Monetary Policy and Regional Inequality*

Oliver de Groot† Sebastian Hauptmeier‡
Fédéric Holm-Hadulla§ Katerina Nikalexi¶

Current version: November 8, 2024

Abstract

We study the impact of monetary policy on regional inequality using granular data on economic activity at the city- and county-level in Europe. We document pronounced heterogeneity in the regional patterns of monetary policy transmission. The output response to monetary policy shocks is stronger and more persistent in poorer regions, with the difference becoming particularly pronounced in the tails of the distribution. Regions in the lower parts of the distribution exhibit hysteresis, consisting of long-lived adjustments in employment and labor productivity in response to the shocks. As a consequence, policy tightening aggravates regional inequality and policy easing mitigates it. Finally we provide a structural interpretation of our results using a New Keynesian Currency Union Model with hysteresis effects.

Keywords: Monetary Policy; Regional Heterogeneity; Local Projections; Endogenous Technological Change; New Keynesian Model
JEL Classification: C32, E32, E52

---

*The analysis presented in this paper has benefited from discussions with Carlo Altavilla, Fabio Canova, Jacopo Cimadomo, Maarten Dossche, Nicola Fuchs-Schündeln, Christophe Kamps, Elias Papaioannou, Giorgio Primiceri, Morten O. Ravn, Hélène Rey, Joao M.C. Santos Silva, Frank Smets, Paolo Surico, Thomas Vlassopoulos and conference participants at the 2021 ASSA Annual Meeting, at the 2020 EEA Congress, London Business School and at the ECB. This study did not receive any funding. The views expressed here are those of the authors and do not necessarily reflect those of the European Central Bank. An earlier version of this study is available as ECB Working Paper No. 2385.

†University of Liverpool Management School & CEPR. Email: oliverdegroot@gmail.com.
Chatham Street, Liverpool, L69 7ZH, United Kingdom.

‡European Central Bank. Email: Sebastian.Hauptmeier@ecb.europa.eu.
Sonnemannstraße 20, 60314 Frankfurt am Main.

§European Central Bank. Email: Federic.Holm-Hadulla@ecb.europa.eu.
Sonnemannstraße 20, 60314 Frankfurt am Main.

¶London Business School. Email: KNikalexi@london.edu. *Corresponding author.*
Regent's Park, London, NW1 4SA, United Kingdom.

1. Introduction

There is an increasingly active debate on whether monetary policy tends to dampen or reinforce economic inequality. Much of this debate has centred on the impact of monetary policy on income, consumption and wealth inequality at the household level (Coibion et al., 2017; Kaplan et al., 2018; Auclert, 2019; Bielecki et al., 2021). An aspect that instead has received little attention so far is that economic inequality typically has a pronounced geographical dimension. In particular, some subnational regions generate significantly and persistently higher per capita incomes than others (Jacobs, 1969; Krugman, 1998); and, in many countries, this aspect of inequality has either intensified over recent decades or halted its previous declining trend.[1]

A natural question is whether these trends have been influenced by the macroeconomic policy mix, within which monetary policy has played a particularly dominant role since the Global Financial Crisis in most economies. However, empirical evidence on the link between monetary policy and regional inequality is sparse and, *a priori*, it is ambiguous whether monetary policy shocks would reduce or aggravate regional inequality.[2]

Against this background, the current paper examines this question based on granular data

[1] See Ganong and Shoag (2017) and Austin, Glaeser and Summers (2018), Iammarino, Rodriguez-Pose and Storper (2018), and Martin et al. (2018) for recent evidence on the United States, the euro area, and the UK, respectively.

[2] Some papers find that economic agents with weak balance sheets are more sensitive to monetary policy. This includes evidence on households (see citations above), firms (Kashyap et al., 1994; Gertler and Gilchrist, 1994), and banks (Peek and Rosengren, 1995; Kashyap and Stein, 2000; Altavilla et al., 2020). To the extent that more vulnerable economic agents cluster geographically, these mechanisms may translate into a stronger responses to monetary policy in less prosperous regions. But, other papers highlight channels that induce lower responsiveness in the weaker parts of the respective distributions for firms (Ottonello and Winberry, 2020) and banks (Boeckx et al., 2017). Further, Dolado et al. (2021) find that monetary policy easing raises labour income inequality between high-skilled and low-skilled workers.

on economic activity at the city- and county-level in the euro area and proposes a theoretical mechanism to rationalize our empirical findings.

The use of city- and county-level data, as the most granular breakdown for which official output statistics are available, has two key advantages for this type of analysis. First, economic disparities are particularly pronounced at this disaggregated level. For instance, the variation of per capita GDP across cities and counties is almost 1.5 times higher than across euro area countries (see Section 3 for details). Accordingly, the disaggregated data provide for a much richer information set to be exploited in analysing the regional transmission of monetary policy. Moreover, it allows us to assess whether monetary policy tends to mitigate and accentuate interregional income disparities – an important question in view of the increased scrutiny that the impact of monetary policy on economic inequality has received in public discourse and the increased recognition that rising inequality in many advanced economies exhibits a pronounced interregional dimension (Coeuré, 2018). At the same time, unlike studies zooming in on specific types of agents (such as households or firms), the regional perspective allows us to still capture the broader interactions within a local economy.

The second advantage is that the regional disaggregation of economic data offers a promising avenue to expand the set of strategies available to identify exogenous variation in monetary policy interest rates. As central banks are equipped with explicit macroeconomic stabilization mandates, monetary policy is by construction endogenous to the state and prospects of the economy. Our identification strategy makes use of the fact that the mandate of the ECB refers to the euro area aggregate level, while the ECB's Governing Council has, since its onset, emphasized that: "*[its] single monetary policy will adopt a euro area-wide perspective; it will not react to specific regional or national developments*" (ECB Governing Council Press Release, 13 October 1998). This euro-area wide focus of monetary policy implies that, controlling for

aggregate conditions, variation in short-term interest rates is likely to be exogenous to GDP at the city- and county-level. Hence, by exploiting economic information at a more granular level than that entering central bank reaction functions, we are able to address the risk of simultaneity bias in estimating the real effects of monetary policy; and our main conclusions remain intact when relaxing the assumption that monetary policy conduct is independent of regional economic disparities, as suggested by Coibion and Goldstein (2012).[3]

To estimate the dynamic impact of exogenous changes in monetary policy across the regional GDP distribution, we combine Jordà (2005)'s local projections method with a subsample analysis, as used by Crouzet and Mehrotra (2020) in the context of firm size classes. The data are based on Eurostat's Nomenclature on Territorial Units for Statistics (NUTS) and broken down to the most disaggregated geographical level at which information on economic activity is available (NUTS3).[4]

Our estimates document a significant and relevant impact of changes in short-term interest rates on regional GDP that strongly intensifies towards the lower end of the distribution. At the sample mean, the response closely resembles the patterns typically found in the literature. After an initial transmission lag, the interest rate coefficient turns significant at a one-year horizon. The impact then further builds up and peaks at the two-year horizon, with the estimates imply-

[3]Similar identification strategies based on sub-national data have been put to productive use in estimating fiscal multipliers (see e.g. Auerbach and Gorodnichenko (2012), Clemens and Miran (2012), Nakamura and Steinsson (2014), Corbi, Papaioannou and Surico (2019), Guren et al. (2020), and Chodorow-Reich (2019a) for a review). But they have received less attention in the monetary economics literature so far.

[4]The subnational administrative setup differs substantially across euro area countries and not all countries have a local layer of government equivalent to what is referred to as a "county" for instance in the United States or in Germany. As we discuss in greater detail in Section 3, the NUTS classification tackles this issue by creating a system of economically coherent regions following harmonized standards across countries. For ease of exposition, we refer to NUTS3 regions as the "city- and county-level" throughout this paper but note that this wording is approximate and needs to be interpreted in conjunction with the more precise definitions presented in Section 3.

ing a decline in average regional GDP of around 2% in response to a 100 basis point exogenous interest rate increase, before fading out over the remainder of the projection horizon. Moving beyond the sample mean, the estimations point to pronounced heterogeneity in the regional patterns of monetary policy transmission. In particular, output responds more strongly in regions with low versus high per capita GDP and this difference becomes particularly pronounced in the tails of the distribution. Further, the same qualitative patterns emerge when, instead of total GDP, we focus on the capital- and labor-intensive sub-sectors of the regional economies while, among these sub-sectors, capital-intensive production is more responsive, in line with standard notions of the interest rate channel of monetary policy.

The most striking feature of our results is that substantial parts of the regional distribution experience long-lived effects of monetary policy shocks on output. In particular, while GDP in the upper part of the distribution returns to its pre-shock level after around three, the response in the lower parts does not reverse over this period, thus pointing to pronounced hysteresis in output. The heterogeneous incidence of hysteresis in turn implies that monetary policy has a long-lasting impact on regional inequality, with tightening shocks aggravating and easing shocks mitigating it.

In terms of anatomy, we find hysteresis to originate from long-lived adjustments in both employment and labor productivity. At the same time, employment hysteresis is more pronounced and more broad-based. In fact, it even extends to the sample mean, as opposed to productivity hysteresis which concentrates in the lower tails of the distribution. As such, our findings confirm labor markets as an important source of hysteresis in the European context. And they add to a growing literature arguing that monetary policy may exert durable impacts on output and employment – in contrast to the common notion of stabilization policies merely smoothing out fluctuations in these variables around some natural levels (Blanchard, 2018; Dupraz, Naka-

mura and Steinsson, 2019; Jordà, Singh and Taylor, 2020; Ravn and Sterk, 2020). Consistent with recent literature, we find hysteresis-prone regions to be characterized by weaker labour markets, a higher share of employment in small firms, lower business dynamism, and a weaker capacity for innovation.

Our findings add a novel perspective to the literature on the geographical incidence of monetary policy. Corsetti et al. (2021) show that the ECB's monetary policy transmits unevenly across euro area economies, whereas Carlino and DeFina (1998) document heterogeneity in monetary policy transmission across (clusters of) US states.[5] At the same time, and notwithstanding the aforementioned benefits, only few papers have studied monetary policy transmission at a richer *regional* disaggregation level. Important exceptions include Francis, Owyang and Sekhposyan (2012) who estimate city-level responses to monetary policy shocks in the US, as well as Fratantoni and Schuh (2003), Di Maggio et al. (2017), Beraja et al. (2018), and Eichenbaum, Rebelo and Wong (2018) who examine the transmission of monetary policy via mortgage markets at similar aggregation levels as the current paper. Our paper is the first to model the impact of monetary policy across the regional GDP distribution of a major advanced economy at the city- and county level; to explicitly tackle the question of how the response differs in richer and poorer places; and thus to shed light on how monetary policy affects regional inequality.

Our final contribution is to provide a structural interpretation of our empirical findings. We start from a stylized classical model with a hysteresis channel in which unemployment has endogenous effects on total factor productivity. In this model, we derive analytical conditions under which low-income regions suffer more persistent fluctuations than high-income regions

[5]For further literature on the US, see *e.g.* Di Giacinto (2003); Owyang and Wall (2009); Beckworth (2010); Furceri et al. (2019); Leahy and Thapar (2019). For literature studying similar geographical units in Europe, see *e.g.* Arnold (2001); Arnold and Vrugt (2002); Rodriguez-Fuentes and Dow (2003); Dow and Montagnoli (2007).

in response to shocks. Building on Hazell et al. (2022) and Gali and Monacelli (2008), we then embed this mechanism in a New Keynesian Currency Union model with heterogeneity across regions and analyze numerically the response to monetary policy shocks. Our model is able to qualitatively replicate our empirical findings. Moreover, it allows us to study: *(i)* how inequality affects the regional inflation response to monetary policy; *(ii)* how it affects the aggregate effects of monetary policy; and *(iii)* how aggregate outcomes would change if regional inequality were to feature in the monetary policy rule.

The remainder of the paper proceeds as follows. The next section presents the empirical model and identification strategy. Section 3 describes the data and highlights some salient stylized facts regarding economic activity at the regional level. Section 4 presents our baseline results, complemented by a range of robustness checks in Appendix B. Section 5 provides a structural interpretation of our empirical results. Section 6 concludes.

2. Empirical setup

To study the dynamic effects of monetary policy on regional output, we apply Jordà (2005)'s local projections method consisting in a set of regressions of the form:

$$y_{i,t+h} = \alpha_i + \beta_h i_t + \gamma_{\mathbf{h}} \mathbf{X}_{\mathbf{i},\mathbf{t}} + \delta_{\mathbf{h}} \mathbf{X}_{\mathbf{j},\mathbf{t}} + \theta_{\mathbf{h}} \mathbf{X}_{\mathbf{k},\mathbf{t}} + \varepsilon_{i,t+h} \tag{1}$$

where the dependent variable $y_{i,t+h}$ denotes real GDP in jurisdiction i and year $t+h$; α_i is a set of region-fixed effects; i_t is the monetary policy-controlled short-term interest rate in year t; $\mathbf{X}_{\mathbf{i},\mathbf{t}}$, $\mathbf{X}_{\mathbf{j},\mathbf{t}}$, and $\mathbf{X}_{\mathbf{k},\mathbf{t}}$ are vectors of time-variant control variables at the local-, country- and euro area-level, respectively; and $\varepsilon_{i,t+h}$ is an error term. All variables enter equation 1 in log-levels, except for the short-term interest rate which enters as a percentage per annum. In the most parsimonious version of the model, the local control variables include the population of each

region; the country controls include GDP of the country j in which region i is located; and the euro area controls include GDP and HICP at the euro area level.

Our main interest is in the impulse response of local output to a change in the short-term interest rate in year t, as captured by the coefficient β_h for each horizon h. The interpretation of the β_h-coefficients as the causal effect of short-term interest rate changes on local output at horizon h relies on the identifying assumption that, controlling for macroeconomic conditions, monetary policy does not respond to economic activity at the city- and county-level.

This assumption is supported by the following considerations. As highlighted in Section 1, the ECB's monetary policy mandate, with its primary objective to maintain price stability, refers to the euro area as a whole. In keeping with this euro-area wide mandate, the ECB's Governing Council has formally defined its price stability objective in terms of the aggregate euro area inflation rate. Vice versa, it has explicitly ruled out regional developments as a determinant of its monetary policy conduct (ECB, 1998). Moreover, given the dearth and long lags in the availability of regionally disaggregated data it is not in a position to factor in that information at a policy-making frequency.[6] Accordingly, by controlling for euro area aggregate inflation and activity in the ECB's reaction function, we are able to partial out the variation in short-term rates that does not reflect the systematic central bank response to the economy. Based on this remaining variation, which effectively amounts to using Taylor-rule residuals to identify exogenous changes in policy-controlled short-term interest rates, we can then estimate the impact of monetary policy shocks on regional GDP. As such, the granular data on economic activity helps us overcome the risk of simultaneity bias, which constitutes one of the key challenges in

[6] In particular, information on economic activity at the NUTS3 level usually becomes available only around two to three years after the period they refer to. This very long lag implies that, at the time policy rates are being set, even the raw data that would allow decision makers to consider regional economic developments are missing.

estimating monetary policy effects on the economy.[7]

The remaining explanatory variables serve to further sharpen this identification approach. The inclusion of country-level GDP further severs the link between the monetary policy reaction function and regional GDP as the dependent variable. The population variable accounts for the large cross-sectional heterogeneity in the size of the regions.[8] And the region-fixed effects allow us to control for a host of other, unobserved, factors that may confound the causal interpretation of our estimates. Moreover, we subject our baseline specification to a range of robustness tests (Appendix B). In the spirit of Romer and Romer (2004), these include the use of real-time ECB forecasts as further macro controls. Further, following Coibion and Goldstein (2012), we present robustness checks in which we relax the identifying assumption that monetary policy conduct is independent of regional economic disparities. Finally, we estimate the model on a significantly longer sample period and with a broader set of macro covariates that may enter central bank rate decisions, while affecting regional economies in a differential manner.

To test for heterogeneous effects of monetary policy, we group the regions according to their position in the per-capita GDP distribution and allow the coefficients in 1 to differ across groups. Specifically, we define a dummy variable D_{it}^{d} for each decile d that is 1 for all regions i whose per capita GDP falls within this decile in year t and zero otherwise. We then interact

[7] The assumption that monetary policy responds to aggregate but not to regional shocks also features prominently in the recent model-based analysis of the aggregate implications of regional business cycles by Beraja et al. (2019), as well as in the conceptual discussion of the increasingly widespread use of regional data in macroeconomics by Chodorow-Reich (2019b).

[8] An alternative would be to directly specify the dependent variables in per-capita terms. Compared to this approach, our choice of specification is somewhat more flexible as it does not restrict the coefficient on the population variable to be equal to one. At the same time, when experimenting with log per-capita GDP as dependent variable, we obtained very similar results.

this dummy with all explanatory variables in equation 1, to obtain decile-specific coefficient estimates (for a similar approach in the context of firm size distributions, see Crouzet and Mehrotra (2020)). As part of the robustness checks, we also deploy the novel quantile fixed effects estimator proposed by Machado and Silva (2019) as an alternative method to test for heterogeneity in monetary policy transmission across the regional GDP distribution.

Inference is based on Driscoll and Kraay (1998) standard errors that account for cross-sectional and temporal dependencies in the data. In terms of lag-length, we follow the heuristic from the first step of the Newey and West (1994) plug-in procedure, which in our data set implies two lags (see Hoechle (2007) for details). When experimenting with higher lag-lengths, however, we found the estimated standard errors to remain largely unaffected.

3. Data and stylized facts

3.1. Sources and definitions

Our analysis relies on a rich dataset of economic and demographic indicators at the sub-national level in Europe. The data are based on Eurostat's Nomenclature of Territorial Units for Statistics (NUTS), which is a hierarchical system for dividing up the economic territory of the EU into four levels. The highest level (NUTS0) corresponds to the nation-state and the lowest (NUTS3), which we use in the ensuing analysis, roughly corresponds to the city- and county-level.

The source of this data is the Annual Regional Database of the European Commission's Directorate General for Regional and Urban Policy (ARDECO), which is maintained and updated by the Joint Research Centre of the European Commission, and is designed to be the regional counterpart to the AMECO database. The NUTS3 data offer the maximum degree of geographical disaggregation for which information on economic activity is available. At the

same time, the coverage in terms of economic variables is relatively limited at the NUTS3 level compared to more aggregated data sets.[9]

Our main regional variables are gross domestic product (GDP) and gross value added (GVA), both deflated to 2015 prices, as well as population and employment. Further, we make use of the breakdown of regional GVA into six sectors of the economy, corresponding to the disaggregation in NACE Rev.2 of: agriculture, forestry and fishing; industry less construction; construction; financial and business services; wholesale, retail, transport, accommodation and food services, information and communication; and non-market services. We merge the sectors of industry and construction to capture the capital-intensive part of regional GVA and the sectors of financial and business services, wholesale, retail, transport, accommodation and food services, information and communication to capture the labor-intensive part.

We complement this regionally disaggregated information with euro-area and country-level variables for real GDP and HICP from the AMECO database of the European Commission. Following the bulk of the related literature, the 3-month Euribor serves as our measure of the policy-controlled short-term interest rate (see, *e.g.*, Coenen et al. (2018), Smets and Wouters (2003), Faust et al. (2003)). We source this variable from the Area Wide Model (AWM) database. Further, we also rely on the AWM database for information on oil prices (expressed in US Dollar per barrel) and on country-specific long-term sovereign interest rates. The nominal effective exchange rate, considering 67 trading partners, is taken from Bruegel.

The baseline sample includes all NUTS3 regions from the eleven initial euro area member states, excluding Luxembourg, over the period 1999-2014 and from Greece over the period 2001-2014. The starting point of the sample corresponds to the year the euro currency was

[9]For instance, data on inflation and unemployment are available only at higher regional aggregation levels; see *e.g.*, Beck, Hubrich and Marcellino (2009), Özyurt and Dees (2015), and Belke, Haskamp and Setzer (2016) for recent studies using these variables.

introduced and the end-point to the last year in which the ECB steered its monetary policy stance primarily via short-term interest rates (see Rostagno et al. (2019) and Appendix B.2).[10] However, in some parts of our analysis, we also draw on longer sample periods, starting prior to the introduction of the euro and extending beyond the introduction of quantitative easing as an alternative tool for the ECB to steer the monetary policy stance (see Appendix B.2 for further detail). Our final baseline sample consists of a panel of 884 NUTS3 regions over 16 years.

