The Geographic Effects of Monetary Policy Shocks *

Juan Herreño — UC San Diego
Mathieu Pedemonte — IADB

This Version: January 22, 2025
First Version: March 2020

Abstract

We estimate the effects of monetary policy shocks across local areas in the US and find substantial variation in their responses. There is a positive covariance of the price and employment effects of monetary policy across regions, and more sensitive regions are those with low per capita income. These patterns are consistent with New Keynesian models of a monetary union where regions have different shares of hand-to-mouth consumers. The model predicts that monetary policy shocks create large differences in consumption and real wages across space, and that heterogeneity across local areas amplifies the aggregate responses to shifts in monetary policy rates.

JEL: E31, E24, E52, E58, F45
Keywords: Heterogeneous Effects of Monetary Policy, Monetary Union, TANK.

*Herreño: jherrenolopera@ucsd.edu. Pedemonte: mathieupe@iadb.org. We thank Hassan Afrouzi, Mark Bils, Corina Boar, Olivier Coibion, Laura Gáti, Yuriy Gorodnichenko, Joe Hazell, James Hamilton, Nir Jaimovich, Ed Knotek, Emi Nakamura, Christina Patterson, Valerie Ramey, David Romer, Sanjay Singh, Jón Steinsson, Fabian Trottner, Nicolas Vincent, Michael Weber, Johannes Wieland, Ines Xavier, and seminar participants at various institutions for useful comments and discussions. We thank Carolina Celis, Michael McMain and Grant Rosenberger for excellent research assistance.

1 Introduction

This paper estimates how the transmission of monetary policy shocks to prices and employment differs across metropolitan areas in the United States and evaluates plausible drivers of economic heterogeneity that can explain our findings.

We use exogenous variation in the stance of monetary policy since 1969, using the Romer and Romer (2004) shocks, extended to 2007 by Wieland and Yang (2020)[1], and a panel of US metropolitan areas. Our analysis uses regionally disaggregated data for employment and consumer prices in the United States. For prices we use the Consumer Price Index (CPI) data for metropolitan areas where the Bureau of Labor Statistics makes data available, and for employment we use the Quarterly Census of Employment and Wages (QCEW) to generate private employment counts.

After a contractionary monetary policy shock, inflation and employment in the US decline but do so at different rates across metropolitan areas, contrary to what textbook models would suggest. Crucially, metropolitan areas that experience larger price declines are the same metropolitan areas that experience larger employment losses. Highly sensitive areas to monetary policy shocks are those with lower average household earnings. These results hold for a variety of consumer expenditure categories, different sources of shocks, and are larger for non-tradeable goods.

Studying the differential effects of monetary policy disruptions across regions requires estimates of the effects on both prices and real quantities to distinguish demand and supply drivers of heterogeneity. Theories that predict heterogeneity in the slope of local Phillips curves across industries or firm types sorted in space predict a negative covariance between price and quantity responses across regions: after a shift in nominal interest rates, prices will adjust by *more* and quantities will react by *less* in regions with steeper supply curves since they are closer to monetary neutrality. Theories that predict heterogeneity in the slope of local demand curves predict a positive covariance between

[1] In Appendix A.7 we consider alternative monetary policy shock series developed by Bu et al. (2021) and Miranda-Agrippino and Ricco (2021). These shocks have different time coverage and, depending on the case, exclude the Volcker disinflation, include the Great Recession and periods with a binding zero lower bound. In material available upon request, we also use the extension of the Romer and Romer (2004) shocks by Acosta (2023).

price and quantity responses across regions: after a shift in nominal interest rates, prices will adjust by *more*, and quantities will react by *more*: in these regions, monetary policy is more powerful and creates larger changes in consumption, output, and real marginal costs, which through a common slope of the Phillips curve, induces larger price effects.

As a pedagogical device, we present a model that speaks to the patterns in the data. Regions in a monetary union are characterized by a differential fraction of hand-to-mouth households, different degrees of price rigidity, and different labor supply elasticities. Our model is a monetary union extension of the Two-Agent New Keynesian (TANK) model in Bilbiie (2008) with additional margins of heterogeneity. Regions with different shares of hand-to-mouth households have differential sensitivities of regional consumption to local real interest rates, and non-Ricardian households may only smooth consumption via their labor supply decisions.

We illustrate that this simple model can reproduce the qualitative regional patterns we estimate in the data with variation in the share of hand-to-mouth households but not with variation in the extent of nominal rigidities. Heterogeneity in demand and supply curve predicts a covariance between price and employment responses of opposite signs. In regions with a higher marginal propensity to consume, quantities react by more, and through the Phillips curve, they generate larger price responses since real marginal costs respond by more to the monetary policy shock even when the slope of the Phillips curve is the same across regions.

In the model, monetary policy has relevant geographical distributional effects in the short run. Contractionary monetary policy shocks induce larger drops in price inflation and employment in regions with a higher share of hand-to-mouth consumers. On top of that, it generates an even larger heterogeneity in consumption and real wages across regions. Local areas with more Ricardian agents can smooth their consumption by importing goods produced in areas with a higher share of hand-to-mouth consumers, which are net exporters. In areas with a higher percentage of hand-to-mouth consumers, real wages drop by more, inducing demand amplification that reduces consumption in equilibrium.

To make the model and data comparable, we use the insight of Patterson (2019), who documents that income is a crucial covariate to explain marginal propensities to consume using data from the United States. Since income is an important determinant of MPCs for which we have available data at the local level and frequency, we compute local projections of employment and prices after monetary policy shocks and decompose them into two determinants: an average effect and a heterogeneous effect by income level at the metropolitan area level. This approach is similar to that advocated by Cloyne et al. (2020b). After a common monetary policy shock, low-income metropolitan areas exhibit larger price and larger employment responses. Metropolitan areas in the bottom 10th percentile of the geographical income distribution face peak employment losses of 2.0 percent after a tightening of 100 basis points. Regions in the top 10th percentile suffer negligible effects after the same shock. The differential effects we estimate are persistent; employment remains depressed for four years after the occurrence of the shock.

Concerning prices, a 100-basis point tightening causes cumulative price responses in metropolitan areas in the 10th percentile of the income distribution to be 50 percent larger compared to the average responses and 50 percent smaller compared to the average effect in regions in the 90th percentile of the income distribution. As a validation exercise, we use CPI data disaggregated by expenditure categories and find consistent results. We find that the prices of goods and services of a wide range of narrow categories react less in high-income areas compared to low-income areas. The differential effects are larger for expenditure categories priced locally, like food away from home, and the differential effects on inflation across metropolitan areas are smaller for highly traded, homogeneous goods, like gasoline. The differential price responses for these highly traded categories are statistically insignificant when we use conservative standard errors.

Our findings do not require us to take a strong stance on the structural driver of marginal propensities to consume across regions with different income levels. Households in regions with lower productivity levels may be closer to their subsistence, or industries with differential sensitivity to national interest rates may sort across space. We illustrate the role of industrial composition by controlling for regional industry employment shares interacted with the monetary policy shock, to illustrate the role of income on top of its importance for factors that are correlated with industrial composition. We show that income is still an important driver of the results. We also include additional demographic and economic controls, like the age structure, debt-to-income ratios and the share of labor income in total household income interacted with the monetary policy shocks, which do not qualitatively affect our results.

We use the TANK model to evaluate the aggregate effects of having regions with different shares of hand-to-mouth households. We find that heterogeneity in the percentage of hand-to-mouth households exacerbates the aggregate effects of monetary policy shocks. The origin of the amplification effect of demand shocks as a function of the share of hand-to-mouth consumers is similar to that explained in Bilbiie (2020). An increase in inequality across space in the US increases the aggregate impact of shifts in the stance of monetary policy on both prices and employment.

To make the connection between our model and empirical analysis tighter we connect our model where the main margin of heterogeneity is the MPC, and our empirical heterogeneous effects by income. We use the Current Population Survey to compute the average income by metropolitan area and Patterson (2019) estimates to back-out the average marginal propensity to consume at the local level. We use our model to back out a share of hand-to-mouth consumers per metropolitan area consistent with the data.

With a cross-sectional measure of MPC heterogeneity across space from external sources at hand, and calibrated parameter values from the literature, we simulate the local and national effects of shifts in the stance of monetary policy when regions are heterogeneous, compared with a monetary union with the same national share of hand-to-mouth consumers and homogeneous regions. We find that the effects on employment are 36% greater and the effects on prices 29% greater in the heterogeneous economy. The same monetary policy shock induces heterogeneous effects on prices, employment, consumption, and real wages. In fact, the consumption responses are even more dispersed than the employment responses since less affected regions import goods from abroad, so more sensitive areas become net exporters.

In Appendix A.7, we present additional identification and aggregation results. We show that the derivatives of the local impulse response functions imply that after a local increase in the dispersion of hand-to-mouth consumers across regions, there is a relative amplification of the employment and price responses of the region with more non-

Ricardian agents compared the region with fewer non-Ricardian agents. We also show that the derivatives of the local projections at the national level imply that an increase in dispersion of hand-to-mouth consumers amplifies the response of monetary policy.

Finally, in the same A.7, we compare the aggregate effects of monetary policy across models with different sources of heterogeneity, named the share of hand-to-mouth consumers, the elasticity of labor supply, the intertemporal elasticity of substitution, and the slope of the Phillips curve. We show that heterogeneity in the share of hand-to-mouth consumers consistently increases the aggregate effects of monetary policy for a wide range of monetary policy reaction functions. None of the other sources of heterogeneity create this effect.

This paper estimates the differential impact of shifts in the stance of monetary policy across metropolitan areas. We illustrate the differential behavior of aggregate local variables to a common monetary policy shock, building on the literature on heterogeneity at the individual-level. The responses of local prices, employment, and consumption illustrate monetary policy transmission across space. The insight we use is that due to the imperfect mobility of labor and local consumption patterns, individual heterogeneity in exposure to monetary policy shocks is aggregated at the local level. A minimal extension of a textbook model predicts that prices and employment react by less in regions with lower marginal propensities to consume, in line with the causal estimates we provide, and predicts that consumption heterogeneity will be even larger due to the amplification of demand shocks at the local level. Moreover, we show that geographic heterogeneity amplifies aggregate responses. A polarized geographic income distribution exacerbates the effects of monetary policy shocks.

Literature Review

This paper is part of a growing literature seeking to understand the distributional effects of monetary policy and its implications. On the empirical front the studies closer to ours are Carlino and Defina (1998) and Neville et al. (2012) that find heterogeneous effects of changes in interest rates across US census regions using VARs. We study the effects of monetary policy on both prices and employment across metropolitan areas using disruptions in the stance monetary policy recovered using narrative methods. Coibion et al. (2017) show that monetary policy affects the distribution of nominal income, and Furceri et al. (2018) find similar effects for a panel of countries. Cloyne et al. (2020a) document heterogeneous effects of monetary policy shocks across households as a function of the financial position of households. Cravino et al. (2018) focus on the heterogeneity of price adjustment as a driver of differential responses of inflation rates across income groups. Andersen et al. (2021) document the effects of monetary policy on several sub-components of income triggered by monetary policy shocks that induce increases in inequality after expansionary shocks. We provide a new data moment, the covariance between price and quantity effects across local geographic areas, to distinguish across competing mechanisms of heterogeneity.

Our work is particularly related to Almgren et al. (2022) who study the heterogeneous effects of monetary policy in the Euro Area on quantities, like output and consumption. They find that countries with a higher share of hand-to-mouth have a higher cumulative drop in output after a monetary policy shock. Different from their work, we highlight that comparing the differential effects of monetary policy shocks on quantities and prices at the same time is key to distinguishing alternative drivers of heterogeneity. We find a positive covariance of the effects of monetary policy shocks on prices and employment, suggesting that heterogeneity in demand effects is a preferred mechanism. Without information about the differential response of prices after a monetary policy shock, it is possible that heterogeneous effects are driven by variations in the slope of the supply curve.

Russ, Shambaugh and Singh (2023) explore heterogeneous sensitivity to macroeconomic variables at the county level. They find persistent differences in county unemployment sensitivity to the aggregate business cycle. Our results complement theirs; we find heterogeneous effects in prices and employment conditional to monetary policy shocks while they study unconditional differences in sensitivity.

Bergman et al. (2022) look at different demographics affected by a monetary policy shock. They find that groups with lower labor market attachment have higher employment growth after expansionary monetary policy shocks when the market is tighter. Using a New Keynesian model with heterogeneous workers, they show that this effect is plausible when there are differences in workers' productivity. In this paper, we focus on the spatial income heterogeneity of the US. This heterogeneity allows us to evaluate not only the effect on employment but also on price indexes. Having employment and prices allows us to have a complete picture of the effects in terms of real income.

The distributional effects of monetary policy and its consequences have been studied in theoretical models. Auclert (2019) and Kaplan et al. (2018) focus on how heterogeneity may change the average effects of monetary policy. Bilbiie (2008) presents a two-agent New-Keynesian model in which hand-to-mouth consumers introduce frictions in determining aggregate quantities. We use a framework similar to that in Bilbiie (2008), extending it to a monetary union with heterogeneity in the presence of hand-to-mouth consumers, and we show that this class of models can rationalize the cross-regional heterogeneous responses of monetary policy shocks in the US.

Households and firms in the economy are affected by aggregate fluctuations differentially as a function of their earnings, balance-sheet positions, or ability to access financial instruments, and households of different characteristics are sorted through space.[2] Imperfect mobility of goods and factors, may amplify or dampen the local effects of aggregate shocks. Local labor markets will fence-in local general equilibrium effects, using the language of Mian et al. (2022), allowing us to measure the importance of the differential effects.

The rest of the paper proceeds in the following way: Section 2 presents the data. Section 3 shows that regions with larger price responses also face larger employment responses to a monetary policy shock. Section 4 presents a monetary union New Keynesian model to illustrate the effect of different drivers of heterogeneity on the relation between price and employment effects. Section 5 assesses empirically the effect of differences in MPCs through income in driving differential impacts of monetary policy shocks. Section 6 shows the implications of monetary policy for geographic inequality according to the model. Section 7 concludes.

[2]For the case of exposure to monetary policy, see Coibion et al. (2017) exploring differences in income inequality; Beraja et al. (2019) and Wong (2021) exploring heterogeneity in balance-sheet positions. See also Doepke and Schneider (2006).

2 Data

To estimate the effects of monetary policy shocks across space, we estimate impulse response functions of inflation and employment using regional data via local projections after a monetary policy shock. We construct a balanced panel for 28 metropolitan areas containing 12-month inflation rates and indicators of real economic activity. Our dataset starts in 1969 and ends in 2007, a restriction of using the Romer and Romer (2004) monetary policy shocks.[3] We use headline CPI inflation as our benchmark and present results for various sub-indexes, including CPI for food, food at home, food away from home, gas, and housing.

Price index data come directly from the Bureau of Labor Statistics (BLS). For our study, the dispersion of economic conditions across space is essential. For that reason, we choose to use city-wide indexes instead of state-wide indexes, such as those produced by Hazell et al. (2022) in order to have more variation in average economic conditions across units of observation. In addition, we will use price indexes for specific consumer categories to illustrate whether our results are heterogeneous by the degree of tradeability, product differentiation, or the degree of nominal rigidities across expenditure categories.

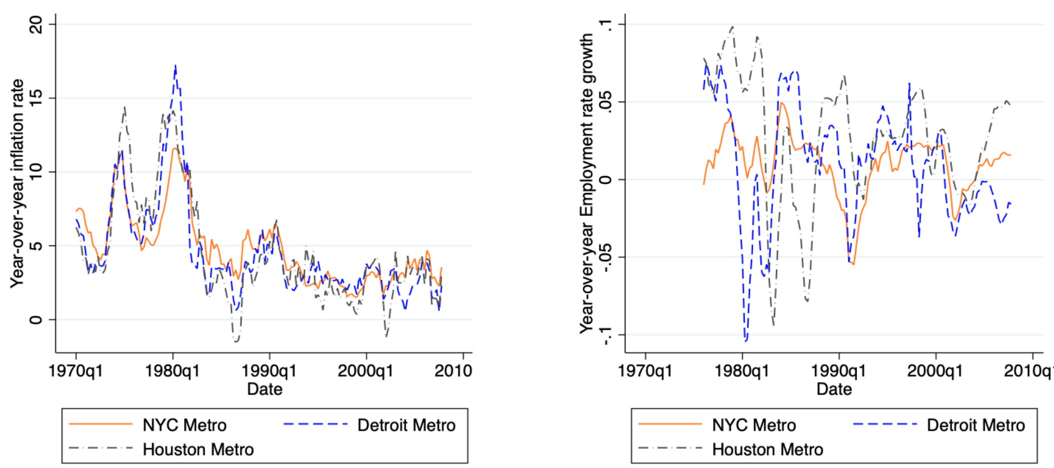
Our main specification focuses on cross-sectional variation across cities, by exploiting interactions of the monetary shock with city characteristics, after controlling for the direct effect of the shock. To illustrate the variation that we will use, we plot the headline CPI inflation for three selected metropolitan areas in the United States, New York-Newark-Jersey City, NY-NJ-PA (area code S12A in the CPI data), the Detroit-Warren-Dearborn, MI (area code S23B), and Houston-The Woodlands-Sugar Land, TX (area code S37B). Figure 1 presents the data, and it is only meant as an illustration of the dispersion in inflation

[3] The metropolitan areas we consider are Boston-Cambridge-Newton (MA-NH), New York-Newark-Jersey City (NY-NJ-PA), Philadelphia-Camden-Wilmington (PA-NJ-DE-MD), Chicago-Naperville-Elgin (IL-IN-WI), Detroit-Warren-Dearborn (MI), Minneapolis-St.Paul-Bloomington (MN-WI), St. Louis (MO-IL), Washington-Arlington-Alexandria (DC-MD-VA-WV), Baltimore-Columbia-Towson (MD), Miami-Fort Lauderdale-West Palm Beach (FL), Atlanta-Sandy Springs-Roswell (GA), Tampa-St. Petersburg-Clearwater (FL), Dallas-Fort Worth-Arlington (TX), Houston-The Woodlands-Sugar Land (TX), Phoenix-Mesa-Scottsdale (AZ), Denver-Aurora-Lakewood (CO), Los Angeles-Long Beach-Anaheim (CA), San Francisco-Oakland-Hayward (CA), Seattle-Tacoma-Bellevue (WA), San Diego-Carlsbad (CA), Urban Hawaii, Urban Alaska, Pittsburgh (PA), Cincinnati-Hamilton (OH-KY-IN), Cleveland-Akron (OH), Milwaukee-Racine (WI), Portland-Salem (OR-WA) and Kansas City (MO-KS).

rates and employment growth rates that is not common across metropolitan areas. The main source of variation we will use is the differential inflation rates that metropolitan areas experienced throughout US business cycles. For example, the Houston metro area experienced a higher inflation rate during the Great Inflation of 1974, the Detroit metro area experienced a higher inflation rate during the 1979 inflation, and both had more pronounced changes in inflation during the 2001 recession compared to New York City.

The employment data come from the Quarterly Census of Employment and Wages (QCEW), which has good geographical coverage. We use county-level data at the quarterly frequency covering private employment since 1975. Since the unit of observation for the employment data is the county, and for prices is the metropolitan area, we create a correspondence between counties in the QCEW and the statistical sampling units created for the CPI, called Primary Sampling Units (PSUs).[4]

Figure 1: Inflation and Employment Across Metropolitan Areas



(a) Inflation

(b) Employment growth

Note: This figure plots the behavior of inflation and employment for three metropolitan areas: New York-Newark-Jersey City, NY-NJ-PA; Detroit-Warren-Dearborn, MI; Houston-The Woodlands-Sugar Land, TX. The top panel shows 12-month headline CPI inflation. The bottom panel shows 12-month employment growth rates at a quarterly frequency.

In a similar way to the treatment we will give to prices in our main regression speci-

[4]Table A.1 in Appendix A.2 shows the correspondence between PSUs in the Price data and the FIPS codes in the QCEW data.

fication, our main employment specifications will soak up any effects on symmetric employment responses triggered by monetary policy shocks. The right panel of Figure 1 illustrates the differential local area business cycles of three metropolitan areas as a matter of example. Houston experienced an employment boom during the early 2000s and a differential employment loss during the late 1980s. Similarly, the Volcker disinflation hit Detroit by more than New York.

We use the Romer and Romer (2004) shocks, extended to 2007 by Wieland and Yang (2020), as our measure of monetary policy shocks.[5] We aggregate monthly shocks at the quarterly frequency. These shocks capture monetary policy changes that are free from the anticipation effects of prices and economic activity inherent to monetary policy decisions. Figure A.1 in the appendix displays the time series of the shock we use. Most of the variation in the Romer and Romer (2004) measure of monetary policy shocks comes from the Volcker disinflation, as pointed out by Coibion (2012). Since the Great Recession, the US policy rule has often been limited by the zero lower bound, which limits the sample period we consider, although we consider robustness to other shocks that use data after the Great Recession.

