Services Deepening and the Transmission of Monetary Policy*

Alessandro Galesi
Banco de España

Omar Rachedi
Banco de España

January 12, 2017

Abstract

The structural transformation from manufacturing to services comes with a process of services deepening: the services share of intermediate inputs rises over time. Moreover, inflation reacts less to monetary policy shocks in countries which are more intensive in services intermediates. We rationalize these facts using a two-sector New Keynesian model where trends in sectoral productivities generate endogenous variations in the Input-Output matrix. Services deepening reduces the response of inflation to monetary policy shocks through a marginal cost channel. Since services prices are very sticky, the rise of services intermediates raises the sluggishness of both sectoral marginal costs and inflation rates.

Key Words: New Keynesian Model, Input-Output Matrix, Structural Change.
JEL Classification Codes: E31, E43, E52, O41.

*Addresses: alessandro.galesi@bde.es (Galesi), omar.rachedi@bde.es (Rachedi). We thank Hafedh Bouakez, Francesco Furlanetto, Patrick Gruning, and participants at several Seminars, Conferences, and Workshops for useful comments and suggestions. The Online Appendix to this paper is available on the web pages of the authors. This paper was previously circulated under the title “Structural Transformation, Services Deepening, and the Transmission of Monetary Policy”. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Banco de España or the Eurosystem.
1 Introduction

Over time advanced economies undergo a process of structural transformation that shifts the economic activity from manufacturing into services. This phenomenon raises the services share of value added, consumption, and employment (Kongsamut et al., 2001; Duarte and Restuccia, 2010; Herrendorf et al., 2013). We document that structural transformation comes with a process of services deepening: both services and manufacturing are becoming more intensive in services intermediate inputs. For instance, in the U.S. in 1947 services intermediate inputs accounted for 62% of the total inputs used by the service sector, and 20% of the inputs used in manufacturing. In 2010 these shares have peaked up to 83% and 35%, respectively.

Although the services deepening moves slowly over time, it generates important implications at the business cycle frequency. Indeed, the rise of services intermediates alters the transmission of monetary policy. We establish this fact by estimating a SVAR model for twenty-five advanced economies. We identify monetary policy shocks with sign restrictions, and find that inflation reacts less to monetary policy shocks in the countries which are more intensive in services intermediate inputs. This relationship holds even when controlling for the sectoral composition of value added.

To explore the implications of services deepening for the transmission of monetary policy, we build a New Keynesian model with two inter-connected sectors. We account for the process of services deepening by considering an Input-Output matrix that changes endogenously over time. This variation is driven by exogenous trends in sectoral productivities. We feed the model with the estimated series of sectoral productivities and calibrate it to match the sectoral reallocation experienced by the U.S. economy from 1947 to 2005. Then, we compare the dynamics of the model around the 1947 and 2005 steady-states. The two equilibria differ only in the levels of sectoral productivities. Throughout our exercise, we keep fixed all
the other parameters, included those of the Taylor rule. In this way, we can ask to what extent the services deepening alone can alter the transmission of monetary policy shocks. The model predicts that over these six decades the contemporaneous response of aggregate inflation to monetary policy shocks decreases by 37%. The dampening in the reaction of inflation raises the real effects of monetary policy, as the response of aggregate output increases by 10%.

Why does services deepening dampen the sensitivity of inflation to monetary policy shocks? The sectoral reallocation of intermediate inputs reduces the responsiveness of aggregate inflation through a *marginal cost channel*: since services prices are much stickier than manufacturing prices, the rise of services intermediates increases the sluggishness of marginal costs. As firms in either sector purchase more and more services intermediate inputs, sectoral marginal costs become stickier. Indeed, moving from the 1947 steady-state to the 2005 steady-state reduces the response of the marginal costs in manufacturing and services to monetary policy shocks by 11% and 7%, respectively. Consequently, also sectoral prices become stickier and less responsive.

In the model the trends in sectoral productivities drive also the structural transformation, intended as the sectoral reallocation of value added from manufacturing into services. Even structural transformation curbs the sensitivity of inflation to monetary policy shocks, although it does so through a different channel. When we shut down the services deepening, the changes in the sectoral composition of value added dampen the response of inflation to monetary policy shocks through a *composition channel*, that tilts the production towards services: aggregate inflation becomes stickier and less responsive to monetary policy shocks although neither sectoral marginal costs nor sectoral inflation rates change their dynamics. This

---

1 There is ample evidence showing that services prices are much stickier than manufacturing prices, e.g. Bils and Klenow (2004), Klenow and Kryvtsov (2008), and Nakamura and Steinsson (2008). The average duration of manufacturing prices is around 3 months, whereas the average duration of services prices ranges between 8 months and 13 months. Section 2.3 reviews the empirical evidence on price stickiness across sectors.
composition channel is consistent with the findings of the literature on multi-sector New Keynesian models, which shows that modeling the heterogeneity in the sectoral composition of value added is crucial to understand the properties of aggregate inflation (Carvalho, 2006; Bouakez et al., 2009, 2011, 2014; Nakamura and Steinsson, 2010; Imbs et al., 2011; Pasten et al., 2016).\footnote{Our results add to this strand of the literature by highlighting that modeling the heterogeneity in the sectoral composition of intermediate inputs is crucial to understand not only the dynamics of aggregate inflation, but also the dynamics of \textit{sectoral} inflation rates.}

Our theoretical framework is a standard Calvo staggered price New Keynesian model with two sectors - services and manufacturing - which are connected by an Input-Output matrix: each sector produces output by using labor and a bundle of services and manufacturing intermediate inputs. The two sectors differ in the durability of the produced good and in the relative use of labor and intermediates. Importantly, the two sectors differ also in the degree of price stickiness.

To account for the process of services deepening, we consider a CES aggregator of services and manufacturing inputs. In the empirically relevant case in which services intermediate inputs and manufacturing intermediate inputs are complementary, an increase in the productivity of manufacturing - relative to the productivity of services - raises firms’ expenditure in services intermediate inputs. This mechanism is the firm-analogue of the Baumol (1967) cost disease channel that Ngai and Pissarides (2007) use to generate endogenous variations in the sectoral shares of consumption and value added. As the complementarity between services and manufacturing consumption is crucial for the results of Ngai and Pissarides (2007), the changes in the Input-Output matrix towards services intermediates hinge on the complementarity between services and manufacturing intermediate inputs.

Our paper adds to the literature that studies how the transmission of monetary

\footnote{The importance of the Input-Output matrix as an amplification mechanism of macroeconomic shocks is pioneered by Long and Plosser (1983), Basu (1995), Foerster et al. (2011), and Acemoglu et al. (2012).}
policy has changed over the recent decades (e.g., Cogley and Sargent, 2001, 2005; Primiceri, 2005; Boivin and Giannoni, 2006; Pancrazi and Vukotic, 2016). The dampening in the response of inflation to monetary policy shocks is consistent with the results of Boivin and Giannoni (2006) and Pancrazi and Vukotic (2016). These papers find that inflation has become less responsive to monetary policy shocks over time, and rationalize this fact through changes in either the stance of monetary policy or the volatility of shocks. We complement this strand of the literature by providing a novel channel that can generate low-frequency movements in the effectiveness of monetary policy. Since in our model both the Taylor rule and the volatility of shocks are constant over time, the variation in the transmission of monetary policy is entirely due to the process of sectoral reallocation.

This paper relates to the structural transformation literature, which shows that advanced economies are experiencing a rise of the services share of GDP, employment, and consumption expenditures (e.g., Kongsamut et al., 2001; Duarte and Restuccia, 2010; Herrendorf et al., 2013). We complement this evidence by documenting that the changes in the sectoral composition of final shares come together with a process of services deepening. This process reminds of the capital deepening emphasized by Acemoglu and Guerrieri (2008). Although capital deepening describes an increase in capital per capita, services deepening highlights the increase in the utilization of services inputs. As Moro (2012, 2015) and Carvalho and Gabaix (2013) discuss how the changes in the sectoral composition of value added affect business cycle fluctuations, we highlight how the changes in the sectoral composition of the Input-Output matrix alter the propagation of monetary policy shocks.
2 Empirical Evidence

2.1 The Process of Services Deepening

The structural transformation literature, such as Kongsamut et al. (2001), Duarte and Restuccia (2010), Buera and Kaboski (2012), and Herrendorf et al. (2013) among others, emphasizes that economies reallocate resources from manufacturing to services as they develop. This literature mainly focuses on the dynamics of value-added, employment, and gross output. Figure 1 plots these shares computed for the U.S. economy from 1947 to 2010. Each share has been trending up since the late 1940’s. The services value added share rises from a value of 69% in 1947 to 86% in 2010. The services shares of employment and gross output display a similar behavior, increasing from 59% and 53% in 1947, up to 88% and 80% in 2010.

Figure 1: U.S. Services Value Added, Employment, and Gross Output Shares.