3.2. Heterogeneity across space and time

A comparison of per capita GDP distributions at the regional and national level demonstrates how the degree of heterogeneity intensifies with greater disaggregation (see Table 1).[11] For instance, in 2014, per capita GDP at the country level ranged from €16,974 in Portugal to €45,062 in Ireland.[12] While substantial, this difference is dwarfed by the dispersion in regional per capita GDP, which ranged from €9,526 for Serres, a region in Northern Greece, to €138,229 in Wolfsburg, a city in the German state of Lower Saxony.

Moreover, this pattern of greater heterogeneity at more granular aggregation levels is not just confined to a few extreme outliers, but it is a general feature of the respective distributions. For instance, the coefficient of variation (CV), computed as the ratio of the standard deviation over the mean of per capita GDP, is almost 1.5 times higher at the regional than at the national

[10] Luxembourg is excluded because it consists of just one NUTS3 region. The sample for Greece starts in 2001 because this is when the country introduced the euro. We also follow the literature in excluding the five NUTS3 regions of the French overseas territories of Guadeloupe, Martinique, French Guiana, La Réunion and Mayotte, along with the Portuguese autonomous regions of the Azores and Madeira (Becker et al., 2010).

[11] Unless otherwise noted, per capita GDP refers to the ratio of real GDP over population throughout the remainder of the paper.

[12] As is well-known, the Irish GDP figures tend to be distorted upwards by the activities of multinational companies. But the difference between the national and regional distribution remains similarly striking when considering the Netherlands, which recorded the second highest per capita GDP in 2014 (of €40,124), as the upper bound for the national distribution.

Table 1: Heterogeneity at different aggregation levels

		Mean	**St. Dev.**	**Min**	**Max**	10^{th}	**Median**	90^{th}
GDP	*Country*	32035	9875	16219	45062	16974	36432	40124
	Region	29896	13442	9526	138229	16893	27800	43036
CI-GVA	*Country*	7093	3155	2366	12676	3295	7108	10144
	Region	7881	6345	458	90859	2609	6873	13030
LI-GVA	*Country*	14665	4442	7821	21326	7970	15440	19389
	Region	12115	6814	3005	69436	6618	10729	18781
Observations	*Country*	11						
	Region	884						

Notes: Figures refer to real GDP and real GVA per capita in 2014 at the NUTS3 (region) or NUTS0 (country) level. Capital-intensive GVA (CI-GVA) is calculated as the ratio of GVA in the sectors of industry and construction over total population. Labor-intensive GVA (LI-GVA) is calculated as the ratio of GVA in the sectors of financial and business services, wholesale, retail, transport, accommodation and food services, information and communication over total population.

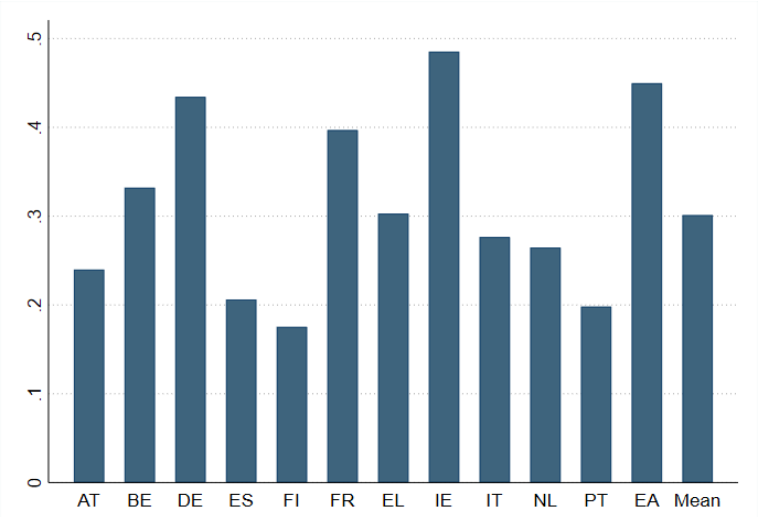
level (see Table 1).

Further, an even more pronounced degree of dispersion emerges for gross value added (GVA), and in particular for the capital-intensive part of GVA, whose CV at the regional level is twice as high as that at the national level. Thus, the dispersion not only in overall activity but also in its composition become more accentuated at the disaggregated level, consistent with the typical patterns of spatial concentration of different types of productive activity (Krugman, 1991; Glaeser et al., 1995).

Importantly, these disparities also arise within the countries. In fact, within-country dispersion in per capita GDP in some cases reaches similar levels as for the euro area as a whole; and the average within-country dispersion in 2014 was more than two thirds of that observed for the euro area (Figure 1). Accordingly, the analysis of regional disparities in the euro area is not just another way of looking at cross-country differences and putting them under a magnifying glass; instead, it addresses an important aspect of the unequal geography of economic activity that transcends the well-studied cross-country perspective.

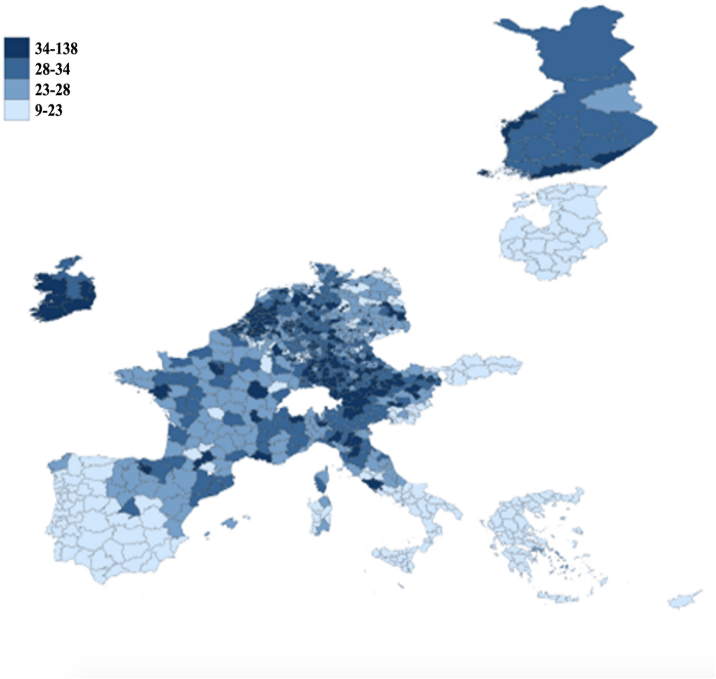
Moreover, notwithstanding some instances of geographical clustering of more and less pros-

Figure 1: Coefficient of variation of regional GDP per capita



Notes: The coefficient of variation (CV) is computed as the ratio of the standard deviation to the mean of all NUTS3 regions within each country in 2014, except for: the bar denoted EA, which refers to the CV over all NUTS3 regions in the sample; and the bar denoted Mean, which refers to the unweighted average of the eleven within-country coefficients of variation displayed in the graph.

Figure 2: Heterogeneity in GDP per capita



Notes: Figures refer to real GDP per capita in 2014 at the NUTS3 level. Different shadings refer to quartiles.

perous regions, the patterns of within-country inequality are fairly nuanced. For instance, in Italy poorer regions tend to cluster in the South, in Germany in the East, and in Spain in the West (Figure 2). As such, some form of core-periphery divide, which is often emphasized in the context of different countries in the euro area, also exists at the regional level within some

of the countries. However, there are also instances of strong disparities arising between regions with a close geographical proximity, as visible from regions in the lower quartiles punctuating clusters of regions in the upper quartiles, and vice versa. Overall, the patterns of heterogeneity within countries are thus fairly diverse, which implies that a less granular subnational disaggregation of the data, such as by provinces or states, would miss important aspects of regional inequality.

Beyond this static perspective on regional inequality, also its dynamics offer interesting observations, especially in the aftermath of the global financial and economic crisis. This crisis triggered a steep fall in euro area economic activity, followed by a double-dip recession, both of which is clearly visible at the mean of the regional per capita GDP distribution (Figure 3). Similar patterns emerge for the median, whereas the dynamics for other parts of the distribution show a nuanced pattern. While the 25^{th} and 75^{th} percentiles have mostly moved in lockstep, and regions in the 25^{th} percentile even recovered more swiftly than those in the 75^{th} percentile, vastly different trajectories emerge for the outer parts of the distribution. For instance, regions in the 90^{th} percentile on average experienced a solid recovery after the 2009 recession, whereas per capita GDP in the 10^{th} percentile just continued drifting down after the crisis and only showed a mild turn-around in 2014. And this drifting-apart of poorer and richer regions even intensifies when considering the wedge between the 5^{th} and 95^{th} percentile.

Taken together, these stylized facts confirm that the city- and county-level data used in the current paper provide a very interesting setting to study heterogeneity in monetary policy transmission. Of course, the basic fact that disparities in economic structures and performance become particularly accentuated at this level is well-established in the urban economics literature (Glaeser, Scheinkman and Shleifer, 1995). But it has so far received little attention in the monetary economics literature. The next section closes this gap by presenting evidence on the

Figure 3: Evolution of average per capita GDP in selected percentiles



Notes: The lines show the normalized percentiles of regional GDP per capita in the total sample of EA11. The percentiles have been normalized to 100 in 2008.

effects of monetary policy on regional inequality.

4. The heterogeneous impact of monetary policy on regional output

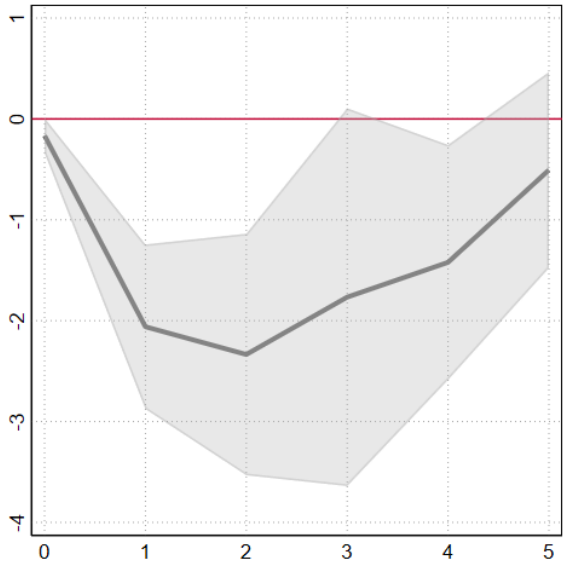
The impulse response functions (IRFs) point to a significant and economically relevant impact of exogenous changes in short-term interest rates on regional GDP and this impact becomes stronger and more persistent when moving towards the lower parts of the distribution. We next describe these findings in greater detail, while providing a range of robustness checks in Appendix B.

4.1. Monetary policy response at the mean

As a starting point, and to benchmark our estimates against the related literature, we first focus on the response at the sample mean. The following patterns emerge for a 100 basis point hike in short-term interest rates (Figure 4). After an initial transmission lag, the downward impact on economic activity turns significant at a one-year horizon. This impact further builds up and peaks at the two-year horizon, with the IRF implying a contraction of around 2% in regional GDP in year $t+2$. The impact then gradually fades out over the remainder of the IRF horizon, with output returning to its initial level.

Overall, the estimates closely resemble the familiar patterns for the output response to monetary policy shocks, as documented in the literature for higher levels of aggregation. For instance, the 2% peak effect on output in response to an exogenous 100 basis point short-term interest rate hike is close to the median reported in prominent papers on the US economy, as reviewed in Ramey (2016) (see Table 1 in that paper). It is also broadly similar to that deriving from the estimated DSGE model of Smets and Wouters (2003), which often serves as a benchmark for the literature on the euro area economy. Moreover, the two-year transmission lag between the policy rate hike and its peak impact, as well as the gradual fading out of the output response in subsequent years is comparable to the patterns in Smets and Wouters (2003). This broad match between the estimated responses is reassuring as it allays potential concerns

Figure 4: Impact of monetary policy on regional output at the mean



Notes: Vertical axis refers to impact of 100 basis point rate hike on regional GDP (in %). Horizontal axis refers to horizon of IRF (in years). Solid lines denote point estimates and shaded areas denote 90% confidence bands.

of time aggregation bias arising from the annual frequency of our data (Marcet, 1991; Hansen and Sargent, 1991).

4.2. Heterogeneity across regions

The sub-sample analysis points to pronounced heterogeneity in the regional patterns of monetary policy transmission: while richer and poorer regions both experience a contraction in output in response to a monetary policy tightening, the strength and persistence of the impact are significantly higher for the latter group. For instance, regions in the upper quartile of the per-capita GDP distribution experience an output contraction of around 2% over the two years following the policy rate increase, followed by a return to the initial level by the end of the horizon (Figure 5, grey line). By contrast, for regions in the lower quartile, the output contraction intensifies over the subsequent two years, to a trough of almost 3% below the initial level, and the recovery remains incomplete by the end of the horizon (blue line).

Figure 5: Impact of monetary policy on regional output: upper vs. lower quartile

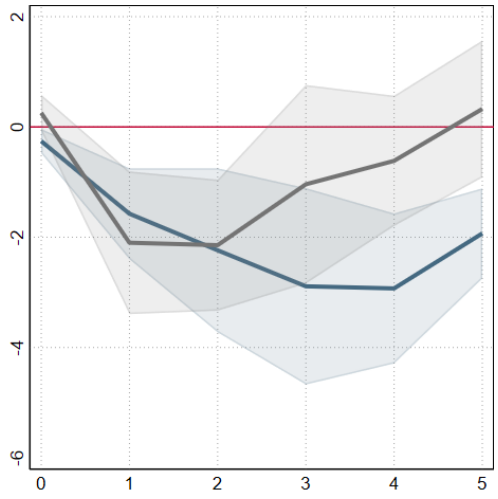
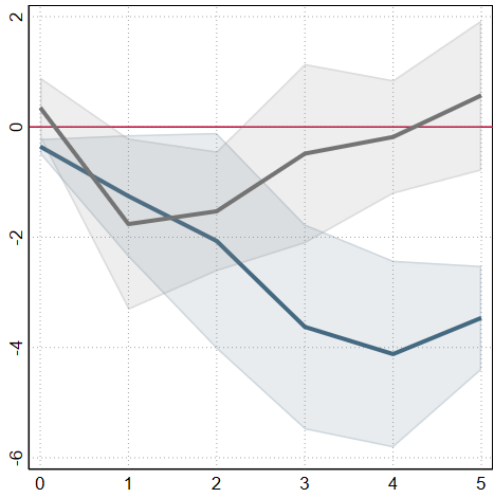


Figure 6: Impact of monetary policy on regional output: bottom versus top decile



Notes: Vertical axis refers to impact of 100 basis point rate hike on regional GDP (in %). Horizontal axis refers to horizon of IRF (in years). Solid lines denote point estimates and shaded areas denote 90% confidence bands. Grey (blue) lines depict the estimates for the higher (lower) quantile-range.

Moreover, these differences in the output response become even more accentuated when considering the outer parts of the distribution. For example, for the lower decile, output reaches a trough of below -4.0%, compared to slightly above -2.0% for the upper decile (Figure 6). Also, the contrast between the quick recovery in the upper part of the distribution and the persistent contractionary effect in the lower part becomes starker for these percentile pairs: while output recovers to its pre-shock levels in $t+3$ in the upper decile, it drifts down further in the lower decile and remains close to the trough by the end of the horizon. Finally, while the confidence intervals in the interquartile comparison overlap over large parts of the horizon, the difference in the outer tails is clearly statistically significant over the later years of the horizon. In the subsequent analysis, we further disentangle these patterns, before providing a theoretical explanation.

4.3. The role of industry structures

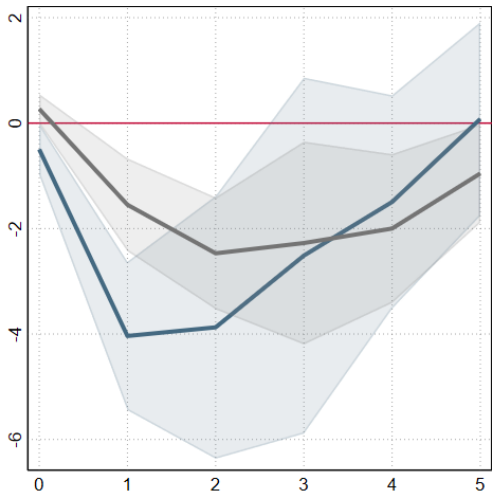
Subnational jurisdictions typically differ markedly in their industry mix and a comprehensive literature has documented that the responsiveness of output to monetary policy shocks

tends to vary across different types of industry (see, e.g., Dedola and Lippi (2005); Peersman and Smets (2005); and Carlino and DeFina (1998) for representative studies). Our data and research design allow us to explore whether these insights are confirmed in our more granular setting and whether regional heterogeneity in industry structures reinforces or mitigates the differential responsiveness to monetary policy shocks.

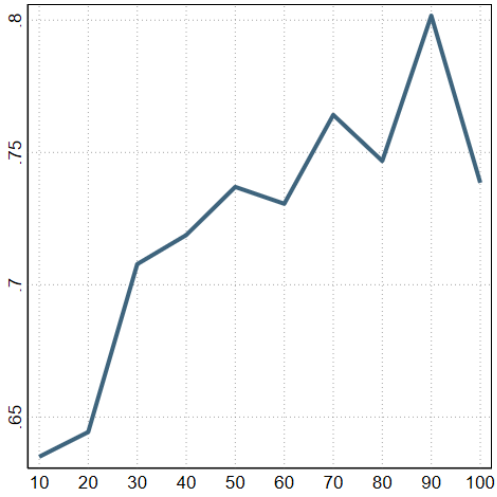
To answer these questions, we again resort to the subdivision of GVA into its capital-intensive and labor-intensive sub-sectors, with the former consisting of construction and industry and the latter of services (see Section 3.2). Based on this breakdown, we conduct two exercises. First, we estimate industry-specific output responses to monetary policy shocks for the sample mean so as to cross-check our priors on the relative interest-rate sensitivity across industry-types. Second, we repeat this exercise for the cross-regional GDP distribution via sub-sample analysis so as to test whether differential responses also arise *for a given type of industry*.

Figure 7: Capital-intensive vs. labor-intensive sectors

Impact of monetary policy on regional output across industries



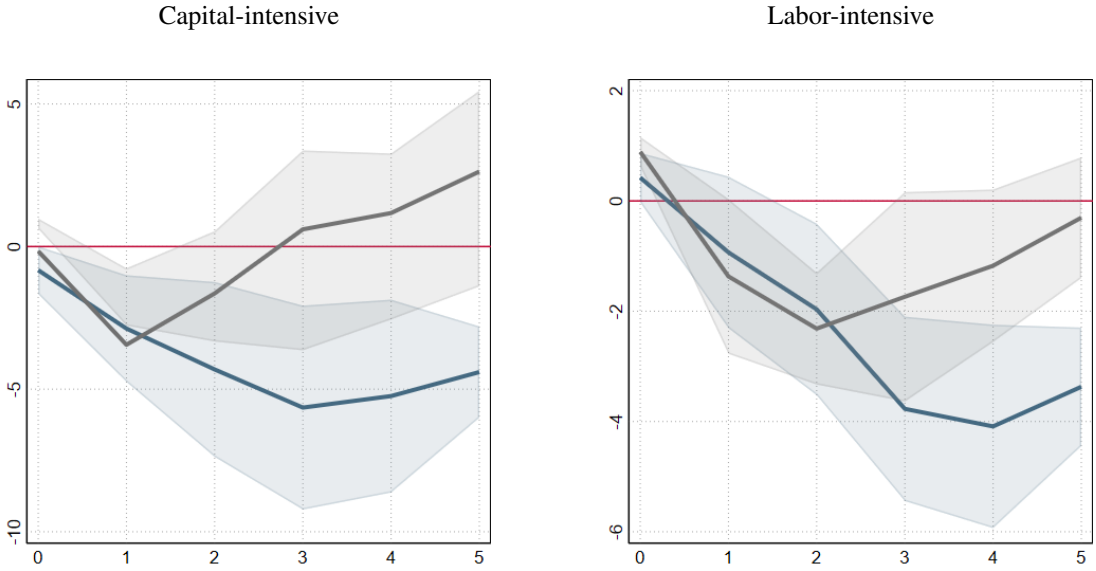
Industry structure across quantiles



Notes: In the left-hand side (LHS) panel, vertical axis refers to impact of 100 basis point rate hike on regional output in the capital-intensive sector (in blue) and in the labor-intensive output (in grey) at the sample mean (in %). Horizontal axis refers to horizon of IRF (in years). Solid lines denote point estimates and shaded areas denote 90% confidence bands. In the right-hand side (RHS) panel, vertical axis refers to the ratio of output in the capital-intensive sector over the output in the labor-intensive sector, averaged over the sample period. Horizontal axis refers to deciles of the regional per-capita GDP distribution.

