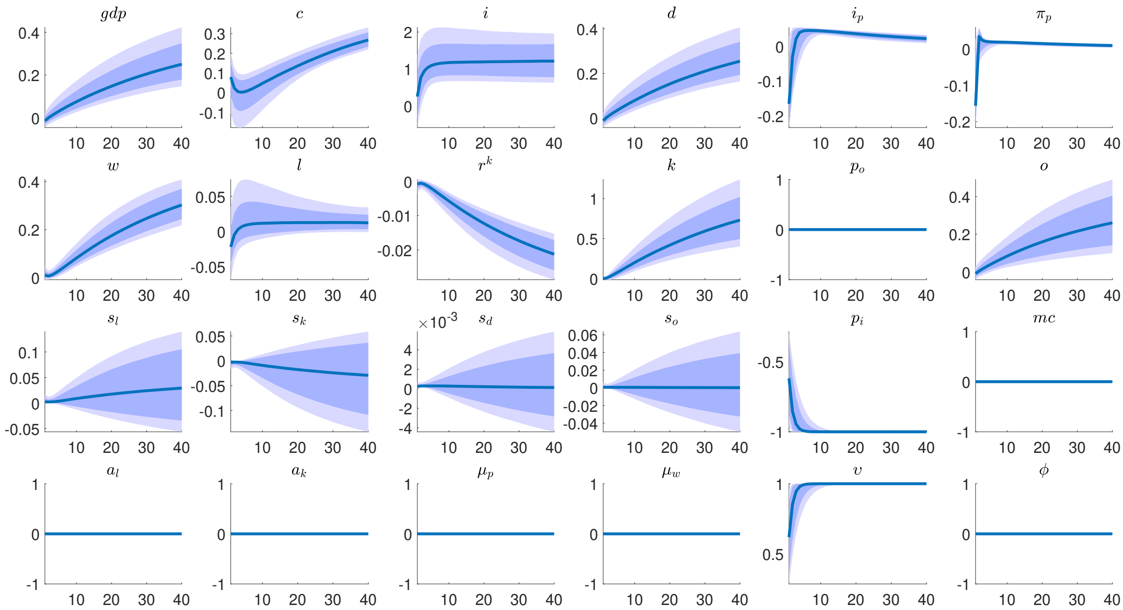
Note: See Figure 1.

Figure 3: Permanent price markup shock



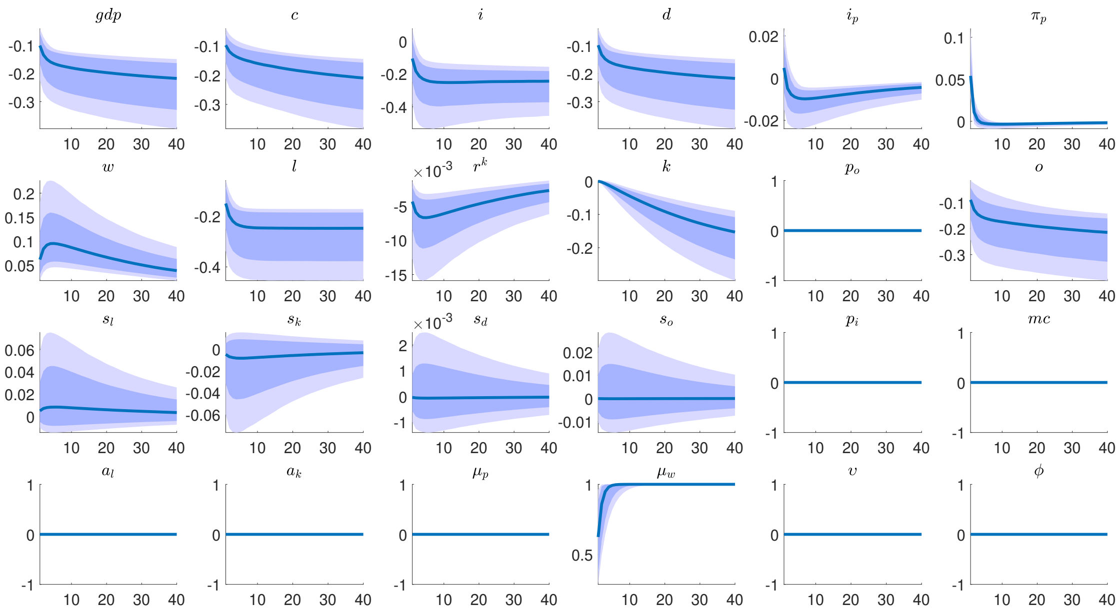
Note: See Figure 1.