![Figure 1: U.S. Services Value Added, Employment, and Gross Output Shares.](image)

We document a novel dimension of the sectoral reallocation towards services. We show that over time both services and manufacturing are becoming more intensive in services inputs. We refer to this new stylized fact as *services deepening*.

Figure 2 shows the dynamics of the share of services inputs in services and
manufacturing for the U.S. economy over the last six decades. Both shares increase over time. The share of services inputs in services rises up to 83% in 2010 from a value of 62% in 1947. Similarly, the share of services inputs in manufacturing rises up to 35% in 2010 from a value of 20% in 1947.

Figure 2: U.S. Services Intermediate Inputs in Services and Manufacturing.

These dynamics are not unique to the U.S. economy. Using data from the World Input-Output Database, which covers a panel of 38 countries over the years 1995 - 2011, we document that many world economies feature a services deepening as they develop. Figure 3 shows the relationship between the share of services inputs in either services or manufacturing and the logarithm of real GDP per capita across countries. In both graphs, we plot each country-year observation, together with the fitted polynomial regression line. The graphs show that countries with higher GDP per-capita feature also higher shares of services inputs in both manufacturing and services.

---

3The countries are Australia, Austria, Belgium, Brazil, Bulgaria, Canada, China, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, India, Indonesia, Italy, Ireland, Japan, Korea, Latvia, Luxembourg, Malta, Mexico, the Netherlands, Poland, Portugal, Romania, Russia, Slovenia, Slovakia,
Figure 3: Services Intermediate Inputs Across Countries.

(a) services Industries

(b) Manufacturing Industries

Note: Panel (a) plots country-year shares of services inputs in services as a function of the logarithm of the real GDP per capita. Panel (b) plots country-year shares of services inputs in manufacturing as a function of the logarithm of the real GDP per capita. The red line indicates the fitted quadratic polynomial regression line. Data come from the World Input-Output database and cover 38 countries from 1995 to 2011.
2.2 Services Deepening and Monetary Policy Shocks

In this Section we document that the transmission of monetary policy changes with the sectoral composition of intermediate inputs. To identify this link, we exploit the cross-sectional variation of sectoral shares across advanced economies.

We build a panel of quarterly output, inflation rates, and interest rates across twenty-five countries. For each country we run a structural VAR on a vector of stationary variables \( [\Delta \log Y_t, \Delta \log P_t, R_t] \), where \( Y_t \) denotes real GDP, \( P_t \) denotes the Consumer Price Index, and \( R_t \) is the nominal short-term interest rate. We estimate the model with four lags, and we identify monetary policy shocks using a sign restriction on impulse-responses. Specifically, we posit that a monetary policy shock raises the nominal interest rate while reducing both real output growth and the inflation rate. The sign restriction is imposed not only on impact but also in the following quarter. Although we impose that inflation raises following an expansionary monetary policy shock, we are completely agnostic about how the response of inflation changes across countries with different sectoral composition of intermediate inputs.

Figure 4 reports the relationship across countries between the contemporaneous response of inflation to an expansionary monetary policy shock and the sectoral composition of intermediate inputs. The Figure shows that there is a strong negative relationship (the correlation equals \(-0.50\)), such that inflation becomes less responsive to monetary policy shocks at higher shares of services intermediate inputs.

---

4In the Online Appendix we use the cross-country data to run panel regressions to show that the share of services inputs increases with GDP per capita, even after controlling for other key characteristics such as financial development, trade openness, and human capital.

5The countries are Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Denmark, Finland, France, Germany, Indonesia, Italy, Ireland, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, Portugal, Spain, Sweden, the United Kingdom, and the United States. The Online Appendix reports all the details and the sources of the data.

6We implement the sign restriction following the algorithm of Rubio-Ramirez et al. (2010), with 5000 draws from the posterior distribution of the reduced form parameters \( B(L) \) and \( \Sigma \) with 5000 rotations each. We normalize each IRF such that the contemporaneous change in the nominal interest rate equals 100 basis points.
puts\textsuperscript{7} If we use the changes in the Input-Output of the US economy from 1947 until 2010 as an example to interpret the economic significance of the relationship of Figure 4, then the response of inflation to a 100 basis point drop in the interest rate should have decreased from 1.24% down to 0.67%, a reduction of roughly 45%.

Figure 4: Intermediate Inputs and the Transmission of Monetary Policy.

A large strand of the literature on multi-sector New Keynesian models (Carvalho, 2006; Bouakez et al., 2009, 2011, 2014; Nakamura and Steinsson, 2010; Imbs et al., 2011; Pasten et al., 2016) argues that the sectoral composition of value added affects the dynamics of inflation. Since the sectoral composition of value added and intermediate inputs are highly correlated across countries, the relationship of Figure 4 could be driven by the variation in the value added shares. To evaluate if this is the case, we study whether the relationship between the sectoral composition of intermediate inputs and the response of inflation to monetary policy shocks holds even when controlling for the sectoral composition of value added.

Table I reports the estimates of a regression in which we estimate the relation-

\textsuperscript{7}The results are robust to the case in which we consider the cumulative response of inflation within four quarters, rather than the contemporaneous response of inflation. In that case, the correlation with the services share of intermediate inputs equals $-0.46$. 
Table 1: Sectoral Shares and the Transmission of Monetary Policy

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>services Share</td>
<td>-1.61***</td>
<td>-1.52***</td>
<td>-1.81**</td>
<td>-1.59*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate Inputs</td>
<td>0.46</td>
<td>0.47</td>
<td>0.77</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>services Share</td>
<td></td>
<td></td>
<td>-1.67</td>
<td>-1.72**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value Added</td>
<td></td>
<td></td>
<td>0.98</td>
<td>0.83</td>
<td>1.24</td>
<td>1.25</td>
</tr>
<tr>
<td>Constant</td>
<td>1.79***</td>
<td>1.74***</td>
<td>2.16***</td>
<td>2.19***</td>
<td>1.55**</td>
<td>1.65**</td>
</tr>
<tr>
<td></td>
<td>0.28</td>
<td>0.23</td>
<td>0.78</td>
<td>0.65</td>
<td>0.67</td>
<td>0.64</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.25</td>
<td>0.21</td>
<td>0.10</td>
<td>0.11</td>
<td>0.26</td>
<td>0.23</td>
</tr>
<tr>
<td>N. Obs.</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

Note: The table reports the estimates of regressions where the dependent variable is the contemporaneous response of inflation to a monetary policy shock across twenty-five countries. In regression (1), the OLS regression considers the services share of intermediate inputs as independent variable. In regression (2), we weight the regression by the logarithm of countries’ GDP. In regression (3), the OLS regression considers the services share of value added as independent variable. In regression (4), we weight the regression by the logarithm of countries’ GDP. In regression (5), the OLS regression considers the services share of intermediate inputs and the services share of value added as independent variables. In regression (6), we weight the regression by the logarithm of countries’ GDP. Robust standard errors clustered at the state level are reported in brackets. *, **, and *** indicate statistical significance at the 10%, 5% and 1%, respectively.

ship between the response of inflation to monetary policy shocks and the services share of intermediate inputs and value added. Columns (1) - (2) shows that inflation becomes less responsive to monetary policy shocks at higher services shares of intermediate inputs. Columns (3) - (4) shows that there is also a negative relationship between the responsiveness of inflation and the services share of value added. Yet, when we regress the response of inflation to monetary policy shocks against both the services share of intermediate inputs and the services share of value added in Columns (5) - (6), the negative relationship between the responsiveness of inflation and the sectoral composition of intermediate inputs keeps holding, whereas the relationship with the sectoral composition of value added vanishes.
To further validate our results, in the Online Appendix we build a panel of data for thirty countries on inflation rates, nominal interest rates, the services share of intermediate inputs, the services share of value added, and other macroeconomic indicators on a sample at the annual frequency from 1995 until 2010. We run panel regressions and find that inflation becomes less correlated with the nominal interest rate at higher services shares of intermediate inputs, even after controlling for other key characteristics such as the services share of value added, GDP per capita, financial development, trade openness, human capital, and also country and time fixed effects. These results provide further suggestive evidence on the link between services deepening and the transmission of monetary policy.

2.3 Price Stickiness Across Sectors

Firms’ price setting behavior differs substantially across sectors. Table 2 reports the average duration of services prices and manufacturing prices, as estimated by Bils and Klenow (2004), Klenow and Kryvtsov (2008), and Nakamura and Steinsson (2008). These authors find that services prices are much stickier than manufacturing prices: the median duration of a manufacturing prices is around 3 months, whereas the median duration of a services price ranges between 8 and 13 months.