Consistent with standard notions of the interest rate channel of monetary policy, the impulse responses point to a significantly greater sensitivity of gross value added (GVA) in the capital-intensive sub-sectors (Figure 7, LHS panel). In particular, capital-intensive output contracts more quickly (reaching its trough already in year $t+1$) and more strongly (with a trough impact roughly double that estimated for labor-intensive output; see blue line). At the same time, capital-intensive output displays a fairly dynamic V-shaped recovery, whereas labor-intensive output remains near its trough for longer and still remains below its initial level through year $t = 5$ (grey line), thus pointing to some mild hysteresis, which we study in greater detail below.

Figure 8: Impact of monetary policy on regional output: bottom versus top decile



Notes: Vertical axis refers to impact of 100 basis point rate hike (in %). Horizontal axis refers to horizon of IRF (in years). Solid lines denote point estimates and shaded areas denote 90% confidence bands. The LHS (RHS) chart refers to the GVA of capital-intensive (labor-intensive) sub-sectors. Grey (blue) lines refer to average effects in top (bottom) deciles.

A natural question is how the differential responsiveness across industry types interacts with our previous results. For instance, if regions in the lower part of the GDP distribution were to also specialize in capital-intensive production, we could not be sure whether their stronger sensitivity to monetary policy shocks reflects the former or the latter characteristic. However, two pieces of evidence speak against industry structure as an explanation for the greater sensitivity of poorer regions. First, they in fact exhibit a lower share of capital-intensive production than

richer regions (Figure 7, RHS panel). Second, even within each type of industry, the differential response across the distribution remains clearly visible: both the capital- and the labor-intensive production contracts more strongly and more persistently in the lower than in the upper part of the distribution (Figure 8). Hence, the inherent differences across more and less prosperous regions also carry over to the industry-breakdown and so does the finding of monetary policy exerting persistent effects on output.

4.4. Sources of hysteresis

The most striking feature of these results is that substantial parts of the regional distribution experience long-lived effects of monetary policy shocks on output. This result clearly contrasts with the common notion of monetary policy as merely causing transitory adjustments in the real economy and, as such, it links to a long-standing debate on potential sources of long-term monetary non-neutrality. This debate has enjoyed a revival in the aftermath of the 2007/2008 financial crisis, which brought renewed urgency to the question whether contractions in aggregate demand may give rise to hysteresis, consisting in lasting declines in the productive capacity of the economy (Yellen, 2016). If so, monetary-policy induced changes in activity may also prove persistent and several recent papers support this conjecture. Most closely related to our work, Jordà, Singh and Taylor (2020) provide empirical evidence of long-lived effects of monetary policy on output in a panel of advanced economies using a local projections framework similar to that adopted here.[13]

An aspect common to these papers is that they tend to adopt a broad perspective regarding the sources of hysteresis, moving beyond the initial emphasis on the labor market in the seminal contribution by Blanchard and Summers (1986) to also consider productivity and capital

[13]For further empirical evidence see, *e.g.*, Blanchard et al. (2015) and for model-based analysis see, *e.g.*, Reifschneider et al. (2015).

Figure 9: Impact of monetary policy on employment and labor productivity in the bottom decile

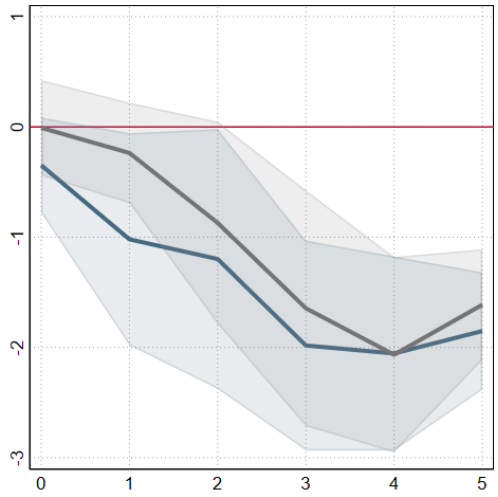
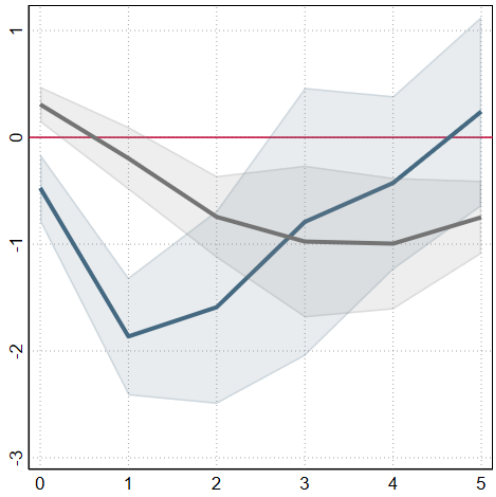


Figure 10: Impact of monetary policy on employment and labor productivity at the mean



Notes: Vertical axis refers to impact of a 100 basis point rate hike on the respective dependent variable (in %), horizontal axis refers to horizon of IRF (in years). In the LHS panel, the IRFs are for regional employment (in grey) and labor productivity (in blue) in the bottom decile of the per-capita GDP distribution; in RHS panel, the IRFs are for regional employment (in grey) and labor productivity (in blue) at the sample mean. Solid lines in the top row denote point estimates and shaded areas denote 90% confidence bands.

accumulation as potential candidates. The debate on which of these potential sources matters (most) is far from settled however. For instance, Jordà, Singh and Taylor (2020) document lasting impacts of monetary policy on the capital stock and on total factor productivity, but not on labor, whereas Blanchard (2018) points to labor markets as an important origin of hysteresis. This question has important implications as to how hysteresis should be modeled and what policy options appear best suited to address it.[14] Here, we seek to add to this debate, by re-running our model separately for employment and labor productivity (with the latter being defined as GDP over employment).

The results point to persistent contractions in employment and labor productivity after a monetary policy shock (Figure 9). Both variables monotonously fall until year $t+4$, to a trough standing around 2% below their initial levels, and only recover slightly thereafter. Since

[14] For instance, consistent with their emphasis on labor market hysteresis, Blanchard and Summers (1986) focus on the wage-setting process in modeling hysteresis. Reifschneider et al. (2015) instead also consider hysteresis in capital deepening and multifactor productivity in their extension of the FRB/US model of the Federal Reserve Board staff.

our data cover a narrower set of variables than those of Jordà, Singh and Taylor (2020), we cannot disentangle whether the latter finding reflects a lasting erosion of the capital stock or of total factor productivity (TFP). However, to be consistent with the estimated responses of employment and labor productivity, the combined adjustment in the capital stock and TFP has to also consist in a long-lived contraction (conversely, if capital and TFP would remain constant, labor productivity would rise, as the drop in employment would raise the capital intensity).

For employment, hysteresis appears as a fairly broad-based phenomenon, applying to large parts of the distribution: even at the sample mean employment remains almost 1% below its initial level by the end of the horizon and only shows tentative signs of recovery after hitting its trough in the preceding year (Figure 10, grey line). Hysteresis in labor productivity by contrast is less pervasive and, at the mean, it displays a full recovery over the horizon (blue line).

The uneven incidence of hysteresis, with monetary policy exerting persistent effects in the lower parts and only transitory ones in the upper parts of the distribution, implies that monetary policy may have a long-lasting impact on regional inequality. That hysteresis concentrates in poorer regions appears consistent with their structural features. First, regions in the lower quantiles of the per-capita GDP distribution exhibit a weaker labour market performance, as indicated by their lower employment rates (Figure 11, blue line). Second, poorer regions host a markedly higher share of employees working in small businesses (Figure 11, grey line). These tend to be more exposed to cyclical fluctuations (Crouzet and Mehrotra, 2020), and they may provide a weaker engine for local agglomeration effects, in turn weighing on local labour productivity (Alfaro et al., 2019). A higher share of fragile firms and low-skilled workers, in turn, implies a stronger vulnerability to policy-induced economic slowdowns (Zens et al., 2020; Duval et al., 2020; Furlanetto et al., 2021). Third, poorer regions are characterised by weaker business dynamism (Figure 12, blue line), which may impede the reallocation of production

Figure 11: Employment rates and small business employment across quantiles

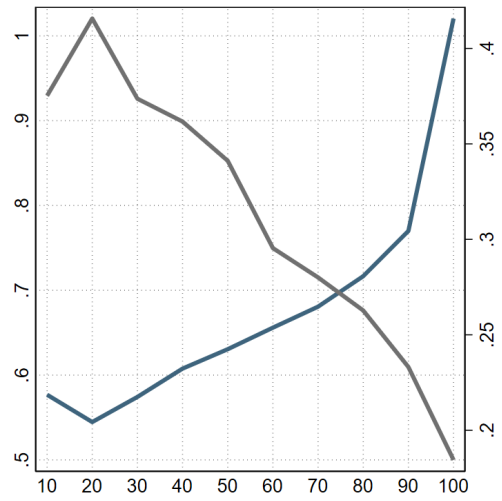
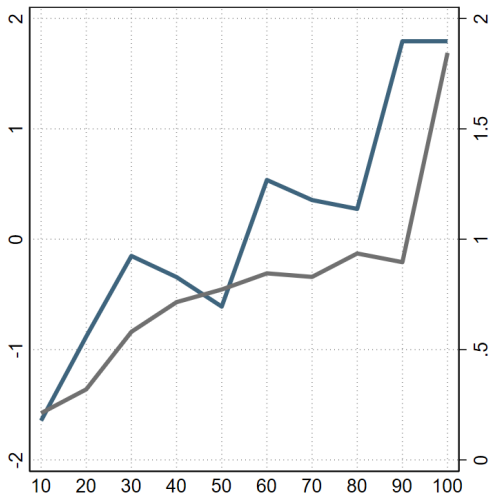


Figure 12: Net business growth and patent applications across quantiles



Source: Eurostat and own computations.
Notes: Horizontal axes refer to deciles of the regional per-capita GDP distribution. Figure 11 shows average regional employment rates over 2002-2014 defined as the ratio of employed persons to the working age population, i.e. the population aged 15-64 years, (blue line, LHS axis) and the average share of regional employment in small businesses with 1-9 employees for the years 2008-2014 (grey line, RHS axis). Figure 12 shows regional net business growth defined as the birth rate minus the death rate of enterprises on average over 2008-2014 (blue line, LHS axis) and the average number of patent applications per 1000 inhabitants by NUTS-3 region for the years 2001 to 2014 (blue line, RHS axis).

factors. Policy-induced cutbacks in activity may hence durably weaken the productive capacity of the local economies and thereby translate into a lower GDP trajectory via endogenous growth effects (Moran and Queralto, 2018; Ikeda and Kurozumi, 2019; Anzoategui et al., 2019; Queralto, 2020). Finally, regions in the lower quantiles of the distribution exhibit a weaker innovative capacity, as measured by patent applications (Figure 12, grey line), which may further dampen their resilience to downturns (Martin, 2012).

5. New Keynesian Currency Union model

In this section, we build a New-Keynesian Currency Union (NKCU) model with endogenous total factor productivity (TFP) to provide a structural interpretation of our empirical findings. The basic model set up is similar to Gali and Monacelli (2008), Bellifemine et al. (2022) and Hazell et al. (2022) but regional total factor productivity (TFP) is endogenously determined by regional economy activity. Motivated by the heterogeneity in innovative capacity documented in Section 4.4, we allow the TFP process to differ across regions. We first show

that the model is consistent with the main empirical findings of this paper–namely, that output, employment and labour productivity responses to monetary policy shocks are more persistent in low-income regions. We then study three counterfactual questions that complement our empirical analysis: One, what are the differences in *inflation* dynamics across high- and low-income regions in response to a monetary tightening? Two, how much does regional heterogeneity affect the transmission of monetary policy at the *aggregate* level? Three, how do aggregate outcomes change if regional inequality were to feature in the monetary policy rule?

5.1. Model environment

The currency union consists of a continuum of small open economies ("regions") of unit mass. We assume the law-of-one-price holds across the union. Since many features of the model are standard, we only sketch the key building blocks here. The full details of the model derivation are in Appendix D.

Households The representative household in region i has additively separable constant-relative-risk-averse preferences over hours worked and a composite consumption basket of home and foreign goods. Furthermore we assume risk sharing across regions by allowing households to trade a complete set of state-contingent claims. What results (in log-deviation space) is a union-wide IS equation of the form

$$\hat{y}_t^* = \mathbb{E}_t \hat{y}_{t+1}^* - \frac{1}{\sigma}\left(\hat{r}_t^* - \mathbb{E}_t \pi_{t+1}^*\right), \tag{2}$$

where $\hat{y}_t^*$, π_t^*, and $\hat{r}_t^*$ are the union-wide level of output, inflation rate, and nominal risk-free interest rate, respectively. To fix notation, $\hat{x}_t \equiv \log(X_t) - \log(X)$, where X_t is the level and X is its deterministic steady state. Parameter σ is the household coefficient-of-relative-risk-aversion.[15]

[15] All structural parameters are common across regions, unless explicitly stated.

Firms, price setting, and endogenous TFP Each region has a unit continuum of firms, each j producing a differentiated good with $Y_t^i(j) = A_t^i \left(N_t^i(j)\right)^{1-\alpha}$ where $N_t^i(j)$ is hours worked and A_t^i is TFP, which is region-specific and taken as given by the firm. Firms set prices as in Calvo (1983). The resulting regional Phillips curve is given by

$$\pi_t^i = \beta \mathbb{E}_t \pi_{t+1}^i + \kappa_y \hat{y}_t^i - \kappa_a \hat{a}_t^i + \kappa_s \hat{s}_t^i, \tag{3}$$

where $\hat{s}_t^i$ is the effective terms of trade for region i, β is the subjective discount factor of households and κ_y, κ_a, and κ_s are functions of other structural parameters. Denote $U_t^i = 1 - N_t^i$ as the unemployment rate. We assume that TFP is endogenous and accumulation follows $A_t^i = \left(A_{t-1}^i\right)^{\rho^i} \left(1 - U_t^i\right)^{\phi}$, where A_t^i is increasing in past TFP and decreasing in the unemployment rate. In other words, unemployment causes hysteresis via a de-accumulation of skills or human capital.[16]

In linking TFP to the level of unemployment, we follow the overwhelming majority of hysteresis-type models in the literature, which relates the endogenous growth engine (for example, capital accumulation, R&D, or learning-by-doing) to the cyclical state of the economy (for example, the level of output, employment, or investment spending). Variations on this theme can be found in Stadler (1986), King et al. (1988), Engler and Tervala (2018), Bianchi et al. (2019), and Cerra et al. (2021). Chodorow-Reich (2019a) and Cerra et al. (2023) provide a comprehensive review of this literature.

We assume structural heterogeneity across regions in the $\rho^i \in [0, 1)$ parameter, motivated by the empirical patterns presented at the end of Section 4.4. In particular, we presented two plausible reasons for why the state of technology may evolve more sluggishly in some regions

[16] To demonstrate the generality of the analytical results to come, Appendix C derives similar analytical results in a Real Business Cycle model in which TFP is related the level of region-wide investment spending.

than in others (Figure 12). First, there is significant variation in business dynamism across regions in Europe (as measured by the net growth rate of new business creation). This means low productivity firms are not replaced as readily by new entrants in some regions, causing the current level of TFP to persist for longer. Second, there is significant variation in R&D activity across regions (as measured by patent applications). This, in turn, means that old technologies are likely to be used for longer in some regions, also increasing the persistence of TFP.

While we assume heterogeneity in ρ^i across regions, the model's ability to explain our empirical findings is not manifest. First, the model needs to predict that a higher ρ^i makes a region poorer. This is not guaranteed as endogenous growth models (as opposed to hysteresis models) attest. Second, a higher ρ^i needs to generate more persistence in *both* output *and* employment. As we discuss below, this co-persistence will depend on the value of σ.

Monetary policy Monetary policy is described by a simple Taylor-type rule

$$\hat{r}_t^* = \phi_\pi \hat{\pi}_t^* + \phi_y y_t^* + u_t, \tag{4}$$

where $u_{t+1} = \rho u_t + \sigma_u \varepsilon_{t+1}$ and ε_{t+1} is drawn from an *iid* standard normal distribution.

Steady state Since our model needs to explain heterogeneity in the level of regional incomes, and since TFP is endogenous, it is important to characterize the steady state of the model. We impose a zero-inflation steady state and set an employment subsidy to ensure the steady state is not distorted. This gives the following system of non-linear equations for steady state TFP,

employment, consumption, and output in a given region:

$$A^i = \left(N^i\right)^{\frac{\phi}{1-\rho^i}}, \tag{5}$$

$$\xi \left(C^i\right)^{\sigma} \left(1-N^i\right)^{-\varphi} = (1-\alpha) A^i \left(N^i\right)^{-\alpha}, \tag{6}$$

$$Y^i = C^i, \tag{7}$$

$$Y^i = A^i \left(N^i\right)^{1-\alpha}, \tag{8}$$

where ξ is the utility weight on hours worked, φ is the Frisch elasticity, and C^i is consumption. The equations are the steady state counterparts of the TFP process, the labour market clearing condition, the aggregate demand condition, and the production function, respectively.

5.2. Analytical results

In order to prove sharp analytical results relating the *level* of income (and unemployment) to the *persistence* of output, employment and labour productivity deviations from steady state, we impose assumptions on some of the model's parameters. First, we assume that prices are fully flexible, reducing the model to a classical monetary model. Second, we impose the parameter restriction $\alpha < \frac{\phi}{1-\rho}$, which ensures that in aggregate, production features decreasing returns to labour (we drop the *i* subscript for notational clarity).

Since prices are fully flexible, the only first-order condition that determines real outcomes is the consumption-leisure trade-off given by $\xi C_t^{\sigma} (1-N_t)^{-\varphi} = (1-\alpha) A_t N_t^{-\alpha}$. The model's equilibrium can then be fully described by two endogenous variables: $\{A_t, N_t\}$, and the following two equations: $\xi N_t (1-N_t)^{-\varphi} = (1-\alpha) A_t^{1-\sigma} N_t^{(1-\alpha)(1-\sigma)}$ and $A_t = A_{t-1}^{\rho} N_t^{\phi}$.

The first proposition regards the comparative static effects of hysteresis on the steady state:

Proposition 1 *A sufficient condition for steady state output, Y_t, to be decreasing in the persistence of TFP, ρ, is $\sigma < 1$ (i.e., an intertemporal elasticity of substitution greater than 1).*

See Appendix C for the proof. The parameter σ measures the strength of the wealth effect of labor supply. When $\sigma < 1$, the substitution effect on labor supply resulting from a lower real wage (e.g., because of a fall in A_t) dominates the positive effect caused by a higher marginal utility of consumption. All else equal, this leads to a decrease in employment. In steady state, the persistence of the TFP process acts as a multiplier to this relationship between labour supply and TFP, resulting in regions with higher TFP persistence experiencing lower incomes.

The second proposition relates hysteresis to the persistence of output, employment, and labour productivity deviations from trend.

Proposition 2 *The conditional autocorrelations of output, employment, and labour productivity are increasing in* ρ *(so long as* $\sigma < 1$*).*

The proof is in Appendix C. The key step here is to show that the dynamics of these variables can be represented as $ARMA(2,1)$ processes, from which it is straightforward to derive analytical expressions for their autocorrelation functions. The persistence of output depends on both the persistence of TFP and labour supply. Again, a key requirement is that $\sigma < 1$. This ensures that labour supply is increasing in TFP (and the real wage). If instead $\sigma > 1$, the household compensates for any (endogenous) fall in TFP by raising labour supply, which would cause employment to return back to steady state faster. Thus, even if TFP is more persistent because of a larger ρ value, the labour supply response would cause the effect on the persistence of output to be ambiguous.

Discussion of σ The empirical literature is divided on whether the intertemporal elasticity of substitution (IES) is above or below unity.[17] An important consideration regarding the value of σ is the relationship between average hours worked and income levels. A σ below one means

[17]For example, Hall (1988) supports an IES close to zero, Basu and Kimball (2002) finds a value around 0.5, and Smets and Wouters (2007) estimated 0.72. In contrast, Bansal and Yaron (2004) chose an IES of 1.5—arguing

the substitution effect on labor supply is larger than the wealth effect, causing labor supply to be increasing in TFP. Bick et al. (2018) find that, within most countries, hours per worker are decreasing in the individual wage, but in the richest countries hours worked are actually flat or increasing in the wage. The latter finding speaks closely to the environment our paper studies–regional variation in a highly developed economy—and more so than aggregate time-series evidence or cross-country studies that combine developed and developing countries.[18]

Finally, the canonical New-Keynesian model, as presented here, has an intensive labor-leisure margin but does not feature involuntary unemployment or an extensive labor force margin. Thus, we follow Hazell et al. (2022) in defining unemployment simply as $U_t = 1 - N_t$. It is reasonable therefore to interpret our model as predicting that high-income regions have a lower unemployment rate (rather than a higher labor supply, per se; for evidence, see Özyurt and Dees, 2015).