Figure 4: Permanent investment specific technology shock



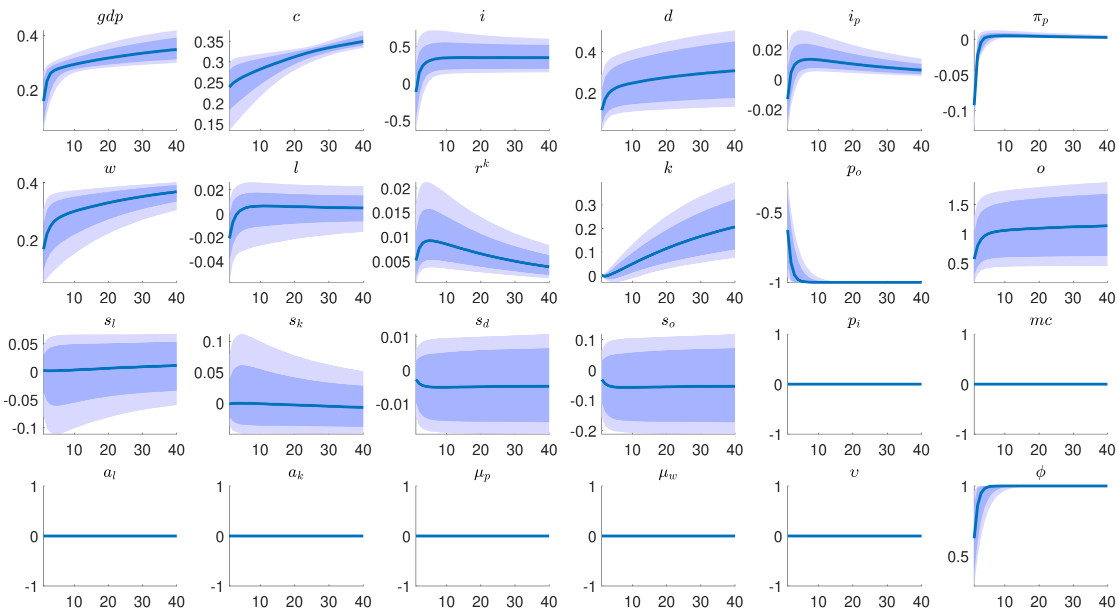
Note: See Figure 1.

Figure 5: Permanent wage markup shock



Note: See Figure 1.