3 Empirical Strategy and Results

In this section, we present our empirical strategy to estimate the causal effect on prices and employment of a monetary policy shock across US metropolitan areas and our estimation results. Our core identification strategy relies on using exogenous shifts to the stance of monetary policy in the United States measured by the Romer and Romer (2004) shocks. We will identify the dynamic causal effects of monetary policy shocks on both employment and prices using local projections with lagged dependent variables as controls (Jorda, 2005; Montiel Olea and Plagborg-Møller, 2021).

The main result of this section comes from running local projections on prices and employment of each individual metropolitan area in the US and showing non-parametrically

[5]Our results are robust to extending further in the 2010s, using, for example, the extension of the Romer and Romer (2004) shocks done by Acosta (2023). Extending the sample with a cost of losing a sample of cities, as the publicly available sample of cities reduced from 28 to 15 in 2007. Because of that, the analysis of this paper goes up to 2007.

that regions in which prices are more sensitive to monetary policy shocks are the same areas where employment is more sensitive to the same shocks. Theories that attach heterogeneity in structural parameters to different regions must confront this fact.

3.1 Prices

We start by estimating the effects of national changes in monetary policy on prices for the average metropolitan area. For a given price index in location i, $\pi_{i,t+h,t-1}$ denotes the cumulative inflation rate between a reference period $t-1$ and $h>0$ periods in the future as

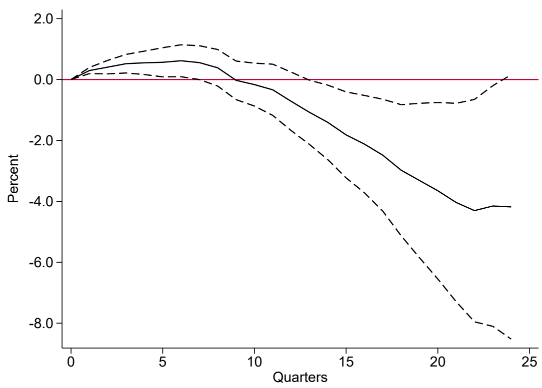
$$\pi_{i,t+h,t-1} = \frac{P_{i,t+h} - P_{i,t-1}}{P_{i,t-1}}.$$

To estimate the effect of a monetary policy shock on prices in the average metropolitan area, we use local projections (Jorda, 2005) method with area fixed effects, formally we run the following set of regressions

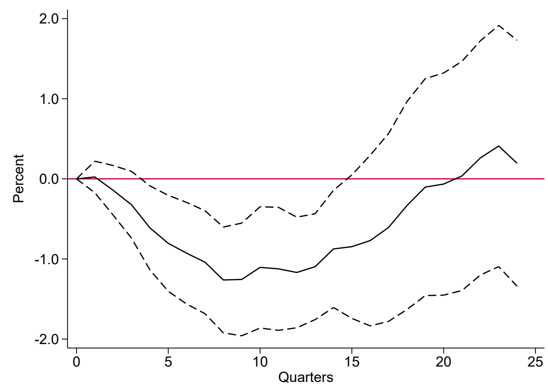
$$\pi_{i,t+h,t-1} = \alpha^h_{p,i} + \sum_{j=0}^{J} \beta^{h,j}_p RR_{t-j} + \sum_{k=0}^{K} \gamma^{h,k}_p \pi_{i,t-1,t-1-k} + \varepsilon^h_{p,i,t+h} \;\; \forall h \in [0,H], \tag{1}$$

where i indexes metropolitan areas, t indexes time, h denotes the number of quarters after the shock, and p denotes that these coefficients and error terms belong to a price regression. The coefficient $\beta^{h,j}_p$ accounts for the cumulative effect of a monetary policy shock j periods ago RR_{t-j}, on inflation $\pi_{i,t+h,t-1}$ h periods in the future. $\alpha^h_{p,i}$ is a metropolitan area fixed effect in the price regression, and $\varepsilon^h_{p,i,t+h}$ is the error term. We cluster standard errors at the metro area and time level. This specification is a panel version of the lag-augmented local projections as in Montiel Olea and Plagborg-Møller (2021).

The terms $\beta^{h,0}$ in equation 1 trace the cumulative impulse response function on prices at horizon h after a monetary policy shock, controlling for permanent city-specific inflation differences, past shocks, and differential time-varying inflation dynamics before the shock. Figure 2a shows the estimated cumulative impulse response function of overall CPI inflation or, equivalently, the impulse response of prices after a monetary policy shock that tightens rates by 1 percentage point.



(a) CPI



(b) Employment

Figure 2: Average Effects of Monetary Policy Shocks on Prices and Employment

Note: The left panel of the figure plots the estimated coefficients of equation (1) for the panel of metropolitan areas. We compute the local projections up to a maximum horizon of $H = 24$, and use eight lags of the dependent variable and the monetary policy shocks as controls ($J = 8$, and $K = 8$). The solid line denotes the estimated coefficients, and the dashed lines represent 90 percent confidence intervals. Standard errors are clustered at the metro area and date level. The right panel of this figure plots the estimated coefficients of equation (2). We use the same values for H, J, K than in the left panel.

Our results are similar to the original Romer and Romer (2004) results obtained by running a regression of national CPI inflation on the monetary policy shock and controls at the aggregate level. The effect of a monetary policy shock on prices is positive and close to zero for the first two years, followed by a sharp decline, reaching a value of -6 percentage points after 20 quarters. Both the point estimate and the standard errors are similar to those obtained using aggregate data.

The conceptual difference between the impulse response functions depicted in figure 2a and the results that would arise from a local projection over aggregate inflation numbers is a difference in weights. In order to compute aggregate inflation, the Bureau of Labor Statistics uses population weights over regional inflation indexes. Instead, our calculations use equal weights over regions. In that sense, our results measure the effect of monetary policy shocks for the average city.

By clustering our standard errors by metropolitan areas, our standard errors also contain information about the heterogeneity in the intensity of the effect of the treatment. In subsequent sections of the paper, we will exploit differences in observable characteristics across metropolitan areas to document heterogeneity in the effects of monetary policy shocks. Before we do so, we document the average effects of monetary policy shocks on employment growth.

3.2 Economic Activity

Linking the empirical evidence on heterogeneity to economic mechanisms of underlying heterogeneity in the class of New Keynesian models requires estimating not only the effects on prices but also the effects on real economic activity. Due to data availability, we focus on employment at the local level.

We run a specification qualitatively similar to equation (1), but with the percentage change of private employment, which we denote by g^e as the dependent variable, given by

$$g^e_{i,t+h,t-1} = \alpha^h_i + \sum_{j=0}^{J} \beta^{h,j}_e RR_{t-j} + \sum_{k=0}^{K} \gamma^{h,k}_e g^e_{i,t,t-k} + \varepsilon^h_{e,i,t+h} \;\; \forall h \in [0, H], \tag{2}$$

where $g^e_{i,t+h,t}$ is the cumulative employment growth in metropolitan area i between time $t-1$ and $t+h$. The rest of the notation is the same as that of equation 1, and the subscript e makes reference to the employment regression.

By estimating $\beta^{h,0}_e$ in equation 2, we trace the average cumulative impulse response function of private employment at different horizons in the average US metropolitan area after a monetary policy shock that tightens rates by one percentage point.

After a monetary policy tightening, there is a negative effect on employment. This effect occurs faster than the effect on prices: After five quarters, we estimate an employment drop that persists for 10 quarters. This effect is significant; the maximum cumulative effect reaches a 1 percent decrease in private employment.

3.3 Metropolitan Area Results

We run local projections for each individual metropolitan area instead of pooling them in a panel specification, to illustrate non-parametrically whether there is comovement in the response of inflation and employment across space.

The comovement of employment and price effects will be informative about the nature of the source of heterogeneity. Heterogeneity in the slopes of the supply block of the model will create a negative comovement of inflation and price responses, while heterogeneity in the demand block of the model will create a positive comovement between price and employment responses.

We run local projections at the individual level on local inflation and local employment growth but allow for arbitrary impulse response functions for each particular city instead of pooling the results together. For prices, the specification we consider takes the form of

$$\pi_{i,t+h,t-1} = \alpha_{0,p} + \sum_{j=0}^{J} \beta_{i,p}^{h,j} RR_{t-j} + \sum_{k=0}^{K} \gamma_{i,p}^{h,k} \pi_{i,t-1,t-1-k} + \varepsilon_{p,i,t+h}^{h} \;\; \forall h \in [0, H], i \in \mathcal{I}, \tag{3}$$

while that of employment takes the following form

$$g_{i,t+h,t-1}^{e} = \alpha_{0,e} + \sum_{j=0}^{J} \beta_{i,e}^{h,j} RR_{t-j} + \sum_{k=0}^{K} \gamma_{i,e}^{h,k} g_{i,t,t-k}^{e} + \varepsilon_{e,i,t+h}^{h} \;\; \forall h \in [0, H] i \in \mathcal{I}, \tag{4}$$

where $\alpha_{0,p}$ and $\alpha_{0,e}$ denote the intercepts of the price and employment equations, respectively, and the β and γ coefficients have the same interpretation as in the previous subsections, with the clarification that they are city-specific coefficients, which we clarify with the i subscript. $\mathcal{I}$ denotes the set of metropolitan areas for which we have data.

The identifying assumption behind equations 3 and 4 is more demanding than the traditional identifying assumption behind local projections with aggregate data. The key added restriction is that the Romer and Romer shocks not only clean anticipation effects of inflation and economic conditions with respect to aggregate variables, but they do so with respect to local variables as well. A violation of this restriction would occur if, for example, the FOMC were more concerned about economic conditions in some regions rather than others. In section 5.2, we run robustness exercises using other sources of shocks.

To present our results, we take the approach suggested by Ramey (2016) of computing ratios of the cumulative responses to summarize the effect of a shock. In particular, we add up the effects on employment 20 quarters after the onset of the shock. For prices, we add up the effects on inflation up until quarter 20.[6]

Figure 3 illustrates the comovement of the impulse responses 20 quarters after the shock for each metropolitan area. The x-axis plots the effects on prices, while the y-axis plots the effects on employment. Each bubble corresponds to one metropolitan area. Cities with larger price effects also exhibit larger employment effects. In Appendix A.4, we show that this positive relationship is statistically significant and remains positive after we consider the underlying standard errors of each point estimate. Additionally, in Figure A.2, in Appendix A.1, we show the same figure with the city names for interested readers.

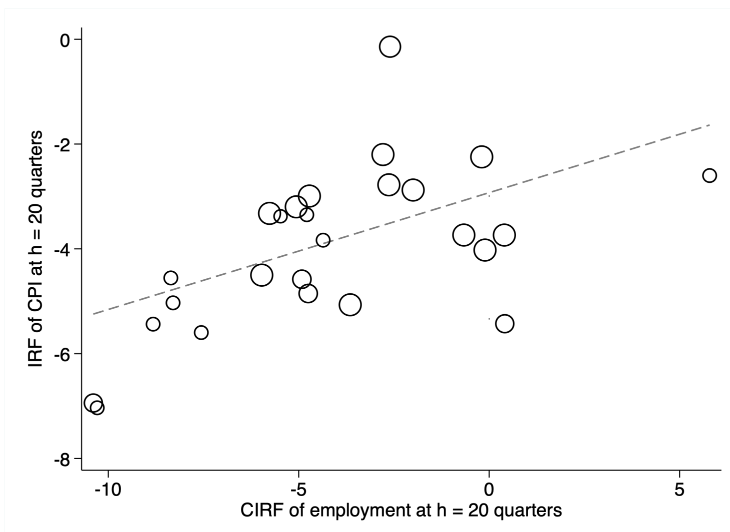


Figure 3: Effect of a Monetary Policy Shock in Employment and Prices for Each City

Note: This figure plots on the y-axis the local projection on local consumer prices of an exogenous monetary policy tightening of 100 basis points 20 quarters after the shock. The x-axis plots the cumulative effect (area under the curve) of local employment 20 quarters after a monetary policy shock of 100 basis points. The units of both axes are percentage points. Each bubble in the scatter plot corresponds to a metropolitan area. The size of each bubble represents the average income per capita of each metropolitan area.

We will use the results of Figure 3 in order to inform the magnitude of the margins of economic heterogeneity that rationalize the heterogeneous responses of local economic

[6] In the case of employment, we compute the total changes of employment relative to the initial employment before the shock. The monetary policy shock is expected to produce a temporary effect on city employment, so we obtain the cumulative IRF. For prices, we take the change after 20 quarters. The shock is expected to produce temporary effects on inflation, leading to a permanent effect on the price level. The empirical measures the effect in the price level, so we obtain the IRF.

conditions to a common monetary policy shock.

In Appendix A.4, we conduct a number of exercises to show that the patterns in Figure 3 are statistically significant. We refer the reader to the details of the appendix, but we summarize the highlights here.

First, we use the standard errors associated with each point in Figure 3 to do a simulation-based exercise in which we perturb the points in Figure 3 and re-estimate its slope. Figure A.12 shows that in 99.6 percent of our simulations, the estimated slope is positive.

Second, we impose a restriction in the system of local projections in order to estimate the slope that rationalizes the data and estimate this slope directly from the microdata.[7] Our exercise is similar in spirit to estimate a Phillips multiplier in the language of Barnichon and Mesters (2021) in a cross-section of regions. Figure A.13 presents the results for different horizons of the impulse responses. The estimate has the interpretation of the reaction of prices to a one percent cumulative effect on employment growth triggered by a monetary policy shock. We find a positive and significant slope coefficient with standard errors clustered by metropolitan area and time. Figure A.13 also shows that our estimate for the slope and its statistical significance is robust to the maximum horizon we use in the computation of the IRFs.

4 Monetary Union TANK Model

The purpose of this section is to present a parsimonious New Keynesian model with as few departures from textbook models as possible that is flexible enough to generate heterogeneity in responses across regions after a monetary policy shock in line with those documented in the data.

In the model, regions are local labor markets without any degree of mobility. Households have standard preferences, although the intertemporal elasticity of substitution and the elasticity of labor supply may vary across space. There is a share of hand-to-mouth households in each region, and this share may change across space. There are firms in each region that produce differentiated varieties subject to region-specific Calvo (1983) frictions.

[7]We thank Jim Hamilton for suggesting this approach.

The model captures other unmodelled margins of heterogeneity insofar as these enter the problem either by changing the sensitivity of local consumption growth to local real interest rates, the sensitivity of producer price inflation to local real marginal costs, or both.

We document that heterogeneity in demand factors, like the differential share of hand-to-mouth consumers, can rationalize our results. Heterogeneity in supply factors, like the heterogeneity in the extent of nominal rigidities, cannot.

4.1 Model Environment

We first present a model of a monetary union in which monetary policy shocks induce differential regional responses. The model has a large tradition in macroeconomics; it is an extension of the TANK model (Bilbiie, 2008) to a monetary union.

The model has two regions: Home (H) and Foreign (F). Each region has two types of households: Ricardian (R) and hand-to-mouth (H) households. Each region is characterized by a differential share of each household type. Aguiar et al. (2020), documents the determinants of being a hand-to-mouth consumer. Heterogeneity in the share of hand-to-mouth consumers will induce differential sensitivity of consumption growth to changes in local real interest rates.

On the supply side, we assume that in principle, the Calvo (1983) parameter could be heterogeneous across regions. On top of the slope of the Phillips curve being different, the forcing variable itself, local real marginal costs, may behave differently as well due to labor immobility across regions, home bias in consumer preferences, and variation in the share of hand-to-mouth households.

Home and Foreign regions are equal in population, an assumption that is not important but reduces notation. The Home region (H) is populated by both Ricardian (HR) and hand-to-mouth households (HH). The share of hand-to-mouth agents in the Home and Foreign regions is denoted by λ_H and λ_F, respectively. Ricardian and hand-to-mouth households in the same region have the same preferences and supply homogeneous labor. Ricardian households save, and own firms and hand-to-mouth households consume their labor income at every point in time. Labor markets are perfectly integrated within a region, and there is no labor mobility across regions.

We present the setting for the Home region, with the understanding that the problem of the Foreign region is analogous. Households have separable preferences for consumption and leisure that take a standard form,

$$U(C_{j,t}, L_{j,t}) = \frac{C_{j,t}^{1-\gamma_H}}{1-\gamma_H} - \psi \frac{L_{j,t}^{1+\alpha_H}}{1+\alpha_H}, \;\; j = \{HH, HR\}$$

Ricardian households maximize their discounted sum of expected utility

$$\max \sum_{t=0}^{\infty} E_0 \beta^t U(C_{HR,t}, L_{HR,t}),$$

subject to a sequence of budget constraints given by

$$B_{HR,t+1} + P_{H,t} C_{HR,t} \leq W_{H,t} L_{HR,t} + B_{HR,t}(1+i_t) + \Pi_{H,t},$$

where $B_{HR,t}$ denote nominal bonds holdings. i_t is the national nominal interest rate common to Home and Foreign regions and set by the central bank. $P_{H,t}$ is the consumer price index in the Home region, $C_{HR,t}$ is the consumption of the Ricardian agent, and $W_{H,t}$ is the nominal wage of the H region. $L_{HR,t}$ denotes hours of work of Ricardian agents. $\Pi_{H,t}$ are the nominal profits of firms in region H.

Hand-to-mouth households maximize the same utility function, but they are subject to a static budget constraint that links labor income to consumption expenditures,

$$P_{H,t} C_{HH,t} \leq W_{H,t} L_{HH,t}.$$

Regional consumption in the home region $C_{H,t}$ aggregates the consumption of both types of households, weighted by their population shares,

$$C_{H,t} = \lambda_H C_{HH,t} + (1-\lambda_H) C_{HR,t}.$$

Households have CES preferences over varieties produced in the Home and Foreign region with an elasticity of substitution ν and potential home bias $\phi \geq 1/2$. Specifically

$$C_{j,t} = \left[\phi^{\frac{1}{\nu}} C_{j,H,t}^{\frac{\nu-1}{\nu}} + (1-\phi)^{\frac{1}{\nu}} C_{j,F,t}^{\frac{\nu-1}{\nu}} \right]^{\frac{\nu}{\nu-1}},$$

with $j = \{HH, HR\}$ and $C_{i,k,t}$ is the consumption of goods produced in region k by agent i, which is a CES aggregate of a continuum of varieties with an elasticity of substitution η,

$$C_{i,k,t} = \left(\int_0^1 C_{i,k,t}(z)^{\frac{\eta-1}{\eta}} dz \right)^{\frac{\eta}{\eta-1}}.$$

The labor supply decisions in the Home region are given by

$$\psi L_{Hj,t}^{\alpha_H} C_{Hj,t}^{\gamma_H} = \frac{W_{Ht}}{P_{Ht}}, \text{ for } j \in [H, R]. \tag{5}$$

For the case of hand-to-mouth households, plugging in the budget constraint and solving for the labor supply yields

$$L_{HHt} = \left(\frac{1}{\psi} \right)^{\frac{1}{\gamma_H + \alpha_H}} \left(\frac{W_{Ht}}{P_{Ht}} \right)^{\frac{1-\gamma_H}{\gamma_H + \alpha_H}}. \tag{6}$$

Equation 6 makes clear that the co-movement of labor supply decisions of hand-to-mouth households and the real wage depends on whether the intertemporal elasticity of substitution is smaller, equal, or greater than 1, a feature of models with hand-to-mouth households with standard preferences. In the case of log-utility, the labor supply of hand-to-mouth households is acyclical. However, for the standard case where $\gamma > 1$, the amount of labor supplied by hand-to-mouth households is countercyclical. In this case, during a recession that lowers the real wage, hand-to-mouth households adjust by supplying more hours of work, the only available means they have to smooth consumption.

There is a continuum of firms in each region producing tradeable varieties. Each firm faces demand coming from Home and Foreign regions. Market clearing in the goods market implies then that production for each variety satisfies consumer demand

$$Y_{H,t}(z) = \lambda_H C_{HH,H,t}(z) + (1-\lambda_H) C_{HR,H,t}(z) + C_{F,t}(z).$$

Firms produce using a production function linear in local labor and are subject to regional productivity shocks, $Y_{Ht}(z) = A_{Ht} L_{Ht}(z)$. Real marginal costs, denoted MC, expressed in terms of domestic prices, are common across firms within a region and equal to $MC_{Ht} = \frac{W_{Ht}}{P_{Ht}} \frac{1}{A_{Ht}}$.