5.3. Regional responses to monetary policy shocks in the NKCU model

Propositions 1 and 2, taken together, are consistent with the main empirical findings of this paper, which show that deviations of output are more persistent in low-income regions. Moreover, the model and propositions are very general. For example, we do not need to specify the shock that drives the economy away from steady state. Neither do we need to invoke nominal frictions nor any assumptions regarding the "small" open economy nature of a region. However, to directly study how monetary policy affects different types of regions, and to examine the aggregate implications of regional heterogeneity, we simulate a calibrated version of the full New Keynesian Currency Union model.

---

that earlier empirical work ignored the effect of time-varying volatility, creating downward bias in the estimates of the IES. Similarly, Van Binsbergen et al. (2012) estimated 1.73.

[18] As an example of the latter, Boppart and Krusell (2020) show evidence of lower labor supply in rich countries and declining labor supply over time as an economy grows.

Calibration The model is calibrated with standard parameter values from the literature. Here, we highlight a few key parameters with details of the calibration in Appendix D.

Each period is calibrated to be a quarter. We assume that all regions are described by a common set of parameters, except ρ. The intertemporal elasticity of substitution (IES) is $1/\sigma = 2$. We assume that half the regions have $\rho = 0$ and the other half have $\rho = 0.925$. Note that while we will label these regions high- and low-income, respectively, these designations are an endogenous outcome of the model. The Frisch elasticity of labour supply, $\tilde{\varphi}$ is set to 1. We calibrate the disutility weight on hours worked, ξ, such that for the high-income regions (with $\rho = 0$), the steady state unemployment rate is 5%. An endogenous outcome of the model is that the low-income regions (with $\rho = 0.925$) have a steady state unemployment rate of 7.5%, or 2.5 percentage points above the high-income region. Furthermore, the calibration implies that steady state income is 13% higher in the high-income region.

The labour share of income, $1 - \alpha = 0.9$. This parameter is set to less that one to ensure that labour productivity is time-varying as in the data. The Calvo parameter, $\theta = 0.72$, implies that firms adjusts prices on average every 3.5 quarters. Finally, the unemployment elasticity of TFP, $\phi = 0.1$, implies that a 1 percentage point rise in the unemployment rate decreases TFP by 10 basis points.

Impulse responses Figure 13 plots the response to an unanticipated union-wide monetary policy tightening. Output falls in all regions, but the depth and, in particular, the persistence of the response differs markedly. While the contraction in high-income regions (red-circle line) is temporary, in the low-income regions (solid-blue line) output remains persistently below steady state. The difference arises from the endogenous TFP channel. As a result of the fall in employment (rise in unemployment), workers lose skills and TFP declines. However, the low-income regions are characterized by hysteresis in that the impact of those declines in

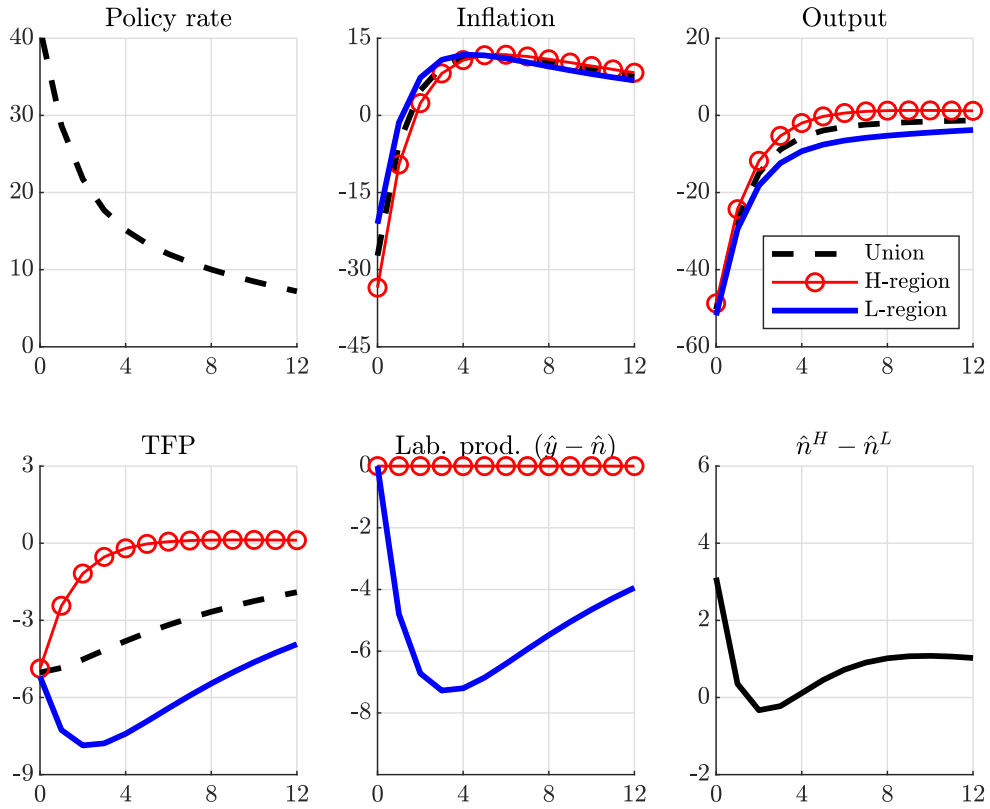
TFP become imbedded. In fact, TFP continues to fall for several quarters after impact in the low-income region as the effect of the high unemployment rate and the persistence of TFP accumulate, leading to the observed hump-shaped response. For the high-income regions, TFP instead begins to recover immediately, tracking the recovery in employment. Like for the output response, both employment and labour productivity are further from steady state in the low-income region, both on impact and thereafter.[19] This is because of the larger fall in the marginal product of labour (i.e., wages) in the low-income regions, which implies a stronger reduction in the opportunity cost of not working.

While our empirical results do not consider the regional inflation effects of a monetary policy shock, the model predicts that inflation falls less in low-income regions. The intuition for this is to think of the scenario in Figure 13 as the combination of a correlated monetary policy and negative technology shock, with the latter being larger in the low-income region. A negative technology shock lowers output and raises inflation, all else equal, by weakening the supply side of the economy. This, in turn, leads inflation to decline by less in the low-income than in the high-income region. In the following quarters, the weakened supply side causes inflation to briefly rise above target. This is true also of high income regions (despite the lack of hysteresis) as they import inflation through tradable goods from low-income regions.

The aggregate effects of heterogeneity Finally, the model allows us to study the aggregate effects of regional heterogeneity. To this end, we compare in Figure 14 the previous simulations (thick-dashed lines) with the response to the monetary policy shock when regional heterogeneity is removed. We consider four different counterfactuals in which either $\rho = 0$, 0.4625, 0.8776, or 0.925 is common across regions. The first and last correspond to setting ρ

[19] The differences in employment outcomes are small across the regions. Hence, the bottom right panel plots the cross-region differences rather than the levels of employment.

Figure 13: Response to a monetary policy shock

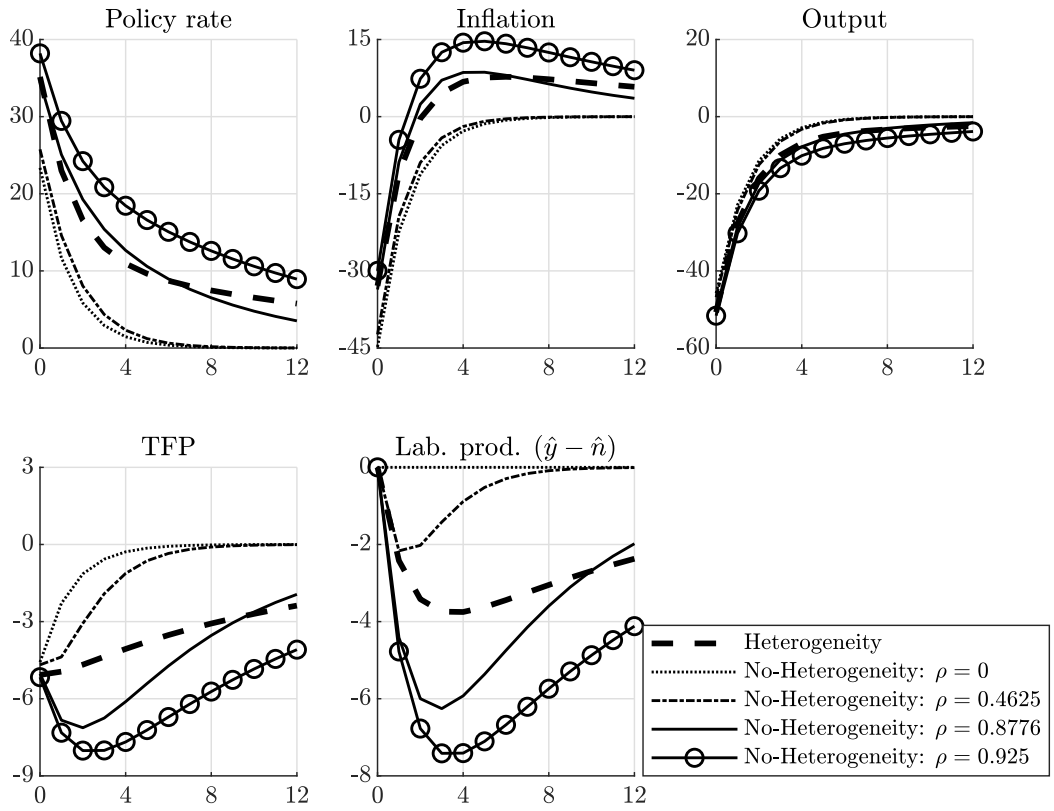


Note: 1 percentage point monetary policy tightening. Variables are basis point deviations from steady state. The policy rate and inflation are annualized.

equal to the value in the high- and low-income regions, respectively. The second sets ρ as an average of the two. The third ($\rho = 0.8776$) is selected such that the Union-wide steady state income level is the same as in the heterogeneous calibration. In all cases, we set the steady state unemployment rate in each region to 6.2% by appropriately adjusting ξ. When heterogeneity across regions is removed and $\rho = 0$, such that there is no hysteresis, inflation falls further on impact and does not overshoot thereafter. As a result, the policy rate rises by less on impact and converges quickly back to steady state. This highlights the fact that in the heterogeneous calibration, the inflation dynamics experienced by the high income regions is in large part driven by the supply side weakness in the low income regions. A comparison of the second and third calibrations, which apply a common ρ equal to the "average ρ" and a ρ that ensures the steady

state Union-wide income level is equal to the heterogeneous Union, respectively, highlights the strong nonlinearity between steady state output and hysteresis in the model. In the former, the aggregate economy displays very little hysteresis. In the latter the economy resembles more closely the aggregate dynamics of the heterogenous Union.

Figure 14: Aggregate effects of heterogeneity



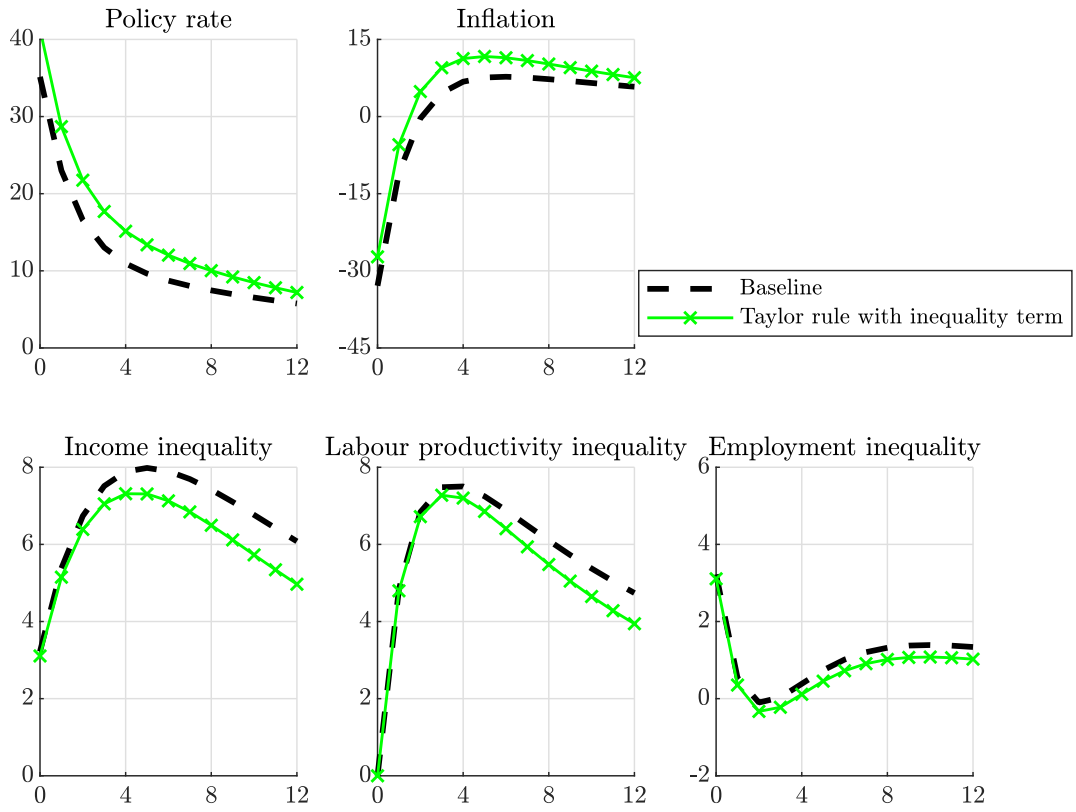
Note: 1 percentage point monetary policy tightening. Variables are basis point deviations from steady state. The policy rate and inflation are annualized.

Responding to inequality Given that tighter monetary policy aggravates regional inequality, we can use the model to explore the effect of monetary policy choosing to respond to a measure of inequality. In particular, we augment the policy rule as follows

$$\hat{r}_t^* = \phi_\pi \hat{\pi}_t^* + \phi_y \hat{y}_t^* - \phi_{in} \left(\hat{y}_t^H - \hat{y}_t^L \right) + u_t, \tag{9}$$

where $\phi_{in} \geq 0$. Figure 15 shows the response to a monetary policy shock when $\phi_{in} = \phi_y = 0.125$. The bottom row shows that the augmented policy rule dampens the rise in both income and employment inequality. However, the trade-off is a higher level of inflation.[20]

Figure 15: Responding to inequality



Note: 1 percentage point monetary policy tightening. Variables are basis point deviations from steady state. The policy rate and inflation are annualized.

6. Conclusion

This paper provides a novel perspective on the regional patterns of monetary policy transmission. Using geographically disaggregated data on economic activity in Europe, we show that the output response to short-term interest rate shocks is significantly more pronounced and

[20] Future research would benefit from a more thorough exploration of the normative monetary policy implications of regional differences in hysteresis.

persistent in poorer than in richer cities and counties. Moreover, while GDP in the upper part of the distribution returns to its pre-shock level after three to five years, the response in the lower parts does not reverse over this period, thus pointing to pronounced hysteresis in output. The heterogeneous incidence of hysteresis implies that monetary policy has a long-lasting impact on regional inequality, with tightening shocks aggravating and easing shocks mitigating it. In terms of anatomy, we find hysteresis to originate from long-lived adjustments in both employment and labor productivity. At the same time, employment hysteresis is more pronounced and more broad-based across the distribution. As such, our findings confirm labor markets as an important source of hysteresis in the European context. From a theoretical perspective, we show that our main findings are consistent with a New Keynesian Currency Union model in which technological progress is endogenous to the regional unemployment rate.

Our paper points to the use of geographically disaggregated data as a promising avenue for further insights into how monetary policy shocks propagate to the economy. First, by resorting to information on economic activity at a more granular level than that entering the central bank reaction function, it offers a novel strategy to identify exogenous changes in monetary policy. Second, by providing empirical estimates for the monetary policy impact on regional inequality, it closes an important gap in the large and growing literature on heterogeneity in monetary policy transmission.

From a policy perspective, our findings underscore the challenges of calibrating monetary policy in heterogeneous economies. In the euro area context, the debate has typically interpreted these challenges as a cross-country phenomenon. As such, they attracted particular attention during the euro area sovereign debt crisis from 2010-12, which was marked by strong cross-country divergence in economic performance and raised concerns as to the suitability of a given aggregate monetary policy stance for individual countries. However, our analysis

demonstrates that the issue runs deeper: interregional heterogeneity becomes more accentuated at more granular geographical levels and this heterogeneity in turn profoundly alters the implications of a given monetary policy stance in different parts of the economy. These implications emerge as particularly relevant in view of our finding that monetary policy exerts durable impacts on output and employment – a finding that contrasts with the common notion of stabilization policies merely smoothing out fluctuations in these variables around their natural levels.

Funding: None received.

References

Alfaro, L., Chen, M. X. and Fadinger, H. (2019). Spatial Agglomeration and Superstar Firms: Firm-Level Patterns from Europe and US, *ECB Forum on Central Banking: 20 Years of European Economic and Monetary Union: Conference Proceedings*, European Central Bank.

Altavilla, C., Canova, F. and Ciccarelli, M. (2020). Mending the Broken Link: Heterogeneous Bank Lending Rates and Monetary Policy Pass-Through, *Journal of Monetary Economics* **110**: 81–98.

Anzoategui, D., Comin, D., Gertler, M. and Martinez, J. (2019). Endogenous Technology Adoption and R&D as Sources of Business Cycle Persistence, *American Economic Journal: Macroeconomics* **11**(3): 67–110.

Arnold, I. J. (2001). The Regional Effects of Monetary Policy in Europe, *Journal of Economic Integration* **16**(3): 399–420.

Arnold, I. J. and Vrugt, E. B. (2002). Regional Effects of Monetary Policy in the Netherlands, *International Journal of Business and Economics* **1**(2): 123–134.

Auclert, A. (2019). Monetary Policy and the Redistribution Channel, *American Economic Review* **109**(6): 2333–67.

Auerbach, A. J. and Gorodnichenko, Y. (2012). Fiscal Multipliers in Recession and Expansion, *in* A. Alesina and F. Giavazzi (eds), *Fiscal Policy After the Financial Crisis*, University of Chicago Press, pp. 63–98.

Austin, B., Glaeser, E. and Summers, L. H. (2018). Saving the Heartland: Place-Based Policies in 21st Century America. *Brookings Papers on Economic Activity, Spring 2018*.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* **59**(4): 1481–1509.
URL: *http://dx.doi.org/10.1111/j.1540-6261.2004.00670.x*

Basu, S. and Kimball, M. (2002). Long-run labor supply and the elasticity of intertemporal substitution for consumption, *manuscript, University of Michigan* .

Beck, G. W., Hubrich, K. and Marcellino, M. (2009). Regional Inflation Dynamics Within and Across Euro Area Countries and a Comparison with the United States, *Economic Policy* **24**(57): 142–184.

Becker, S. O., Egger, P. H. and Von Ehrlich, M. (2010). Going NUTS: The Effect of EU Structural Funds on Regional Performance, *Journal of Public Economics* **94**(9-10): 578–590.

Beckworth, D. (2010). One Nation Under the Fed? The Asymmetric Effects of US Monetary Policy and its Implications for the United States as an Optimal Currency Area, *Journal of Macroeconomics* **32**(3): 732–746.

Belke, A., Haskamp, U. and Setzer, R. (2016). Regional Bank Efficiency and its Effect on Regional Growth in "Normal" and "Bad" Times, *Economic Modelling* **58**: 413–426.

Bellifemine, M., Couturier, A. and Jamilov, R. (2022). The Regional Keynesian Cross, *Economics Series Working Papers 995*, University of Oxford, Department of Economics.

Beraja, M., Fuster, A., Hurst, E. and Vavra, J. (2018). Regional Heterogeneity and the Refinancing Channel of Monetary Policy, *Quarterly Journal of Economics* **134**(1): 109–183.

Beraja, M., Hurst, E. and Ospina, J. (2019). The Aggregate Implications of Regional Business Cycles, *Econometrica* **87**(6): 1789–1833.

Bianchi, F., Kung, H. and Morales, G. (2019). Growth, slowdowns, and recoveries, *Journal of Monetary Economics* **101**: 47–63.