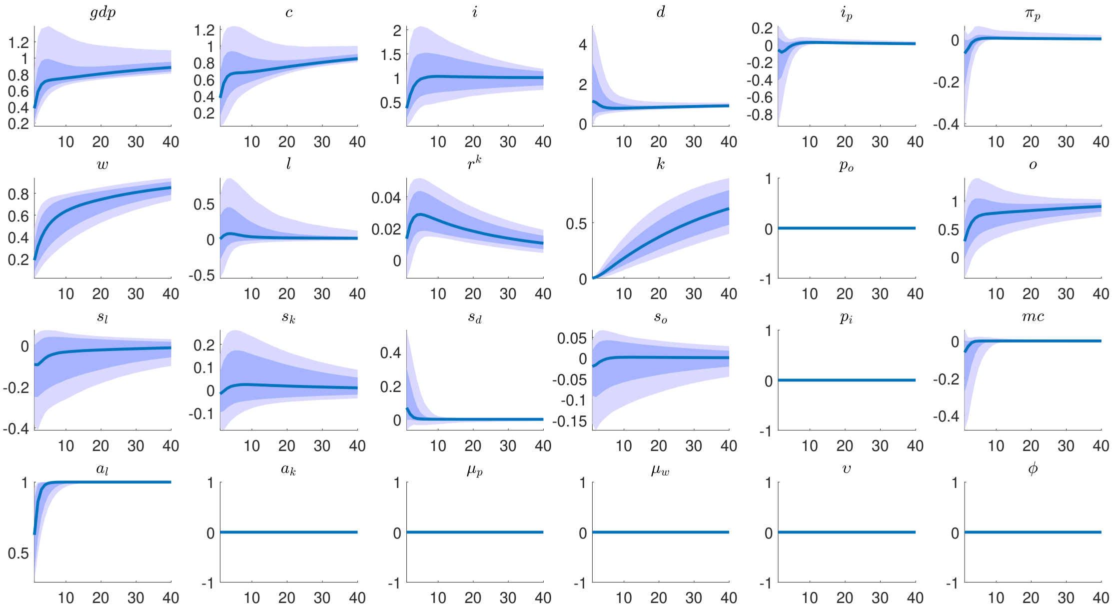
Figure 6: Permanent offshoring shock



Note: See Figure 1.

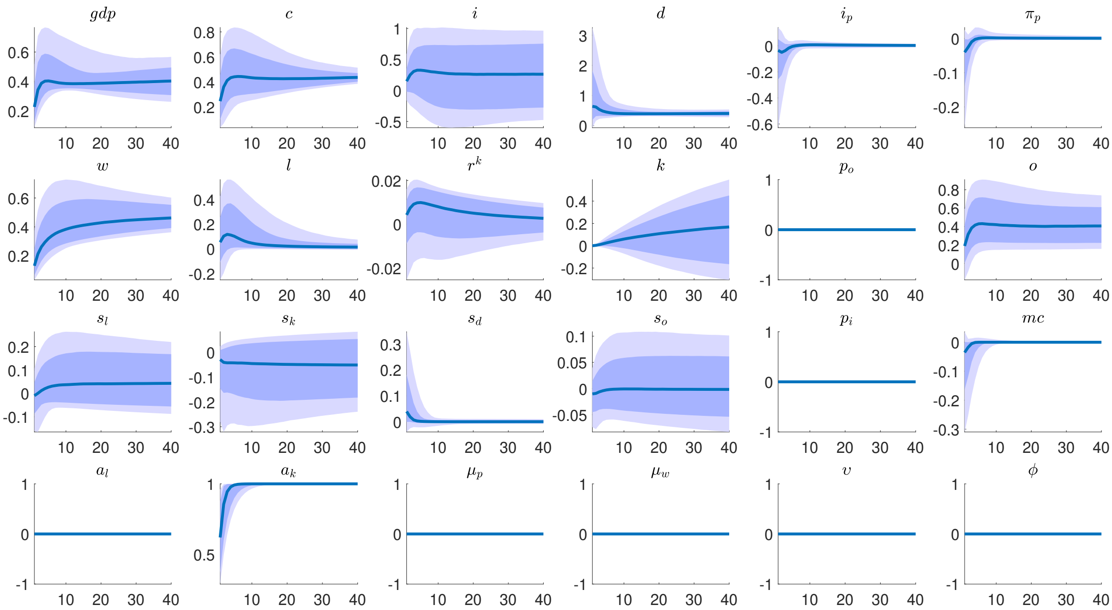
B IMPULSE RESPONSES FROM THE NEW KEYNESIAN MODEL