The price-setting problem of these firms is standard. Firms change their prices freely with probability $(1-\theta_H)$, and must keep their prices unchanged with probability θ_H, as in Calvo (1983). Firms choose to set prices equal to a markup over the weighted discounted sum of nominal marginal costs whenever they have the chance to do so. Up to first-order approximation, the optimal price-setting rule consists of a price $\bar{p}_{Ht}$ that depends on regional prices, real marginal costs, the discount factor β, and the probability that firms may not adjust their prices θ_H. In particular, reset prices are characterized by

$$\bar{p}_{Ht} = (1-\beta\theta_H) \sum_{k=0}^{\infty} (\beta\theta_H)^k \, \mathbb{E}_t \left[mc_{H,t+k} + p_{H,t+k} \right]. \tag{7}$$

The Phillips curve in the Home and foreign region has a slope κ_H, and κ_F, respectively, given by

$$\pi_{Ht} = \beta \mathbb{E}_t \pi_{H,t+1} + \kappa_H mc_{Ht} \tag{8}$$

$$\pi_{Ft} = \beta \mathbb{E}_t \pi_{F,t+1} + \kappa_F mc_{Ft} \tag{9}$$

where $mc_{j,t}$ is the average marginal cost in region j and $\kappa_H = \frac{(1-\theta_H\beta)(1-\theta_H)}{\theta_H}$ is a coefficient that captures the extent of nominal rigidities. The slope of the Phillips curve for the Foreign region is symmetric as a function of θ_F and the common discount factor β.

The risk-sharing condition states that consumption of the Ricardian households in the

Home and Foreign regions obey the following relationship,

$$(C_{HR,t})^{\gamma_H} (C_{FR,t})^{-\gamma_F} \vartheta_0 = \frac{P_{F,t}}{P_{H,t}}$$

.

where ϑ_0 is a constant that takes the value of 1 in the special case where Home and Foreign regions are equally productive in the long run. In the general case, ϑ_0 captures the current expectations of price and quantity differentials in the infinite future.

There is a single central bank for the monetary union that sets an interest rate i_t according to a monetary policy reaction function that takes as inputs national inflation and output, and a monetary policy shock ε_t,

$$i_t = \phi_\pi(\pi_{Ht} + \pi_{Ft}) + \phi_y(y_{Ht} + y_{Ft}) + \varepsilon_t.$$

Parameterization

Our benchmark parameterization follows a standard textbook calibration of the standard parameters in the model, which we summarize in Table A.3 in the Appendix. The two parameters not included in the table are λ, the share of hand-to-mouth consumers, and θ_H, θ_F, the frequency of price changes in the home and foreign regions. We will do comparative statics for these parameters to understand the effects of their heterogeneity in the response to monetary policy shocks across space.

Heterogeneity in λ and positive comovement of inflation and employment responses

To provide intuition on the effect of increasing the difference in the share of hand-to-mouth consumers, we start by fixing $\theta_H = \theta_F = 0.75$, a common value in the literature, and solve the model for a set of values for $\lambda_H \in [0, 0.5]$, while keeping λ_F fixed at 0. We simulate a 100 basis point interest rate tightening in the model and compute the on-impact responses of employment and prices in each region.

Figure 4 shows the relative effect of a monetary policy shock on prices and employment between the Home and Foreign regions. We will present the result of these alternative models using a series of scatterplots. The x-axis of each scatterplot will show the present value of the impulse response function of prices in the Home region relative to the present value of the impulse response of prices in the Foreign region. The y-axis will be analogous but for the employment responses rather than for prices. Each point in the scatterplot will correspond to a model with a different value for the parameter of interest in the Home region. We keep the calibration for the Foreign region fixed.

The main message of 4 is that heterogeneity in the share of hand-to-mouth consumers will generate, in equilibrium, a positive relation between the causal effects of monetary policy on employment and on prices. Regions with a higher share of hand-to-mouth consumers will suffer larger employment losses and larger price declines after the same shock.

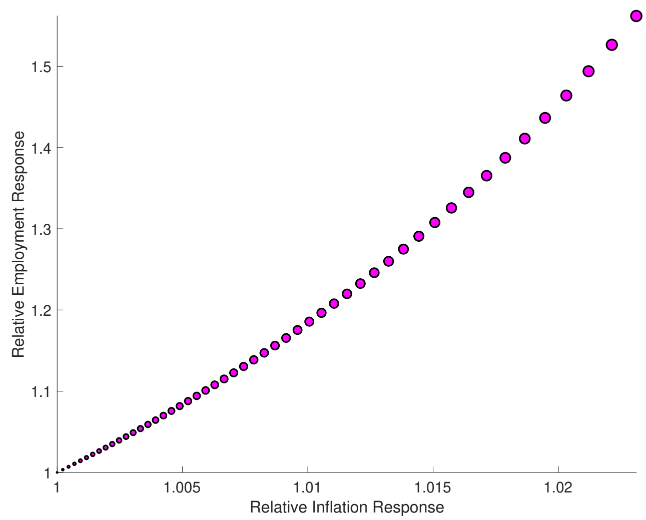


Figure 4: Relative Price and Employment Responses - Fraction of Hand-to-Mouth Consumers

Note: This figure shows the relative behavior of regional prices, on the x-axis, and employment, on the y-axis, after a national monetary policy shock. The source of regional heterogeneity is the share of hand-to-mouth households (λ). Relative inflation and employment are computed as the ratio between the discounted cumulative impulse response functions of each variable in the Home region divided by the analogous object in the Foreign region. A value of 1 means that the Home and Foreign regions have responses of the same magnitude in present value. Each point of the scatterplot represents the solution of a model with a different value of λ. The size of the marker represents how large is the heterogeneity in parameters across regions. The calibrations that underlie the figure are presented in Appendix A.5.

We now move to a model where each region is populated by Ricardian agents ($\lambda = 0$), and there is dispersion between the extent of nominal rigidities across regions, $\kappa_H < \kappa_F$. We focus on this alternative to illustrate the effects of a driver of heterogeneity on the slope of the supply block of the model, the Phillips curve.

Figure 5 shows the results. It makes clear that when regions are heterogeneous due to the steepness of local supply curves, regions with prices that are more sensitive to demand shocks are those with employment being less sensitive to the same demand shock. Intuitively, variation in the slope of the Phillips curve creates differences in the extent of monetary non-neutrality, which in a cross-section of regions generates a negative covariance between the effects of a monetary policy shock on prices and on employment. This finding is the opposite of what we find in the empirical section; regions with larger price responses have larger real responses as well.

We present results from two-region New Keynesian models of an open economy in which geographical heterogeneity arises from different alternative mechanisms, including the elasticity of labor supply and the intertemporal elasticity of substitution. We set the fraction of hand-to-mouth households λ to zero. We will present the main takeaways of these exercises in this section, although the figures and details on the calibration of the models are relegated to Appendix A.6.

The exercise we will perform will be analogous to our main exercise in the previous section. For each economic mechanism highlighted above, we will compare the impulse response of inflation and employment of Home and Foreign economies to a monetary policy shock. Home and Foreign economies are symmetric except for one particular dimension (elasticity of labor supply, elasticity of intertemporal substitution) that we will vary. Each of these margins of heterogeneity will induce differential impulse responses across regions.

Figure A.14 in the appendix considers other possibilities. The first alternative we consider is that the driver of heterogeneity is differences in labor supply elasticities. Variation in the elasticity of labor supply across regions induces changes in marginal costs. So although the sensitivity of inflation to real marginal costs is the same across regions with different elasticities of labor supply, the reaction of inflation to demand shifts will be different across regions.

This intuition explains why the left panel of Figure A.14 is qualitatively similar to Figure 5. The frequency of price changes and the elasticity of labor supply affect the slope of the Phillips curve. So, models in which these margins drive regional heterogeneity imply

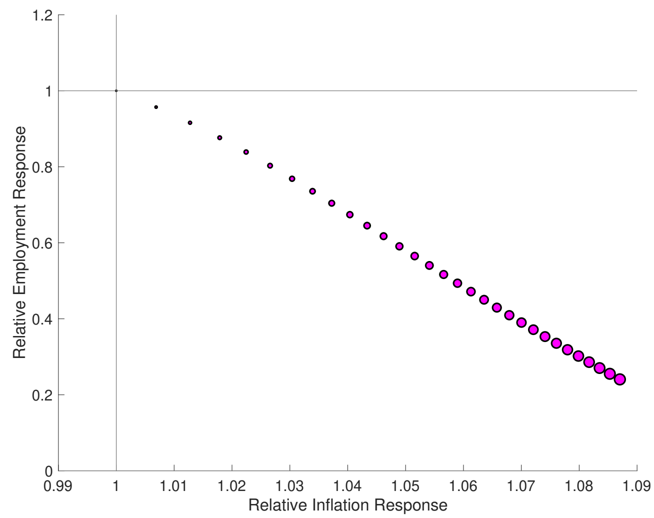


Figure 5: Relative Price and Employment Responses - Phillips curve

Note: This figure shows the relative behavior of regional prices, on the x-axis, and employment, on the y-axis, after a national monetary policy shock. The source of regional heterogeneity is variation in the extent of nominal rigidities. Relative inflation and employment are computed as the ratio between the discounted cumulative impulse response functions of each variable in the Home region divided by the analogous object in the Foreign region. A value of 1 means that Home and Foreign regions have responses of the same magnitude in present value. Each point of the scatterplot represents the solution of a model with different variations in the extent of nominal rigidities. The size of the marker represents how large the heterogeneity in parameters is across regions. The calibrations that underlie the figure are in Appendix A.6.

that economies in which inflation is more sensitive to monetary policy shocks should be closer to monetary neutrality.

A final alternative is that regional heterogeneity is driven by differences in the intertemporal elasticity of substitution. The case of the intertemporal elasticity of substitution is a priori less evident, since variation in this margin will introduce cross-sectional changes in the intertemporal IS curve and in the Phillips curve via changes in the behavior of real marginal costs when using separable preferences.

Figure A.14, right panel, shows that cross-sectional variation in the intertemporal elasticity of substitution creates a pattern counter to the ones we have presented before and in line with those in the data. In fact, the monetary union TANK model we presented before aims to introduce the same variation as reduced-form heterogeneity in intertemporal elasticity of substitution across regions. By placing a fraction of the population out of their Euler equation, the TANK model changes the effective intertemporal elasticity of substitution.

The covariance of the regional response of prices and employment to a monetary policy shock is sufficient to distinguish supply and demand margins of heterogeneity but is not enough to distinguish across different drivers of demand effects. In that sense, we cannot distinguish whether in the data the variation is driven by the share of hand-to-mouth consumers, or by households with different elasticities of intertemporal substitution. However, Aguiar et al. (2020) show that these two margins are correlated in the data.

There are certainly more margins of heterogeneity that one may consider. To the extent that these margins map into either differential elasticities of the Euler equation or differential elasticities of the Phillips curve, our analysis covers those additional margins of heterogeneity. Margins of heterogeneity that create dispersion in the slope of the Euler equation (the sensitivity of local consumption growth to local interest rates) can explain our results. Margins of heterogeneity that create differences in the slope of local Phillips curves (the sensitivity of local inflation to changes in local demand) cannot.

5 Heterogeneous effects of Monetary Policy

To make the connection between the data and the model sharper, We link income and marginal propensities to consume using evidence by Patterson (2019) that documents that income is the most important determinant of variation in marginal propensities to consume (MPC) and our model that makes the argument that differences in MPCs generate variation in the cross-section of metropolitan areas in line with the data.

On top of being an important covariate behind MPCs, income data is available for our sample of metropolitan areas. Other alternative drivers for MPCs, such as the stock of liquid assets, would also be suitable for this exercise, but we do not have as reliable data as we do for income.

We use a transformed measure of real personal income per capita to rank local areas. We deflate nominal income per capita using national CPI to avoid a mechanical correlation between regional real income per capita and regional inflation. Then, we regress real personal income per capita on time fixed effects and use the residual as our normalized measure of income. The interpretation of this residual is the difference in income between a specific city with respect to the average income across cities in our sample for a given year.[8]

We focus on the heterogeneous effects of monetary policy shocks across local economic areas in the United States. We start by estimating local projections for each individual location, computing the cumulative effect on prices of monetary policy shocks 8, 12, 16, and 20 quarters after the onset of the shock. To show our results systematically, we plot our estimated effects in Figure 6, as a function of the income of each city expressed in thousands of dollars of the year 2000.

There is substantial heterogeneity across space and horizons in Figure (6). Two years after the shock (left top panel), the effects on prices of monetary policy shocks are small. Three years after the shock (top right panel), poorer cities have accumulated a 2 percent price drop, while cities with higher income levels have experienced none. Four and five years after the shock, peak effects of the shocks materialize, with cumulative declines in prices of 2.5 percentage points after 4 years and meaningful heterogeneity that correlates with city-average income levels.

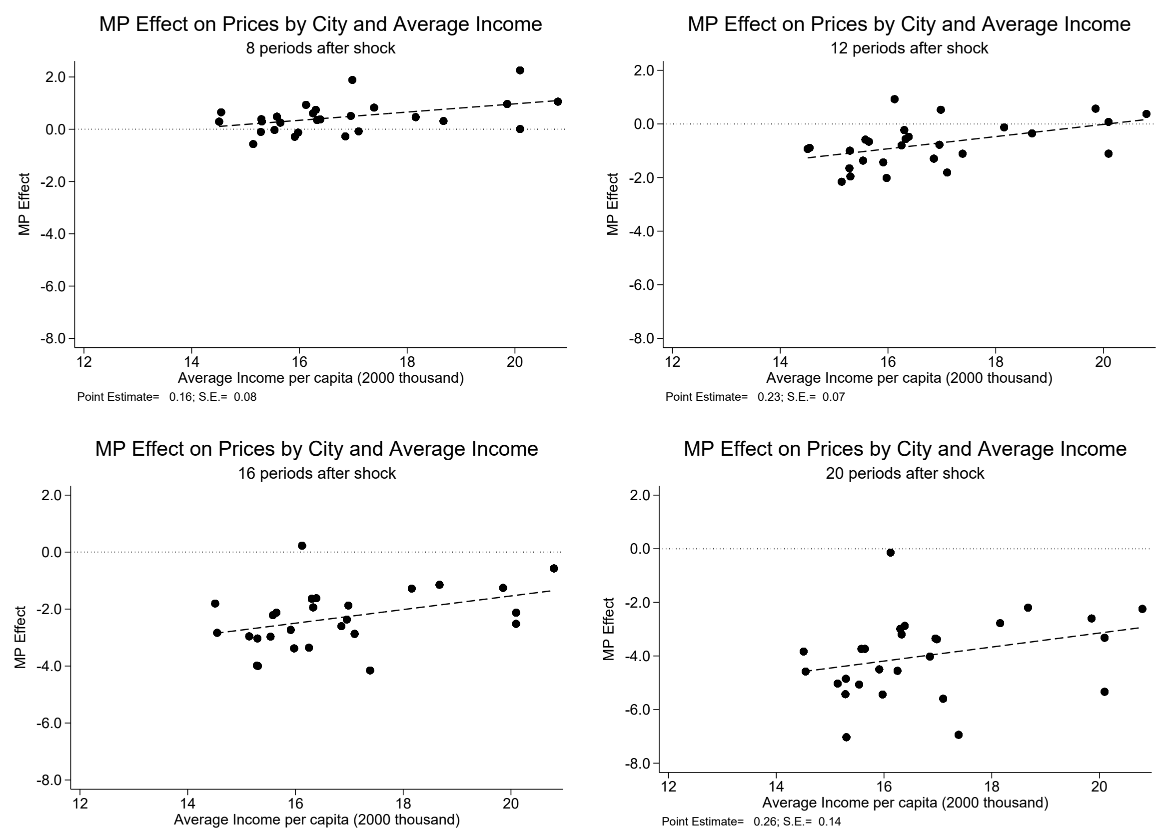
Figure (6) presents the heterogeneity of the estimates across regions, but fails to give a sense of their economic size, or their statistical significance. Intuitively, each point in the scatter plot above does not transmit information about the standard errors associated with the estimation of each local projection. However, it is reassuring that at each horizon, there is a positive relation between income and the size of price responses after monetary policy shock, which dictates our specification choices going forward.[9]

We extend equation 1 to account for regional heterogeneity in terms of real income per capita, which we estimate by running a regression of local inflation rates on the monetary policy shocks, interactions between the monetary policy shock and real relative income per capita, and local area controls that are included in the information set at time t. Our

[8]The decision to deflate income by the CPI avoids introducing heteroskedasticity in the data as the dispersion measured in current values increases through time. Our results are robust to not deflating nominal income by aggregate prices but using the residuals of a regression of nominal income on time fixed effects. Our results are also robust to deflate by local CPI, as shown in Figure A.3, in Appendix A.1. However, the interpretation of deflating by local CPI is not to make income comparable across regions since local CPIs do not play the role of price parities across space but to account for differential trends in inflation across metropolitan areas.

[9]These results should be interpreted as the effects on the price level, so even if inflation returns to its pre-existing rate, the price level is permanently changed, as predicted by standard theories.

Figure 6: Effect of Monetary Policy Shock on Prices - CPI by Cities



Note: The figure shows the results of equation (1) for each individual metropolitan area. We use $J = 8$, and $K = 8$. The upper-left panel plots cumulative effects over 8 quarters, the upper-right panel 12 quarters, the lower-left panel 16 quarters, and the lower-right panel 20 quarters.

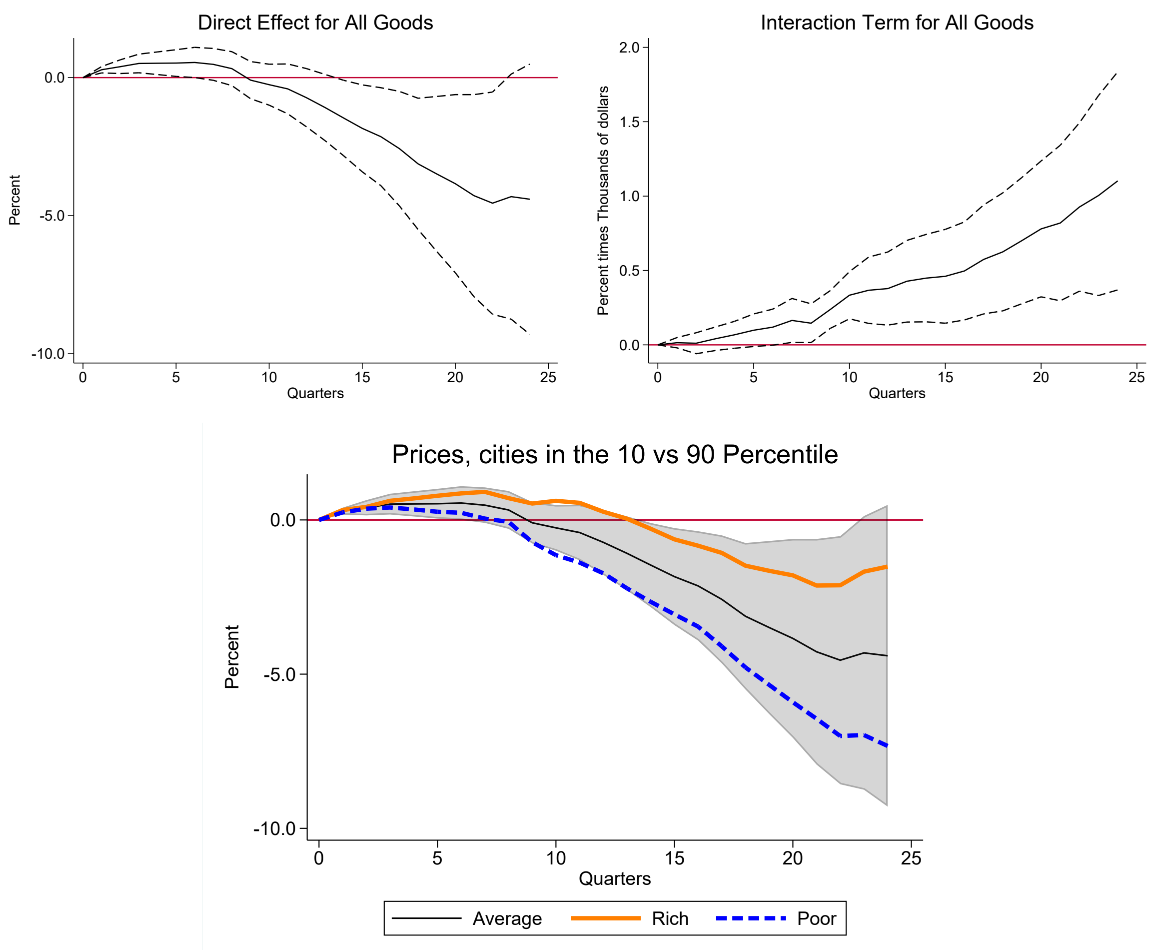
specification uses the Blinder-Oaxaca decomposition on local projections as in Cloyne et al. (2020b), applied to a panel setting. Formally, we estimate,

$$\pi_{i,t+h,t} = \alpha^h_{i,p} + \sum_{j=0}^{J} \beta^{h,j}_p RR_{t-j} + \sum_{j=0}^{J} \gamma^{h,j}_p RR_{t-j} \times RPIPC_{i,t-j-1} + \sum_{j=0}^{J} X'_{i,t-j}\theta^{h,j}_p + \varepsilon^h_{p,i,t+h}, \quad (10)$$

$\forall h \in [0, H]$ with $X_{i,t-j} = [RPIPC_{i,t-j-1} \;\; \pi_{i,t,t-j}]$, where $RPIPC_{i,t}$ is the relative personal income per capita in city i at time t, and π and RR represent the same objects as before.