Bick, A., Fuchs-Schündeln, N. and Lagakos, D. (2018). How do hours worked vary with income? cross-country evidence and implications, *American Economic Review* **108**(1): 170–199.

Bielecki, M., Brzoza-Brzezina, M. and Kolasa, M. (2021). Intergenerational Redistributive Effects of Monetary Policy, *Journal of the European Economic Association* **20**(2): 549–580. **URL:** *https://doi.org/10.1093/jeea/jvab032*

Blanchard, O. (2018). Should We Reject the Natural Rate Hypothesis?, *Journal of Economic Perspectives* **32**(1): 97–120.

Blanchard, O., Cerutti, E. and Summers, L. (2015). Inflation and Activity, Two Explorations and their Monetary Policy Implications, *NBER WP No. 21726* .

Blanchard, O. J. and Summers, L. H. (1986). Hysteresis and the European Unemployment Problem, *NBER Macroeconomics Annual* **1**: 15–78.

Boeckx, J., Dossche, M. and Peersman, G. (2017). Effectiveness and Transmission of the ECB's Balance Sheet Policies, *International Journal of Central Banking* **2013**(1): 297–333.

Boppart, T. and Krusell, P. (2020). Labor supply in the past, present, and future: a balanced-growth perspective, *Journal of Political Economy* **128**(1): 118–157.

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

Carlino, G. and DeFina, R. (1998). The Differential Regional Effects of Monetary Policy, *Review of Economics and Statistics* **80**(4): 572–587.

Cerra, M. V., Hakamada, M. and Lama, M. R. (2021). Financial crises, investment slumps, and slow recoveries, *IMF Working Paper* .

Cerra, V., Fatás, A. and Saxena, S. C. (2023). Hysteresis and business cycles, *Journal of Economic Literature* **61**(1): 181–225.

Chodorow-Reich, G. (2019a). Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?, *American Economic Journal: Economic Policy* **11**(2): 1–34.

Chodorow-Reich, G. (2019b). Regional Data in Macroeconomics: Some Advice for Practitioners. *NBER WP No. 26501*.

Clemens, J. and Miran, S. (2012). Fiscal Policy Multipliers on Subnational Government Spending, *American Economic Journal: Economic Policy* **4**(2): 46–68.

Coenen, G., Karadi, P., Schmidt, S. and Warne, A. (2018). The New Area-Wide Model II: an Extended Version of the ECB's Micro-Founded Model for Forecasting and Policy Analysis with a Financial Sector. *ECB WP No. 2200*.

Coeuré, B. (2018). The Local Impact of the ECB's Monetary Policy. *European Central Bank*. Retrieved on 14 November 2019 on: <https://www.ecb.europa.eu/press/key/date/2018/html/ecb.sp181004.en.html>.

Coibion, O. and Goldstein, D. (2012). One for Some or One for All? Taylor Rules and Interregional Heterogeneity, *Journal of Money, Credit and Banking* **44**(2-3): 401–431.

Coibion, O., Gorodnichenko, Y., Kueng, L. and Silvia, J. (2017). Innocent Bystanders? Monetary Policy and Inequality, *Journal of Monetary Economics* **88**: 70–89.

Corbi, R., Papaioannou, E. and Surico, P. (2019). Regional transfer multipliers, *The Review of Economic Studies* **86**(5): 1901–1934.

Corsetti, G., Duarte, J. B. and Mann, S. (2021). One Money, Many Markets, *Journal of the European Economic Association* **20**(1): 513–548.
URL: *https://doi.org/10.1093/jeea/jvab030*

Crouzet, N. and Mehrotra, N. R. (2020). Small and Large Firms over the Business Cycle, *American Economic Review* **110**(11): 3549–3601.

Dedola, L. and Lippi, F. (2005). The Monetary Transmission Mechanism: Evidence from the Industries of Five OECD Countries, *European Economic Review* **49**(6): 1543–1569.

Dhaene, G. and Jochmans, K. (2015). Split-Panel Jackknife Estimation of Fixed-Effect Models, *The Review of Economic Studies* **82**(3): 991–1030.

Di Giacinto, V. (2003). Differential Regional Effects of Monetary Policy: a Geographical SVAR Approach, *International Regional Science Review* **26**(3): 313–341.

Di Maggio, M., Kermani, A., Keys, B. J., Piskorski, T., Ramcharan, R., Seru, A. and Yao, V. (2017). Interest Rate Pass-Through: Mortgage Rates, Household Consumption, and Voluntary Deleveraging, *American Economic Review* **107**(11): 3550–88.

Dolado, J. J., Motyovszki, G. and Pappa, E. (2021). Monetary Policy and Inequality under Labor Market Frictions and Capital-Skill Complementarity, *American Economic Journal: Macroeconomics* **13**(2): 292–332.

Dow, S. C. and Montagnoli, A. (2007). The Regional Transmission of UK Monetary Policy, *Regional Studies* **41**(6): 797–808.

Driscoll, J. C. and Kraay, A. C. (1998). Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data, *Review of Economics and Statistics* **80**(4): 549–560.

Dupraz, S., Nakamura, E. and Steinsson, J. (2019). A Plucking Model of Business Cycles. *NBER WP No. 26351*.

Duval, R., Hong, G. H. and Timmer, Y. (2020). Financial Frictions and the Great Productivity Slowdown, *The Review of Financial Studies* **33**(2): 475–503.

ECB (1998). A Stability-Oriented Monetary Policy Strategy for the ESCB. *European Central Bank*. Retrieved on 14 November 2019 on: <https://www.ecb.europa.eu/press/key/date/1998/html/sp981110.en.html>.

ECB (2015). The Governing Council's Expanded Asset Purchase Programme. *European Central Bank Economic Bulletin, Issue 1/2015, Box 1*.

Eichenbaum, M., Rebelo, S. and Wong, A. (2018). State-Dependent Effects of Monetary Policy: The Refinancing Channel. *NBER WP No. 25152*.

Engler, P. and Tervala, J. (2018). Hysteresis and fiscal policy, *Journal of Economic Dynamics and Control* **93**: 39–53.

Faust, J., Rogers, J. H., Swanson, E. and Wright, J. H. (2003). Identifying the Effects of Monetary Policy Shocks on Exchange Rates using High Frequency Data, *Journal of the European Economic Association* **1**(5): 1031–1057.

Francis, N., Owyang, M. T. and Sekhposyan, T. (2012). The Local Effects of Monetary Policy, *The BE Journal of Macroeconomics* **12**(2): 1–38.

Fratantoni, M. and Schuh, S. (2003). Monetary Policy, Housing, and Heterogeneous Regional Markets, *Journal of Money, Credit and Banking* **35**(4): 557–589.

Furceri, D., Mazzola, F. and Pizzuto, P. (2019). Asymmetric Effects of Monetary Policy Shocks Across US States, *Papers in Regional Science* **98**(5): 1861–1891.

Furlanetto, F., Lepetit, A., Robstad, Ω., Rubio Ramírez, J. and Ulvedal, P. (2021). Estimating Hysteresis Effects. CEPR Discussion Paper No. 16558.

Gali, J. and Monacelli, T. (2008). Optimal monetary and fiscal policy in a currency union, *Journal of International Economics* **76**(1): 116–132.

Ganong, P. and Shoag, D. (2017). Why Has Regional Income Convergence in the U.S. Declined?, *Journal of Urban Economics* **102**: 76 – 90.

Gertler, M. and Gilchrist, S. (1994). Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms, *The Quarterly Journal of Economics* **109**(2): 309–340.

Glaeser, E. L., Scheinkman, J. and Shleifer, A. (1995). Economic Growth in a Cross-Section of Cities, *Journal of Monetary Economics* **36**(1): 117–143.

Guren, A., McKay, A., Nakamura, E. and Steinsson, J. (2020). What Do We Learn From Cross-Regional Empirical Estimates in Macroeconomics? *NBER WP No. 26881*.

Hall, R. E. (1988). Intertemporal substitution in consumption, *Journal of political economy* **96**(2): 339–357.

Hansen, L. and Sargent, T. (1991). Two Difficulties in Interpreting Vector Autoregressions, *in* L. Hansen and T. Sargent (eds), *Rational Expectations Econometrics*, Westview Press, Boulder London.

Hazell, J., Herreno, J., Nakamura, E. and Steinsson, J. (2022). The slope of the phillips curve: evidence from us states, *The Quarterly Journal of Economics* **137**(3): 1299–1344.

Hoechle, D. (2007). Robust Standard Errors for Panel Regressions With Cross-Sectional Dependence, *The Stata Journal* **7**(3): 281–312.

House, C. L., Proebsting, C. and Tesar, L. L. (2020). Regional Effects of Exchange Rate Fluctuations, *Journal of Money, Credit and Banking* **52**(S2): 429–463.

Iammarino, S., Rodriguez-Pose, A. and Storper, M. (2018). Regional Inequality in Europe: Evidence, Theory and Policy Implications, *Journal of Economic Geography* **19**(2): 273–298.

Ikeda, D. and Kurozumi, T. (2019). Slow Post-Financial Crisis Recovery and Monetary Policy, *American Economic Journal: Macroeconomics* **11**(4): 82–112.

Jacobs, J. (1969). *The Economy of Cities*, Vintage Books.

Jordà, O. (2005). Estimation and Inference of Impulse Responses by Local Projections, *American Economic Review* **95**(1): 161–182.

Jordà, Ò., Singh, S. R. and Taylor, A. M. (2020). The Long-Run Effects of Monetary Policy, Federal Reserve Bank of San Francisco.

Kaplan, G., Moll, B. and Violante, G. L. (2018). Monetary Policy according to HANK, *American Economic Review* **108**(3): 697–743.

Kashyap, A. K., Lamont, O. A. and Stein, J. C. (1994). Credit Conditions and the Cyclical Behavior of Inventories, *The Quarterly Journal of Economics* **109**(3): 565–592.

Kashyap, A. K. and Stein, J. C. (2000). What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?, *American Economic Review* **90**(3): 407–428.

King, R. G., Plosser, C. I. and Rebelo, S. T. (1988). Production, growth and business cycles: Ii. new directions, *Journal of Monetary Economics* **21**(2-3): 309–341.

Koenker, R. W. and Bassett, G. (1978). Regression Quantiles, *Econometrica* **46**(1): 33–50.

Krippner, L. (2015). *Zero Lower Bound Term Structure Modeling: A Practitioner's Guide*, Springer.

Krugman, P. (1991). Increasing Returns and Economic Geography, *Journal of Political Economy* **99**(3): 483–499.

Krugman, P. (1998). Space: the Final Frontier, *Journal of Economic Perspectives* **12**(2): 161–174.

Lancaster, T. (2000). The Incidental Parameter Problem Since 1948, *Journal of Econometrics* **95**(2): 391–413.

Lane, P. R. and Stracca, L. (2018). Can Appreciation be Expansionary? Evidence from the Euro Area, *Economic Policy* **33**(94): 225–264.

Leahy, J. V. and Thapar, A. (2019). Demographic Effects on the Impact of Monetary Policy. *NBER WP No. 26324*.

Lemke, W. and Vladu, A. L. (2017). Below the Zero Lower Bound: a Shadow-Rate Term Structure Model for the Euro Area. *ECB WP No. 1991*.

Machado, J. A. and Silva, J. S. (2019). Quantiles via Moments, *Journal of Econometrics* **213**(1): 145–173.

Marcet, A. (1991). Time Aggregation of Econometric Time Series, *in* L. Hansen and T. Sargent (eds), *Rational Expectations Econometrics*, Westview Press, Boulder and London.

Martin, R. (2012). Regional economic resilience, hysteresis and recessionary shocks, *Journal of economic geography* **12**(1): 1–32.

Martin, R., Sunley, P., Gardiner, B., Evenhuis, E. and Tyler, P. (2018). The City Dimension of the Productivity Growth Puzzle: the Relative Role of Structural Change and Within-Sector Slowdown, *Journal of Economic Geography* **18**(3): 539–570.

Moran, P. and Queralto, A. (2018). Innovation, Productivity, and Monetary Policy, *Journal of Monetary Economics* **93**: 24–41.

Nakamura, E. and Steinsson, J. (2014). Fiscal Stimulus in a Monetary Union: Evidence from US Regions, *American Economic Review* **104**(3): 753–92.

Newey, W. K. and West, K. D. (1994). Automatic Lag Selection in Covariance Matrix Estimation, *The Review of Economic Studies* **61**(4): 631–653.

Ottonello, P. and Winberry, T. (2020). Financial Heterogeneity and the Investment Channel of Monetary Policy, *Econometrica* **88**(6): 2473–2502.

Owyang, M. T. and Wall, H. J. (2009). Regional VARs and the Channels of Monetary Policy, *Applied Economics Letters* **16**(12): 1191–1194.

Özyurt, S. and Dees, S. (2015). Regional Dynamics of Economic Performance in the EU: to What Extent Spatial Spillovers Matter? *ECB Working Paper No. 1870*.

Peek, J. and Rosengren, E. S. (1995). Bank Lending and the Transmission of Monetary Policy, *Conference Series: Federal Reserve Bank of Boston*, Vol. 39, Federal Reserve Bank of Boston, pp. 47–68.

Peersman, G. and Smets, F. (2005). The Industry Effects of Monetary Policy in the Euro Area, *The Economic Journal* **115**(503): 319–342.

Queralto, A. (2020). A Model of Slow Recoveries from Financial Crises, *Journal of Monetary Economics* **114**: 1–25.

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

Ravn, M. O. and Sterk, V. (2020). Macroeconomic Fluctuations with HANK and SAM: an Analytical Approach, *Journal of the European Economic Association* **19**(2): 1162–1202. **URL:** *https://doi.org/10.1093/jeea/jvaa028*

Reifschneider, D., Wascher, W. and Wilcox, D. (2015). Aggregate Supply in the United States: Recent Developments and Implications for the Conduct of Monetary Policy, *IMF Economic Review* **63**(1): 71–109.

Rodriguez-Fuentes, C. and Dow, S. (2003). EMU and the Regional Impact of Monetary Policy, *Regional Studies* **37**(9): 969–980.

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

Rostagno, M., Altavilla, C., Carboni, G., Lemke, W., Motto, R., Saint-Guilhem, A. and Yiangou, J. (2019). A Tale of Two Decades: The ECB's Monetary Policy at 20. *ECB WP No. 2346*.

Sims, C. A. (1992). Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy, *European Economic Review* **36**(5): 975–1000.

Smets, F. and Wouters, R. (2003). An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area, *Journal of the European Economic Association* **1**(5): 1123–1175.

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

Stadler, G. W. (1986). Real versus monetary business cycle theory and the statistical characteristics of output fluctuations, *Economics Letters* **22**(1): 51–54.

Van Binsbergen, J. H., Fernández-Villaverde, J., Koijen, R. S. and Rubio-Ramírez, J. (2012). The term structure of interest rates in a dsge model with recursive preferences, *Journal of Monetary Economics* **59**(7): 634–648.

Wu, J. C. and Xia, F. D. (2017). Time-Varying Lower Bound of Interest Rates in Europe, *Chicago Booth Research Paper* (17-06).

Yellen, J. L. (2016). Macroeconomic Research After the Crisis. *The Elusive 'Great' Recovery: Causes and Implications for Future Business Cycle Dynamics,* 60^{th} *Annual Economic Conference*, Board of Governors of the Federal Reserve System Speeches.

Zens, G., Böck, M. and Zörner, T. O. (2020). The Heterogeneous Impact of Monetary Policy on the US Labor Market, *Journal of Economic Dynamics and Control* **119**: 103989.

Appendices For Online Publication

Appendix A. Construction of forward-looking variables

In one of our robustness checks, we augment the baseline specification with forward-looking variables for GDP growth and HICP inflation (see Appendix B). These variables are constructed from the quarterly ECB staff macroeconomic projections. These projections are the natural choice among a variety of alternative sources that provide publicly available macroeconomic projections for the euro area (including for instance other policy institutions or private forecasters). While these projections are a staff-level exercise that does not necessarily have to fully match the Governing Council's assessment of the economic outlook, they are a key element of the Governing Council's information set.

We choose the last projection year as the reference point because we consider it as the most suitable approximation of the medium-term orientation: while the ECB has clarified that the definition of the "medium-term" may vary depending on the state of the economy and has refrained from defining it in terms of a specific time horizon, existing communication indicates that the outlook for the final projection year carries particular weight in its policy deliberations.

For most of the sample period, the horizon of the ECB projection vintages in year t stretches until year $t+2$. The ECB changed this convention in December 2016: since then, the forecast horizon for the fourth quarterly vintage of each year has extended to year $t+3$. For the sake of consistency, and given this change took place only after the end of our baseline sample, we abstract from it in constructing the forward-looking control variables. In summary, we hence calculate the forward-looking inflation variable as:

$$\pi^{e}_{t+2|t} = \frac{1}{4}(\pi^{e}_{t+2|MAR,t} + \pi^{e}_{t+2|JUN,t} + \pi^{e}_{t+2|SEP,t} + \pi^{e}_{t+2|DEC,t})$$

where $\pi^{e}_{t+2|t}$ is the expected inflation rate for year $t+2$ entering the central bank information set in year t, $\pi^{e}_{t+2|MAR,t}$ is the expected annual inflation rate in year $t+2$ according to the March ECB staff macroeconomic projections of year t, and $\pi^{e}_{t+2|JUN,t}$, $\pi^{e}_{t+2|SEP,t}$, and $\pi^{e}_{t+2|DEC,t}$ are the corresponding values according to the June, September, and December projections of year t. The forward-looking variable for the rate of real economic growth is calculated in analogous fashion.

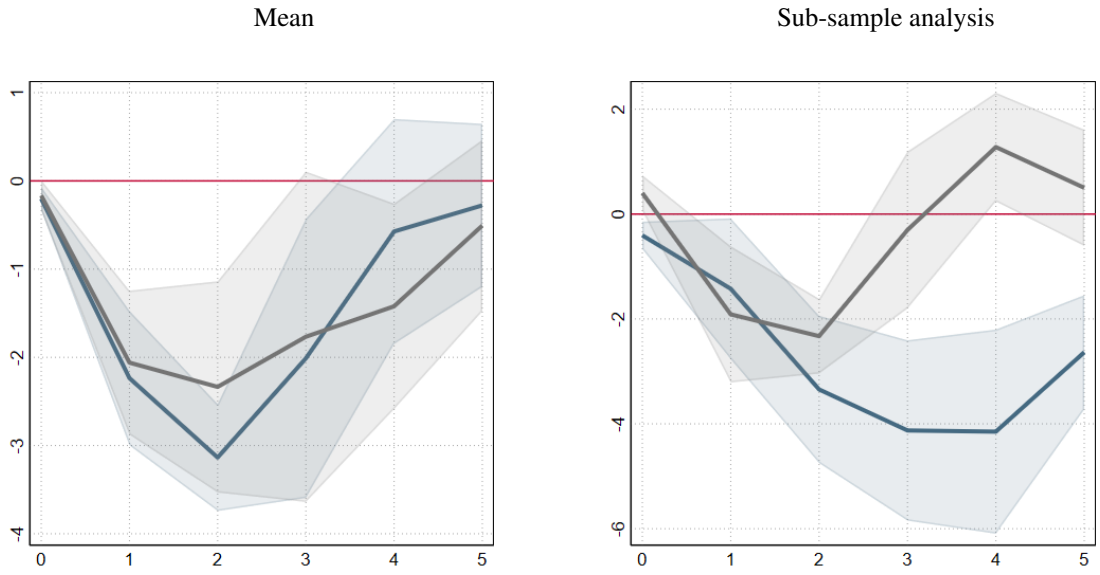
Appendix B. Robustness

Appendix B.1. Additional covariates

To check the robustness of our main findings, we first extend the list of controls in equation 1 with forward-looking variables that may shape the ECB's monetary policy deliberations in real-time, as proposed by Sims (1992) and Romer and Romer (2004). To this end, we rely on the quarterly ECB staff macroeconomic projections for real GDP growth and HICP inflation. As a rule, we use the final year of the respective projection horizon as the relevant reference point and compute the average across the four projection vintages of each year (see Appendix A for detail). The resultant estimates are very similar to our baseline (Figure B.16). For the sample mean, the confidence intervals of the two specifications overlap throughout the IRF horizon; and, for the sub-sample analysis, the alternative specification confirms the weaker and less persistent response of economic activity in the upper than in the lower parts of the distribution.