Why services prices are much stickier than manufacturing prices? The longer duration of prices in the service sector can be due to multiple factors: (i) the share of labor in the gross output of services is almost twice as large as the share of labor in manufacturing. Since wages are very sticky, the higher labor share of services could imply that services prices adjust less frequently; (ii) since services are largely non-tradable whereas manufacturing goods are largely tradable, services are characterized by a lower degree of price competition. The higher mark-up

---

Table 2: Price Duration Across Sectors

<table>
<thead>
<tr>
<th></th>
<th>Sector</th>
<th>Duration in Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bils and Klenow (2004)</td>
<td>Services</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>3.4</td>
</tr>
<tr>
<td>Nakamura and Steinsson (2008)</td>
<td>Services</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>3.8</td>
</tr>
</tbody>
</table>

could allow services firms to adjust less frequently their prices; (iii) durables goods are characterized by a frequent product turnover which is likely to increase the frequency of price adjustments in the manufacturing sector.

The scope of this paper is not to micro-found the asymmetry in the duration of prices across sectors, but rather to evaluate its implications through the lenses of the process of sectoral reallocation. For this reason, we consider a model with an exogenous price stickiness that differs across sectors. All the potential factors that can rationalize the different duration of sectoral prices are captured in a reduced form by differences in the parameter of price stickiness.

3 The Model

The economy is a version of a cashless Calvo (1983) staggered price New Keynesian model. We consider two sectors - services and manufacturing - which are connected through an Input-Output structure: each sector produces output by using labor and a bundle of intermediate manufactured goods and services.
The two sectors differ in the durability of the consumption good and the relative use of labor, services intermediate inputs, and manufacturing intermediate inputs. Importantly, the two sectors differ also in the degree of price stickiness.

### 3.1 Household

The economy is populated by an infinitely-lived representative household that has preferences over the consumption of services $C_s^t$, the consumption of durable manufactured goods $D_t$, and labor $N_t$. The lifetime utility of the household equals

$$
\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left( \frac{C_t^{1-\sigma}}{1-\sigma} - \psi \frac{N_t^{1+\gamma}}{1+\gamma} \right)
$$

with

$$
C_t = \left[ \omega^\frac{1}{\nu} \times C_t^s \frac{\nu-1}{\nu} + (1 - \omega)^\frac{1}{\nu} \times D_t^\frac{\nu-1}{\nu} \right]^\frac{\nu}{\nu-1}
$$

where $\beta$ is the discount factor, $\sigma$ is the inverse of the intertemporal elasticity of substitution, $\gamma$ is the inverse of the Frisch elasticity of labor supply, $\omega$ denotes the share of overall services in the CES aggregator, and $\nu$ is the elasticity of substitution across services and manufactured goods.

The stock of manufactured goods $D_t$ follows the law of motion

$$
D_t = (1 - \delta) D_{t-1} + C_t^m - \frac{\chi^2}{2} \left( \frac{C_t^m}{D_{t-1}} - \delta \right)^2 D_{t-1}
$$

where $C_t^m$ denotes the purchase of new manufactured goods at time $t$, $\delta$ is the depreciation rate, and the last term captures convex adjustment costs, which depend on the parameter $\chi$.

---

Erceg and Levin (2006) show that modeling the durability of the final good is crucial to understand the transmission of monetary policy. In the model, the non-services component of consumption is entirely composed of non-durable goods, while in the data the share of non-durables in the consumption basket exceeds 50% of the non-services component. To account for this fact, in the calibration we set the depreciation rate to the weighted average depreciation rate of the manufacturing sector.
The household maximizes lifetime utility \( U \) subject to the budget constraint

\[
P_t^s C_t^s + P_t^m C_t^m + B_t = W_t N_t + (1 + i_{t-1}) B_{t-1} + \Pi_t^s + \Pi_t^m + T_t.
\]

The household buys \( C_t^s \) services at the nominal price \( P_t^s \) and \( C_t^m \) manufactured goods at the nominal price \( P_t^m \). The household also invests in a one-period bond \( B_t \) which yields a nominal interest rate \( i_t \). In addition, the household earns a nominal labor income \( W_t N_t \), and receives nominal profits from services firms \( \Pi_t^s \) and manufacturing firms \( \Pi_t^m \), and a lump-sum nominal transfer \( T_t \).

The first-order conditions of the household’s problem read

\[
U_{C_t^s} = \frac{\omega_1^s \nu C_t^s - 1 \nu C_t^s 1 \nu - \sigma P_t^s}{P_t^s},
\]

(5)

\[
U_{C_t^s} = \beta \mathbb{E}_t \left[ (1 + i_t) U_{C_{t+1}^s} \right],
\]

(6)

\[
\psi N_t^s = \frac{U_{C_t^s} W_t}{P_t^s},
\]

(7)

\[
q_{D,t} = \left[ 1 - \chi \left( \frac{C_{t+1}^m}{D_t - \delta} \right) \right]^{-1}
\]

(8)

\[
q_{D,t} = \left( 1 - \omega \right) \frac{1}{U_{C_t^s}} D_t - \frac{1}{U_{C_t^s}} + \beta \mathbb{E}_t \left[ \frac{U_{C_{t+1}^s}}{U_{C_t^s}} q_{D,t+1} \left\{ (1 - \delta) + \ldots \right. \right.
\]

\[
\ldots + \chi \left( \frac{C_{t+1}^m}{D_t} - \delta \right) \frac{C_{t+1}^m}{D_t} - \chi \left( \frac{C_{t+1}^m}{D_t} - \delta \right)^2 \right\} \right]\]

(9)

where Equation (5) determines the marginal utility of consumption of services \( U_{C_t^s} \). Equation (6) defines the inter-temporal Euler condition for the consumption in services. Equation (7) governs the optimal supply of labor, while Equation (8) defines \( q_{D,t} \) the relative price of the durable manufactured good in terms of overall consumption, that is, the marginal utility of adding one further unit to the stock of durable goods \( D_t \). Equation (9) defines the inter-temporal Euler condition for
the consumption of the durable manufactured goods, which takes into account the convex adjustment costs.

3.2 Final Goods Firm

As in standard New Keynesian models, the production side is split in two levels: in both the service sector and the manufacturing sector there is a competitive final goods firm and a continuum of intermediate goods firms. In particular, we consider a unit measure of services intermediate goods firms indexed by $i \in [0, 1]$, whose value-added equals $Y_{i,t}^s$, and a unit measure of manufacturing intermediate goods firms indexed by $j \in [0, 1]$, whose value-added equals $Y_{j,t}^m$.

In the service sector, the competitive final goods firms aggregate the different varieties produced by the continuum of intermediate goods firms using the CES function

$$Y_t^s = \left( \int_0^1 Y_{i,t}^s \frac{\varepsilon - 1}{\varepsilon} di \right)^{\frac{\varepsilon}{\varepsilon - 1}}.$$  \hspace{1cm} (10)

Analogously, in the manufacturing sector, the competitive final goods firms use the CES function

$$Y_t^m = \left( \int_0^1 Y_{j,t}^m \frac{\varepsilon - 1}{\varepsilon} dj \right)^{\frac{\varepsilon}{\varepsilon - 1}}.$$  \hspace{1cm} (11)

The parameter $\varepsilon$ denotes the elasticity of substitution across different varieties of intermediate goods. The elasticity of substitution is constant across sectors.

Final good producers are perfectly competitive and take as given the price of the final good in each sector, $P_t^s$, $P_t^m$, and the price of each of the intermediates in each sector, $P_{i,t}^s$, $P_{i,t}^m$. As a result, the demand of each intermediate variety is isoelastic:

$$Y_{i,t}^s = \left( \frac{P_{i,t}^s}{P_t^s} \right)^{-\varepsilon} Y_t^s$$  \hspace{1cm} (12)

and

$$Y_{j,t}^m = \left( \frac{P_{j,t}^m}{P_t^m} \right)^{-\varepsilon} Y_t^m.$$  \hspace{1cm} (13)
The combination of the isoelastic demand functions (12) and (13) with perfect competition among final goods producers makes the nominal price of the final goods in each sector to be combination of the nominal price of the intermediate varieties

\[ P^s_t = \left( \int_0^1 P^s_{i,t} \frac{1}{1-\varepsilon} d\varepsilon \right) \]

and

\[ P^m_t = \left( \int_0^1 P^m_{j,t} \frac{1}{1-\varepsilon} d\varepsilon \right) \].