The marginal effect of a monetary policy shock that occurs in period t on inflation in city i, h periods after the shock is given by $\beta^{h,0}_p + \gamma^{h,0}_p RPIPC_{i,t-1}$. Since our income control does not vary with h, we do not use any variation in real income per capita caused by the monetary policy shock. Instead, we use pre-existing differences across metropolitan areas at the onset of the shock.

Figure 7: Effect of Monetary Policy and Income Heterogeneity



Note: The top left and right panel of the figure shows the estimated coefficient $\hat{\beta}_p^h$ and $\hat{\gamma}_p^h$ from equation 10, respectively. We use $H = 24$, $J = 8$, and $K = 8$. Relative income per capita is denominated in 2000 dollars. The dashed lines show 90 percent intervals. Standard errors are clustered at the metropolitan area and time level. The bottom panel shows the point estimates of the impulse response for notional metropolitan areas in the 10th and 90th percentiles of the income distribution, together with the average response coming from the top left panel. The 90th percentile of the distribution is USD 3,060 higher than the average annual income, and the 10th percentile is USD 2,105 lower than the average annual income.

The top left panel of Figure 7 shows the impulse response of prices for a city of average income. Due to the normalization of real income per capita, the identity of the average city may change at different points in time. The interpretation of the top interaction term in the right panel is the additional effect on prices experienced by a city with a real income that is $1000 (in the year 2000) higher than average after a monetary policy shock of 1 percentage point. The main takeaway of the right panel is that a contractionary monetary policy shock causes a smaller decline in prices in high-income metropolitan areas compared to those suffered in low-income areas. The differential effects are economically sizable; a city with an income per capita that is $1000 higher than the average gets one percentage point less cumulative inflation after a monetary policy shock of one hundred basis points after twenty quarters.

To illustrate further the economic relevance of our estimated heterogeneous effects, the bottom panel of Figure 7 shows the effect for cities in the 10th percentile of the income distribution versus cities in the 90th percentile, giving a sense of the quantitative importance of our result throughout the geographical distribution of income. A monetary policy shock of the same size causes an effect on prices almost 50 percent larger for cities in the 10th percentile of the distribution compared to the average and 50 percent milder in the richer 90th percentile compared to the average. Among cities as rich as those in the 90th percentile of the income distribution, we fail to detect negative effects of monetary policy shocks on prices.

Although the effects for headline CPI are appealing, headline prices are not free of shortcomings. Since regions can vary in their expenditure weights, it could be the case that our results emerge from differences in weights rather than differences in the prices of different categories. The comparison of the sub-components of the CPI allows us to dig deeper into the mechanism behind our main results.

Our results hold across goods with a differential degree of tradeability, with larger differential effects for consumer categories that are closer to being non-traded. Figure A.4 in Appendix A.1 shows our estimated impulse responses for "food at home,", a category with a substantial tradeable component, and "food away from home," a category with a large non-tradeable component. In Appendix A.1, Figure A.6 shows similar results for "housing," which also has a large non-tradeable component due to the relevance of shelter in that consumption category. Figure A.4 is in line with the intuition that the relative effects in the right panel should be larger for consumption categories that have a larger non-tradeable component to them since, intuitively, consumption and pricing of those goods depend on local economic conditions more than for the case of tradeable goods.

We provide results for gasoline, a highly tradeable, homogeneous, flexible-price good, which we show in Figure A.5. Gasoline has very flexible prices (see Nakamura and Steinsson (2008)), with a frequency of price change of once every month. Its price change behavior is dominated by national and world events, implying that our heterogeneous results as a share of the average results must be smaller. This is what we find: prices react less in regions with higher average income, and using conservative standard errors, the effects are insignificant. We take these results as indicative that our findings are not driven by particular regional differences in particular aspects of a small set of consumer expenditure categories.

5.1 Economic Activity

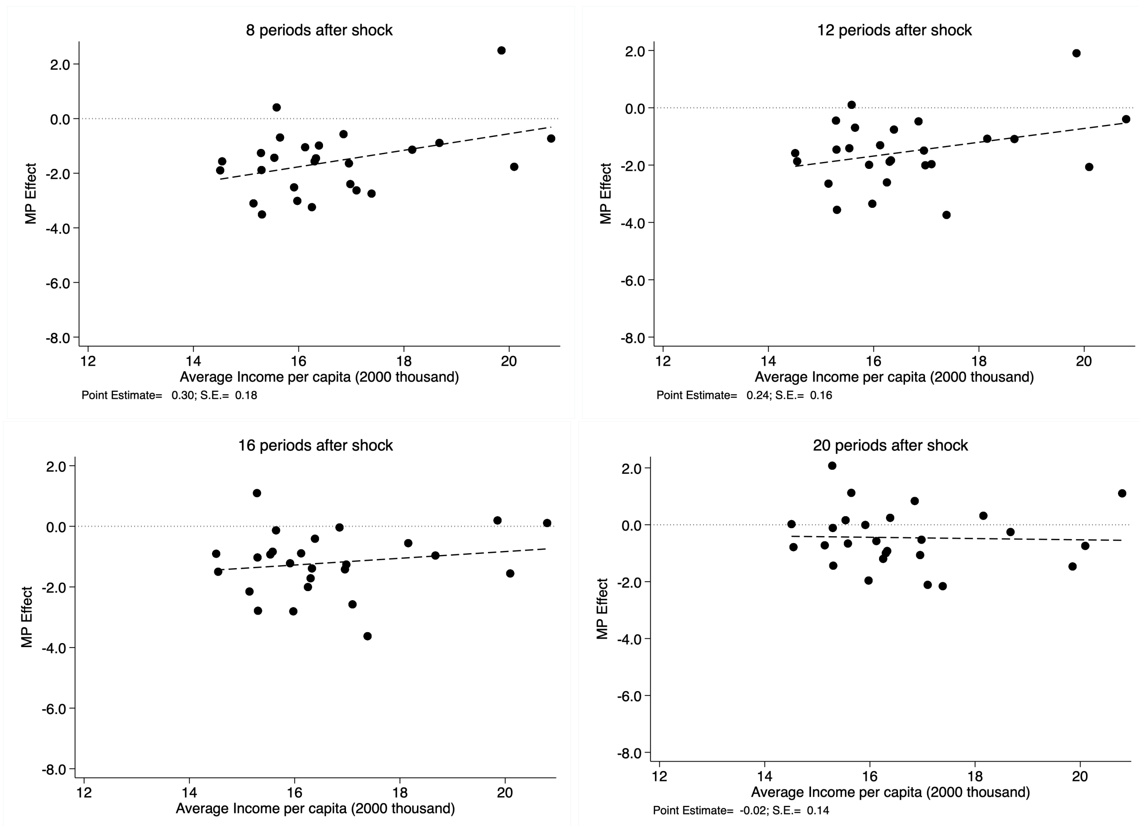
We now present analogous results for employment at the local level. We start by running local projections for each city and sorting these cities by their average income levels. Figure 8 plots the results 8, 12, 16, and 20 quarters after a shock that tightens rates by 1 percent.

Qualitatively similar to in Section 3, the effect in most of local markets is faster compared to the behavior of the impulse response for prices. Negative effects kick in 8 quarters after the shock. Lower-income areas have, on average, larger negative employment effects. We can see that this pattern stays there after 12 quarters but starts to dissipate after. The real effects of monetary policy dissipate 20 quarters after the shock, meaning that metropolitan areas return to their employment level prior to the shock.[10]

We estimate local projections with heterogeneous effects on the panel of metropolitan areas, following our approach of interacting the Romer and Romer (2004) shock with the pre-existing metro area real personal income per capita. The upper panel of Figure A.3 presents the direct and interaction effects. We estimate a significant effect of the interaction term that dampens the negative effects for richer cities. The interaction term

[10]That the slope of the effect of employment as a function of income reaches zero means that employment goes back to its pre-shock value in levels.

Figure 8: Effect of Monetary Policy Shock on Employment by Metropolitan Area

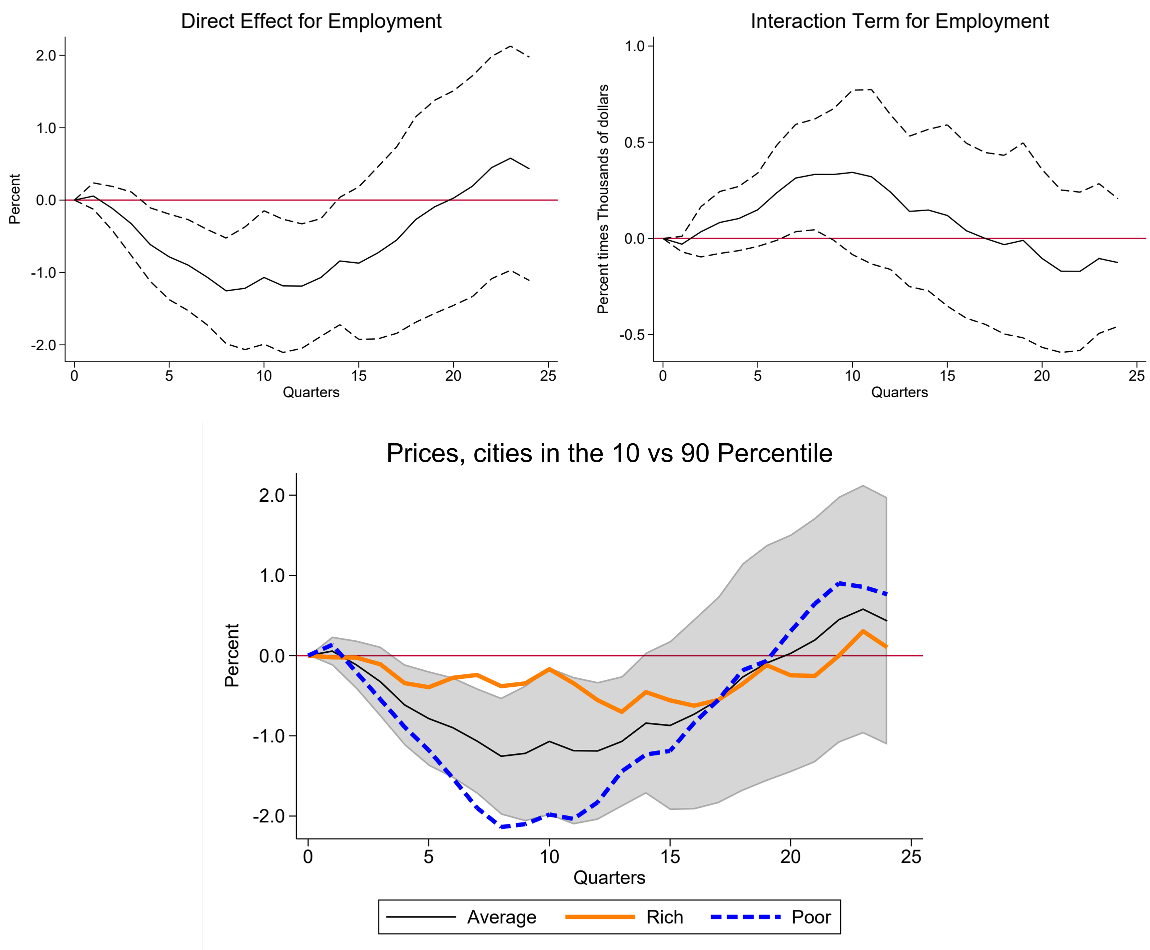


Note: The figure shows the results of equation (1) for each individual metropolitan area and employment growth as the dependent variable. We use $J = 8$, and $K = 8$. The upper-left panel plots cumulative effects over 8 quarters, the upper-right panel 12 quarters, the lower-left panel 16 quarters and the lower-right panel 20 quarters.

goes in the opposite direction of the direct effect; higher-income areas have smaller relative employment declines when the direct effect is negative. On average, high-income metropolitan areas experience smaller improvements when employment starts to recover. These results together imply smaller causal effects on employment to monetary policy shocks in high-income areas.

The lower panel of Figure A.3 shows the effect for a city in the 10th percentile of real relative income versus a city in the 90th percentile. Our results indicate that poor cities shape the national profile of employment effects. We do not find significant employment effects for areas with income as high as those in the 90th percentile of the geographic income distribution. Metro areas with income as low as those in the 10th percentile of the distribution have employment losses two times as large as those observed on average.

Figure 9: Effect of Monetary Policy Shock and Income Heterogeneity for Employment



Note: The top left and right panel show the estimated coefficients $\hat{beta}^h$ $\hat{gamma}^h$, respectively when the left-hand side variable in equation (10) for private employment. We use $H = 24$, $J = 8$ and $K = 8$. The dashed lines show 90 percent intervals. Standard errors are clustered at the city and time level. The lower panel shows the point estimates $\beta^h + \gamma^h RPIPC_{i,t+h}$ of equation (10) for metropolitan areas in the 90th and 10th percentiles of the geographic income distribution along with the average effects from the top left panel. The 90th percentile of the employment distribution is 4,755 USD (in 2000 dollars) higher than the average annual income, while the 10th is 3,596 USD (in 2000) lower than the average annual income.

Figure A.3 shows that the effects of monetary policy shocks during the first 15 quarters are negligible in high-income areas, while the peak response for a low-income area is roughly 2 percent, which reverts after 15 quarters. Low-income metropolitan areas drive the national effects: the effects of metropolitan areas with higher income are small throughout the horizon of the impulse response function.

5.2 Robustness

The main heterogeneous results use the Romer and Romer (2004) shock and heterogeneous results by relative personal income per capital of a given metropolitan area. In this section, we explore robustness of these results to other forms of heterogeneity and other sources of monetary policy shocks.

A natural candidate as a source of heterogeneity is to include differences in industrial composition across local areas. Sectors might be heterogeneous in their exposure to interest rate changes, or changes in aggregate demand within the set of metropolitan areas from which the price data comes, which are large, urban areas. A natural question is whether focusing on sectoral heterogeneity is sufficient to understand the differential effects of monetary policy we documented.

Even if cities might have a distinct industrial composition, it is unclear whether average income is a function of industrial composition or the other way around. In Section 4 we show that only a certain family of models can explain our results. That would reject models with industrial heterogeneity due to, for example, the heterogeneity in the frequency of price changes, in the pass-through of marginal costs to prices in the flexible price equilibrium via heterogeneity in the slope of demand curves, or the elasticity of marginal costs to quantity changes via heterogeneity in the shape of production functions, or firm-specific input markets. Additionally, industries might sort across cities due to the demographic characteristics of the population, or workers might migrate to a city due to its industrial composition. That discussion is beyond the scope of this paper. It is important to highlight that the metropolitan areas that the BLS samples are large, complex, and financially developed. We do not include any data on small cities or rural areas.

Therefore, in order to evaluate the importance of industrial composition, we need to impose some discipline in the possible form in which industries must be heterogeneous to explain our results. They must affect the demand block of the model primarily, not the supply block. The main form of industry heterogeneity we think is plausible is heterogeneity in the durability of locally produced goods coupled with home bias, such that local households in more durable-producing regions consume more durables that are more easily intertemporal substituted.

Because of this, the role of industrial composition is not clear. To evaluate its importance, we extend our main regression 10 by including as a control time-fixed effect interacted with lagged local sectoral employment shares. Figure A.8 presents the results. The heterogeneous effects are qualitatively similar to our benchmark specification and still significant, highlighting the relevance of the regional dimension of the data. To unpack the employment shares that are important in generating our result, Figure A.7 in Appendix A.1 shows the results of including one sector share at a time.

Additionally, Figure A.9 in Appendix A.1 shows results including other potential local heterogeneities that can explain the results, such as access to financial markets. We include time-fixed effects interacted with the share of labor income, the average debt level of households, and the age structure. The figure shows that the effect of the interaction is almost unchanged with these controls.

Another potential concern is that the shock in Romer and Romer (2004) identification assumption relies on the Greenbook forecast capturing anticipation effects on inflation and output. A reasonable concern to have is that the FOMC, at the same time, reacts differentially to future expected trends in some regions relative to others, and that *aggregate* Greenbook forecasts do not appropriately capture these *differential* expected future trends at the local level.

The concern is that while the Romer and Romer (2004) shock controls for information about the expected future trends of the national economy included in the information set of the FOMC, this shock might not clean anticipation effects about local economies. We test for this possibility and we find that the Romer and Romer (2004) is not predictable by local inflation rates. We also use other shocks related to monetary policy surprises. One is the series developed by Bu et al. (2021) and the second by Miranda-Agrippino and Ricco (2021). Results are presented in Appendix A.3. The direct effects of monetary policy shocks are lower for richer cities, which is the same we found using the Romer and Romer (2004) shock.

Using these shocks also allows us to evaluate our main effect on a sample size that covers the period after the Great Recession. We see that the results are robust to that extension of the period. In addition, these results are obtained with data from the 90s, excluding the Volcker disinflation period, which is one of the main sources of variation of the Romer and Romer (2004) shock according to Coibion (2012).

6 Aggregate Implications

Up to this section we have mostly discussed the cross-sectional implications of heterogeneity in the parameters of the model across regions. In this section, we discuss the aggregate implications of them. In this version of the monetary union model, and in others such as Nakamura and Steinsson (2014), workers can't move from one region to the other. This means that shocks can have different local implications if there is heterogeneity in some parameters. This heterogeneity can have aggregate implications depending on the nature of the heterogeneity and the policy reaction. In some cases, monetary policy targets a certain level of inflation and output, washing up the aggregate effects of local heterogeneity. The relationship between the effects of the local heterogeneity and the weights the monetary authority puts on the output gap and inflation will influence the capacity of the central bank to reduce the aggregate effects of the heterogeneity.

In Appendix A.7, we show that heterogeneity in the IES, labor supply elasticity, slope of the Phillips curve, and share of hand-to-mouth have potential aggregate implications. But some of them work in a direction where monetary policy can alleviate them, such as the IES, and others exacerbate the trade-off between output and inflation, such as the slope of the Phillips curve. In the case of the share of hand-to-mouth, the effects are exacerbated because of the amplifying effect that non-ricardian agents have in local demand Bilbiie (2008). In that case, if the heterogeneity is big enough, the central bank would not have enough room to mitigate the aggregate effects. Figure A.15 in Appendix A.7 shows how different monetary policy rules can reduce or amplify the aggregate effects in output or prices. Additionally, A.7 shows how the heterogeneity in different parameters can have aggregate implications in different versions of the model.

In Section 5, we showed that the average relative income of a city is a relevant margin of heterogeneity for the local effects of monetary policy shocks on employment and prices. We showed our results are consistent with a model of a monetary union where regions differ in their share of hand-to-mouth (HtM) households. Aguiar et al. (2020) and Patterson (2019) show a large negative correlation of HtM (or high MPC consumers) with income at the individual level.[11]

We use estimates of the relationship between income and MPCs produced by Patterson (2019) to characterize the average MPCs across cities in the US. Figure A.10 shows the evolution of MPCs for US cities since 1986 and their distribution. The median of the distribution has been relatively stable over time, with a slight decrease in recent years, but there is substantial heterogeneity across US cities.

This section explores the implications of the heterogeneity of regionaIn particular, the MPC out of transitory income shock for hand-to-mouth consumers is equal to 1 since they consume all their income.ifts in the stance of monetary policy. We will run counterfactuals that vary the dispersion in the share of HtM households across locations keeping the national share of HtM households constant. We will use the model presented in Section 4.1 to back out the relevance of geographical heterogeneity in determining aggregate outcomes.

We impute the relationship between MPCs and income to individual earnings data from the CPS using estimates by Patterson (2019). We have a panel of MPCs for 177 metropolitan areas from 1986 to 2020.[12] We extend our model to include share of hand-to-mouth in both regions (λ_i), and compute the 90th and 10th percentiles of the distribution of hand-to-mouth to each region using the MPC estimates. In particular, the MPC out of transitory income shock for hand-to-mouth consumers is equal to 1, since the consume all their income. In the case of Ricardians consumers, such a shock would induce a direct effect equal to $(1 - \beta)$. Then, after taking a stance on the time-preference parameter β, we can obtain a share of HtM λ_i. Specifically, $MPC_i = \lambda_i + (1 - \lambda_i) * (1 - \beta)$ or

[11] Kaplan et al. (2014) also refer to wealthy hand-to-mouth. Regarding those agents, Aguiar et al. (2020) result indicates that hand-to-mouth households that are illiquid (as opposed to being low net worth), do have higher income than the rest of hand-to-mouth households, but they do not have, on average, high income. In that sense, our TANK model draws a similar mapping between income and HtM compared to other models that incorporate more heterogeneity.

[12] The start date is determined by changes in the geographical sampling of the CPS and our intention to have a balanced panel of metropolitan areas.

$\lambda_i = \frac{MPC_i - (1-\beta)}{\beta}$.[13]

We use the parameter values summarized in table A.3. We simulate the model using two regions keeping the national average λ constant, but varying its geographical dispersion. Table 1 shows the results of the simulations.