Figure 7: Permanent labor augmenting productivity shock



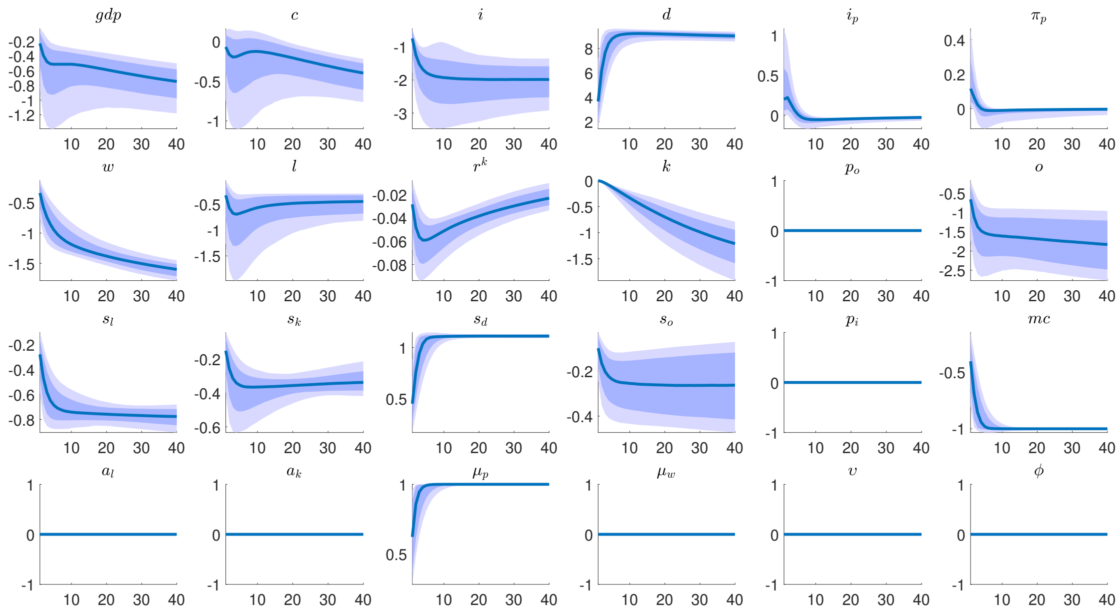
Note: Median (solid line), 90%, and 68% credible bands based on 10000 draws. Income shares and interest rates are expressed in percentage point deviations from initial values. Remaining variables are in percentage deviations. The shocks are shown in the last row.

Figure 8: Permanent capital augmenting productivity shock



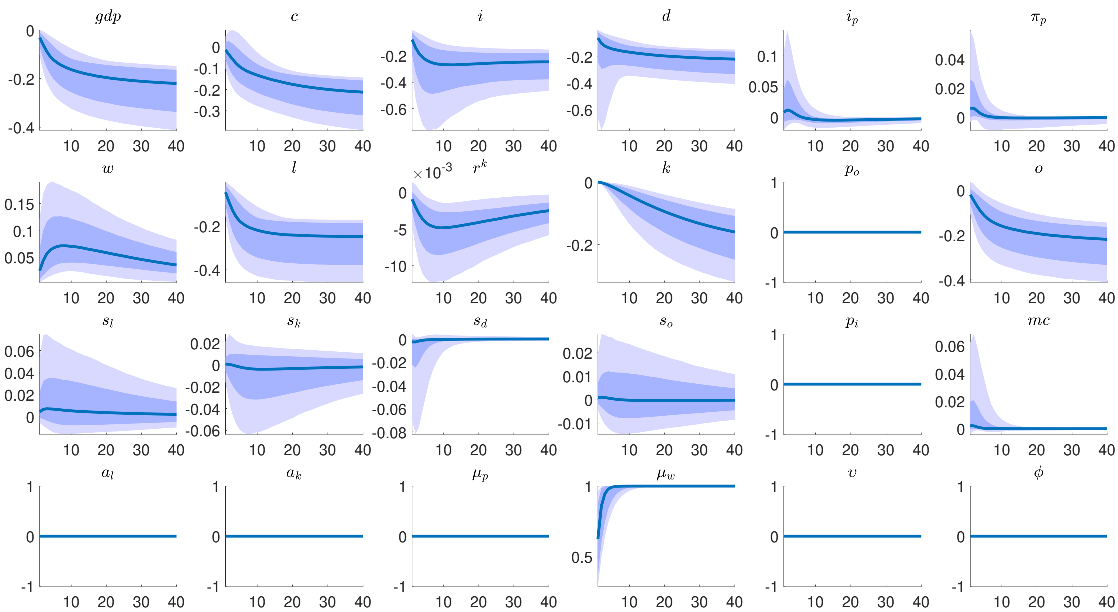
Note: See Figure 7.