### 3.3 Intermediate Goods Firm

In the service sector, each individual intermediate goods firm \( i \) produces gross output using labor \( N^s_{i,t} \) and intermediate inputs \( I^s_{i,t} \) as follows

\[ G^s_{i,t} = A^s_t \left( N^s_{i,t} \right)^{\alpha_s} \left( I^s_{i,t} \right)^{1-\alpha_s} \] (16)

where \( A^s_t \) denotes the level of services productivity, which grows over time following an exogenous trend with a growth rate \( \zeta_s \), such that \( A^s_t = (1 + \zeta_s) A^s_{t-1} \), and \( \alpha_s \) is the share of labor in gross output. In the manufacturing sector, each individual intermediate goods firms \( j \) produces gross output with a technology

\[ G^m_{j,t} = A^m_t \left( N^m_{j,t} \right)^{\alpha_m} \left( I^m_{j,t} \right)^{1-\alpha_m} \] (17)

where \( A^m_t \) denotes the level of manufacturing productivity, which grows over time following an exogenous trend with a growth rate \( \zeta_m \), such that \( A^m_t = (1 + \zeta_m) A^m_{t-1} \), and \( \alpha_m \) is the share of labor in gross output. We allow the shares of labor in the production function \( \alpha_s \) and \( \alpha_m \) to differ across sectors to capture the fact that, in the data, services are more labor-intensive than manufacturing.

The intermediate inputs are modeled as a bundle of services inputs and manu-
facturing inputs. We consider a CES aggregator for both sectors:

\[ I_{s,t}^s = \left[ \frac{1}{\omega_s} \left( S_{s,i,t}^s + \bar{s}_s \right)^{\mu - 1} + (1 - \omega_s)^{\frac{1}{\mu}} \left( M_{s,i,t}^s \right)^{\mu - 1} \right]^{\frac{\mu}{\mu - 1}} \]  

(18)

and

\[ I_{m,j,t}^m = \left[ \frac{1}{\omega_m} \left( S_{m,j,t}^m + \bar{s}_m \right)^{\mu - 1} + (1 - \omega_m)^{\frac{1}{\mu}} \left( M_{m,j,t}^m \right)^{\mu - 1} \right]^{\frac{\mu}{\mu - 1}}. \]  

(19)

The variable \( S_{s,i,t}^s \) denotes the services intermediate inputs used in the production of gross output by the firm \( i \) in the service sector at time \( t \). Instead, \( S_{m,j,t}^m \) denotes the services intermediate inputs used in the manufacturing sector. Analogously, \( M_{s,i,t}^s \) and \( M_{s,j,t}^m \) are the manufacturing intermediate inputs that are used in the production of gross output in the service sector and manufacturing sector, respectively.

The parameters \( \omega_s \) and \( \omega_m \) denote the weight of services inputs in total services intermediates and manufacturing intermediates, respectively. We also add a non-homothetic component in each aggregator. This component changes across sectors: it equals \( \bar{s}_s \) in the aggregator of services intermediates and \( \bar{s}_m \) in the aggregator of manufacturing intermediates. Following the interpretation of Kongsamut et al. (2001), these parameters capture in a reduced form the amount of services inputs produced in-house by firms. The empirical evidence of Berlingieri (2014) on the marketization of firms’ services justifies our modeling choice.\(^{10}\) Finally, \( \mu \) defines the elasticity of substitution across services and manufacturing intermediates.

Each intermediate is a CES aggregator that compounds different varieties into a single input. The services inputs used in the service sector equal

\[ S_{s,i,t}^s = \left[ \int_0^1 S_{s,i,t}^s \frac{1}{\xi - 1} d\xi \right]^{\frac{\xi - 1}{\xi - 1}} \]  

(20)

\(^{10}\)The non-homothetic parameters are not crucial for our results. The only rationale for these components is to allow the model to be able to match the entire variation in the Input-Output matrix observed in the data during the calibration exercise. Yet, in the Online Appendix, we show that most of the variation of the Input-Output matrix is driven by changes in sectoral productivities. The Online Appendix reports also a version of the model without the non-homotheticities and show that even in this case the changes in the Input-Output matrix - due to the variation in sectoral productivities - would anyway dampen the response of inflation.
where $S_{i,t}^{s}$ denotes the services inputs produced by the services firm $l$ and used by the services firm $i$ at time $t$. The services inputs used in the manufacturing sector are

$$S_{j,t}^{m} = \left[ \int_{0}^{1} S_{i,j,t}^{m} \frac{\varepsilon-1}{\varepsilon} \, dl \right]^{\frac{\varepsilon}{\varepsilon-1}}. \tag{21}$$

Analogously, we define the manufacturing inputs used in both sectors as

$$M_{i,t}^{s} = \left[ \int_{0}^{1} M_{i,j,t}^{s} \frac{\varepsilon-1}{\varepsilon} \, dl \right]^{\frac{\varepsilon}{\varepsilon-1}} \tag{22}$$

and

$$M_{j,t}^{m} = \left[ \int_{0}^{1} M_{i,j,t}^{m} \frac{\varepsilon-1}{\varepsilon} \, dl \right]^{\frac{\varepsilon}{\varepsilon-1}}. \tag{23}$$

In each case, the elasticity of substitution is $\varepsilon$ as in the aggregator of the different varieties of the final goods.

Equations (18) and (19) jointly define the Input-Output matrix that links services and manufacturing. As pointed out in the literature, the role of the Input-Output matrix is a crucial ingredient for many reasons. First, it captures quantitatively relevant features of the data (Basu, 1995). Second, it amplifies the persistence of the effects arising from sectoral productivity and monetary policy shocks (Huang and Liu, 2005; Bouakez et al., 2009). Third, it allows monetary policy shocks to generate a positive co-movement in the reaction of sectoral output (Bouakez et al., 2011). Finally, it is an important determinant of the process of the structural change of an economy, which in turn affects the dynamics of key macroeconomic variables (Moro, 2012; Moro, 2015).

As we show in detail in Section 4.1, in our environment the Input-Output matrix varies endogenously over time as a function of relative sectoral productivities and the amount of output produced in the economy. As either the manufacturing productivity rises relative to the services productivity or aggregate output increases, the shares of services inputs used in manufacturing and services rise. This endoge-
nous time-variation in the inter-sectoral network structure pins down the process of services deepening.

Given the production function (16) and (17), the cost minimization problem of intermediate firms implies the following first-order conditions for the optimal amount of labor to hire and the optimal amount of total intermediate inputs to purchase

\[ N_{s,t}^s = \alpha_s MC_{s,t}^s \frac{G_{s,t}}{w_t} \]  
\[ I_{s,t}^s = \frac{(1 - \alpha_s) MC_{s,t}^s G_{s,t}^s}{P_{t}^{I_s}} \]  
\[ N_{m,t}^m = \alpha_m MC_{m,t}^m \frac{G_{m,t}^m}{w_t} \]  
\[ I_{m,t}^m = \frac{(1 - \alpha_m) MC_{m,t}^m G_{m,t}^m}{P_{t}^{I_m}} \]

where \(MC_{s,t}^s\) and \(MC_{m,t}^m\) denote the real marginal cost for services and manufacturing intermediate goods producers, respectively. The price of the bundle of services intermediate inputs and the price of the bundle of manufacturing inputs equal respectively

\[ P_{t}^{I_s} = \left(\omega_s P_t^{s,1-\mu} + (1 - \omega_s) P_t^{m,1-\mu}\right)^{\frac{1}{1-\mu}} \]  
\[ P_{t}^{I_m} = \left(\omega_m P_t^{s,1-\mu} + (1 - \omega_m) P_t^{m,1-\mu}\right)^{\frac{1}{1-\mu}}. \]

Given the intermediates aggregators (18) and (19), the first-order conditions on the decision between services intermediates and manufacturing intermediates read

\[ S_{s,t}^s = \omega_s \left(\frac{P_t^{s}}{P_{t}^{I_s}}\right)^{-\mu} I_{s,t}^s - \bar{s}_s \]  
\[ M_{s,t}^s = (1 - \omega_s) \left(\frac{P_t^{m}}{P_{t}^{I_s}}\right)^{-\mu} I_{s,t}^s \]
\[ S^m_{j,t} = \omega_m \left( \frac{P^s_t}{P^m_{t}} \right)^{-\mu} I^m_{j,t} - \bar{s}_m \]  

(32)

\[ M^m_{j,t} = (1 - \omega_m) \left( \frac{P^m_{t}}{P^m_{t}} \right)^{-\mu} I^m_{j,t}. \]  

(33)

With respect to the price setting, we assume that firms face a Calvo staggered price mechanism. In each period, a fraction \( \phi_s \) of services intermediate goods producers and a fraction \( \phi_m \) of manufacturing intermediate goods producers cannot reset prices, and maintain the price of the previous period. The fractions \( \phi_s \) and \( \phi_m \) are constant over time. Consequently, the optimal price setting problem of an intermediate goods producer in the service sector consists of maximizing the expected discounted stream of real dividends expressed in terms of the price of manufactured goods

\[
\max_{P^s_{i,t}} \mathbb{E}_t \sum_{r=t}^{\infty} \beta^r \phi^r_s \frac{U_{C^s_{i,t+r}}}{U_{C^s_t}} \frac{\Pi^s_{i,t+r} (P^s_{i,t})}{P^m_{t+r}}
\]  

(34)

where \( \Pi^s_{i,t} (P^s_{i,t}) \) denotes the nominal profits of the \( i \)-th firm in the service sector which sets its price to \( P^s_{i,t} \):

\[
\Pi^s_{i,t} (P^s_{i,t}) = P^s_{i,t} G^s_{i,t} - W_t N^s_{i,t} - P^s_{t} S^s_{i,t} - P^m_t M^s_{i,t}.
\]  

(35)

Each firm maximizes the expected discounted stream of dividends of keeping in the future the current price \( P^s_{i,t} \) with a probability \( \phi_s \).