Our identification strategy rests on the assumption that, after controlling for macro-economic conditions, policy rates are independent of regional developments. As a second robustness check, we test this assumption and assess its relevance for our main conclusions. The analysis is motivated by Coibion and Goldstein (2012) who find that measures of economic dispersion across US states do emerge as relevant determinants of the Federal Reserve's interest

Figure B.16: Inclusion of expected HICP inflation and GDP growth



Notes: Vertical axis refers to impact of 100 basis point rate hike on regional GDP (in %). Horizontal axis refers to horizon of IRF (in years). Solid lines denote point estimates and shaded areas denote 90% confidence bands. LHS-panel shows mean estimates for the baseline (in grey) and the modified model (in blue). RHS-panel shows the corresponding estimates from the sub-sample analysis for the bottom ten percent (in blue) and top ninety percent (in grey) of the per-capita GDP distribution.

rate decisions in estimated Taylor rules. Following Coibion and Goldstein (2012), we alternate between three dispersion measures, which we include as additional controls in equation 1: *(i)* the population-weighted standard deviation in per-capita GDP; *(ii)* the difference between per-capita GDP in the 80^{th} and the 20^{th} percentile of the regional distribution; and *(iii)* the equivalent measure for the 90^{th} and 10^{th} percentile.[21] For the sample mean, all three specifications yield a similar a pattern as the baseline (Figure B.17). Likewise, the sub-sample analysis again confirms the uneven transmission of monetary policy to poorer and richer regions and, for the specification including the 90^{th} to 10^{th} percentile difference, the IRF for richer regions

[21] While Coibion and Goldstein (2012) base their dispersion measures on unemployment rates, the use of GDP-based measures appear preferable in our setting because GDP is the main dependent variable of interest also in the other parts of our analysis. In constructing the weighted standard deviation, we also experimented with GDP- instead of population-weights, but found the results to be unaffected by this choice. To lower the hurdle for rejecting our H_0 that regional dispersion does not affect policy rate setting, the regressions make use of the full time series dimension of the data, ranging from 1990 to 2018. Since euro area inflation and growth forecasts are unavailable for the pre-euro period, we estimate the policy rule based on realised values for these variables.

even turns insignificant.

Controlling for regional dispersion in the local projections

Figure B.17: Mean

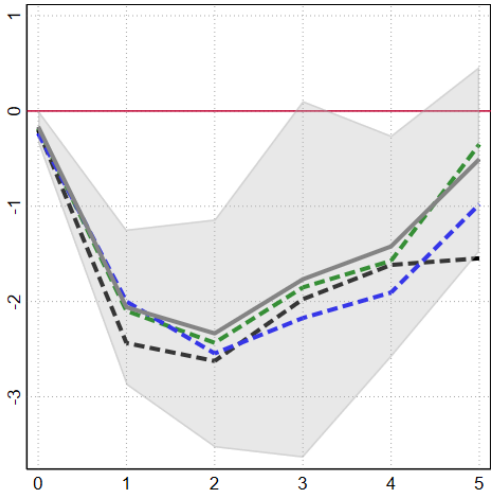


Figure B.18: Sub-sample analysis with population-weighted standard deviation

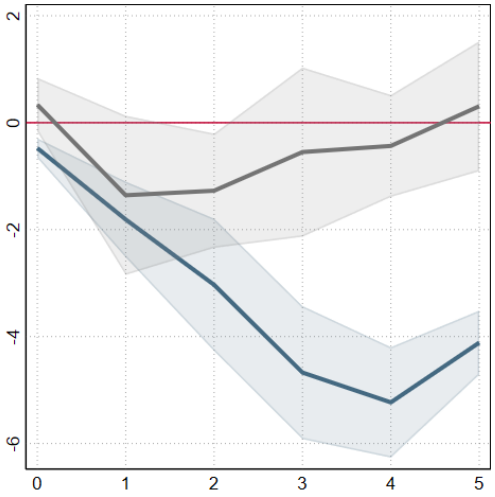


Figure B.19: Sub-sample analysis with 80^{th}-20^{th} percentile difference

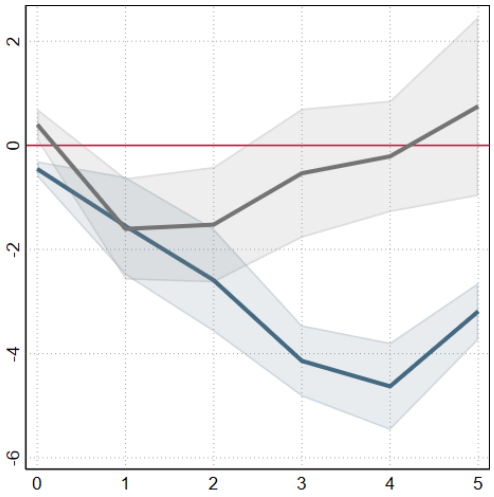
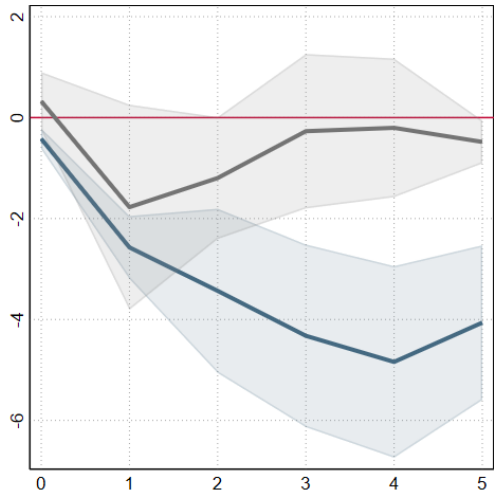


Figure B.20: Sub-sample analysis with 90^{th}-10^{th} percentile difference

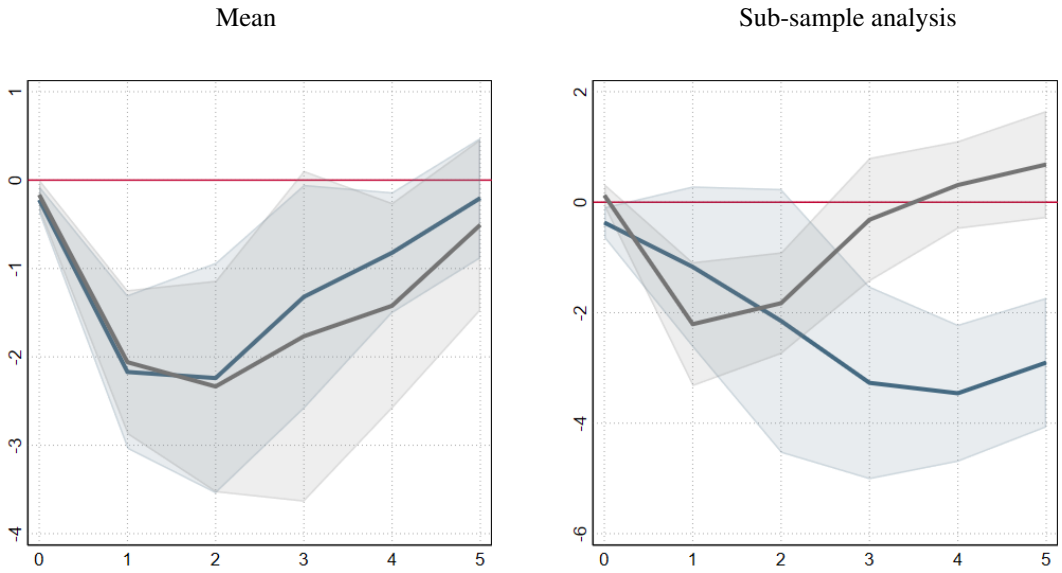


Notes: Upper LHS-panel shows the baseline and the specifications including the population-weighted standard deviation, 80^{th}-20^{th} percentile difference, and 90^{th}-10^{th} percentile difference in blue, green, and black, respectively. The other panels show the estimates from the sub-sample analysis for the bottom ten percent (in blue) and top ninety percent (in grey) of the per-capita GDP distribution for specifications including the respective regional dispersion measure as an additional control. For further detail, see notes to Figure B.16.

To guard against other sources of omitted variable bias,we run another robustness check including global oil prices and exchange rates against major trading partners among the controls. These variables routinely feature in ECB policy communication and they appear as potentially important controls at a disaggregated level where the typical spatial concentration of different

types of economic activity may render certain regions more and others less responsive to these types of shocks (see, *e.g.* House, Proebsting and Tesar (2020) for a recent analysis documenting differential effects of exchange rate fluctuations across US regions).[22] The main conclusions from our baseline estimations also carry over when including these additional regressors (Figure B.21). For the mean, the IRFs display very similar patterns up to horizon $h = 3$. After that point, the model including the additional variables yields a somewhat faster recovery of the output losses, but the difference to the baseline is again statistically insignificant. The sub-sample analysis, in turn, confirms the differential response across the GDP-distribution, characterized by a significantly stronger and more persistent contraction in economic activity in the lower than in the upper parts.

Figure B.21: Controlling for external factors



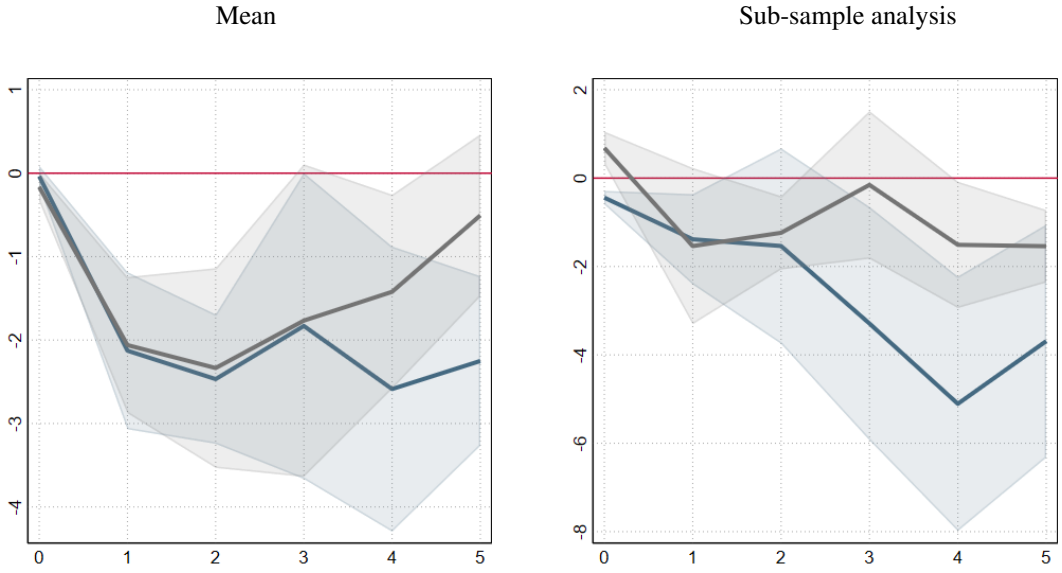
See notes to Figure B.16.

Finally, we account for serial correlation in the residuals of the (shadow) short-term interest rates (Figure B.22). As these residuals serve to identify exogenous policy changes, such serial

[22] Similar effects appear plausible also for the euro area; for instance, Lane and Stracca (2018) document heterogeneity in exchange rate pass through across euro area countries and, given the large differences in economic structures at the NUTS3 level, such heterogeneity is likely to arise also here.

correlation may cast doubt on their being unanticipated. We therefore follow Ramey (2016) in extending the set of explanatory variables by the lagged residual of the policy indicator. The resultant IRFs point to a similar contraction at the mean as in the baseline, albeit again with a somewhat greater persistence. The sub-sample analysis also confirms the pronounced heterogeneity across different parts of the distribution, although with some nuances compared to the baseline in that the IRF for the upper decile follows a slight jigsaw pattern and in the last projection year still exhibits a mild contraction.

Figure B.22: Inclusion of lagged residuals



See notes to Figure B.16.

Appendix B.2. Alternative sample periods

The results shown so far have been estimated on our baseline sample running from 1999 through 2014. The sample starting date reflects our focus on euro area monetary policy, which has been assigned to the ECB since then. The endpoint is motivated by a structural break in monetary policy conduct that arose in 2015. Until then, the ECB had primarily steered its monetary policy stance via short-term interest rates.[23] But with short-term interest rates

[23] While the ECB had also resorted to other, 'non-standard', monetary policy measures since 2010, these had mainly served to preserve the smooth transmission of short-term interest rate changes, rather than acting as inde-

approaching their lower bound, the ECB introduced a large-scale asset purchase program in January 2015, explicitly motivated as ensuring a further monetary policy easing (ECB, 2015). The presence of multiple stance-related policy instruments, in turn, is difficult to capture in our baseline empirical setup.

As part of the current set of robustness checks, we extend the sample period to include observations preceding the introduction of the euro and following the introduction of large-scale asset purchases. While this requires us to impose some additional assumptions in the construction of the main variables (see below), it yields two main benefits. First, it better aligns the estimation period with the concept of hysteresis, which generally refers to adjustments in the economy over a medium- to longer-term horizon. Second, an earlier starting point raises the number of observations dating back to the period prior to the Global Financial Crisis (GFC), which escalated in 2008 and may have triggered structural shifts in key economic relationships.

The starting point for the long sample is 1990, the year in which the then-participating countries entered into the process of monetary unification ('Stage One of EMU'). To ensure consistency with the previous research design, we back-cast some of the key variables in equation 1, including the policy-controlled short-term interest rate i_t, with which we measure the euro area monetary policy stance. For the years prior to 1999, when national central banks were still able to set their own policy rates, we approximate this euro-area wide policy stance with the GDP-weighted average of the respective short-term interest rates at national level.[24]

pendent instruments to alter the stance.

[24] Specifically, we use the nominal short-term interest rates from the AMECO database of the European Commission. The short-term interest rate is predominantly the three-month interbank rate, though this might vary slightly between countries during given years. As an alternative to the GDP-weighted average of national short-term interest rates, we also experimented with a back-casting procedure that simply uses the German short-term rate as the measure of the euro area monetary policy stance over the period prior to euro introduction – a choice that may be motivated with the benchmark role of German Bundesbank for the policy stance of other European

Likewise, we back-cast the euro area HICP with its GDP-weighted national average and euro area GDP with the sum of national GDPs over the pre-euro period. The latest sample endpoint we consider is 2019, just before the Covid-19 crisis struck. For the period beyond 2014, we replace the 3-month Euribor with a 'shadow' short-term interest over the last three years of the sample. As such, we follow a recent literature that proposes shadow rates as a proxy of the monetary policy stance in the presence of lower bound constraints and non-standard measures (Krippner, 2015; Wu and Xia, 2017; Lemke and Vladu, 2017).[25]

Based on these additional data, we test our baseline findings on four alternative sample periods: 1990-2019 to cover the full time-series dimension; 1990-2014 to exclude the quantitative-easing period starting in 2015; 1990-2007 to exclude the GFC and its aftermath; and 1999-2019 to raise the number of data-points since the start of the euro.

For all alternative sample definitions, we find the monetary policy tightening to trigger significant contraction in regional output on average, with GDP hitting its trough in the second or third year after the shock (Figure B.23). Compared to the euro-area sample period, shown by the dashed blue line, the contraction is slightly shallower in the initial parts of the IRF horizon, but then proves more persistent.[26] Further, the estimates over the longer sample again clearly document differential effects of monetary policy across the distribution, with regions in the lower parts experiencing a more pronounced and long-lasting contraction in response to a tightening shock (Figure B.24).

Finally, the estimates over the period 1999-2019 exhibit very similar patterns as in the base-

central banks in the run-up to euro introduction. But we found this alternative choice to yield almost identical results to that based on GDP-weighted averages.

[25] In terms of implementation, we rely on the shadow rate developed by Lemke and Vladu (2017) and add the cumulative changes in that rate to the 2014 level of the short-term interest rate deployed in our baseline estimations.

[26] While the point estimates remain statistically significant in $h = 5$, they converge back to zero when extending the IRF horizon for the long sample.

Figure B.23: Backcasted sample (1990-2014, blue) vs. pre-crisis sample (1990-2007, dashed grey) vs. long sample (1990-2019, dashed black) vs. baseline sample (1999-2014, dashed blue)

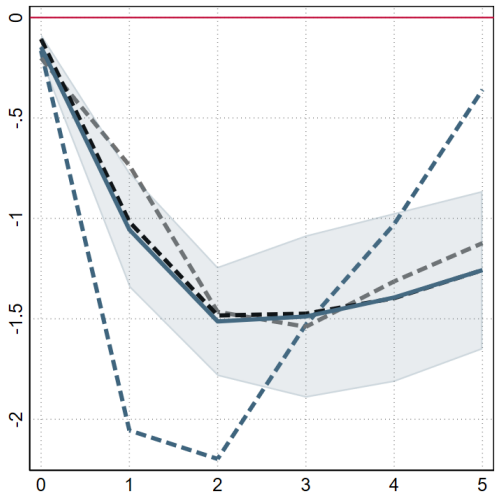


Figure B.24: Backcasted sample (1990-2014): bottom vs. top deciles for GDP

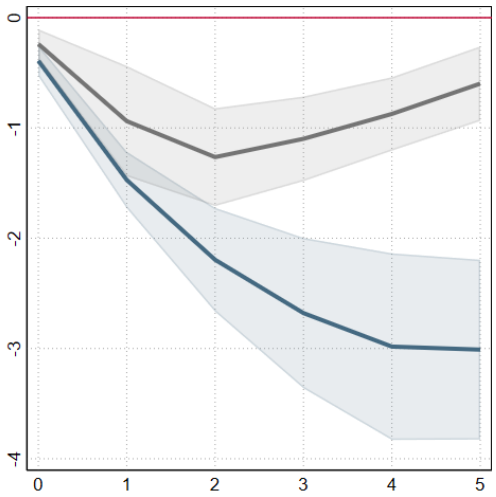


Figure B.25: Pre-crisis sample (1990-2007): bottom vs. top deciles

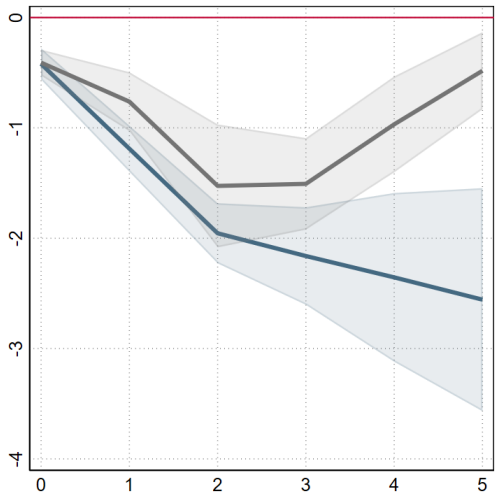
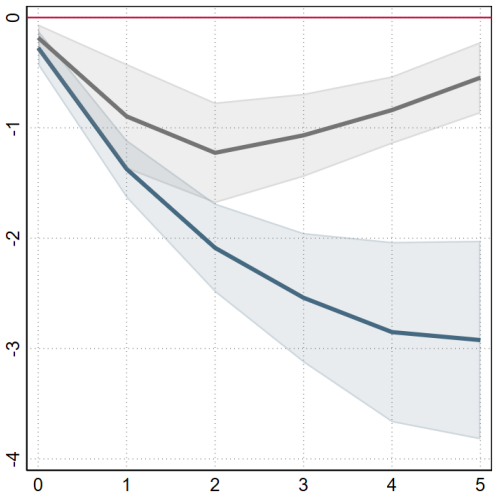


Figure B.26: Long sample (1990-2019): bottom vs. top deciles



Notes: Vertical axis refers to impact of a 100 basis point rate hike on the respective dependent variable (in %), horizontal axis refers to horizon of IRF (in years). In top-LHS, IRFs are for GDP during the backcasted period 1990-2014 (in blue), the pre-crisis period 1990-2007 (in dashed grey), the full sample period 1990-2019 (in black dashed) and the baseline period 1999-2014 (in dashed blue); in top-RHS, IRFs are for GDP during the baclcasted period 1990-2014 at the bottom (in blue) and top decile (in grey) of the per-capita GDP distribution. Bottom-LHS (RHS) show the corresponding IRFs for for the pre-crisis period (1990-2007) (long sample period, 1990-2019). Solid lines denote point estimates and shaded areas denote 90% confidence bands.

line and the confidence intervals overlap throughout the IRF horizon (Figure B.27). Moreover, the extended sample estimates again point to a heterogeneous footprint of monetary policy. Whereas richer regions recover from the monetary-policy induced GDP contraction after a few years, the impact on poorer regions proves persistent and still remains statistically significant by the end of the horizon (Figure B.28).