Figure 9: Permanent price markup shock



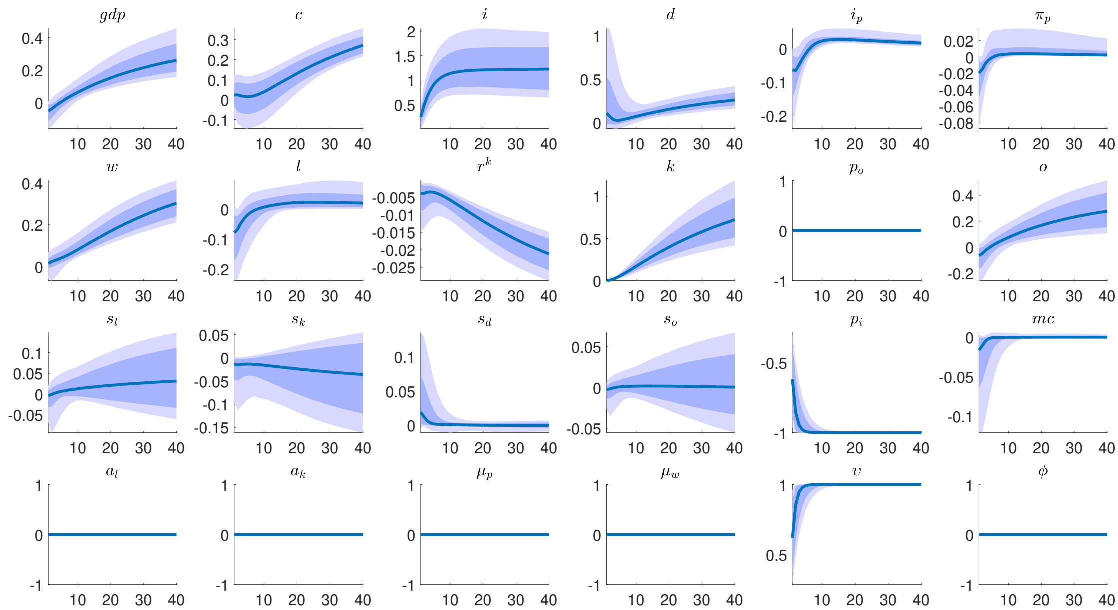
Note: See Figure 7.