Table 1: Simulation of Heterogeneous and Homogeneous Monetary Union

	Heterogeneity			Homogeneity		
	Region 1	Region 2	Aggregate	Region 1	Region 2	Aggregate
Share of HtM	70.2	57.9	64.0	64.0	64.0	64.0
Employment	-1.739	-0.440	-1.090	-0.799	-0.799	-0.799
Consumption	-2.174	-0.005	-1.090	-0.799	-0.799	-0.799
Real Wage	-3.334	-0.298	-1.816	-1.331	-1.331	-1.331
Inflation	-0.197	-0.097	-0.147	-0.114	-0.114	-0.114

Note: This table shows the effect on impact of a monetary policy shock of 1 percentage points on employment, inflation, consumption, and the real wage. We introduce the same experiment for economies with heterogeneity in the share of hand-to-mouth consumers, and without heterogeneity in hand-to-mouth consumers. Both economies have an average share of hand-to-mouth consumers of 64%. Columns 2 to 4 (heterogeneity) show the effect of the shock in an economy with heterogeneous values of HtM across regions. We show the results for each region (columns 2 and 3) and the aggregate economy (column 4). Columns 5 to 7 show the same effects, but for an economy where regions have the same share of hand-to-mouth consumers. All the numbers are shown in percentage points.

Table 1 contains two main messages. The first one, is that heterogeneity is very important to understand the transmission of monetary policy to different aggregates. In standard textbook models, the reaction of employment and consumption to a monetary policy shock are equivalent, and that equivalence still holds in our economy at the aggregate level (the second and third row of the *Aggregate* columns contain the same numbers). However, the heterogeneity in hand-to-mouth consumers we use, generates significant dispersion in the responses of consumption relative to production at the local level. After a common monetary policy shock, consumption for households in Region 2 is almost neutral, while consumption in Region 1 contracts more than their production. The response of real wages in Region 1 is more than 10 times higher than that in Region 1. There is an

[13] In Appendix A.8, we use the model and simulated method of moments to match the slope between employment and price responses shown in Figure 3, and the cross-sectional dispersion of cumulative prices and employment effects of a monetary policy shock. This procedure delivers a similar level of dispersion in the share of HtM across cities compared to the method used in this Section.

important disparity of inflation across space.

Hand-to-mouth consumers use their labor supply as their only available means to smooth consumption. In our parameterization, HtM households do not adjust their labor supply, while Ricardian agents reduce their hours worked as the real wage falls. Declines in economic activity introduce additional downward pressure on the real wage in regions with a higher share of hand-to-mouth consumers in equilibrium. Since consumption falls more than production in Region 1, there is a reallocation of consumption from Region 1 into Region 2. The effect on prices is relatively smaller, which is a result of our assumption of having only tradable goods that are relatively substitutable.

The second message of Table 1 is that heterogeneity in MPCs amplifies the response of the aggregate economy to monetary policy. Amplification arises due to the non-linear effects of the share of hand-to-mouth consumers described in Bilbiie (2020). After a contractionary monetary policy shock, Ricardian agents reduce consumption and labor supply, reducing real wages in the local region. The effect on real wages makes hand-to-mouth (HtM) consumers reduce their spending as they consume exclusively from their labor income. The reduction in local wages, common for a given region by our assumption of integrated local labor markets, produces an additional decrease in demand in the local economy that depends on the share of hand-to-mouth households. This additional effect reduces marginal costs, increasing profits and producing an income effect.[14]

This effect depends critically on the labor supply elasticity (determined by α in our model), and it is non-linear in the share of hand-to-mouth consumers. The higher the share of HtM, the higher the effect in absolute value and at an increasing rate. Because of this non-linearity, the average effect is also larger in absolute value when there is a region with a higher share of HtM compared to the average. Therefore, the higher the dispersion of HtM, the higher the effect will be. Heterogeneity across regions amplifies the effect of monetary policy on both employment and prices.

[14] See Bilbiie (2008) for details on the conditions for this equilibrium.

7 Conclusions

This paper documents the differential regional effects on real and nominal variables of monetary policy shocks in the US. We find that cities that experience larger price effects also experience larger employment effects. The positive covariance of price and employment effects is significant and robust to include the variation of individual estimates. We evaluate a set of economic mechanisms typically discussed in the New Keynesian literature to document which are consistent with our results. We propose a model in which a different fraction of hand-to-mouth consumers characterizes regions. By affecting the sensitivity of consumption to real interest rates, the model rationalizes the larger employment and price responses we estimate in the data. Models with variation in intertemporal elasticities of substitution can also explain our results. On the contrary, models in which differential slopes of the Phillips curve characterize regions fail to rationalize our findings since they would imply lower employment responses in areas with higher price responses.

More sensitive regions tend to have lower income, and income is a key covariate behind MPC variation. We estimate that monetary policy shocks induce larger effects on prices and employment in low-income metropolitan areas. The price results hold for overall prices and a wide range of consumer expenditure categories.

The effects we estimate are economically large and suggest an important challenge for the monetary authority since the power of its main tool varies across regions. This challenge is compounded for the case in which regions have differential exposure to the underlying shocks, as in trade shocks (Autor et al., 2016), or government spending shocks (Nakamura and Steinsson, 2014).

Our results highlight the potential role of fiscal policy in generating the same aggregate effects as those induced by monetary policy, but with different local effects, as studied in the literature on equivalence results between monetary and fiscal policies (Wolf, 2021). Along that same line, the results of this paper highlight the potential complementary role of fiscal policy in correcting undesirable distributional effects of monetary policy.

References

Acosta, Miguel, "A new measure of central bank transparency and implications for the effectiveness of monetary policy," *International Journal of Central Banking*, 2023, *19* (3), 49–97.

Aguiar, Mark A., Mark Bils, and Corina Boar, "Who are the Hand-to-Mouth?," Working paper 26643, National Bureau of Economic Research August 2020.

Almgren, Mattias, José-Elías Gallegos, John Kramer, and Ricardo Lima, "Monetary policy and liquidity constraints: Evidence from the euro area," *American Economic Journal: Macroeconomics*, 2022, *14* (4), 309–340.

Andersen, Asger Lau, Niels Johannesen, Mia Jorgensen, and Jose-Luos Peydro, "Monetary Policy and Inequality," Working Paper 1761, Department of Economics and Business, Universitat Pompeu Fabra March 2021.

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

Autor, David H., David Dorn, and Gordon H. Hanson, "The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade," *Annual Review of Economics*, 2016, *8*, 205–240.

Barnichon, Regis and Geert Mesters, "The phillips multiplier," *Journal of Monetary Economics*, 2021, *117*, 689–705.

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra, "Regional Heterogeneity and the Refinancing Channel of Monetary Policy," *The Quarterly Journal of Economics*, 2019, *134* (1), 109–183.

Bergman, Nittai, David A. Matsa, and Michael Weber, "Inclusive Monetary Policy: How Tight Labor Markets Facilitate Broad-Based Employment Growth," Working Paper 29651, National Bureau of Economic Research January 2022.

Bilbiie, Florin O., "Limited asset markets participation, monetary policy and (inverted) aggregate demand logic," *Journal of Economic Theory*, 2008, *140* (1), 162–196.

_ , "The New Keynesian cross," *Journal of Monetary Economics*, 2020, *114*, 90–108.

Bu, Chunya, John Rogers, and Wenbin Wu, "A unified measure of Fed monetary policy shocks," *Journal of Monetary Economics*, 2021, *118*, 331–349.

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

Carlino, Gerald and Robert Defina, "The Differential Regional Effects of Monetary Policy," *The Review of Economics and Statistics*, 1998, *80* (4), 572–587.

Cloyne, James, Clodomiro Ferreira, and Paolo Surico, "Monetary Policy when Households have Debt: New Evidence on the Transmission Mechanism," *The Review of Economic Studies*, 2020, *87* (1), 102–129.

_ , Oscar Jorda, and Alan Taylor, "Decomposing the Fiscal Multiplier," *Federal Reserve Bank of San Francisco Working Paper, 2020-12*, 2020.

Coibion, Olivier, "Are the Effects of Monetary Policy Shocks Big or Small?," *American Economic Journal: Macroeconomics*, April 2012, *4* (2), 1–32.

_ , Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia, "Innocent Bystanders? Monetary policy and inequality," *Journal of Monetary Economics*, 2017, *88*, 70–89.

Cravino, Javier, Ting Lan, and Andrei A. Levchenko, "Price stickiness along the income distribution and the effects of monetary policy," *Journal of Monetary Economics*, 2018, *110*, 19–32.

Doepke, Matthias and Martin Schneider, "Inflation and the Redistribution of Nominal Wealth," *Journal of Political Economy*, 2006, *114* (6), 1069–1097.

Furceri, Davide, Prakash Loungani, and Aleksandra Zdzienicka, "The effects of monetary policy shocks on inequality," *Journal of International Money and Finance*, 2018, *85*, 168–186.

Hazell, Jonathon, Juan Herre no, Emi Nakamura, and Jón Steinsson, "The Slope of the Phillips Curve: Evidence from US States," *The Quarterly Journal of Economics*, 2022.

Jorda, Oscar, "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, March 2005, *95* (1), 161–182.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante, "Monetary Policy According to HANK," *American Economic Review*, 2018, *108* (3), 697–743.

_ , Giovanni L. Violante, and Justin Weidner, "The Wealthy Hand-to-Mouth," Working Paper 20073, National Bureau of Economic Research August 2014.

Mian, Atif, Andrés Sarto, and Amir Sufi, "Estimating Credit Multipliers," 2022.

Miranda-Agrippino, Silvia and Giovanni Ricco, "The transmission of monetary policy shocks," *American Economic Journal: Macroeconomics*, 2021, *13* (3), 74–107.

Nakamura, Emi and Jón Steinsson, "Five Facts about Prices: A Reevaluation of Menu Cost Models," *The Quarterly Journal of Economics*, 2008, *123* (4), 1415–1464.

_ and _ , "Fiscal stimulus in a monetary union: Evidence from U.S. regions," *American Economic Review*, 2014, *104* (3), 753–92.

Neville, Francis, Owyang Michael T., and Sekhposyan Tatevik, "The Local Effects of Monetary Policy," *The B.E. Journal of Macroeconomics*, March 2012, *12* (2), 1–38.

Olea, José Luis Montiel and Mikkel Plagborg-Møller, "Local projection inference is simpler and more robust than you think," *Econometrica*, 2021, *89* (4), 1789–1823.

Patterson, Christina, "The Matching Multiplier and the Amplification of Recessions," Unpublished Manuscript, Northwestern University 2019.

Ramey, Valerie A, "Macroeconomic shocks and their propagation," *Handbook of macroeconomics*, 2016, *2*, 71–162.

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

Russ, Katheryn, Jay C Shambaugh, and Sanjay R Singh, "Currency Areas, Labor Markets, and Regional Cyclical Sensitivity," Technical Report, National Bureau of Economic Research 2023.

Uhlig, Harald, "A Toolkit for Analysing Nonlinear Dynamic Stochastic Models Easily," *Ramon Marimon and Andrew Scott, eds, Computational Methods for the Study of Dynamic Economies*, 1999, pp. 30–61.

Wieland, Johannes F and Mu-Jeung Yang, "Financial Dampening," *Journal of Money, Credit and Banking*, 2020, *52* (1), 79–113.

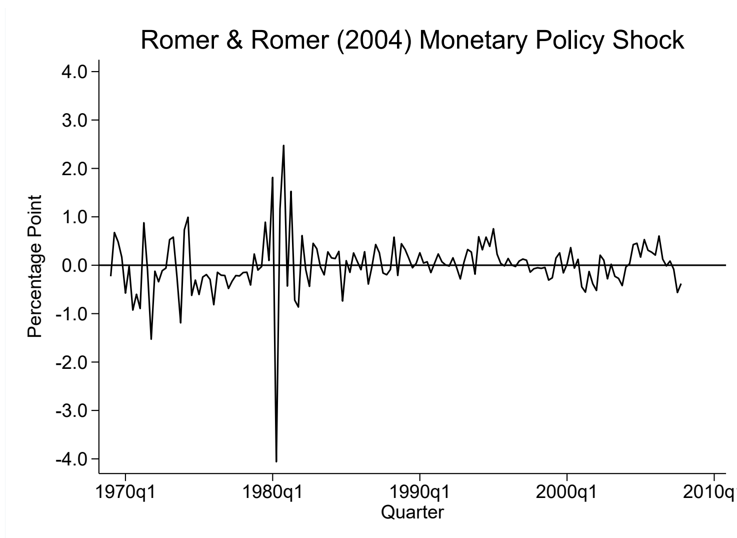
Wolf, Christian, "Interest rate cuts vs. stimulus payments: An equivalence result," Working paper 29193, National Bureau of Economic Research August 2021.

Wong, Arlene, "Refinancing and the Transmission of Monetary Policy to Consumption," Working paper 2021-57, Princeton University 2021.

A Appendix

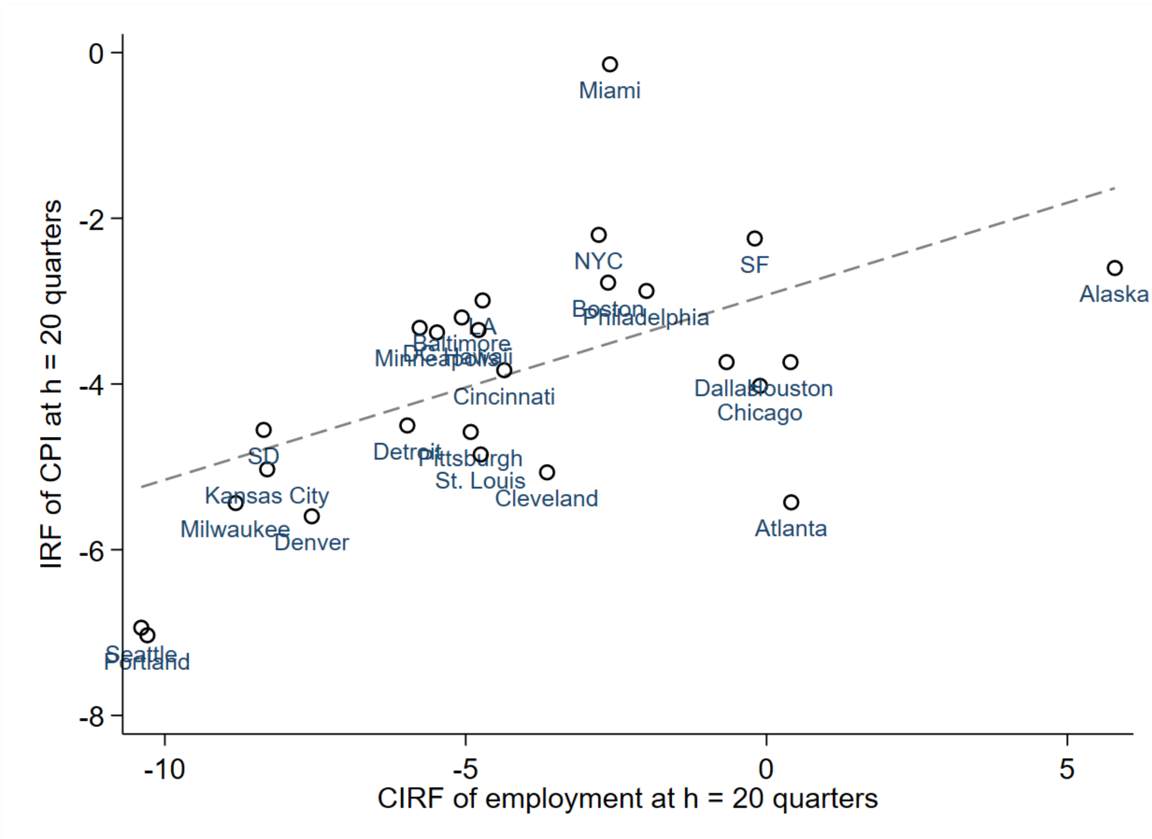
A.1 Additional Figures

Figure A.1: Romer and Romer (2004) Monetary Policy Shock



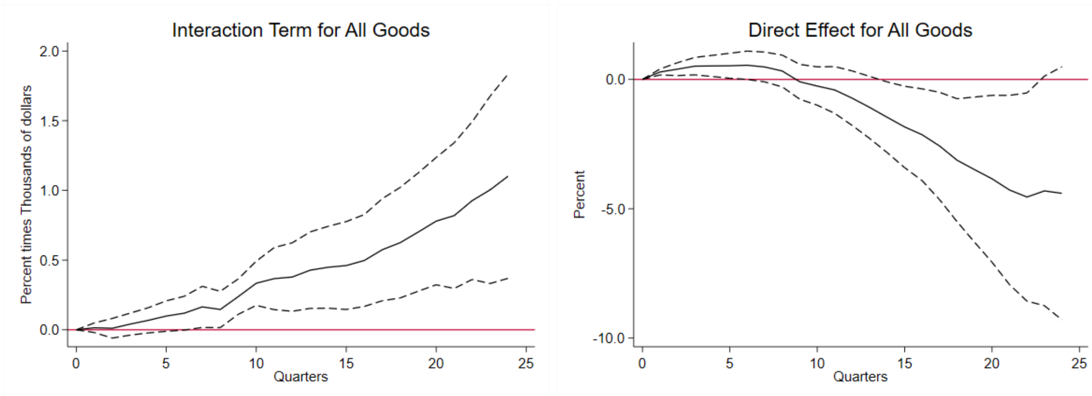
Note: This figure plots the Romer and Romer (2004) monetary policy shocks extended by Wieland and Yang (2020) aggregated at a quarterly level. We aggregate monetary policy shocks at a quarterly frequency by computing a sum of the monthly-level shocks.

Figure A.2: Effect of a Monetary Policy Shock in Employment and Prices for Each City



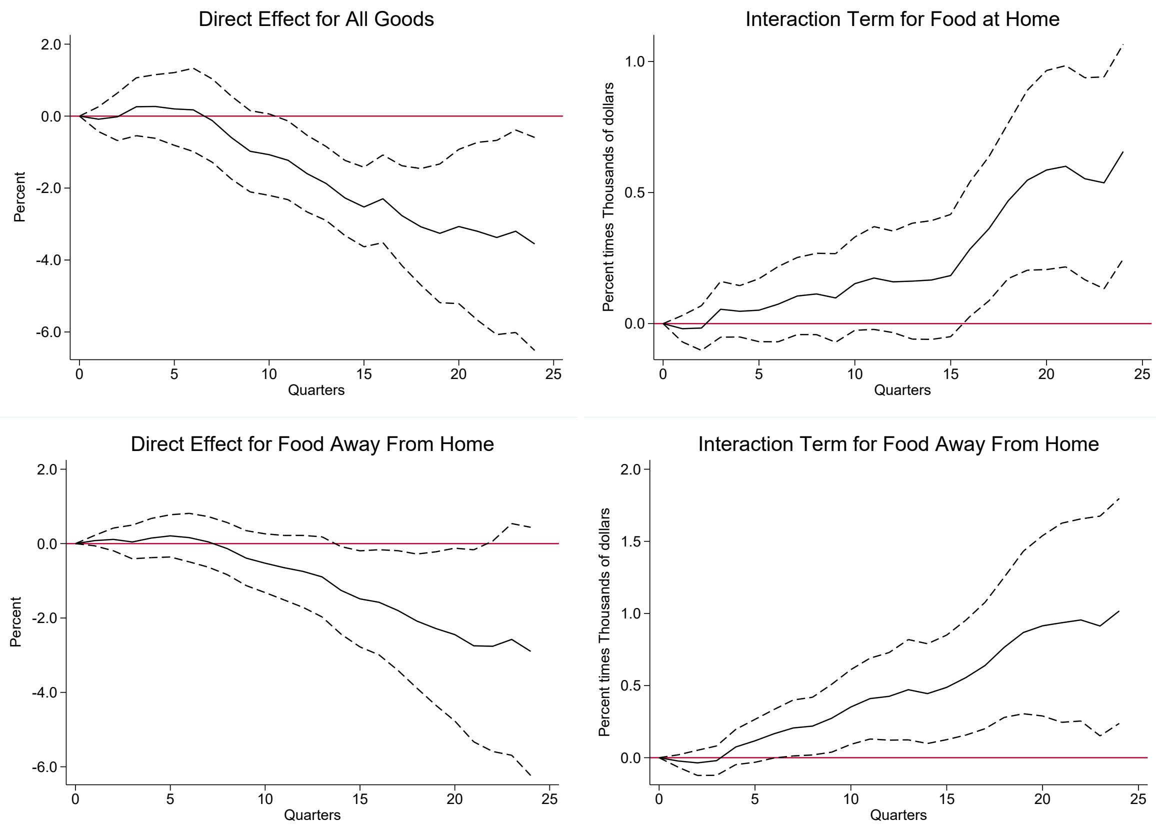
Note: This figure plots on the y-axis the local projection on local consumer prices of an exogenous monetary policy tightening of 100 basis points 20 quarters after the shock. The x-axis plots the cumulative effect (area under the curve) of local employment 20 quarters after a monetary policy shock of 100 basis points. The units of both axes are percentage points. Each bubble in the scatter plot corresponds to a metropolitan area. The size of each bubble has the name of the main city of each of the metropolitan areas.