Figure B.27: Extended sample (1999-2019, blue) vs. baseline sample (1999-2014, grey)

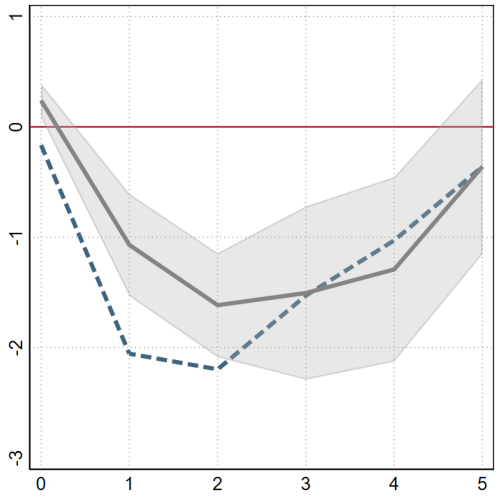
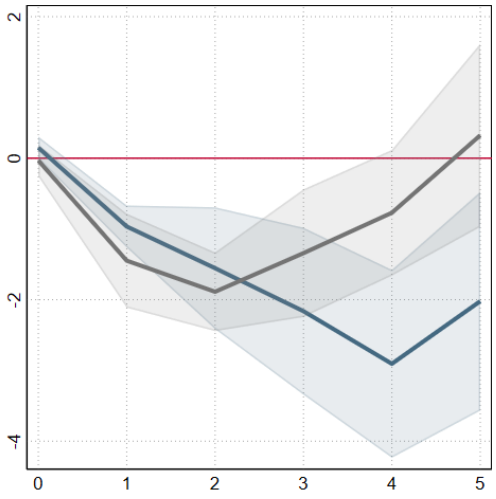


Figure B.28: Extended sample (1999-2019): bottom vs. top deciles for GDP



Notes: Vertical axis refers to impact of a 100 basis point rate hike on regional GDP (in %), horizontal axis refers to horizon of IRF (in years). On the LHS, grey line and shaded areas refer to the baseline sample period 1999-2014. Blue line and shaded areas refer to the sample period 1999-2019. Solid lines denote point estimates and shaded areas denote 90% confidence bands.

Appendix B.3. Quantile local projections

As an alternative method to analyse how the response of regional GDP to monetary policy shocks differs across the (conditional) GDP distribution, we also deployed quantile estimation techniques. The seminal approach proposed by Koenker and Bassett (1978) in principle provides a flexible way to model the conditional distribution of the dependent variable. At the same time, it is not ideally suited for panel data models with fixed effects when the cross-section is large relative to the time dimension, in which case the estimates are prone to incidental parameter problems (see, e.g., Lancaster, 2000).[27] To address this issue, we employ the quantile regression approach proposed by Machado and Silva (2019), which enables us to control for unobserved heterogeneity while estimating quantile-specific coefficients of the covariates in

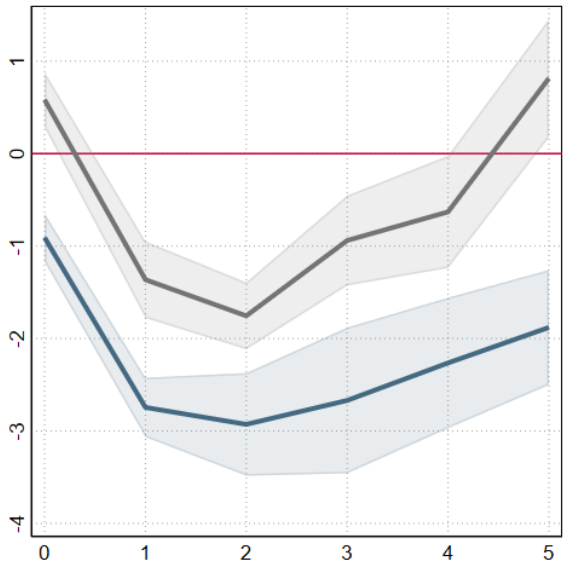
[27] A related issue arises with regard to the interpretation of the coefficients. In our application, the regions populating the upper and lower parts of the GDP distributions *conditional on the fixed effects* may be very different than those in the respective unconditional distributions. For instance, it is possible for relatively prosperous regions to temporarily end up in the bottom part of the distribution in years in which their GDP exhibits a transitory drop relative to its sample mean. However, our intention is to sort the cross-sectional units according to the more persistent aspect of regional inequality, which requires an alternative approach.

our model via location- and scale-functions.[28]

The quantile estimates confirm the pronounced heterogeneity in the regional patterns of monetary policy transmission. For the interdecile range, notable differences emerge in terms of both, the peak and the persistence of the impacts (Figure B.29). Output, in both parts of the distribution, hits its trough in the second year after the shock, but the point estimates point to a deeper contraction at the lower than at the upper decile. Moreover, this gap widens over the horizon, as output in the upper decile rebounds, whereas output in the lower decile recovers at a much slower pace and fails to return to its initial level by the end of the horizon. As such, the quantile estimates deliver patterns that are consistent with our baseline specification.

[28] In the quantile regressions, we again account for potential error correlation across space and time. To this end, we resort to two-way clustering, given the Driscoll-Kraay correction used for the mean regressions is not available for the quantile estimator. As suggested by Machado and Silva (2019), we also tested whether our conclusions may be biased due to the large ratio of cross-sectional units relative to time periods via the split-sample jackknife estimator developed by Dhaene and Jochmans (2015); this also left our results qualitatively unaffected and, if anything, led to somewhat more accentuated heterogeneity in the responses across regions.

Figure B.29: Impact of monetary policy on regional output: 10^{th} vs. 90^{th} percentile



Notes: Vertical axis refers to impact of 100 basis point rate hike on regional GDP (in %). Horizontal axis refers to horizon of IRF (in years). Solid lines denote point estimates and shaded areas denote 90% confidence bands.

Appendix C. Analytical results

Appendix C.1. Proof of proposition 1

Steady state output (in log-levels) is given by

$$y = \left(\frac{\phi}{1-\rho} + 1 - \alpha \right) n. \tag{C.1}$$

Hence

$$\frac{\partial y}{\partial \rho} = \frac{\phi}{(1-\rho)^2} n + \left(\frac{\phi}{1-\rho} + 1 - \alpha \right) \frac{\partial n}{\partial \rho}. \tag{C.2}$$

The first term is negative since $0 < N < 1$, which means $n < 0$. Thus, $\frac{\partial y}{\partial \rho}$ is unambiguously negative if $\frac{\partial n}{\partial \rho} < 0$.

Steady state employment is given by

$$\xi(1-N)^{-\varphi} = (1-\alpha)N^{\left(\frac{\phi}{1-\rho}+1-\alpha\right)(1-\sigma)-1}. \tag{C.3}$$

Taking logs

$$\log\xi - \varphi\log(1-N) = \log(1-\alpha) + \left(\left(\frac{\phi}{1-\rho}+1-\alpha\right)(1-\sigma)-1\right)\log N. \tag{C.4}$$

Total differentiation with respect to ρ yields

$$\varphi\frac{1}{1-N}\frac{\partial N}{\partial\rho} = \frac{\phi}{(1-\rho)^2}(1-\sigma)\log N + \left(\left(\frac{\phi}{1-\rho}+1-\alpha\right)(1-\sigma)-1\right)\frac{1}{N}\frac{\partial N}{\partial\rho}. \tag{C.5}$$

Rearranging gives

$$\frac{\partial N}{\partial\rho} = \frac{\frac{\phi}{(1-\rho)^2}(1-\sigma)\log N}{\varphi\frac{1}{1-N} - \left(\left(\frac{\phi}{1-\rho}+1-\alpha\right)(1-\sigma)-1\right)\frac{1}{N}}. \tag{C.6}$$

Since we imposed the condition $\alpha > \phi/(1-\rho)$ to ensure non-increasing returns to scale, the term $\left(\left(\frac{\phi}{1-\rho}+1-\alpha\right)(1-\sigma)-1\right)$ is negative and the denominator is positive. Thus, since $0 < N < 1$ (i.e., $\log N < 0$), we have that $\frac{\partial N}{\partial\rho} < 0$ so long as $\sigma < 1$. This completes the proof.

Appendix C.2. Proof of proposition 2

The labour market clearing and TFP equations in log-deviations from steady state are given by

$$0 = (1-\sigma)\hat{a}_t - \mathcal{A}\hat{n}_t, \tag{C.7}$$

$$\hat{a}_t = \rho\hat{a}_{t-1} + \phi\hat{n}_t, \tag{C.8}$$

where $\hat{x}_t \equiv x_t - x$ denotes log-deviations from steady state, $\mathcal{A} \equiv 1 + \tilde{\varphi} - (1-\alpha)(1-\sigma) > 0$ as long as $\sigma < 1$, and where $\tilde{\varphi} \equiv \frac{N}{1-N}\varphi$.

Correlation: Since $\hat{n}_t = \frac{1-\sigma}{\mathcal{A}}\hat{a}_t$, it follows that $\text{corr}(n_t, a_t) > 0$ as long as $\sigma < 1$.

Persistence: First, suppose that employment, $\hat{n}_t$, is perturbed by a non-TFP shock, u_t, such that $\hat{n}_t = \mathcal{B}\hat{a}_t + u_t$, where u_t follows an $AR(1)$ process $u_{t+1} = \theta u_t + \varepsilon_{t+1}$, where ε_{t+1} is an *iidN* innovation and $\mathcal{B} \equiv \frac{1-\sigma}{\mathcal{A}}$. Next, we can rewrite the equation for $\hat{a}_t$ and the AR(1) process using lag operators as $\hat{a}_t = \frac{\phi}{1-\rho L}\hat{n}_t$ and $u_t = \frac{1}{1-\theta L}\varepsilon_t$, respectively. Rewriting in terms of n_t, we get

$$n_t = \frac{\mathcal{B}\phi}{1-\rho L} n_t + \frac{1}{1-\theta L} e_t, \tag{C.9}$$

$$(1-\rho L)(1-\theta L) n_t = (1-\theta L)\mathcal{B}\phi n_t + (1-\rho L) e_t, \tag{C.10}$$

$$n_t = \frac{\rho + \theta - \theta\mathcal{B}\phi}{1-\mathcal{B}\phi} n_{t-1} - \frac{\rho\theta}{1-\mathcal{B}\phi} n_{t-2} + \tilde{e}_t - \rho\tilde{e}_{t-1}, \tag{C.11}$$

$$= s_1 n_{t-1} + s_2 n_{t-2} + \tilde{e}_t + c\tilde{e}_{t-1}, \tag{C.12}$$

where $s_1 \equiv \frac{\rho+\theta-\theta\mathcal{B}\phi}{1-\mathcal{B}\phi}$, $s_2 \equiv -\frac{\rho\theta}{1-\mathcal{B}\phi}$, and $c \equiv -\rho$. Thus, n_t follows an $ARMA(2,1)$ process.[29] Using standard results, it is straightforward to calculate the first-order autocorrelation of an $ARMA(2,1)$ process:

$$\text{ac}(\cdot) = \frac{(1 + c(s_1 + c))s_1 + \left(1 - s_2^2\right)c}{s_1(1+s_2)c + (1 + c(s_1+c))(1-s_2)}. \tag{C.13}$$

When $\rho = 0$, the autocorrelation of employments follows the exogenous process: $\text{ac}(n_t) = \theta$.

[29] Since the autocorrelation function is independent of the variance of ε_t, we have renormalized this variable and denoted it as $\tilde{\varepsilon}_t$ to avoid unnecessary terms.

Next we calculate the partial derivative with respect to ρ near $\rho = 0$:

$$\left.\frac{\partial ac(n_t)}{\partial \rho}\right|_{\rho=0} = \mathcal{B}\frac{\left(1-\theta^2\right)\phi}{1-\mathcal{B}\phi}. \tag{C.14}$$

Thus, $\left.\frac{\partial ac(n_t)}{\partial \rho}\right|_{\rho=0} > 0$ (i.e., employment becomes more persistent when ρ increases) if $\mathcal{B} > 0$, which requires $\sigma < 1$.

Using the equation for output, $y_t = a_t + (1-\alpha)n_t$, and labour productivity, $lp_t = y_t - n_t = a_t - \alpha n_t$, we can also write y_t and lp_t as $ARMA(2,1)$ processes. Specifically

$$y_t = \frac{\theta+\rho-\theta\mathcal{B}\phi}{1-\mathcal{B}\phi}y_{t-1} - \frac{\rho\theta}{1-\mathcal{B}\phi}y_{t-2} + \tilde{\varepsilon}_t - \frac{\rho}{1+\frac{\phi}{1-\alpha}}\tilde{\varepsilon}_{t-1}, \tag{C.15}$$

$$lp_t = \frac{\theta+\rho-\theta\mathcal{B}\phi}{1-\mathcal{B}\phi}lp_{t-1} - \frac{\rho\theta}{1-\mathcal{B}\phi}lp_{t-2} + \varepsilon_t - \frac{\rho}{1+\frac{\phi}{-\alpha}}\varepsilon_{t-1}. \tag{C.16}$$

In both cases, $\mathrm{ac}(y_t) = \mathrm{ac}(lp_t) = \theta$ when $\rho = 0$. The partial derivatives with respect to ρ near $\rho = 0$ are given by

$$\left.\frac{\partial ac(y_t)}{\partial \rho}\right|_{\rho=0} = \frac{1+\mathcal{B}(1-\alpha)}{1-\alpha+\phi}\frac{\left(1-\theta^2\right)\phi}{1-\mathcal{B}\phi}, \tag{C.17}$$

$$\left.\frac{\partial ac(lp_t)}{\partial \rho}\right|_{\rho=0} = \frac{1-\mathcal{B}\alpha}{-\alpha+\phi}\frac{\left(1-\theta^2\right)\phi}{1-\mathcal{B}\phi}, \tag{C.18}$$

$$\tag{C.19}$$

For output, the key term is $1+\mathcal{B}(1-\alpha)$. The requirement for $\left.\frac{\partial ac(y_t)}{\partial \rho}\right|_{\rho=0} > 0$ is not as strong as for employment. Only if $\mathcal{B}$ is sufficiently negative (caused by an extremly large σ) would the counterbalancing force of increased labour supply in response to falling TFP cause output hysteresis to fall as ρ increases. For labour productivity, the key term is $\frac{1-\mathcal{B}\alpha}{-\alpha+\phi}$.

Appendix C.3. Real Business Cycle (RBC) example

This example is similar to the one in the main text, with the key difference that we employ a real business cycle (RBC) and generate hysteresis by relating TFP to the level of region-wide investment spending.

Time, t, is discrete and a region i is populated by an infinitely lived representative household with preferences over consumption, c_t, given by $\mathbb{E}_t \sum_{t=0}^{\infty} \beta^t \log(c_t)$, where $\beta \in (0,1)$ is the discount factor. The household resource constraint is given by $c_t + i_t = y_t = A_t k_{t-1}^{\alpha}$, where A_t is total factor productivity (TFP), i_t is investment, k_t is the capital stock, and $\alpha \in (0,1)$. Capital evolves according to $k_t = (1-\delta) k_{t-1} + i_t$ although we restrict attention to the case of full capital depreciation in every period (in which case $\delta = 1$). TFP is assumed to be endogenous and increasing in aggregate (region-level) investment spending, I_t, which individual households take as given. Concretely, we assume the evolution of TFP is given by $A_t = A_{t-1}^{\rho_i} I_{t-1}^{\phi}$, where $\rho_i, \phi \in (0,1)$. Note that ρ_i is again the only region-specific structural parameter in the set-up of the model. All other parameters are common across regions. Finally, we impose the parameter restriction $\frac{\phi}{1-\rho_i} + \alpha < 1$ which ensures stationarity.

The first-order condition of the household is given by $1 = \beta \mathbb{E}_t (c_t / c_{t+1}) A_{t+1} \alpha k_t^{\alpha-1}$. In equilibrium, all household variables correspond to their aggregate values. Hence $x_t = X_t$ for $x_t = \{c_t, i_t, y_t, k_t\}$. Furthermore, it is straightforward to show via guess and verify that, in equilibrium, investment and consumption are fixed fractions ($\alpha\beta$ and $1 - \alpha\beta$, respectively) of output. Thus, the evolution of the economy is fully described by the joint dynamics of the capital stock, $\hat{k}_t = \log(\alpha\beta) + \rho_i \hat{a}_{t-1} + (\phi + \alpha) \hat{k}_{t-1}$, and TFP, $\hat{a}_t = \rho_i \hat{a}_{t-1} + \phi \hat{k}_{t-1}$, where $\hat{x} = \log(X)$. In log-deviations from steady state, $\tilde{k}_t = \rho_i \tilde{a}_{t-1} + (\phi + \alpha) \tilde{k}_{t-1}$ and $\tilde{a}_t = \rho_i \tilde{a}_{t-1} + \phi \tilde{k}_{t-1}$, where $\tilde{x}_t = \hat{x}_t - \hat{x}$.

We are now ready to present our two key propositions. The first concerns the steady state of

the model and the second concerns the persistence of the model dynamics when transitioning back to steady state.

Proposition 3 *Steady state output is decreasing in* ρ_i

Proposition 4 *The unconditional autocorrelation of output is increasing in* ρ_i

The two propositions, taken together, are consistent with the main finding of this paper that deviations of output are more persistent in Low income regions. Note that our model and the propositions that follow are very general. We do not need to specify the shock that drives the economy away from steady state.

Proof of Proposition 3. The steady state of TFP is given by $\hat{a} = \frac{\phi}{1-\rho_i}\hat{k}$. Substituting into the capital accumulation equation and rearranging gives $\hat{k} = \frac{1}{1-\alpha-\frac{\phi}{1-\rho_i}} \log(\alpha\beta)$. Substituting into the production function gives steady state output as $\hat{y} = \frac{\alpha(1-\rho_i)+\phi}{(1-\alpha)(1-\rho_i)-\phi} \log(\alpha\beta)$. It is straightforward to show that $\partial\hat{k}/\partial\rho_i, \partial\hat{y}/\partial\rho_i < 0$. ■

Proof of Proposition 4. First, rewrite the TFP evolution equation with lag operators as follows: $\tilde{a}_t = \frac{\phi}{(1-\rho_i L)}\tilde{k}_{t-1}$. Next, substitute this expression into the capital accumulation equation and expand so that it becomes an $AR(2)$ in $\tilde{k}_t$ only, as follows:

$$\begin{aligned}
\tilde{k}_t &= \frac{\phi}{(1-\rho_i L)}\tilde{k}_{t-1} + \alpha\tilde{k}_{t-1}, \\
(1-\rho_i L)\,\tilde{k}_t &= \phi\tilde{k}_{t-1} + (1-\rho_i L)\,\alpha\tilde{k}_{t-1}, \\
\tilde{k}_t &= (\rho_i + \phi + \alpha)\,\tilde{k}_{t-1} - \rho_i\alpha\tilde{k}_{t-2}.
\end{aligned}$$

A standard result of an $AR(2)$ process of the form $x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2}$ is a first-order autocorrelation given by $\text{ac}(x_t) \equiv \frac{\mathbb{E}(x_t x_{t-1})}{\mathbb{E}(x_t^2)} = \frac{\theta_1}{1-\theta_2}$. Hence, $\text{ac}(\tilde{k}_t) = \frac{\rho_i+\phi+\alpha}{1+\rho_i\alpha}$. It is straightforward to show that $\partial\text{ac}(\tilde{k}_t)/\partial\rho_i > 0$. Since $\tilde{y}_t = \tilde{k}_t$, it follows that $\partial\text{ac}(\tilde{y}_t)/\partial\rho_i > 0$. ■

Appendix D. A currency union of small-open economy regions

The model follows closely Gali and Monacelli (2008). The currency union consists of a continuum of small open economies ("regions") of unit mass. We assume the regions share the same structural parameters, except one—the persistence of endogenous total factor productivity changes. Furthermore, we assume the law-of-one-price holds across the union and that households can trade a complete set of state-contingent claims.

Households The representative household in region i has preferences given by

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{\left(C_t^i\right)^{1-\sigma}}{1-\sigma} + \xi \frac{\left(1-N_t^i\right)^{1-\varphi}}{1-\varphi}, \tag{D.1}$$

where N_t^i denotes hours worked and C_t^i is a composite consumption index, constructed of the following objects:

$$C_t^i \equiv \frac{\left(C_{i,t}^i\right)^{1-\omega}\left(C_{F,t}^i\right)^{\omega}}{(1-\omega)^{1-\omega}\omega^{\omega}}, \quad C_{F,t}^i \equiv \exp \int_0^1 \log\left(C_{f,t}^i\right) df, \quad C_{f,t}^i \equiv \left(\int_0^1 C_{f,t}^i(j)^{\frac{\varepsilon-1}{\varepsilon}} dj\right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{D.2}$$

where $C_{F,t}^i$ is a consumption index of imported goods and $C_{f,t}^i$ is an index of goods produced by region f and consumed by region i, where $j \in [0,1]$ denotes the type of good. The parameter ω is an index of openness (where decreasing ω reflects greater home bias), and $\varepsilon > 1$ is the elasticity of substitution between varieties in a given region.