Similarly, an intermediate goods producer in the manufacturing sector maximizes

\[
\max_{P^m_{j,t}} \mathbb{E}_t \sum_{r=t}^{\infty} \beta^r \phi^r_m \frac{U_{C^m_{j,t+r}}}{U_{C^m_t}} \frac{\Pi^m_{j,t+r} (P^m_{j,t})}{P^m_{t+r}}
\]  

(36)

where \( \Pi^m_{j,t} (P^m_{j,t}) \) denotes the nominal profits

\[
\Pi^m_{j,t} (P^m_{j,t}) = P^m_{j,t} G^m_{j,t} - W_t N^m_{j,t} - P^s_{t} S^m_{j,t} - P^m_t M^m_{j,t}.
\]  

(37)
Then, the optimal reset price for a services firm $P^*_t$ and the optimal reset price for a manufacturing firm $P^*_t$ equal respectively

$$P^*_t = \frac{\mathbb{E}_t \sum_{r=t}^{\infty} \beta^r \phi_s \frac{U_{C_s}^t}{U_{C_s}^t} \lambda_s \frac{P^*_t}{P^*_t + r} (P^*_t + S^*_t + S^m_t)}{\varepsilon - 1} \mathbb{E}_t \sum_{r=t}^{\infty} \beta^r \phi_s \frac{U_{C_s}^t}{U_{C_s}^t} \lambda_s \frac{P^*_t}{P^*_t + r} (P^*_t + S^*_t + S^m_t)} \varepsilon (Y^*_t + S^*_t + S^m_t + S^m_t)$$

and

$$P^*_t = \frac{\mathbb{E}_t \sum_{r=t}^{\infty} \beta^r \phi_m \frac{U_{C_m}^t}{U_{C_m}^t} \lambda_m \frac{P^*_t}{P^*_t + r} (P^*_t + S^*_t + S^m_t)}{\varepsilon - 1} \mathbb{E}_t \sum_{r=t}^{\infty} \beta^r \phi_m \frac{U_{C_m}^t}{U_{C_m}^t} \lambda_m \frac{P^*_t}{P^*_t + r} (P^*_t + S^*_t + S^m_t)} \varepsilon (Y^*_t + S^*_t + S^m_t + S^m_t).$$

The Calvo friction together with the optimal price setting conditions derived above imply that sectoral prices evolve as

$$P^*_t = \left[ (1 - \phi_s) \left( P^*_t \right)^{1-\varepsilon} + \phi_s \left( P^*_t \right) \right]^{1-\varepsilon}, \quad (38)$$

and

$$P^*_t = \left[ (1 - \phi_m) \left( P^*_t \right)^{1-\varepsilon} + \phi_m \left( P^*_t \right) \right]^{1-\varepsilon}. \quad (39)$$

### 3.4 Closing the Model

In any time period, the aggregate nominal GDP equals the sum of the nominal services value added and the nominal manufacturing value added, that is

$$P_t Y_t = P^*_s Y^*_t + P^*_m Y^*_t.$$  \quad (40)

To compute the series of real GDP we follow the same definition used by NIPA. We set the real aggregate GDP by fixing constant base year-prices. In particular, we normalize the base-year prices to one, such that the aggregate real GDP reads

$$Y_t = \bar{P}^*_s Y^*_t + \bar{P}^*_m Y^*_t = Y^*_t + Y^*_t.$$  \quad (41)
The aggregate price level is then derived as a GDP deflator, obtained by computing the ratio of nominal GDP to real GDP

\[
P_t = \frac{P_t^s Y_t^s + P_t^m Y_t^m}{Y_t}
\]

so that the aggregate inflation rate equals \(1 + \pi_t = \frac{P_t}{P_{t-1}}\).

To close the model, we define the Taylor rule as

\[
\frac{1 + i_t}{\bar{i}} = \left(\frac{1 + \bar{i}}{1 + \bar{i}}\right)^{\rho_i} \left[ (1 + \pi_t)^{\phi_\pi} (x_t)^{\phi_y} \right]^{1 - \rho_i} \exp (\epsilon^r_t)
\]

where \(x_t\) defines the output gap, that is \(x_t = \log \left(\frac{Y_t}{Y_t^{FLEx}}\right)\), where \(Y_t^{FLEx}\) is real GDP of the economy with fully flexible prices, \(\bar{i}\) denotes the steady-state interest rate, \(\rho_i\) captures the degree of the inertia in the nominal interest rate, \(\phi_\pi\) and \(\phi_y\) define the elasticity at which monetary authorities adjust the interest rate to movements in the current inflation rate and output gap, respectively, and \(\epsilon^r_t\) is an IID monetary policy shock such that \(\epsilon^r_t \sim IID N(0, \sigma^r)\).

Finally, in a symmetric equilibrium the market-clearing conditions for final services and final manufactured goods are

\[
C_t^s = Y_t^s = \left[ \int_0^1 Y_{i,t}^{s_1} \right]^{\frac{\bar{\rho}}{\epsilon - 1}}
\]

and

\[
C_t^m = Y_t^m = \left[ \int_0^1 Y_{j,t}^{m_1} \right]^{\frac{\bar{\rho}}{\epsilon - 1}}
\]

The market-clearing conditions for the intermediate firms posit that gross output should be allocated between the intermediate inputs provided to either sector and the different varieties of final goods provided to the final good firms, that is

\[
\int_0^1 G_{i,t}^s di = \int_0^1 S_{i,t}^s di + \int_0^1 S_{j,t}^m dj + \int_0^1 Y_{i,t}^s di
\]
The last market-clearing condition refers to the labor market, and states that the total labor supplied by the households equals the sum of the amount of labor demanded by the service sector and the manufacturing sector,

\[ N_t = N^s_t + N^m_t = \int_0^1 N^s_{i,t} \, di + \int_0^1 N^m_{j,t} \, dj. \]

This market-clearing condition assumes that labor can freely relocate across sectors.

### 4 Workings of the Model

#### 4.1 A Characterization of Services Deepening

We propose a mechanism that generates an endogenous variation of the Input-Output matrix through two channels:

(i) a Baumol (1967) cost disease channel that raises the share of services inputs when the relative productivity of manufacturing rises, as in Ngai and Pissarides (2007);

(ii) the non-homotheticities in the intermediate aggregators, which raise the share of services inputs when output increases.

In this Section, we characterize analytically the process of services deepening. For ease of exposition we make the following assumptions: (i) the weight given to services inputs in the CES aggregator of intermediates is the same across sectors, that is, \( \omega_s = \omega_m = \omega \); (ii) the gross output of both sectors is just a function of productivities and intermediate inputs, that is, \( \alpha_s = \alpha_m = 0 \). Given these assumptions, in the steady-state the shares of services inputs in services and manufacturing equal respectively.
\[ \frac{P^s S^s}{P^{I^s} I^s} = \frac{\omega Z^{1-\mu}}{\omega Z^{1-\mu} + (1 - \omega)} - \frac{s_s}{A_s} \left( \omega + \frac{1 - \omega}{Z^{1-\mu}} \right) \]  
(44)

and

\[ \frac{P^s S^m}{P^{I^m} I^m} = \frac{\omega Z^{1-\mu}}{\omega Z^{1-\mu} + (1 - \omega)} - \frac{s_m}{A^m} \left( \omega + \frac{1 - \omega}{Z^{1-\mu}} \right) \]  
(45)

where \( Z = A^m / A^s \) denotes the relative productivity of manufacturing. Equations (44) and (45) show that the shares of services inputs in both sectors depend on two components. The first one highlights the contribution of the Baumol disease channel and the second one captures the variation due to the non-homothetic components.