Figure A.3: Effect of Monetary Policy Shock and Income Heterogeneity Using Local Prices



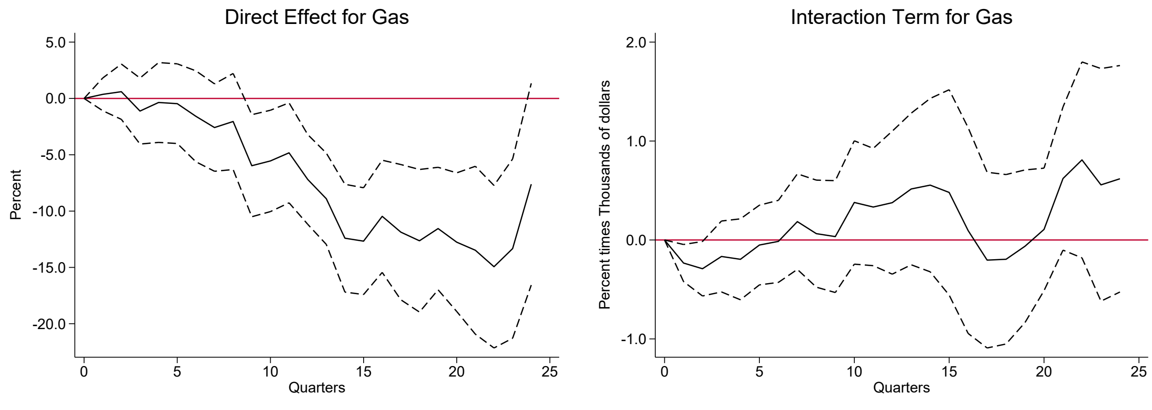
Note: The top left and right panel show the estimated coefficients $\hat{\beta}^h$ $\hat{\gamma}^h$, respectively when the left-hand side variable in equation (10) for private employment. We use $H = 24$, $J = 8$ and $K = 8$. The dashed lines show 90 percent intervals. Standard errors are clustered at the city and time level.

Figure A.4: Monetary Policy Shocks and Income Heterogeneity - By Tradeability



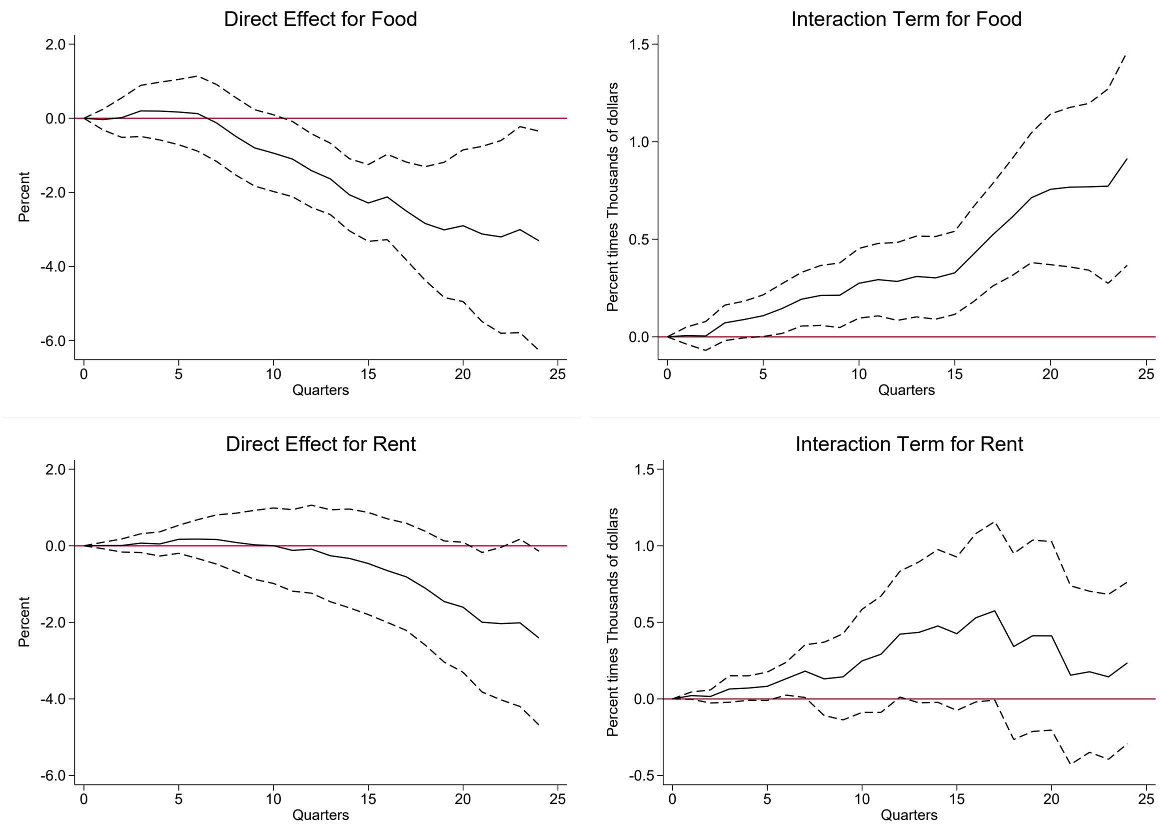
Note: The left panel shows the β^h coefficient and the right panel shows the γ^h coefficient of equation (10) for Food Away From Home. We use $H = 24$, $J = 8$, and $K = 8$. The dashed lines show 90 percent intervals. Standard errors are clustered at the city level.

Figure A.5: Effect of Monetary Policy Shock and Income Heterogeneity for Gas



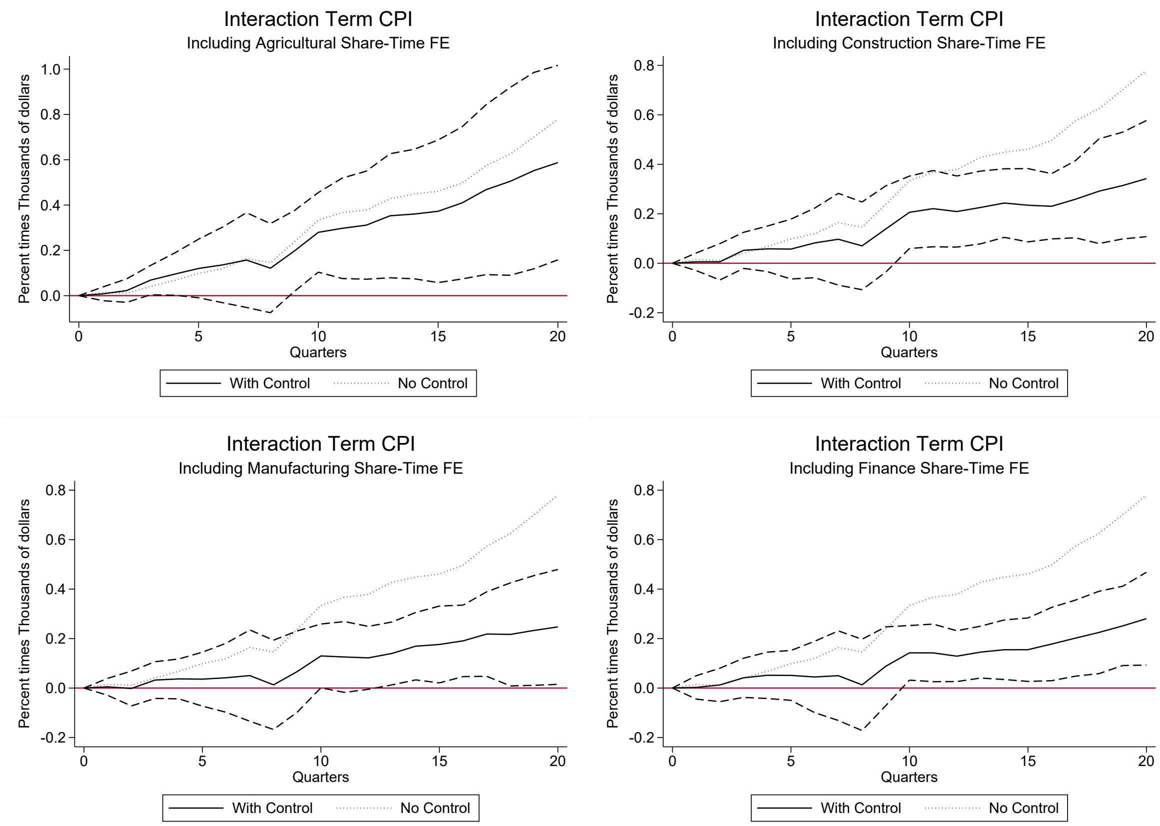
Note: The left panel shows the β^h coefficient and the right panel shows the γ^h coefficient of equation (10) for gasoline (regular).We use $H = 24$, $J = 8$, and $K = 8$. The dashed lines show 90 percent intervals. Standard errors are clustered at the city level.

Figure A.6: Effect on Narrow Price Indexes



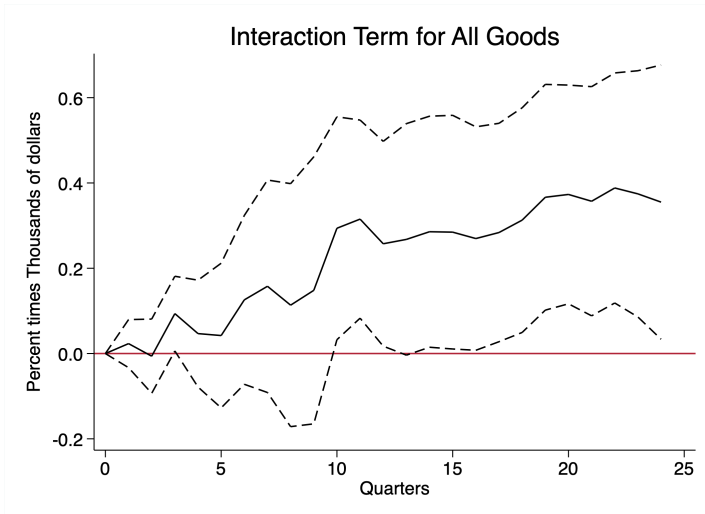
Note: The left panel shows the β^h coefficient and the right panel shows the γ^h coefficient of equation (10) for different price indexes. We use $H = 20$, $J = 8$ and $K = 8$. The dashed lines show 90 percent intervals. Standard errors are clustered at the city level.

Figure A.7: Effect with Sectoral-Time FE



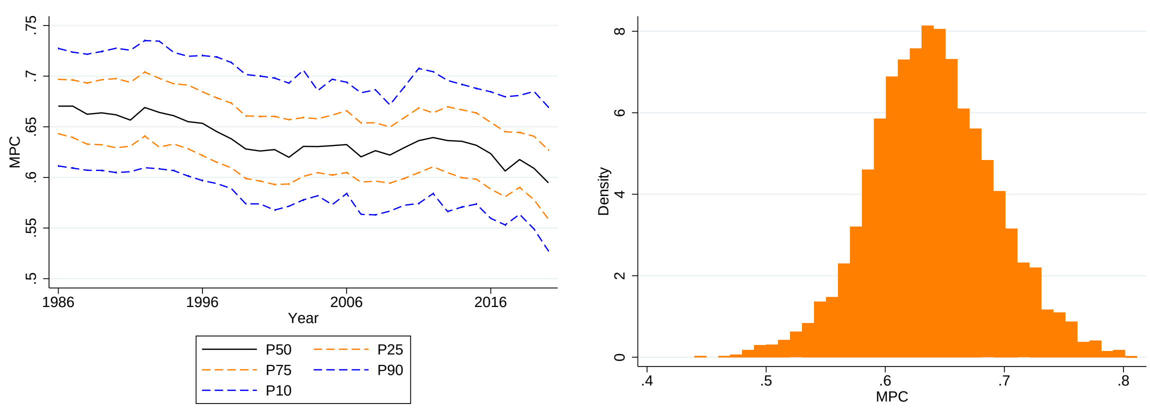
Note: Each figure shows the baseline regression for CPI inflation, controlling by a time fixed effect interacted by the share of employment in the sector indicated in each graph for each city. Agriculture is sector SIC A. Construction is sector SIC C. Manufacturing is sector SIC D and Finance is sector SIC H. We use $H = 20$, $J = 8$ and $K = 8$. The dashed lines show 90 percent intervals. Standard errors are clustered at the city and time level. The dot line shows the baseline regression result.

Figure A.8: Effect with Controls



Note: The figure shows the baseline regression for CPI inflation, controlling by a time fixed effect interacted by the share of employment in agriculture (sector SIC A), construction (sector SIC C), manufacturing is sector (SIC D), and the finance is sector (SIC H). We use $H = 20$, $J = 8$ and $K = 8$. The dashed lines show 90 percent intervals. Standard errors are clustered at the city and time level. The dot line shows the baseline regression result.

Figure A.10: Distribution of MPCs in the US over Time



Note: These figures show the distribution of the marginal propensity to consume across US metropolitan areas and over time. We use the estimates from Patterson (2019) and compute them for each metropolitan area at every period of time. The left panel shows the evolution over time for the mean (solid black), 25th and 75th percentile (orange dashed) and 10th and 90th percentile (blue dashed) between 1986 and 2020. The right panel is a histogram that shows the complete distribution of values and their density for all periods of time and year.

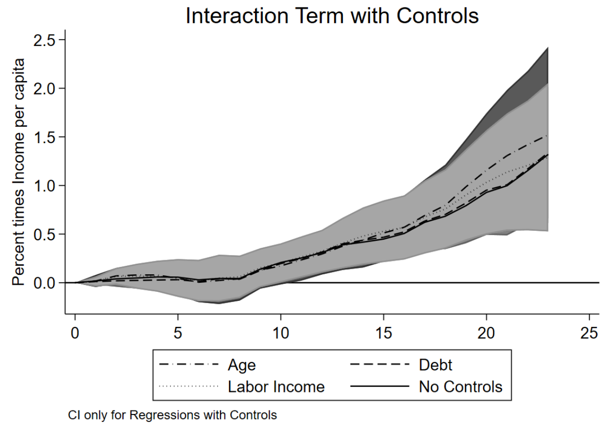


Figure A.9: Interaction Effect of CPI with Controls

Note: The figure shows the baseline regression for CPI inflation, controlling by a time fixed effect interacted by the share of labor income to total household income at the city level (CPS, dotted line). The second variable is the debt to income ratio coming from FRBNY Consumer Credit Panel (dashed). Finally we control by the average age of the city residents (Census, dash-dot). All variables are normalized to zero taking the residual from a regression using a time fixed effect. We use $H = 20$, $J = 8$ and $K = 8$. The dashed lines show 90 percent intervals. Standard errors are clustered at the city and time level. The dot line shows the baseline regression result.

A.2 Correspondence CPI and QCEW

To merge the CPI and employment data, we get the counties according to the FIPS code that match the PSU zones. The PSU zones have changed over time, so we take the larger set of counties, as adding or removing counties would change employment as well. We keep the numbers of counties constant over the sample. Table A.1 shows the correspondence, with the PSU codes and name and FIPS codes.

Table A.1: Commuting zone and equivalent FIPS codes

PSU 18	PSU 98	Name	FIPS			
S11A	A103	Boston-Cambridge-Newton (MA-NH)	25009	25025	25013	23031
			25017	33015	25027	9015
			25021	33017	33011	
			25023	25005	33013	
S12A	A101	New York-Newark-Jersey City (NY-NJ-PA)	34003	34031	36061	42103
			34013	34035	36071	34021
			34017	34037	36079	34041
			34019	34039	36081	9001
			34023	36005	36085	9005
			34025	36027	36087	9007
			34027	36047	36103	9009
			34029	36059	36119	
S12B	A102	Philadelphia-Camden-Wilmington(PA-NJ-DE-MD)	10003	34015	42045	34009
			24015	34033	42091	34011
			34005	42017	42101	
			34007	42029	34001	
S23A	A207	Chicago-Naperville-Elgin (IL-IN-WI)	17031	17089	17197	18127
			17037	17093	18073	55059
			17043	17097	18089	17091
			17063	17111	18111	
S23B	A208	Detroit-Warren-Dearborn, (MI)	26087	26125	26049	26161
			26093	26147	26091	
			26099	26163	26115	
S24A	A211	Minneapolis-St. Paul-Bloomington (MN-WI)	27003	27053	27123	27163
			27019	27059	27139	27171
			27025	27079	27141	55093
			27037	27095	27143	55109
S24B	A209	St. Louis (MO-IL)	17005	17117	29071	29189
			17013	17119	29099	29510
			17027	17133	29113	28149
			17083	17163	29183	29055
S35A		Washington-Arlington-Alexandria (DC-MD-VA-WV)	11000	51510	51061	51179
			24009	51013	51630	51187
			24017	51043	51107	51685
			24021	51047	51153	54037
			24031	51600	51157	
			24033	51610	51177	
S35E		Baltimore-Columbia-Towson (MD)	24003	24510	24025	24035
			24005	24013	24027	

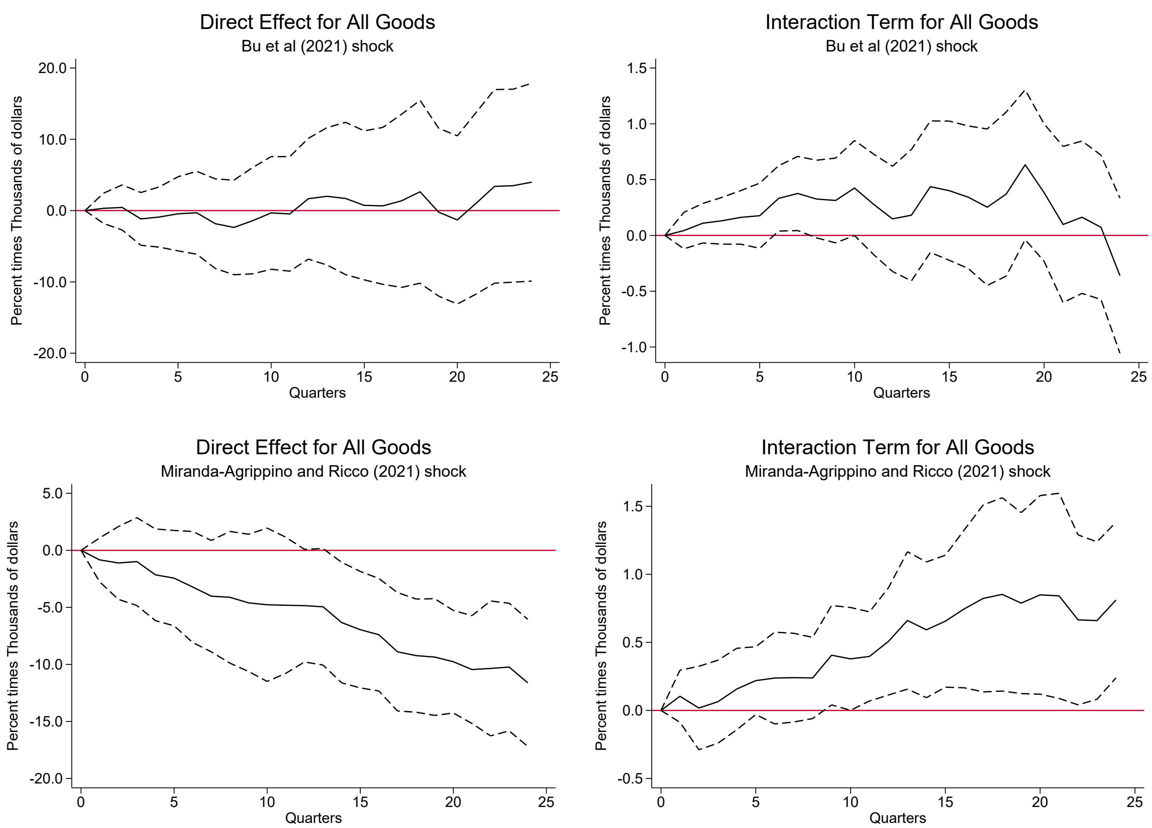
Table A.2: Commuting zone and equivalent FIPS codes (cont)

PSU 18	PSU 98	Name	FIPS			
S35B	A320	Miami-Fort Lauderdale-West Palm Beach (FL)	12011	12025	12086	
S35C	A319	Atlanta-Sandy Springs-Roswell (GA)	13013	13085	13149	13227
			13015	13089	13151	13231
			13035	13097	13159	13247
			13045	13113	13171	13255
			13057	13117	13199	13297
			13063	13121	13211	
			13067	13135	13217	
			13077	13143	13223	
S35D	A321	Tampa-St. Petersburg-Clearwater (FL)	12053	12057	12101	12103
S37A	A316	Dallas-Fort Worth-Arlington (TX)	48085	48221	48367	48497
			48113	48231	48397	
			48121	48251	48425	
			48139	48257	48439	
S37B	A318	Houston-The Woodlands-Sugar Land (TX)	48015	48157	48291	
			48039	48167	48339	
			48071	48201	48473	
S48A	A429	Phoenix-Mesa-Scottsdale (AZ)	4013	4021		
S48B	A433	Denver-Aurora-Lakewood (CO)	8001	8019	8039	8093
			8005	8031	8047	8013
			8014	8035	8059	8123
S49A		Los Angeles-Long Beach-Anaheim (CA)	6037	6059		
S49C		Riverside-San Bernardino-Ontario(CA)	6065	6071		
S49B	A422	San Francisco-Oakland-Hayward (CA)	6001	6075	6085	6097
			6013	6081	6087	
			6041	6055	6095	
S49D	A423	Seattle-Tacoma-Bellevue (WA)	53033	53061	53035	
			53053	53029	53067	
S49E	A424	San Diego-Carlsbad (CA)	6073			
S49F	A426	Urban Hawaii	15003			
S49G	A427	Urban Alaska	2020	2170		
	A104	Pittsburgh (PA)	42003	42019	42125	
			42007	42051	42129	
	A213	Cincinnati-Hamilton (OH-KY-IN)	18029	21077	39015	39165
			18115	21081	39017	
			21015	21117	39025	
			21037	21191	39061	
	A210	Cleveland-Akron (OH)	39007	39055	39093	39133
			39035	39085	39103	39153
	A212	Milwaukee-Racine (WI)	55079	55101	55133	
			55089	55131		
	A425	Portland-Salem (OR-WA)	41005	41047	41053	41071
			41009	41051	41067	53011
	A214	Kansas City (MO-KS)	20091	20209	29049	29165
			20103	29037	29095	29177
			20121	29047	29107	

A.3 Other Shocks

In this Appendix, we run regression (10) for prices, with the interaction on income using different sources of shock. We use the Bu et al. (2021) shock and the Miranda-Agrippino and Ricco (2021) shock. The Bu et al. (2021) is available from 1994 to 2017 in the case of our sample and the Miranda-Agrippino and Ricco (2021) from 1990 to 2015. We plot the direct and indirect effect.