The household maximizes (D.1) subject to the budget constraint given by

$$\int_0^1 P_t^i(j) C_{i,t}^i(j) dj + \int_0^1 \int_0^1 P_t^f(j) C_{f,t}^i(j) dj df + \mathbb{E}_t\left(Q_{t+1} D_{t+1}^i\right) = D_t^i + W_t^i N_t^i - T_t^i, \tag{D.3}$$

where $P_t^f(j)$ is the price of good j produced by region f; D_{t+1}^i denotes a random variable

indicating the number of assets purchased in period t to be delivered in each state of period $t+1$ and Q_{t+1} is the period t price of an asset that pays one unit in a particular state in $t+1$ divided by the time-t conditional probability of that state occurring. W_t^i are wages and T_t^i are lump-sum taxes and dividends (in nominal terms).

The optimal allocation of expenditure yields demand curves given by

$$\frac{C_{f,t}^i(j)}{C_{f,t}^i} = \left(\frac{P_t^f(j)}{P_t^f}\right)^{-\varepsilon}, \frac{C_{f,t}^i}{C_{F,t}^i} = \left(\frac{P_t^f}{P_t^*}\right)^{-1}, \tag{D.4}$$

where $P_t^f \equiv \left(\int_0^1 P_t^f(j)^{1-\varepsilon} dj\right)^{\frac{1}{1-\varepsilon}}$ is both country f's domestic price index and the price index for the goods imported from region f by country i, and $P_t^* \equiv \exp\int_0^1 \log\left(P_t^f\right) df$ is the union-wide price index. It follows that $P_t^f C_{f,t}^i = \int_0^1 P_t^f(j) C_{f,t}^i(j)\, dj$ and $P_t^* C_{F,t}^i = \int_0^1 P_t^f C_{f,t}^i df$. The optimal allocation of expenditures between domestic and imported good is given by

$$\frac{C_{i,t}^i}{C_t^i} = (1-\omega)\left(\frac{P_t^i}{P_{c,t}^i}\right)^{-1} \quad \text{and} \quad \frac{C_{F,t}^i}{C_t^i} = \omega\left(\frac{P_t^*}{P_{c,t}^i}\right)^{-1}, \tag{D.5}$$

where $P_{c,t}^i \equiv \left(P_t^i\right)^{1-\omega}(P_t^*)^{\omega}$ denotes the Consumer Price Index (CPI) for region i. Total consumption expenditure in region i is given by $P_{c,t}^i C_t^i = P_t^i C_{i,t}^i + P_t^* C_{F,t}^i$. Thus, the period budget constraint can be rewritten as

$$P_{c,t}^i C_t^i + \mathbb{E}_t\left(Q_{t+1} D_{t+1}^i\right) = D_t^i + W_t^i N_t^i - T_t^i. \tag{D.6}$$

The respective optimality conditions for households' labour/leisure and consumption/savings

decisions in region i are given by

$$\xi \left(C_t^i\right)^{\sigma} \left(1 - N_t^i\right)^{-\varphi} = \frac{W_t^i}{P_{c,t}^i}, \tag{D.7}$$

$$\beta \left(\frac{C_{t+1}^i}{C_t^i}\right)^{-\sigma} \frac{1}{\Pi_{c,t+1}^i} = Q_{t+1}, \tag{D.8}$$

where $\Pi_{c,t}^* = P_{c,t}^i / P_{c,t-1}^i$. Taking expectations of (D.8) gives the conventional Euler equation

$$\beta R_t^* \mathbb{E}_t \left(\frac{C_{t+1}^i}{C_t^i}\right)^{-\sigma} \frac{1}{\Pi_{c,t+1}^i} = 1, \tag{D.9}$$

where $R_t^* = 1/\mathbb{E}_t Q_{t+1}$.

Next, define the bilateral terms of trade between region i and f as $S_{f,t}^i \equiv P_t^f / P_t^i$ and the effective terms of trade for region i as $S_t^i = P_t^* / P_t^i = \exp \int_0^1 \log\left(S_{f,t}^i\right) df$. The CPI and domestic price levels are related by $P_{c,t}^i = P_t^i \left(S_t^i\right)^{\omega}$ and hence $\Pi_{c,t}^i = \Pi_t^i \left(\Delta S_t^i\right)^{\omega}$. For the currency union as a whole $P_{c,t}^* = P_t^*$ and hence $\Pi_{c,t}^* = \Pi_t^*$.

International risk sharing Under the assumption of complete markets for state-contingent securities, an analogous first-order condition to (D.8) for region f is given by

$$\beta \left(\frac{C_{t+1}^f}{C_t^f}\right)^{-\sigma} \frac{1}{\Pi_{c,t}^f} = Q_{t,t+1}. \tag{D.10}$$

Combining first-order conditions gives

$$\frac{C_t^i}{C_t^f} = \frac{C_{t+1}^i}{C_{t+1}^f} \frac{P_{c,t+1}^i}{P_{c,t+1}^f} \frac{P_{c,t}^i}{P_{c,t}^f}. \tag{D.11}$$

This has the form $x_t = x_{t+1} y_{t+1} / y_t$. Rearranging gives $x_t = x_{t-1} y_{t-1} / y_t$. Iterating backwards

yields $x_t = x_0 y_0 / y_t$. Hence, we can rewrite (D.11) as

$$\frac{C_t^i}{C_t^f} = \nu_{i,f} \frac{P_{c,t}^f}{P_{c,t}^i} = \nu_{i,f} \left(S_{f,t}^i \right)^{1-\omega}. \tag{D.12}$$

By assuming symmetric initial conditions, $\nu_{i,f} = \nu = 1$. Integrating over f, we obtain

$$C_t^i = C_t^* \left(S_t^i \right)^{1-\omega}, \tag{D.13}$$

where $C_t^* \equiv \exp \int_0^1 \log C_t^f \, df$.

Firms Each region has a unit continuum of firms, each producing a differentiated good with $Y_t^i(j) = A_t^i \left(N_t^i(j) \right)^{1-\alpha}$ where A_t^i is region-specific. Real marginal cost is given by

$$MC_t^i(j) = \left(\frac{1-\tau^i}{1-\alpha} \right) \frac{W_t^i}{P_t^i} \frac{1}{A_t^i \left(N_t^i(j) \right)^{-\alpha}}. \tag{D.14}$$

where τ^i is a constant employment subsidy.

The aggregate output index for region i is given by

$$Y_t^i \equiv \left[\int_0^1 Y_t^i(j)^{\frac{\varepsilon-1}{\varepsilon}} \, dj \right]^{\frac{\varepsilon}{\varepsilon-1}}, \tag{D.15}$$

and the amount of labour hired is given by

$$N_t^i \equiv \int_0^1 N_t^i(j) \, dj = \left(\frac{Y_t^i}{A_t^i} \right)^{\frac{1}{1-\alpha}} \int_0^1 \left(\frac{P_t^i(j)}{P_t} \right)^{-\frac{\varepsilon}{1-\alpha}} dj. \tag{D.16}$$

Price setting Firms set prices as in Calvo (1983), where $1-\theta$ is the probability that a firms can set a new price. The fact that each individual region is open does not affect the form of the equation relating domestic inflation to real marginal cost. As such, the dynamics of domestic

inflation in terms of real marginal costs in region i is given by

$$\hat{\pi}_t^i = \beta \mathbb{E}_t \hat{\pi}_{t+1}^i + \lambda \hat{mc}_t^i, \tag{D.17}$$

where $\hat{mc}_t^i = mc_t^i - \mu$ denotes the log-deviation of real marginal costs from its steady state, and $\lambda = \frac{(1-\beta\theta)(1-\theta)}{\theta}\Theta$, where $\Theta \equiv \frac{1-\alpha}{1-\alpha+\alpha\varepsilon}$.

Aggregate dynamics The clearing of the market for good j in region i is given by

$$\begin{aligned}
Y_t^i(j) &= C_{i,t}^i(j) + \int_0^1 C_{f,t}^i(j)\,df \\
&= \left(\frac{P_t^i(j)}{P_t^i}\right)^{-\varepsilon}\left(C_{i,t}^i + \int_0^1 C_{i,t}^f df\right), \\
&= \left(\frac{P_t^i(j)}{P_t^i}\right)^{-\varepsilon}\left((1-\omega)\left(\frac{P_t^i}{P_{c,t}^i}\right)^{-1} C_t^i + \omega\int_0^1 \left(\frac{P_t^i}{P_{c,t}^f}\right)^{-1} C_t^f df\right), \\
&= \left(\frac{P_t^i(j)}{P_t^i}\right)^{-\varepsilon}\left((1-\omega)\left(S_t^i\right)^{\omega} C_t^i + \omega\int_0^1 \left(S_t^i\right)^{\omega}\left(S_{f,t}^i\right)^{1-\omega} C_t^f df\right), \\
&= \left(\frac{P_t^i(j)}{P_t^i}\right)^{-\varepsilon} C_t^i\left(S_t^i\right)^{\omega},
\end{aligned} \tag{D.18}$$

and hence the market clearing condition is $Y_t^i = C_t^i\left(S_t^i\right)^{\omega}$.

Endogenous total factor productivity Finally, we arrive at the key feature that distinguishes regions. Denote $U_t^i = 1 - N_t^i$ as the unemployment rate. We assume that total factor productivity (TFP) accumulation follows $A_t^i = \left(A_{t-1}^i\right)^{\rho^i}\left(1-U_t^i\right)^{\phi}$, where A_t^i is increasing in past TFP and decreasing in the unemployment rate. In other words, unemployment causes a de-accumulation of skills or human capital.

Equilibrium dynamics Next, we rewrite the model in log-deviations, where $x \equiv \log(X)$ and

$\hat{x}_t \equiv x_t - x$. The labour supply condition and Euler equation are then

$$\sigma \hat{c}_t^i + \tilde{\varphi} \hat{n}_t^i = \hat{w}_t^i - \hat{p}_{c,t}^i, \tag{D.19}$$

$$\hat{c}_t^i = \mathbb{E}_t \hat{c}_{t+1}^i - \frac{1}{\sigma} \left(\hat{r}_t^* - \mathbb{E}_t \pi_{t+1}^i \right), \tag{D.20}$$

where $\tilde{\varphi} \equiv \varphi \frac{N^i}{1-N^i}$. The CPI and domestic price levels and inflation rates are related by

$$\hat{p}_{c,t}^i = \hat{p}_t^i + \omega \hat{s}_t^i, \quad \hat{\pi}_{c,t}^i = \hat{\pi}_t^i + \omega \Delta \hat{s}_t^i. \tag{D.21}$$

International risk sharing ensures

$$\hat{c}_t^i = \hat{c}_t^* + (1-\omega) \hat{s}_t^i. \tag{D.22}$$

Equilibrium variations in $\int_0^1 f\left(P_t^i(j)/P_t^i\right) dj$ are of second-order. Thus, up to a first-order approximation, the following relationships hold:

$$\hat{mc}_t^i = \hat{w}_t^i - \hat{p}_t^i - \hat{a}_t^i + \alpha \hat{n}_t^i, \tag{D.23}$$

$$\hat{y}_t^i = \hat{a}_t^i + (1-\alpha) \hat{n}_t^i. \tag{D.24}$$

The market clearing condition is

$$\hat{y}_t^i = \hat{c}_t^i + \omega s_t^i = \hat{c}_t^* - \left(\hat{p}_t^i - \hat{p}_t^* \right). \tag{D.25}$$

Integrating over (D.25) yields

$$\hat{y}_t^* = \hat{c}_t^*. \tag{D.26}$$

Thus, the union-wide IS curve is given by

$$\hat{y}_t^* = \mathbb{E}_t \hat{y}_{t+1}^* - \frac{1}{\sigma}\left(\hat{r}_t^* - \mathbb{E}_t \pi_{t+1}^*\right). \tag{D.27}$$

Marginal cost can be rewritten as follows:

$$\begin{aligned}
\hat{mc}_t^i &= \hat{w}_t^i - \hat{p}_t^i - \hat{a}_t^i + \alpha \hat{n}_t^i, \\
&= \left(\hat{w}_t^i - \hat{p}_{c,t}^i\right) + \left(\hat{p}_{c,t}^i - \hat{p}_t^i\right) - \hat{a}_t^i + \alpha \hat{n}_t^i, \\
&= \sigma \hat{c}_t^i + \tilde{\varphi} \hat{n}_t^i + \omega s_t^i - \hat{a}_t^i + \alpha \hat{n}_t^i, \\
&= \sigma\left(\hat{y}_t^i - \omega \hat{s}_t^i\right) + \left(\tilde{\varphi} + \alpha\right) \hat{n}_t^i + \omega \hat{s}_t^i - \hat{a}_t^i, \\
&= \sigma\left(\hat{y}_t^i - \omega \hat{s}_t^i\right) + \left(\frac{\tilde{\varphi} + \alpha}{1 - \alpha}\right)\left(\hat{y}^i - \hat{a}_t^i\right) + \omega \hat{s}_t^i - \hat{a}_t^i, \\
&= \left(\sigma + \frac{\tilde{\varphi} + \alpha}{1 - \alpha}\right) \hat{y}^i - \left(\frac{1 + \tilde{\varphi}}{1 - \alpha}\right) \hat{a}_t^i + (1 - \sigma)\, \omega \hat{s}_t^i.
\end{aligned} \tag{D.28}$$

Hence, the Phillips curve is given by

$$\hat{\pi}_t^i = \beta \mathbb{E}_t \hat{\pi}_{t+1}^i + \kappa_y \hat{y}_t^i - \kappa_a \hat{a}_t^i + \kappa_s \hat{s}_t^i, \tag{D.29}$$

where $\kappa_y \equiv \lambda\left(\sigma + \frac{\tilde{\varphi}+\alpha}{1-\alpha}\right)$, $\kappa_a \equiv \lambda\left(\frac{1+\tilde{\varphi}}{1-\alpha}\right)$ and $\kappa_s \equiv \lambda(1-\sigma)\,\omega$.

Monetary policy Monetary policy is described by a simple Taylor-type rule

$$\hat{r}_t^* = \phi \hat{\pi}_t^* + u_t, \tag{D.30}$$

where $u_{t+1} = \rho u_t + \sigma_u \varepsilon_{t+1}$ and ε_{t+1} is drawn from an *iid* standard normal distribution.

Equilibrium system Suppose there are two types of region: High (H) and Low (L) income regions, with mass w and $1 - w$, respectively. The full set of 13 endogenous variables are:

$\left\{\hat{\pi}_t^i, \hat{y}_t^i, \hat{a}_t^i, \hat{n}_t^i, \hat{s}_t^i\right\}$ and $\{\hat{r}_t^*, \hat{\pi}_t^*, \hat{y}_t^*\}$. The full system of 13 equations is given by:

1. Regional equations for $i = H, L$:

$$\pi_t^i = \beta \mathbb{E}_t \pi_{t+1}^i + \kappa_y^i \hat{y}_t^i - \kappa_a^i \hat{a}_t^i + \kappa_s \hat{s}_t^i, \tag{D.31}$$

$$\hat{y}_t^i = \hat{y}_t^* + \hat{s}_t^i, \tag{D.32}$$

$$\hat{y}_t^i = \hat{a}_t^i + (1-\alpha)\,\hat{n}_t^i, \tag{D.33}$$

$$\hat{a}_t^i = \rho^i \hat{a}_{t-1}^i + \phi \hat{n}_t^i. \tag{D.34}$$

2. Definition

$$\Delta \hat{s}_t^H = \hat{\pi}_t^* - \hat{\pi}_t^H. \tag{D.35}$$

3. Aggregation equations

$$\hat{\pi}_t^* = w \hat{\pi}_t^H + (1-w)\,\hat{\pi}_t^L, \tag{D.36}$$

$$\hat{y}_t^* = w \hat{y}_t^H + (1-w)\,\hat{y}_t^L. \tag{D.37}$$

4. Union wide equation

$$\hat{y}_t^* = \mathbb{E}_t \hat{y}_{t+1}^* - \frac{1}{\sigma}\left(\hat{r}_t^* - \mathbb{E}_t \pi_{t+1}^*\right). \tag{D.38}$$

$$\hat{r}_t^* = \phi_\pi \hat{\pi}_t^* + \phi_y \hat{y}_t^* + u_t. \tag{D.39}$$

Steady state We impose zero-inflation steady state $\Pi_c^i = \Pi = 1$ and set the employment

subsidy, τ^i to ensure the steady state is not distorted.[30] Then:

$$A^i = \left(N^i\right)^{\frac{\phi}{1-\rho^i}}, \tag{D.40}$$

$$\xi \left(C^i\right)^{\sigma} \left(1-N^i\right)^{-\varphi} = (1-\alpha) A^i \left(N^i\right)^{-\alpha}, \tag{D.41}$$

$$Y^i = C^i, \tag{D.42}$$

$$Y^i = A^i \left(N^i\right)^{1-\alpha}. \tag{D.43}$$

Combining these steady-state conditions, we obtain a non-linear expression in N^i:

$$\left(N^i\right)^{1+\left(\frac{\phi}{1-\rho^i}-\alpha\right)(\sigma-1)} \left(1-N^i\right)^{-\varphi} = \frac{1-\alpha}{\xi}. \tag{D.44}$$

Calibration The calibration is presented in Table D.2. The subjective discount factor, $\beta = 0.99$, implies a steady state annual real interest rate of 4%. The intertemporal elasticity of substitution (IES), $1/\sigma = 2$. An IES greater than 1 ensures that the wealth effect of changes in TFP does not dominate the substitution effect on labour supply of a change in the real wage. Hence, the effect of endogenously falling TFP is exacerbated by an additional fall in employment. For the high-income region, we calibrate the model such that the steady state unemployment rate is 5% ($1 - N^H = 0.05$) and the Frisch elasticity of labour supply, $\tilde{\varphi}^H = 1$. This fixes the preference parameters φ and ξ. The labour share of income, $1 - \alpha = 0.9$. This parameter is set to less than one, despite production being absent capital or other variable inputs, because this ensures that labour productivity is time-varying as in the data. The elasticity of substitution across goods (in a given region), $\varepsilon = 3.2$, implies a steady state mark up of 1.45. The Calvo parameter, $\theta = 0.72$, implies that firms adjusts prices on average every 3.5 quarters. House-

[30] Since we do not study optimal monetary policy, this subsidy has no impact on the equilibrium dynamics.

holds are assumed to have a home bias in consumption, with $\omega = 0.25$. We assume that there are only two types of region (high- and low-income regions), with an equal fraction of both ($w = 0.5$). Moreover, we assume one region type experiences hysteresis ($\rho^L = 0.925$) whereas the other does not ($\rho^H = 0$). Note that while we label these regions high- and low-income, these designations are an endogenous outcome of the model. The unemployment elasticity of TFP, $\phi = 0.1$, implies that a 1 percentage point rise in the unemployment rate decreases TFP by 10 basis points.

Table D.2: Calibration

Parameter	Description	Value
β	Subjective discount factor	0.99
$1/\sigma$	Intertemporal elasticity of substitution	2
φ	Utility curvature on leisure	0.053
ξ	Utility weight on leisure	0.810
$1-\alpha$	Labour share of income	0.9
ε	Elasticity of substitution	3.2
θ	Calvo parameter	0.72
ω	Home bias	0.25
w	Share of High income regions	0.5
ϕ	Unemployment elasticity of TFP	0.1
ρ^H	TFP persistence (High income region)	0
ρ^L	TFP persistence (Low income region)	0.925
ϕ_π	Monetary policy inflation coefficient	1.2
ϕ_y	Monetary policy output coefficient	0.125
ρ_u	Monetary policy shock persistence	0.5

Numerical steady state Given the calibration presented above, the steady state outcomes of the model are presented in Table D.3. Steady state unemployment in the high-income region is calibrated to be 5.0%. The consequences of hysteresis is that the Low income region has a higher unemployment rate (at 7.5%) and the income of a high-income region is 13% higher than a low-income region.

Table D.3: Steady state

Moment	Y^H/Y^L	U^H	U^L
Value	1.13	5.0%	7.5%