Figure 10: Permanent wage markup shock



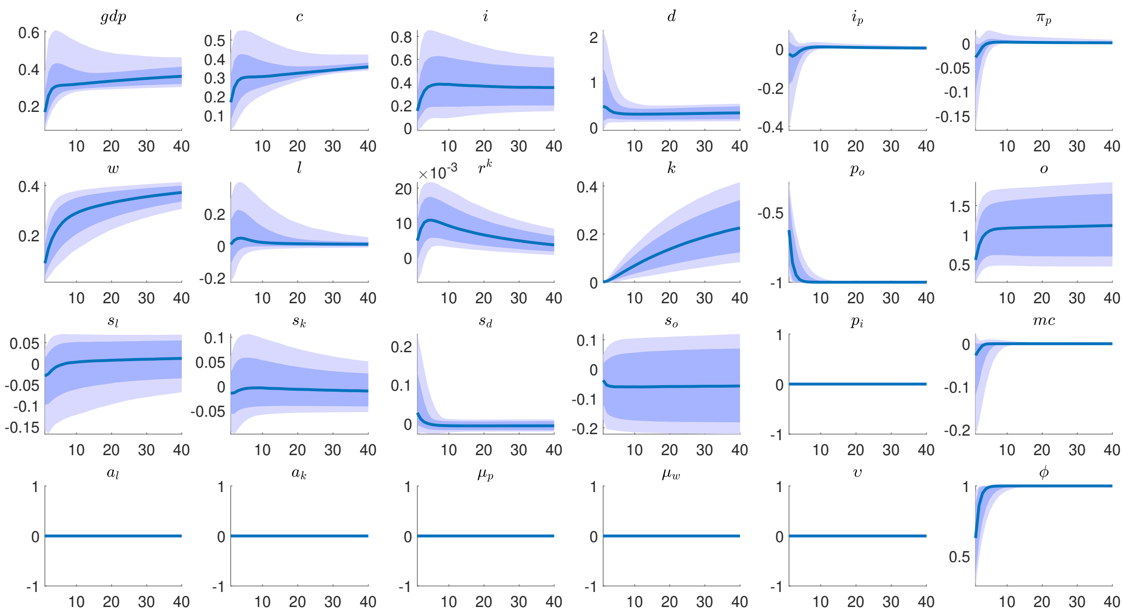
Note: See Figure 7.

Figure 11: Permanent investment specific technology shock



Note: See Figure 7.

Figure 12: Permanent offshoring shock



Note: See Figure 7.