Figure A.11: Effect of Monetary Policy and Income Heterogeneity with Alternative Shocks



Note: The top left and right panel of the figure shows the estimated coefficient $\hat{\beta}^h$ and $\hat{\gamma}^h$ from equation 10, respectively using the Bu et al. (2021) shock. The bottom left and righ panel use the Miranda-Agrippino and Ricco (2021) shock. We use $H = 24$, $J = 8$, and $K = 8$. The relative income per capita numbers are year 2000 dollars. The dashed lines show 90 percent intervals. Standard errors are clustered at the metropolitan area and time level.

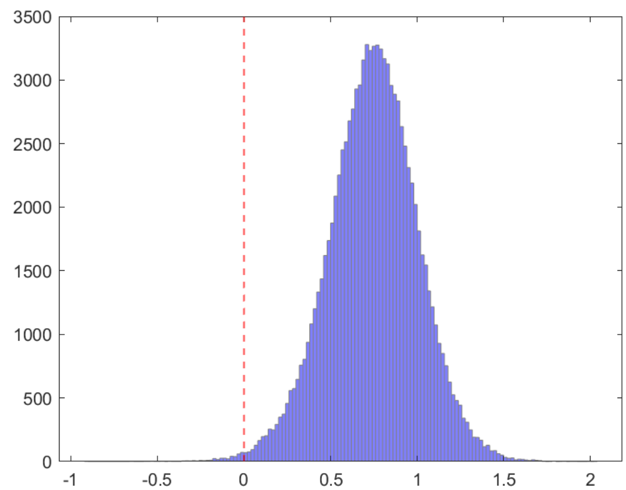
We can see that, despite the direct effect, the interaction term shocks that the effect is milder or more positive for the richer cities, as with the Romer and Romer (2004) shock.

A.4 Robustness Positive Relationship Between Price and Employment result

Figure 3 uses point estimate results of equations 4 and 3. However, Figure 3 does not take into account that each point in the scatter plot is estimated with uncertainty. In this section, we perform robustness exercises to confirm the positive slope, considering the uncertainty around the coefficients.

The plot is built with 26 coefficients for CPI and employment. We assume normal distributions for each coefficient and independence across coefficients. We simulate 100,000 random draws of the coefficients using the standard errors underlying each estimate. For each draw, we run the same regression as in Figure 3 and collect the slope coefficient analogous to the dotted line in Figure 3. Figure A.12 shows the histogram of the estimated slopes. We find that 99.6 percent of the draws give as a result a positive relationship between price and employment effects.

Figure A.12: Result of a Regression for Simulated Coefficients of City Employment and Price Regressions



Note: The figure is an histogram of the coefficients from 100,000 regressions of the city level effect of a monetary policy shock on prices and employment, where those coefficients are built using the all sample point estimate, and the standard deviation of those coefficients. Then, we simulate coefficients independently, using random draws assuming a normal distribution.

Additionally, in this section we formally test the slope of Figure 3 by estimating the relative effect of a monetary policy on inflation relative to the effect on employment.

The local projection of local cumulative inflation on a monetary policy shock takes the form of

$$\pi_{i,t+h,t-1} = \alpha^{h}_{p,i} + \sum_{j=0}^{J} \beta^{h,j}_{p,i} RR_{t-j} + \sum_{k=0}^{K} \gamma^{h,k}_{p} \pi_{i,t-1,t-1-k} + \varepsilon^{h}_{p,i,t+h} \;\; \forall h \in [0, H], \tag{11}$$

where we allow for the effect of the monetary policy shocks on prices to be different for each metropolitan area, see the notation $\beta^{h,j}_{p,i}$.

Similarly the local projection of cumulative employment growth on the monetary policy shock is given by

$$\sum_{\tau=0}^{h} g^{e}_{i,t+\tau,t-1} = \alpha^{h}_{i,e} + \sum_{j=0}^{J} \beta^{h,j}_{e,i} RR_{t-j} + \sum_{k=0}^{K} \gamma^{h,k}_{e} g^{e}_{i,t,t-k} + \varepsilon^{h}_{e,i,t+h} \;\; \forall h \in [0,H], \tag{12}$$

where again, we allow the impact of a monetary policy shock on employment to be different across regions, and notice that the left hand side variable is the area below the curve of the cumulative employment changes.

We add the additional constraint that we want to estimate, a linear relation between the causal effects of the monetary policy shock on prices relative to the causal effect of those same shocks on employment. Formally, we want to estimate for the coefficient φ such that,

$$\sum_{j=0}^{J} \beta^{h,j}_{p,i} RR_{t-j} = \varphi_h \times \left(\sum_{j=0}^{J} \beta^{h,j}_{e,i} RR_{t-j} \right). \tag{13}$$

By replacing equation 13 on equation 11, and replacing equation 12, we find

$\pi_{i,t+h,t-1} = \alpha^{h}_{p,i} + +\varphi_h \alpha^{h}_{i,e} + \varphi_h \sum_{\tau=0}^{h} g^{e}_{i,t+\tau,t-1} - \varphi_h \sum_{k=0}^{K} \gamma^{h,k}_{e} g^{e}_{i,t,t-k} + \sum_{k=0}^{K} \gamma^{h,k}_{p} \pi_{i,t-1,t-1-k} + \varepsilon^{h}_{p,i,t+h} - \varphi_h \varepsilon^{h}_{e,i,t+h} \;\; \forall h \in [0,H],$

which we can represent in a more concise way as

$$\pi_{i,t+h,t-1} = \alpha^{h}_{2s,i} + \varphi_h \sum_{\tau=0}^{h} g^{e}_{i,t+\tau,t-1} - \sum_{k=0}^{K} \gamma^{h,k}_{e,2s} g^{e}_{i,t,t-k} + \sum_{k=0}^{K} \gamma^{h,k}_{2s,p} \pi_{i,t-1,t-1-k} + \varepsilon^{h}_{2s,i,t+h} \;\; \forall h \in [0,H], \tag{14}$$

and we can estimate using the monetary policy shocks as instruments for $\sum_{\tau=0}^{h} g^{e}_{i,t+\tau,t-1}$.

The results for the estimation are in Figure A.13. The figure shows in the y-axis estimates of φ_h for each horizon h between 1 and $H = 20$. In other words, each point represents a slope for a given horizon in a plot similar to Figure 3. The orange area shows the 95% confidence interval. Standard errors are clustered at the city and time dimension.

A.5 TANK Monetary Union

In this appendix we present the log-linearized equations that characterize the model explained in Section 4.1. In the following equations, lower case represents deviation from

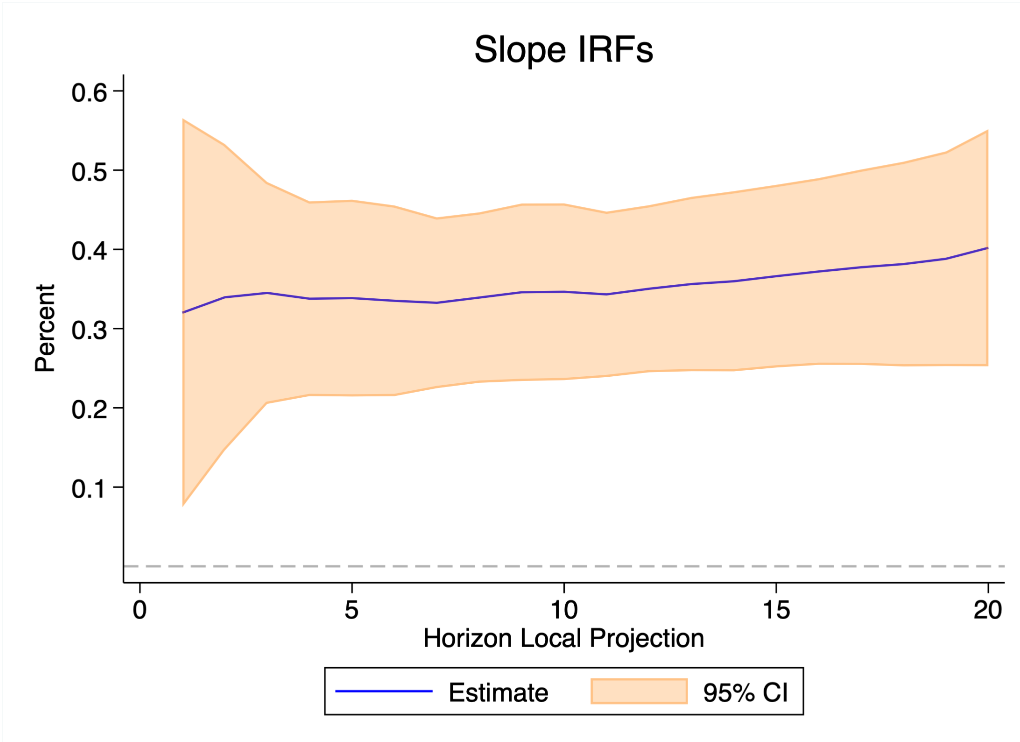


Figure A.13: Slope between the impulse responses of inflation and employment

the steady state, other than for the case of the price index $P_{j,t}$ and the inflation of the price index $\Pi_{j,t}$, to differentiate it from the price of the good produced in j, $p_{j,t}$ and the price inflation $\pi_{j,t}$.

$$\pi_{H,t} = \kappa mc_{H,t} + \beta \pi_{H,t+1}$$

$$\pi_{F,t} = \kappa mc_{F,t} + \beta \pi_{F,t+1}$$

$$c_{HR,t} = -\frac{1}{\gamma}(i_t - \Pi_{H,t+1}) + c_{HR,t}$$

$$c_{HH,t} = w_{H,t} - P_{H,t} + l_{HH,t}$$

$$-\gamma c_{HR,t} + \gamma c_{F,t} = P_{H,t} - P_{F,t}$$

$$i_t = \phi_\pi(\Pi_{H,t} + \Pi_{F,t}) + \phi_y(y_{H,t} + y_{F,t}) + e_t$$

$$P_{H,t} = \phi p_{H,t} + (1-\phi) p_{F,t}$$

$$P_{F,t} = \phi p_{F,t} + (1-\phi) p_{H,t}$$

$$\Pi_{H,t} = P_{H,t} - P_{H,t-1}$$

$$\Pi_{F,t} = P_{F,t} - P_{F,t-1}$$

$$\pi_{H,t} = p_{H,t} - p_{H,t-1}$$

$$\pi_{F,t} = p_{F,t} - p_{F,t-1}$$

$$mc_{H,t} = \alpha y_{H,t} + (\gamma - (1/\nu))c_{H,t} + (1/\nu)(\lambda c_{HH,H,t} + (1-\lambda)c_{HR,H})$$

$$mc_{F,t} = \alpha y_{F,t} + (\gamma - (1/\nu))c_{F,t} + (1/\nu)c_{FF,t}$$

$$y_{H,t} = \lambda l_{HH,t} + (1-\lambda)l_{HR,t}$$

$$\gamma c_{HR,t} + \alpha l_{HR,t} = w_{H,t} - P_{H,t}$$

$$\gamma c_{HH,t} + \alpha l_{HH,t} = w_{H,t} - P_{H,t}$$

$$-c_{FF,t} + c_{FH,t} = \nu(p_{F,t} - p_{H,t})$$

$$-c_{HH,H,t} + c_{HH,F,t} = \nu(p_{H,t} - p_{F,t})$$

$$-c_{HR,H,t} + c_{HR,F,t} = \nu(p_{H,t} - p_{F,t})$$

$$c_{H,t} = \lambda c_{HH,t} + (1-\lambda)c_{HR,t}$$

$$c_{HH,t} = \phi c_{HH,H,t} + (1-\phi)c_{HH,F,t}$$

$$c_{HR,t} = \phi c_{HR,H,t} + (1-\phi)c_{HR,F,t}$$

$$c_{F,t} = \phi c_{FF,t} + (1-\phi)c_{FH,t}$$

$$y_{H,t} = \lambda\phi c_{HH,H,t} + (1-\lambda)\phi c_{HR,H,t} + (1-\phi)c_{FH,t}$$

$$y_{F,t} = \phi c_{FF,t} + \lambda(1-\phi)c_{HH,F,t} + (1-\lambda)(1-\phi)c_{HR,F,t}$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + e_t$$

Table A.3: Parameterization

Parameter	Explanation	Value
β	Discount factor	0.99
γ	Intertemporal elasticity of substitution	1
α	Inverse labor supply elasticity	2/3
η	Elasticity of substitution among local varieties	4
ν	Elasticity of substitution between Home and Foreign varieties	3
θ	Price stickiness	0.75
π_π	Taylor rule coefficient on inflation	1.5
π_y	Taylor rule coefficient on output	0.5
ϕ	Home bias coefficient	0.85
ρ	Monetary policy shock persistence	0

Note: This table presents the calibration of our model for every parameter except for θ and λ, which we vary in our main exercise.

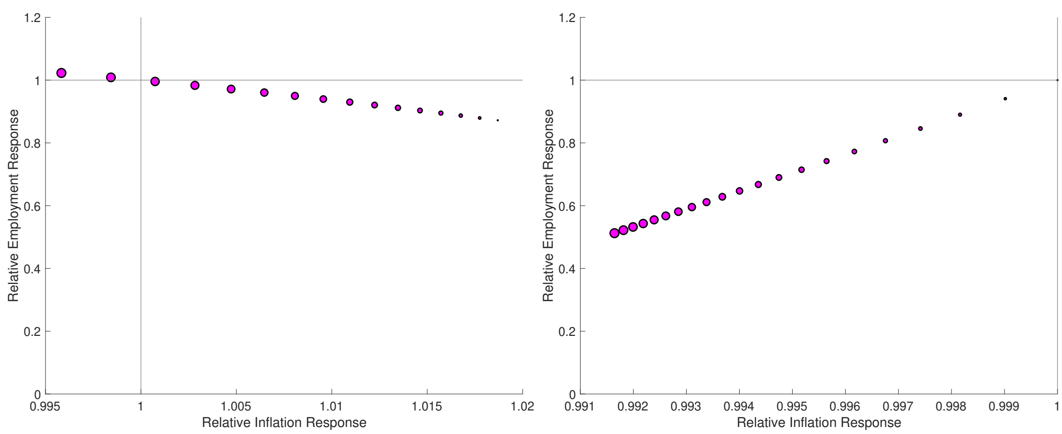


Figure A.14: Relative Price and Employment Responses - Labor Supply and IES

Note: These figures show the relative behavior of regional prices, on the x-axis, and employment, on the y-axis, after a national monetary policy shock. The source of regional heterogeneity is variation in the elasticity of labor supply (left panel) and the intertemporal elasticity of substitution (right panel). Relative inflation and employment are computed as the ratio between the discounted cumulative impulse response functions of each variable in the Home region divided by the analogous object in the Foreign region. A value of 1 means that Home and Foreign regions have responses of the same magnitude in present value. Each point of the scatterplot represents the solution of a model with different variations in the extent of nominal rigidities, labor supply or intertemporal elasticity of substitution. The calibrations that underlie the figure are in Appendix A.6.

A.6 Alternative New Keynesian Models

We simplify the model used in Section 4. In this case, we assume $\lambda = 0$, but we allow for regional heterogeneity in the parameters of the model. The model is characterized by the following equations:

$$\pi_{Ht} = \beta \mathbb{E}_t \pi_{H,t+1} + \kappa_H mc_{Ht} \tag{15}$$

$$\pi_{Ft} = \beta \mathbb{E}_t \pi_{F,t+1} + \kappa_F mc_{Ft} \tag{16}$$

with

$$mc_{Ht} = \alpha_H y_{H,t} + \left(\gamma_H - \frac{1}{\nu}\right) C_{H,t} + \left(\frac{1}{\nu}\right) C_{H,H,t} \tag{17}$$

$$mc_{Ft} = \alpha_F y_{F,t} + \left(\gamma_F - \frac{1}{\nu}\right) C_{F,t} + \left(\frac{1}{\nu}\right) C_{F,F,t} \tag{18}$$

where $C_{k,j,t}$ is the consumption of region k on region j good in time t. Since here $\lambda = 0$, there are only Ricardian agents; then the IS curve is characterized by:

$$C_{H,t} = -\frac{1}{\gamma_H} \left(i_t - E_t \Pi_{H,t+1}\right) + E_t C_{H,t+1} \tag{19}$$

For region F, we replace that condition with the risk-sharing condition (does not really matter which one we replace).

$$\gamma_H C_{H,t} - \gamma_F C_{F,t} = P_{F,t} - P_{H,t} \tag{20}$$

Finally, we have a national monetary policy rule that symmetrically weights both regions:

$$i_t = \phi_\pi(\pi_{Ht} + \pi_{Ft}) + \phi_y(y_{Ht} + y_{Ft}) + \varepsilon_t.$$

In Section 4, we allow for differences in the intertemporal elasticity of substitution γ_i, extent of nominal rigidities κ_i and the elasticity of labor supply α_i.

The values for α and γ we consider are values between 1 and 3. The values for θ that we consider are between 0.6 and 0.9. The benchmark values for these parameters for the Foreign region, which we keep fixed, are $\alpha = 1$, $\gamma = 1$, and $\theta = 0.75$.

A.7 Amplification/Dampening of Local Responses in the Aggregate

A.7.1 Aggregate Implications of Regional Heterogeneity

In this section we explore the aggregate implications of regional heterogeneity in the main parameters of the model. We start by doing it by generating simulation of the model and see the impact effect of monetary policy in the aggregate economy, after modifying the baseline calibration of the model to allow for heterogeneous inter-temporal elasticity of substitution, labor supply elasticity, slope of the Phillips curve and share of hand-to-mouth consumers. For each of these parameters we consider simulations where we generate heterogeneity. In the case of α we consider the baseline ($\alpha = 2/3$), then $\alpha = 3/5$ and finally $\alpha = 1/2$. For γ we consider the baseline ($\gamma = 1$), then $\gamma = 0.9$ and finally $\gamma = 0.7$. For κ we consider the baseline ($\kappa = 0.1$), then $\kappa = 0.09$ and finally $\kappa = 0.7$. Finally, for the share of HtM λ, we consider the baseline ($\lambda = 0$), then $\lambda = 0.3$ and finally $\lambda = 0.5$. These values are somewhat arbitrary, but the objective is to show how the affect the aggregate effect and in which direction they affect both prices and output.

Figure A.15 shows the results of the different simulations. We also consider different parameters for the Taylor Rule coefficient on output.