First, let us abstract from the non-homothetic components by setting \( \bar{s}_s = \bar{s}_m = 0 \). In this case, the shares of services inputs are the same across sectors and depend on the weight of services inputs in the CES aggregators \( \omega \), the relative productivity of manufacturing \( Z \), and the elasticity of substitution between services inputs and manufacturing inputs \( \mu \). As long as manufacturing and services intermediates are relatively poor substitutes (i.e., \( \mu < 1 \)), an increase in the relative productivity of manufacturing raises the share of services inputs:

\[ \frac{\partial \left( \frac{P^s S^s}{P^{I^s} I^s} \right)}{\partial Z} = \frac{\partial \left( \frac{P^s S^m}{P^{I^m} I^m} \right)}{\partial Z} = \frac{\partial \left( \frac{\omega Z^{1-\mu}}{\omega Z^{1-\mu} + (1 - \omega)} \right)}{\partial Z} > 0. \]

This derivative captures the Baumol cost disease channel, according to which relative productivities affect the sectoral allocation of inputs by changing the relative price across sectors. The services share increases with the relative productivity \( Z \) only if \( \mu < 1 \). Although there is no estimate of \( \mu \), the literature has extensively documented that the analogous elasticity in the consumption bundle of the households is well below unity. Duarte and Restuccia (2010) consider an elasticity of substitution of 0.8 while Herrendorf et al. (2013) estimate a value as low as 0.002. We conjecture that also manufacturing inputs and services inputs are weak substitutes because firms cannot easily replace a consultant with some manufacturing goods.
Hence, we think that the empirically relevant value of the elasticity of substitutions between services and manufacturing inputs is $\mu < 1$.

Now let us abstract from the changes in relative productivities by setting $Z = 1$, such that $A^s = A^m = A$. In this case, the share of services inputs in services equals

$$\frac{P^s S^s}{\bar{I}^s I^s} = \omega - \bar{s}_s$$

and the share of services inputs in manufacturing equals

$$\frac{P^s S^m}{\bar{I}^m I^m} = \omega - \bar{s}_m.$$

These conditions posit that the shares of services inputs in both sectors are inversely related to the non-homothetic components $\bar{s}_s$ and $\bar{s}_m$. Nevertheless, when productivity rises, the negative contribution of the non-homothetic components vanishes over time. In this way, an increase in output leads to a switch towards services inputs even in the absence of movements in the relative productivities.

### 4.2 The Role of Sectoral Productivities

Our mechanism hinges on the exogenous trends in sectoral productivities. As the productivity of manufacturing grows faster than the productivity of services, as it happens in the data, then the Input-Output matrix undergoes an endogenous process of sectoral reallocation towards services intermediate inputs. In this Section, we propose a simple numerical illustration of the effects of the changes in sectoral productivities on both the sectoral reallocation of intermediate inputs and the transmission of monetary policy.

The Panels (a) and (b) of Figure 5 report the model implications on the services share of intermediate inputs used by the service sector and the services share of intermediate inputs used by the manufacturing sector as a function of the relative
sectoral productivity, measured as the ratio between the productivity of manufacturing over the productivity of services. Hence, the relative sectoral productivity rises when the productivity of manufacturing increases more than the productivity of services. To highlight the role of the CES intermediate aggregator for the endogenous variation of the Input-Output matrix, in each Panel we show the predictions of both the benchmark model and a counterfactual model without services deepening, in which the intermediate aggregator is a Cobb-Douglas function.

Consistently with the previous analytical characterization of services deepening, Panels (a) and (b) show that the services share of intermediate inputs in either sector rises with the relative sectoral productivity. In the case we consider a Cobb-Douglas aggregator for the intermediate inputs (i.e., when $\mu = 1$), then the process of services deepening is shut down and the sectoral shares of the Input-Output matrix are constant over time, independently of the level of sectoral productivities.

What are the implications of the changes in sectoral productivities for the transmission of monetary policy? Panels (c) and (d) report the responses of services inflation and manufacturing inflation to a monetary policy shock as a function of the relative sectoral productivity. In the model with services deepening, the rise in the relative sectoral productivity dampens the response of both services inflation and manufacturing inflation to a monetary policy shock. Instead, when we shut down the services deepening, the responses of sectoral inflation rates hardly change.\(^1\)

How does the services deepening affect the dynamics of sectoral inflation rates? Panels (e) and (f) show the responses of services marginal costs and manufacturing marginal costs to a monetary policy shock as a function of the relative sectoral productivity. The figures show that the rise in the relative sectoral productivity

---

\(^1\)As the model without services deepening still features a structural transformation that shifts the sectoral shares of value added, the rise of services value added alters the degree of strategic complementarities across sectors and changes the responses of sectoral inflation rates. Yet, these changes are very small. In the following Section, we show that indeed the strategic complementarities channel of structural transformation is quantitatively negligible.
Figure 5: Relative Sectoral Productivity and the Response to Monetary Policy Shocks.

(a) services Intermediates Share in Services
(b) services Intermediates Share in Manufacturing

(c) Services Inflation
(d) Manufacturing Inflation

(e) Services Marginal Cost
(f) Manufacturing Marginal Cost

Note: Panel (a) plots the model prediction on the share of services intermediate inputs used by the service sector as a function of the relative sectoral productivity, measured as the ratio between the productivity of manufacturing over the productivity of services. Panel (b) plots the model prediction on the share of services intermediate inputs used by the manufacturing sector as a function of the relative sectoral productivity. Panel (c) plots the model prediction on the response of services inflation to an expansionary monetary policy shock as a function of the relative sectoral productivity. Panel (d) plots the model prediction on the response of manufacturing inflation to an expansionary monetary policy shock as a function of the relative sectoral productivity. Panel (e) plots the model prediction on the response of services marginal costs to an expansionary monetary policy shock as a function of the relative sectoral productivity. Panel (f) plots the model prediction on the response of manufacturing marginal costs to an expansionary monetary policy shock as a function of the relative sectoral productivity. In each Panel, the continuous line indicates the predictions of a model with services deepening, whereas the dashed line indicates the predictions of a model without services deepening.
curbs substantially the responsiveness of sectoral marginal costs to a monetary policy shock. Again, without services deepening the behavior of the marginal costs barely change. These plots highlight that services deepening propagates through a marginal cost channel that alters the dynamics of each sectoral price. As both sectors increase the use of services inputs, marginal costs become stickier and less responsive to shocks. Then, the changes in the dynamics of marginal costs spill over to the behavior of sectoral prices.

Figure 6 plots the response of aggregate inflation to a monetary policy shock for different values of the relative sectoral productivity. As the relative productivity rises, the increase in the services share of intermediate inputs dampens the responsiveness of aggregate inflation.

Figure 6: The Response of Aggregate Inflation to Monetary Policy Shocks.

Nevertheless, when we shut down the changes in the Input-Output matrix, the model still experiences a reduction in the responsiveness of aggregation inflation, even though to a lesser extent. Indeed, also in the absence of services deepening, the model features a structural transformation in value added shares, which shifts the
economic activity towards services, that is the sector with the highest degree of price rigidity and the lowest responsiveness to monetary policy shocks. Consequently, the rise of services value added raises the stickiness of aggregate inflation, which then becomes less responsive to monetary policy shocks. This result points out that the sectoral reallocation in value added affects the transmission of monetary policy through a composition channel, which raises the relative importance of the sector with the highest degree of price rigidity in the economy, without altering the behavior of each individual sectoral inflation rate.

Overall we find that services deepening alters the transmission of monetary policy through a marginal cost channel. This novel channel adds to the literature on multi-sector New Keynesian models, that studies the link between the sectoral composition of the economy and the dynamics of inflation (Carvalho, 2006; Bouakez et al., 2009, 2011, 2014; Nakamura and Steinsson, 2010; Imbs et al., 2011; Pasten et al., 2016). This strand of the literature shows that modeling the heterogeneity in the sectoral composition of value added is crucial to understand the properties of aggregate inflation. Instead, we highlight that modeling the heterogeneity in the sectoral composition of intermediate inputs is crucial to understand not only the dynamics of aggregate inflation, but also the dynamics of sectoral inflation rates. The next Section quantifies the importance of the marginal cost channel of services deepening.

5 Quantitative Analysis

5.1 Calibration

In the calibration exercise, we group the parameters of the model in two sets. The first set of parameters is calibrated as in standard New Keynesian models, with the unique difference that we consider two inter-connected sectors. The second set
of parameters is calibrated to match the changes in the sectoral composition of intermediate inputs experienced by the U.S. economy from 1947 to 2005. In this way, the model can fully account for the process services deepening observed in the U.S. economy over the last six decades. In all the exercises, we fix one period of the model to coincide with a quarter.

Table 3 reports the values of the first set of parameters. We choose the parameters of the utility function such that the inverse of the elasticity of intertemporal substitution is $\sigma = 2$ and the inverse of the Frisch elasticity is $\gamma = 1$, which are standard values in the literature. The relative disutility of labor is set to $\psi = 15.7828$ to match a steady-state labor of $\bar{N} = 0.33$. We set the time discount parameter to $\beta = 0.995$ to have a steady-state annual inflation rate of 2%. We set the elasticity of substitution between services and manufacturing in consumption to $\nu = 0.4$. Although this value is slightly lower than the 0.8 used in Duarte and Restuccia (2007), our choice is anyway much higher than the value of $\nu = 0.002$ that Herrendorf et al. (2013) find in consumption expenditure data. We set the elasticity of substitution between services and manufacturing inputs as $\mu = 0.4$ to equalize the elasticity of substitution on the consumption side.