C IMPULSE RESPONSES, VARIANCE AND HISTORICAL DECOMPOSITIONS FROM THE VAR MODEL

Figure 13: Baseline IRFs

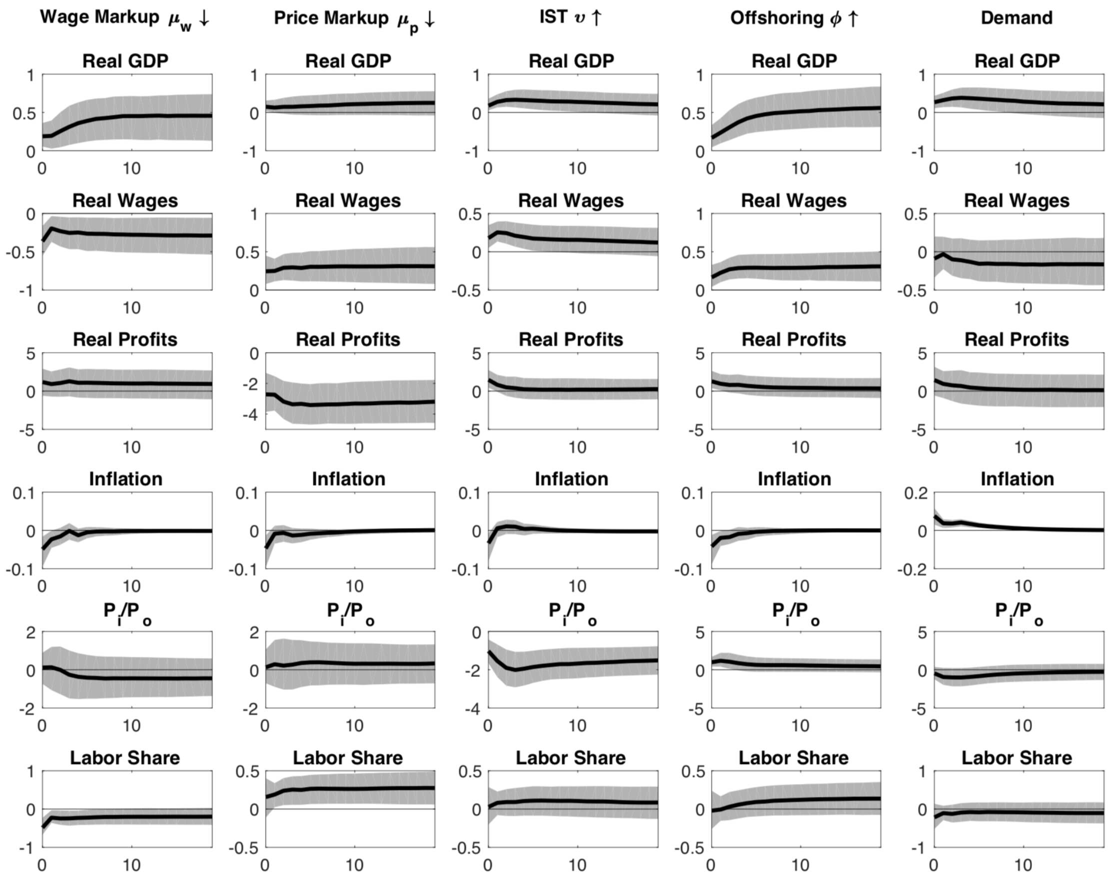


Figure 14: Baseline - variance decomposition

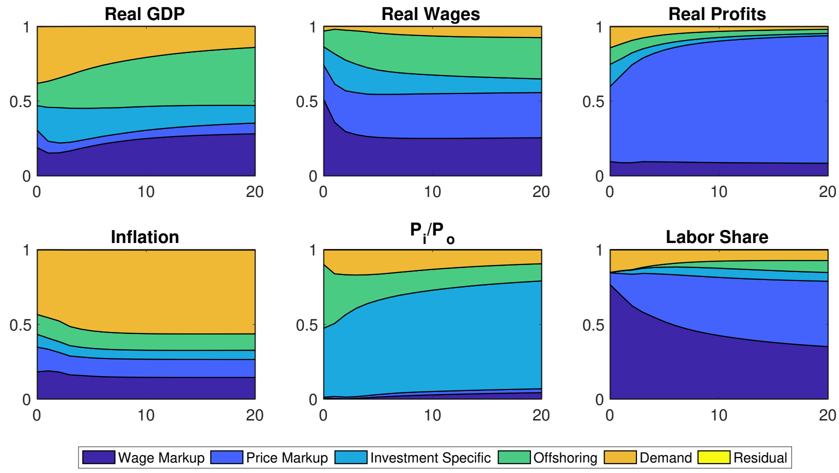