To set the depreciation rate of the manufactured good, first we consider the fact that from 1947 to 2005 around 60% of the manufactured goods were durable while the remaining 40% were non-durable. These shares are rather constant over time. We weight the depreciation rates of durables and non-durables with their shares in total manufacturing and find that the implied average quarterly depreciation rate of manufactured goods equals $\delta = 0.154$.

The adjustment cost parameter is calibrated to match the relative contemporaneous response of manufacturing output with respect to services output to a monetary policy shock. To do so, we run a SVAR model with services inflation, services output, manufacturing inflation, manufacturing output, and nominal interest rates.
Table 3: Calibrated Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Target/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity Intertemporal Substitution</td>
<td>$\sigma = 2$</td>
<td>Standard value</td>
</tr>
<tr>
<td>Inverse Frisch elasticity</td>
<td>$\gamma = 1$</td>
<td>Standard value</td>
</tr>
<tr>
<td>Relative Disutility of Labor</td>
<td>$\psi = 15.7828$</td>
<td>Steady-State Labor $= 0.33$</td>
</tr>
<tr>
<td>Elasticity of Substitution Consumption</td>
<td>$\nu = 0.4$</td>
<td>Standard value</td>
</tr>
<tr>
<td>Elasticity of Substitution Inputs</td>
<td>$\mu = 0.4$</td>
<td>Elasticity of Substitution Consumption</td>
</tr>
<tr>
<td>Time discount</td>
<td>$\beta = 0.995$</td>
<td>Steady-State Annual Interest Rate $= 0.02$</td>
</tr>
<tr>
<td>Depreciation Rate Manufacturing</td>
<td>$\delta = 0.1$</td>
<td>Avg. Depreciation of Manufactured Goods</td>
</tr>
<tr>
<td>Adjustment Cost</td>
<td>$\chi = 2.4602$</td>
<td>Output Response to Monetary Policy Shock</td>
</tr>
<tr>
<td>Elasticity of Substitution Intermediate Goods</td>
<td>$\varepsilon = 6$</td>
<td>Standard value</td>
</tr>
<tr>
<td>Labor Share Services</td>
<td>$\alpha_s = 0.5279$</td>
<td>Intermediates Share Services</td>
</tr>
<tr>
<td>Labor Share Manufacturing</td>
<td>$\alpha_m = 0.2902$</td>
<td>Intermediates Share Manufacturing</td>
</tr>
<tr>
<td>Calvo Frequency Services</td>
<td>$\phi_s = 0.7788$</td>
<td>Duration of Prices of 12 Months</td>
</tr>
<tr>
<td>Calvo Frequency Manufacturing</td>
<td>$\phi_m = 0.3679$</td>
<td>Duration of Prices of 3 Months</td>
</tr>
<tr>
<td>Interest Rate Inertia</td>
<td>$\rho_i = 0.8$</td>
<td>Standard Value</td>
</tr>
<tr>
<td>Taylor Parameter Inflation</td>
<td>$\phi_\pi = 1.5$</td>
<td>Standard Value</td>
</tr>
<tr>
<td>Taylor Parameter Output Gap</td>
<td>$\phi_\pi = 0.2$</td>
<td>Standard Value</td>
</tr>
<tr>
<td>Standard Deviation Monetary Policy Shock</td>
<td>$\sigma_{\epsilon} = 0.1$</td>
<td>Standard Value</td>
</tr>
</tbody>
</table>
For the sectoral variables, we consider data on sectoral personal consumption expenditures from the BEA from 1947Q1 to 2005Q4. We identify monetary policy shocks using sign restrictions by assuming that a monetary policy shock raises on impact (and the following quarter) the nominal interest rate, and decreases sectoral outputs and inflation rates. We find that the contemporaneous response of manufacturing output is 2.435 times as large as the one of services output. The model matches this value with an adjustment cost parameter that equals $\chi = 2.4602$.

We set both the elasticity of substitution across intermediate goods in both the service sector and manufacturing sector to $\varepsilon = 6$, which is the value estimated by Christiano et al. (2005).

With respect to the calibration of the parameters of the Cobb-Douglas technology, we discipline our choices using data from the 35-sector Jorgenson (2007) database on sectoral value added and sectoral intermediate inputs value over the period 1947 to 2005. First, we derive from the data a model consistent definition of gross output, which sums the labor compensation to the compensation of intermediate inputs. Then, we compute the average share of intermediates in gross output over the sample period for both sectors. For manufacturing, the share of intermediates on value added equals 0.7098. This value implies a labor share for manufacturing that equals $\alpha_m = 0.2902$. For services, the share of intermediates on value added equals 0.4721, which implies a labor share of $\alpha_s = 0.5279$.

For the price friction, we follow the evidence of Bils and Klenow (2004), Klenow and Kryvtsov (2008), and Nakamura and Steinsson (2008), who point out that the median duration of prices in the service sector ranges between 8 months and 13 months, while manufactured goods have a much lower duration of 3.2 months. Accordingly, we set $\phi_s = 0.7788$ for services firms, which implies an average duration of 12 months, and $\phi_m = 0.3679$ for manufacturing firms, which implies an average
duration of 3 months.\footnote{The implied aggregate duration of prices, once averaging sectoral durations using the equilibrium value added shares, equals 7 months in the 1947 equilibrium and 9 months in the 2005 equilibrium, in line with the values typically estimated in the literature.}

We set the parameters of the Taylor rule following the estimates of Clarida et al. (2000), which have become standard reference values in the literature. We set the inertia in nominal interest rate to $\rho_i = 0.8$ and the Taylor rule coefficients of inflation and output gap are set to $\phi_\pi = 1.5$ and $\phi_y = 0.2$, respectively. Finally, the standard deviations of the monetary policy shocks is set to $\sigma_\epsilon = 0.1$.

Then, we report in Table 4 the parameters that discipline the process of services deepening in the model. There are nine parameters to be calibrated: the weight of services in the consumption aggregator $\omega$, the weight of services in the aggregator of services inputs $\omega_s$, the value of in-house production of services inputs in services $\bar{s}_s$, the weight of services in the aggregator of manufacturing inputs $\omega_m$, the value of in-house production of services inputs in manufacturing $\bar{s}_m$, the levels of the productivity of manufacturing $A^m$ and services $A^s$ in the 1947-model and in the 2005-model.

First, we compute gross output productivities using data from the 35-sector Jorgenson (2007) dataset.\footnote{To avoid that the noise on the size of mark-ups could affect the estimates of productivities, we introduce in the model a fixed production subsidy which offsets firms’ market power. This feature alters only the steady-state of the model and not its dynamics. In this way, we can derive in the data the equivalent version of gross output productivities defined in the model and abstract from the computation of mark-ups.} Then we normalize the productivities in 1947 to unity. This procedure yields 2005 productivity values of $A^s = 1.2675$ for services and $A^m = 1.6941$ for manufacturing. We calibrate the remaining five parameters to match five moments: the share of services value added in 1947 (68.9%), the share of services inputs in services in 1947 (61.5%), the share of services inputs in services in 2005 (83.6%), the share of services inputs in manufacturing in 1947 (20.0%), and the share of services inputs in manufacturing in 2005 (33.6%). In this way, we find $\omega = 0.4677$, $\omega_s = 0.8688$, $\bar{s}_s = 0.0321$, $\omega_m = 0.4729$, and $\bar{s}_m = 0.0167$. Although
Table 4: Calibration Parameters Structural Transformation & Services Deepening

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight Services Consumption Aggregator</td>
<td>$\omega = 0.4677$</td>
</tr>
<tr>
<td>Weight Services Inputs for Services</td>
<td>$\omega_s = 0.8688$</td>
</tr>
<tr>
<td>Weight Services Inputs for Manufacturing</td>
<td>$\omega_m = 0.4729$</td>
</tr>
<tr>
<td>In-House Production services Inputs in Services</td>
<td>$\bar{s}_s = 0.0321$</td>
</tr>
<tr>
<td>In-House Production services Inputs in Manufacturing</td>
<td>$\bar{s}_m = 0.0167$</td>
</tr>
<tr>
<td>Services Productivity in 2005</td>
<td>$A^s = 1.2675$</td>
</tr>
<tr>
<td>Services Productivity in 1947 = 1</td>
<td></td>
</tr>
<tr>
<td>Manufacturing Productivity in 2005</td>
<td>$A^m = 1.6941$</td>
</tr>
<tr>
<td>Manufacturing Productivity in 1947 = 1</td>
<td></td>
</tr>
</tbody>
</table>

the model is not calibrated to match the process of structural change, it succeeds in explaining around half of the rise in the services share of GDP (i.e., in the model the services share of GDP rises from 68.9% to 76.1%, while in the data it rises from 68.9% to 85.2%).

5.2 The Transmission of Monetary Policy

To understand the effects of sectoral reallocation on the transmission of monetary policy, we compare the response of inflation to monetary policy shocks, over the steady-states in 1947 and 2005. The two equilibria differ only in the levels of sectoral productivities. Throughout our exercise, we keep fixed all the other parameters, included those of the Taylor rule. In this way, we can ask to what extent structural change and services deepening alone can alter the transmission of monetary policy shocks.\textsuperscript{14} To disentangle the relative role of structural change and services deepening.

\textsuperscript{14}We consider the monetary policy of the 1947 steady-state as described by a Taylor rule to define a counterfactual economy, in which everything is similar to the 2005 steady-state but for both the level of sectoral
ening, we also compare the same steady-states in a counterfactual economy which abstracts from the process of services deepening. In this case, the sectoral shares of intermediate inputs are constant over time and calibrated to the average shares observed between 1947 and 2005.15

Table 5 reports the reaction of inflation, marginal costs, and output to an expansionary monetary policy shock - a drop of 100 basis points in the nominal interest rate - in 1947 and 2005 (respectively columns 1 and 2), the ratio of 2005-response over 1947-response (column 3), and the same ratio for the counterfactual economy which does not feature services deepening (column 4).

Table 5 shows that prices in services respond less than in manufacturing: the response of manufacturing inflation is thrice as large as the response of services inflation, both in 1947 and in 2005. Consequently, when moving from the 1947 steady-state to the 2005 one, the rise of services reduces the reaction of aggregate inflation by about 37%, from 1.03% to 0.64%. Instead, when we shut down the process of services deepening, by considering an economy with a constant Input-Output matrix, the reduction of the response of aggregate inflation equals just 26%. Hence, services deepening accounts for one third of the reduction in the response of aggregate inflation to a monetary policy shock.

The process of sectoral reallocation also reduces the contemporaneous response of sectoral inflation rates. Comparing the two steady-states, the reaction of services inflation shrinks by 7%, and the reaction of manufacturing inflation decreases by 10%. The drop in the reaction of sectoral prices is mirrored by the fall in the reactions of sectoral marginal costs, which decrease by about 7% in services and 11% in manufacturing. Without services deepening, the responses of sectoral inflation

---

15 The average share of services inputs used by the service sector from 1947 to 2005 equals 69.8%. For manufacturing, the analogous average share equals 27.1%.
<table>
<thead>
<tr>
<th></th>
<th>Benchmark Economy</th>
<th>Counterfactual Economy without Services Deepening</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_t )</td>
<td>1.03%</td>
<td>0.64%</td>
</tr>
<tr>
<td>( \pi_t^s )</td>
<td>0.56%</td>
<td>0.52%</td>
</tr>
<tr>
<td>( \pi_t^m )</td>
<td>1.65%</td>
<td>1.48%</td>
</tr>
<tr>
<td>( MC_t^s )</td>
<td>2.26%</td>
<td>2.11%</td>
</tr>
<tr>
<td>( MC_t^m )</td>
<td>1.16%</td>
<td>1.03%</td>
</tr>
<tr>
<td>( Y_t )</td>
<td>1.89%</td>
<td>2.09%</td>
</tr>
<tr>
<td>( Y_t^s )</td>
<td>1.43%</td>
<td>1.48%</td>
</tr>
<tr>
<td>( Y_t^m )</td>
<td>3.32%</td>
<td>3.59%</td>
</tr>
</tbody>
</table>

Note: The entries report the contemporaneous response (in log-deviations from the steady-state) of each variable to a drop of 100 basis points in the nominal interest rate. \( \pi_t \) denotes the aggregate inflation rate, \( \pi_t^s \) is the inflation rate of services, \( \pi_t^m \) is the inflation rate of manufacturing, \( MC_t^s \) is the real marginal cost of services, \( MC_t^m \) is the real marginal cost of manufacturing, \( Y_t \) denotes the aggregate output growth, \( Y_t^s \) is the output growth of services and \( Y_t^m \) is the output growth of manufacturing. “Benchmark Economy” refers to the economy with both structural transformation and services deepening. “Model 1947” refers to the equilibrium calibrated to the shares of services in intermediates observed in the U.S. in 1947. “Model 2005” refers to the equilibrium calibrated to the shares of services in intermediates observed in the U.S. in 2005. “Model 1947” and “Model 2005” differ just in the values of sectoral productivities. “Counterfactual Economy without Services Deepening” refers to the equilibrium in which the intermediates aggregators are Cobb-Douglas functions and there is no non-homothetic component, that is, \( \mu = 1 \) and \( \bar{s_s} = \bar{s_m} = 0 \). We calibrate the shares of services inputs in services and manufacturing to the average values observed between 1947 and 2005. The levels of TFP in the counterfactual economy are model-consistent.
barely change. These results support the quantitative relevance of the marginal cost channel of services deepening, through which the sectoral reallocation of the Input-Output matrix towards services intermediate inputs reduces the responsiveness of both sectoral marginal costs and sectoral inflation rates to monetary policy shocks.

When looking at the response of output, Table 5 displays that monetary policy shocks nowadays have slightly larger real effects. Over the two steady-states, the reaction of aggregate output rises by 10%, while the reaction of services output and manufacturing output increases by 3% and 8%, respectively. Abstracting from the services deepening halves the increase in the real effects of monetary policy.

Why do monetary policy shocks have larger real effects? Sectoral reallocation shifts economic activities towards services, whose prices are much stickier than manufacturing prices. Hence, the rise of services raises the average duration of prices in the economy. As prices get stickier, inflation reacts less to shocks, and therefore quantities have to react by a larger amount. This mechanism explains why monetary policy shock have larger effects on output, at the cost of a lower influence on the dynamics of inflation.\footnote{The asymmetry in the degree of price stickiness across sectors is crucial for our results. Without it, a shift from manufacturing to services would actually increase the response of inflation to a monetary policy shock. Since the share of intermediate inputs in the gross output of the manufacturing sector is as twice as large than in the service sector, the rise of services would raise the flexibility of prices, by reducing the mechanism of strategic complementarities emphasized by Nakamura and Steinsson (2010). The Online Appendix shows the effects of the sectoral reallocation in an economy in which price stickiness is common across sectors.}

Are the results of our model in line with our empirical evidence on the relationship between the response of inflation to monetary policy shocks and the sectoral composition of intermediate inputs? To validate the quantitative implications of our theory, we compare the relationship between the contemporaneous response of aggregate inflation and the services share of intermediate inputs across countries in the data and in the model. On the one hand, we take the estimates of the SVAR exercise of Section 2.2. On the other hand, we compute the response of inflation to a monetary policy shock in the model in a series of steady-states with different lev-
els of sectoral productivities and therefore different services shares of intermediate inputs.

Figure 7 compares the response of inflation to a monetary policy shock in the data and in the model, for given shares of services intermediate inputs. The quantitative results of the model are successfully in line with the magnitude of the estimates from the data on the relationship between the sectoral composition of intermediate inputs and the response of inflation to monetary policy shocks. The model can account for the low responsiveness of inflation in a country highly intensive in services intermediate inputs, like Luxembourg, while being able at the same time to explain the high sensitivity of inflation in a country very intensive in manufacturing intermediate inputs, like Korea. The consistency of the model with our empirical estimates gives further support and validation to the quantitative appeal of the marginal cost channel of services deepening.

Figure 7: Intermediate Inputs and the Transmission of Monetary Policy: Data vs. Model.

The figure shows the relationship between the contemporaneous response of aggregate inflation and the services share of intermediate inputs across countries in the data (the country identifiers) and in the model (the continuous line).
6 Conclusion

In this paper we document that advanced economies are experiencing a process of services deepening that raises the services share of intermediate inputs. In addition, we show that inflation becomes less responsive to monetary policy shocks in countries which are more intensive in services inputs.

To rationalize these facts, we build a New Keynesian model with two sectors connected by an Input-Output matrix, which changes endogenously over time. We use the model to understand the effects of the sectoral reallocation experienced by the U.S. from 1947 to 2005. We find that over these six decades the reaction of aggregate inflation to monetary policy shocks decreases by 37%, whereas the response of aggregate output rises by 10%. The reduction in the responsiveness of inflation is driven by services deepening, through a marginal cost channel that affects the behavior of firms in either sector. As services and manufacturing increase the intensity of services inputs, marginal costs becomes stickier. Thus, even sectoral prices become stickier and less reactive to monetary policy shocks.

Overall our results show that the process of services deepening alters the response of inflation and output to monetary policy shocks. A recent strand of the literature has emphasized a reduction in the responsiveness of inflation to monetary policy shocks over the recent decades. The conventional view rationalizes this fact with changes in either the stance of monetary policy or the volatility of shocks. We provide a novel channel that can generate low-frequency movements in the effectiveness of monetary policy. Since in our model both the Taylor rule parameters and the volatility of shocks are constant over time, the variation in the transmission of monetary policy is entirely due to the process of sectoral reallocation.
References