Figure 15: Baseline - historical decomposition

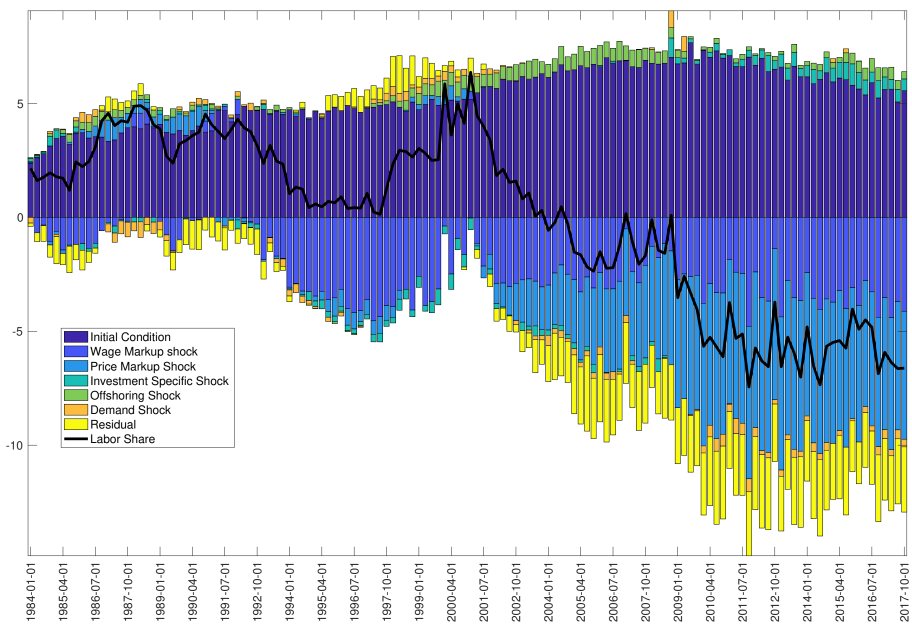


Figure 16: Including automation shock - IRFs

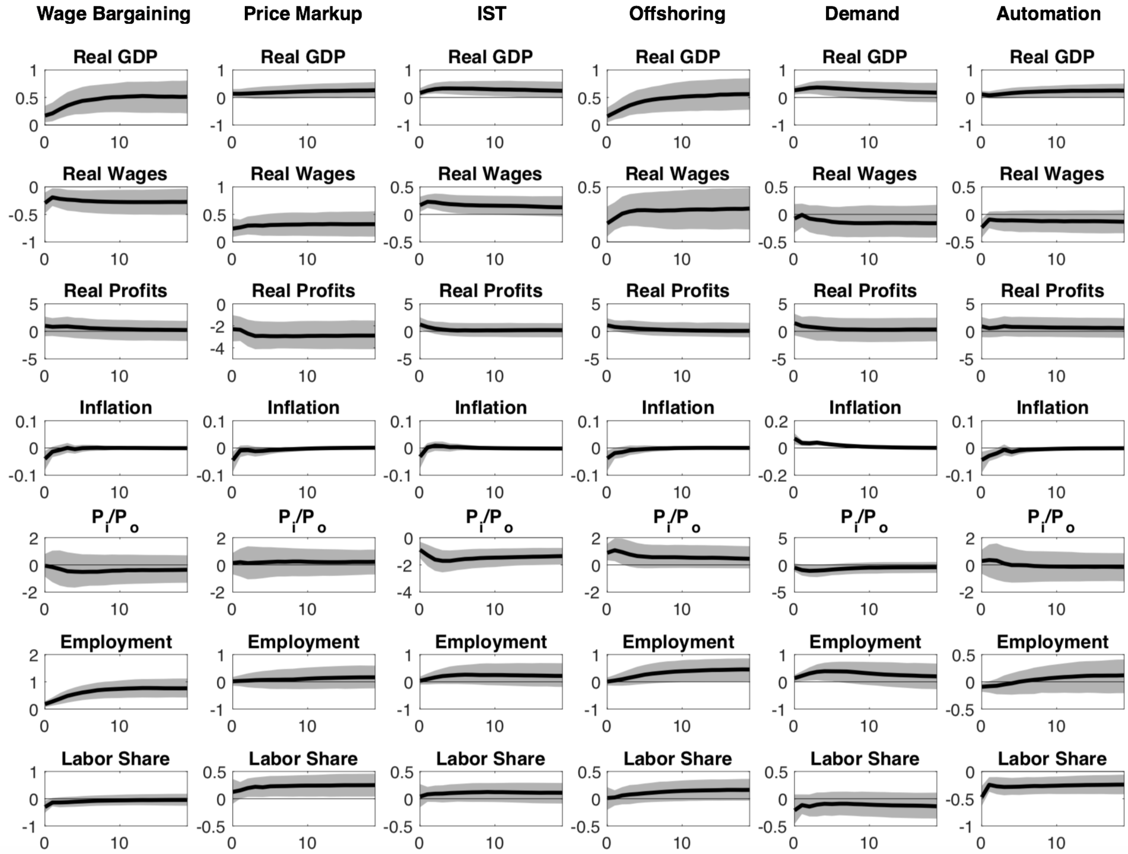


Figure 17: Including automation shock - variance decomposition