As in the across cities differences, the inter-temporal elasticity of substitution and the share of HtM create amplification in the aggregate: the more heterogeneity in those variables, the most cost in terms of output and prices for a given monetary policy rule. This also implies that the central bank can design a policy design that reduces the cost in terms of those variables. We also see that heterogeneity in the labor supply elasticity has small

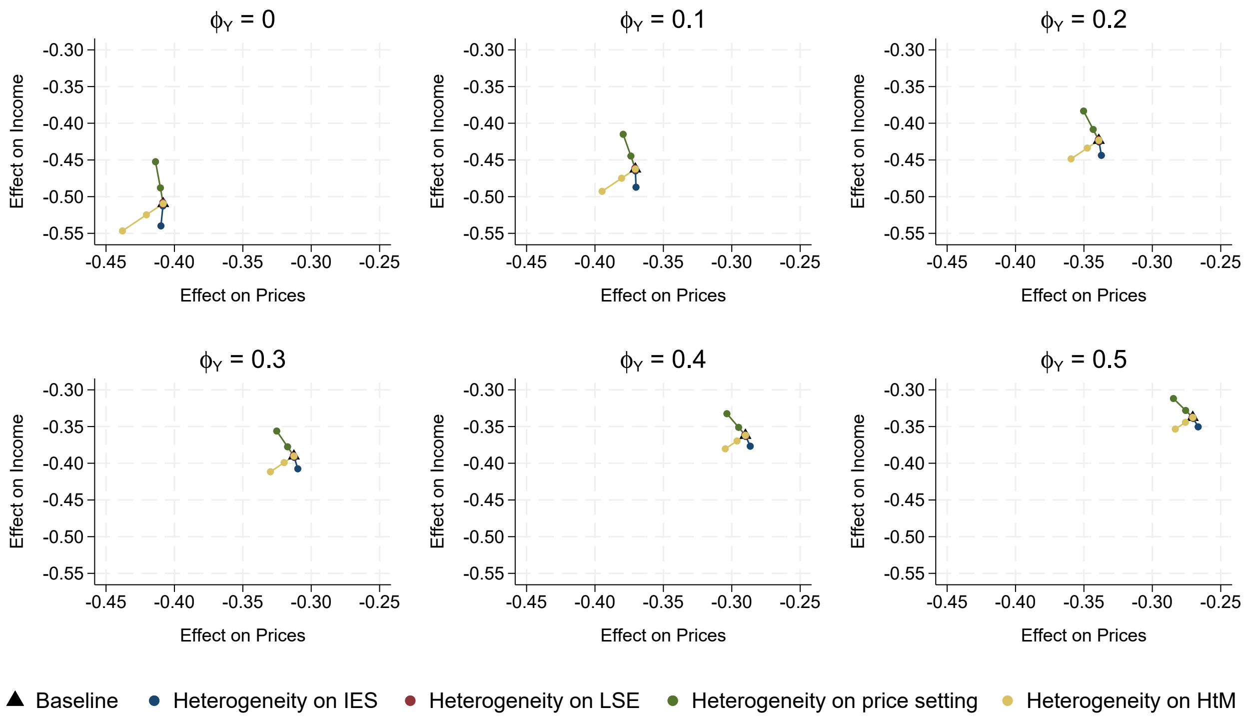


Figure A.15: Aggregate Implications of Regional Heterogeneity

aggregate implications and that the slope of the Phillips curve has sizable aggregate, but with dampening, in that sense, changes in the monetary policy rule can reduce the overall economic effects effect, but the trade-off between output and prices can be exacerbated. We can also see that for the amplifiers the aggregate consequences are lower when the monetary policy weight changes.

We can also see that the inter-temporal elasticity of substitution heterogeneity slope changes. This is because that parameter acts in both the demand and the supply side, through the marginal cost. Our results show that eventually the supply side consequences dominate, given a certain monetary policy reaction. Finally, in the case of the dampening heterogeneity (price setting), the changes in the monetary policy only changes the trade-off between output and prices, but the aggregate consequences are of similar magnitude, as the direction that the heterogeneity acts makes that the monetary authority can't accommodate both output and prices.

A.7.2 Derivation of the system of equations of Model without HtM

We aim to characterize a simple model with “demand” and “supply” heterogeneity and obtain closed-form solutions. We start from log-linearized expressions. The regions are heterogeneous in terms of their supply elasticities λ and their demand elasticities σ. We assume GHH preferences in this block so that σ does not show up in any transformation of the Phillips curve, and the separation of these two structural parameters into supply and demand elasticities is transparent.

The initial block we start from is the following 21 system equation:

$$\pi_{Ht} = \beta \mathbb{E}_t \pi_{H,t+1} + \lambda_H mc_{Ht} \tag{21}$$

$$\pi_{Ft} = \beta \mathbb{E}_t \pi_{F,t+1} + \lambda_F mc_{Ft} \tag{22}$$

$$c_t = \mathbb{E}_t c_{t+1} - \frac{1}{\sigma_H}(i_t - \mathbb{E}_t \pi_{t+1}) \tag{23}$$

$$c_t^* = \mathbb{E}_t c_{t+1}^* - \frac{1}{\sigma_H}(i_t - \mathbb{E}_t \pi_{t+1}^*) \tag{24}$$

$$\pi_t = \phi \pi_{Ht} + (1-\phi)\pi_{Ft} \tag{25}$$

$$\pi_t^* = \phi \pi_{Ft} + (1-\phi)\pi_{Ht} \tag{26}$$

$$mc_{Ht} = \frac{1}{\varphi} l_t - p_{Ht} \tag{27}$$

$$mc_{Ft} = \frac{1}{\varphi} l_t^* - p_{Ft} + q_t \tag{28}$$

$$l_t = y_t \tag{29}$$

$$l_t^* = y_t^* \tag{30}$$

$$i_t = \frac{\psi}{2}(\pi_t + \pi_t^*) + \epsilon_t \tag{31}$$

$$y_t = \phi c_{Ht} + (1-\phi) c_{Ht}^* \tag{32}$$

$$y_t^* = (1-\phi) c_{Ft} + \phi c_{Ft}^* \tag{33}$$

$$c_t = \phi c_{Ht} + (1-\phi) c_{Ft} \tag{34}$$

$$c_t^* = \phi c_{Ft} + (1-\phi) c_{Ht} \tag{35}$$

$$c_{Ht} = c_t - \eta(p_{Ht}) \tag{36}$$

$$c_{Ft} = c_t - \eta(p_{Ft}) \tag{37}$$

$$c_{Ht}^* = c_t^* - \eta(p_{Ht} - q_t) \tag{38}$$

$$c_{Ft}^* = c_t^* - \eta(p_{Ft} - q_t) \tag{39}$$

$$p_{Ht} = \log P_{Ht} - \log P_t \tag{40}$$

$$p_{Ft} = \log P_{Ft} - \log P_t \tag{41}$$

$$q_t = p_t^* - p_t \tag{42}$$

We will try now to reduce the dimensionality of this system.

The first step is to realize that due to the definition of the relative price, which is defined as differences with respect to the local price level, then the following relations hold:

$$p_{Ht} - p_{H,t-1} = \pi_{Ht} - \pi_t \tag{43}$$

$$p_{Ft} - p_{F,t-1} = \pi_{Ft} - \pi_t \tag{44}$$

$$0 = \phi p_{Ht} + (1-\phi) p_{Ft} p_t^* = \phi(p_{Ft} + p_t) + (1-\phi)(p_{Ht} + p_t) \tag{45}$$

$$q_t = \phi p_{Ft} + (1-\phi) p_{Ht} \tag{46}$$

We replace labor for production in the system and the relations we just derived, reducing it further to 17 equations. We then replace the marginal cost into the Phillips curves. The definition of CPI inflation enters into the Euler equations and the monetary policy rule. The monetary policy rule only enters into the Euler equations. Therefore we will replace them into the Euler equation and reduce the system further. Then, to simplify the system further we will work with the two Euler conditions. In their simplest form the local Euler equation takes the form of:

$$c_t = \mathbb{E}_t c_{t+1} - \frac{1}{\sigma_H}(i_t - \mathbb{E}_t \pi_{t+1}) \tag{47}$$

$$\tag{48}$$

We can iterate this equation forward and we will use a couple of relations. The first one states that conditional on the information set of period t, $\mathbb{E}_t c_{t+\infty} = 0$. Moreover in the long run PPP applies, so $\mathbf{E}_t(p_{t+\infty} - p^*_{t+\infty}) = 0$. The iterated forward Euler equation looks like

$$c_t = \mathbb{E}_t c_{t+\infty} - \frac{1}{\sigma_H}\mathbb{E}_t \sum_{j=0}^{\infty}(i_{t+j} - \pi_{t+j+1}). \tag{49}$$

The last infinite sum has some interesting properties. The cumulated sum of inflation rates is just the "long" inflation rate. Specifically,

$$\sum_{j=0}^{\infty} \pi_{t+j+1} = p_{t+\infty} - p_t. \tag{50}$$

Using this property and monetary neutrality in the long run, the iterated forward Euler equation takes the form

$$c_t = -\frac{1}{\sigma_H}\mathbb{E}_t \sum_{j=0}^{\infty} i_{t+j} + \frac{1}{\sigma_H}\mathbb{E}_t(p_{t+\infty} - p_t) \tag{51}$$

By symmetry, for the foreign economy:

$$c_t^* = -\frac{1}{\sigma_F}\mathbb{E}_t \sum_{j=0}^{\infty} i_{t+j} + \frac{1}{\sigma_F}\mathbb{E}_t(p_{t+\infty}^* - p_t^*). \tag{52}$$

From the local economy relation, solve for the infinite interest rate sum

$$\mathbb{E}_t \sum_{j=0}^{\infty} i_{t+j} = -\sigma_H c_t + \mathbb{E}_t(p_{t+\infty} - p_t), \tag{53}$$

and replace it into the relation for the foreign economy

$$c_t^* = -\frac{1}{\sigma_F}\left(-\sigma_H c_t + \mathbb{E}_t(p_{t+\infty} - p_t)\right) + \frac{1}{\sigma_F}\mathbb{E}_t(p_{t+\infty}^* - p_t^*), \tag{54}$$

and simplify:

$$c_t^* = \frac{\sigma_H}{\sigma_F}c_t + \frac{1}{\sigma_F}\mathbb{E}_t(p_{t+\infty}^* - p_t^* - (p_{t+\infty} - p_t)), \tag{55}$$

which using PPP in the long run:

$$c_t^* = \frac{\sigma_H}{\sigma_F} c_t - \frac{1}{\sigma_F}(p_t^* - p_t), \tag{56}$$

or

$$c_t^* = \frac{\sigma_H}{\sigma_F} c_t - \frac{1}{\sigma_F} q_t, \tag{57}$$

We will use this last risk-sharing condition equation in lieu of the foreign Euler equation. Carrying the definition of the price indexes and the definition of consumption bundles is redundant, so we will rewrite the system dropping the consumption bundle definitions. We will plug the demand curves into the only place they appear, the market clearing conditions for local and foreign output. Using the definition of the relation of the relative prices, we can replace away p_F from the system using the relation. After working with this model, we get

$$\pi_{Ht} = \beta \mathbb{E}_t \pi_{H,t+1} + \lambda_H \left(\frac{1}{\varphi} y_t - p_{Ht} \right) \tag{58}$$

$$\pi_{Ft} = \beta \mathbb{E}_t \pi_{F,t+1} + \frac{\lambda_F}{\varphi} \left((1-\phi) + \frac{\sigma_H}{\sigma_F}\phi \right) c_t + \left(\lambda_F + \frac{\phi}{1-\phi} \frac{2\phi - 1}{\sigma_F} + 2\eta\phi \right) p_{Ht} \tag{59}$$

$$c_t = \mathbb{E}_t c_{t+1} - \frac{1}{\sigma_H}((\frac{\psi}{2}(\pi_{Ht} + \pi_{Ft}) + \epsilon_t) - \mathbb{E}_t(\phi \pi_{Ht+1} + (1-\phi)\pi_{Ft+1}))) \tag{60}$$

$$y_t = \left(\phi + (1-\phi)\frac{\sigma_H}{\sigma_F} \right) c_t + \left(\frac{2\phi - 1}{\sigma_F} - 2\eta\phi \right) p_{Ht} \tag{61}$$

$$p_{Ht} - p_{H,t-1} = (1-\phi)(\pi_{Ht} - \pi_{Ft}) \tag{62}$$

A.7.3 Solution method

We will use the Uhlig (1999) method. The method consists on writing the model in terms the following system:

$$G_1 \mathbb{E}_t z_{t+1} + G_2 z_t + G_3 z_{t-1} + G_4 \epsilon_t = 0 \qquad \mathbb{E}_t \epsilon_{t+1} = 0 \tag{63}$$

where ϵ denotes a vector of shocks, and z denotes a vector of endogenous variables.

The method starts by making a guess about the behavior of z, which in this case would be that z follows an autoregressive process, given by

$$z_t = P z_{t-1} + Q \epsilon_t, \tag{64}$$

and accordingly

$$\mathbb{E}_t z_{t+1} = P z_t + Q \mathbb{E}_t \epsilon_{t+1} = P z_t. \tag{65}$$

Replacing this relationship into the original system of equations yields

$$G_1 \mathbb{E}_t P z_t + G_2 z_t + G_3 z_{t-1} + G_4 \epsilon_t = 0 \tag{66}$$

$$G_1 P (P z_{t-1} + Q \epsilon_t) + G_2 (P z_{t-1} + Q \epsilon_t) + G_3 z_{t-1} + G_4 \epsilon_t = 0 \tag{67}$$

$$G_1 P^2 z_{t-1} + G_1 P Q \epsilon_t + G_2 P z_{t-1} + G_2 Q \epsilon_t + G_3 z_{t-1} + G_4 \epsilon_t = 0 \tag{68}$$

$$(G_1 P^2 + G_2 P + G_3) z_{t-1} + (G_1 P Q + G_2 Q + G_4) \epsilon_t = 0 \tag{69}$$

$$\tag{70}$$

and Q and P must be such that

$$(G_1 P^2 + G_2 P + G_3) = 0 \tag{71}$$

$$(G_1 P + G_2) = -G_4 \tag{72}$$

Usually, these solution method is applied over an already calibrated model, but nothing precludes the possibility of finding analytic expressions for P and Q. In particular, notice that these two matrices fully describe the impulse response function of the variables in z after a shock to ϵ. In particular, the entries in Q will characterize the on-impact response of z.

A.7.4 Model with HtM consumers

$$\pi_{Ht} = \beta \mathbb{E}_t \pi_{H,t+1} + \kappa mc_{Ht} \tag{73}$$

$$\pi_{Ft} = \beta \mathbb{E}_t \pi_{F,t+1} + \kappa mc_{Ft} \tag{74}$$

$$c_{R,t} = \mathbb{E}_t c_{R,t+1} - \frac{1}{\sigma}(i_t - \mathbb{E}_t \pi_{t+1}) \tag{75}$$

$$c^*_{R,t} = \mathbb{E}_t c^*_{R,t+1} - \frac{1}{\sigma}(i_t - \mathbb{E}_t \pi^*_{t+1}) \tag{76}$$

$$mc_{H,t} = \frac{1}{\varphi} l_{R,t} - p_{Ht} \tag{77}$$

$$mc_{F,t} = \frac{1}{\varphi} l^*_{R,t} - p_{Ft} + q_t \tag{78}$$

$$mc_{H,t} = \frac{1}{\varphi} l_{NR,t} - p_{Ht} \tag{79}$$

$$mc_{F,t} = \frac{1}{\varphi} l^*_{NR,t} - p_{Ft} + q_t \tag{80}$$

$$c_{NR,t} = \frac{1+\varphi}{\varphi} l_{NR,t} \tag{81}$$

$$c^*_{NR,t} = \frac{1+\varphi}{\varphi} l^*_{NR,t} \tag{82}$$

$$c_t = \lambda_H c_{NR,t} + (1-\lambda_H) c_{R,t} \tag{83}$$

$$c^*_t = \lambda_F c^*_{NR,t} + (1-\lambda_F) c^*_{R,t} \tag{84}$$

$$\pi_t = \phi \pi_{Ht} + (1-\phi) \pi_{Ft} \tag{85}$$

$$\pi^*_t = \phi \pi_{Ft} + (1-\phi) \pi_{Ht} \tag{86}$$

$$l_t = \lambda l_{NR,t} + (1-\lambda) l_{R,t} \tag{87}$$

$$l^*_t = \lambda l^*_{NR,t} + (1-\lambda) l^*_{R,t} \tag{88}$$

$$l_t = y_t \tag{89}$$

$$l^*_t = y^*_t \tag{90}$$

$$i_t = \frac{\psi}{2}(\pi_t + \pi^*_t) + \epsilon_t \tag{91}$$

$$y_t = \phi c_{Ht} + (1-\phi) c^*_{Ht} \tag{92}$$

$$y^*_t = (1-\phi) c_{Ft} + \phi c^*_{Ft} \tag{93}$$

$$c_t = \phi c_{Ht} + (1-\phi) c_{Ft} \tag{94}$$

$$c_t^* = \phi c_{Ft}^* + (1-\phi) c_{Ht}^* \tag{95}$$

$$c_{Ht} = c_t - \eta(p_{Ht}) \tag{96}$$

$$c_{Ft} = c_t - \eta(p_{Ft}) \tag{97}$$

$$c_{Ht}^* = c_t^* - \eta(p_{Ht} - q_t) \tag{98}$$

$$c_{Ft}^* = c_t^* - \eta(p_{Ft} - q_t) \tag{99}$$

$$p_{Ht} = \log P_{Ht} - \log P_t \tag{100}$$

$$p_{Ft} = \log P_{Ft} - \log P_t \tag{101}$$

$$q_t = p_t^* - p_t \tag{102}$$

Due to GHH preferences, it is obvious that labor supplied by Ricardian and non-Ricardian households is the same. Therefore $l_t = l_{R,t} = l_{NR,t}$. We will also replace the production function, so that everything is a function of output and not labor. Similar to the previous model, we will also use that due to the definition of the relative price, which is defined as differences with respect to the local price level. We obtain:

$$\pi_{Ht} = \beta \mathbb{E}_t \pi_{H,t+1} + \kappa \left(\frac{1}{\varphi} y_t - p_{Ht} \right) \tag{103}$$

$$\pi_{Ft} = \beta \mathbb{E}_t \pi_{F,t+1} + \kappa \left(\frac{1}{\varphi} y_t^* + p_{Ht} \right) \tag{104}$$

$$c_{R,t} = \mathbb{E}_t c_{R,t+1} - \frac{1}{\sigma} (\left(\frac{\psi}{2} (\pi_{Ht} + \pi_{Ft}) + \epsilon_t \right) - \mathbb{E}_t \left(\phi \pi_{H,t+1} + (1-\phi) \pi_{F,t+1} \right)) \tag{105}$$

$$y_t = \nu_H \left(\phi (1 - \lambda_H) + (1-\phi)(1-\lambda_F) \right) c_{R,t} + \frac{1-\phi}{\phi} (\nu_F - 1) \frac{\nu_H}{\nu_F} y_t^* + \nu_H \left((1-\phi)\iota - 2\eta\phi \right) p_{H,t} \tag{106}$$

$$y_t^* = \nu_F \left((1-\phi)(1-\lambda_H) + \phi (1-\lambda_F) \right) c_{R,t} + \frac{\nu_F}{\nu_H} (\nu_H - 1) \frac{(1-\phi)}{\phi} y_t + \nu_F \phi (\iota + 2\eta) p_{H,t} \tag{107}$$

$$p_{Ht} - p_{H,t-1} = \pi_{Ht} - (\phi \pi_{Ht} + (1-\phi) \pi_{Ft}), \tag{108}$$

for a constant

$$\nu_H = \left(1 - \frac{\phi \lambda_H (1+\varphi)}{\varphi} \right)^{-1} \tag{109}$$

It would be of course possible to reduce this system forward by plugging y^* in the Home resource constraint and the Foreign Phillips curve, but we will not do that.

A.7.5 Model with HtM consumers and separable preferences

In log-linear form the labor supply curve for Ricardian households is given by:

$$w_t - p_t = \frac{1}{\varphi} l_{R,t} + \sigma c_{R,t}, \tag{110}$$

which in terms of the real marginal cost for local firms

$$mc_{Ht} = \frac{1}{\varphi} l_{R,t} + \sigma c_{R,t} - p_{H,t}. \tag{111}$$

For non-ricardian households the added constraint that consumption expenditures are equal to labor income implies that

$$w_t - p_t = l_{NR,t} \left(\frac{1 + \sigma\varphi}{\varphi(1 - \sigma)} \right) \tag{112}$$

$$mc_{Ht} = \left(\frac{1 + \sigma\varphi}{\varphi(1 - \sigma)} \right) l_{NR,t} - p_{H,t}. \tag{113}$$

Therefore, manipulating these equations and aggregating them with weights λ_H for NR households, and $1 - \lambda_H$ for R households, gives rise to a single equation for local real marginal costs.

Let me introduce a new constant $\tilde{\varphi} = \left(\frac{1+\sigma\varphi}{\varphi(1-\sigma)} \right)$.

Therefore the two labor supply equations are:

$$l_{R,t} = \varphi(mc_{Ht} - \sigma c_{R,t} + p_{H,t}) \tag{114}$$

$$l_{NR,t} = \tilde{\varphi}(mc_{Ht} + p_{H,t}), \tag{115}$$

and multiplying the first equation by $1 - \lambda_H$ and the second equation by λ_H, and adding up, and using that aggregate labor in the local economy is given by $l_t = \lambda_H l_{R,t} + (1 - \lambda_H) l_{NR,t}$.

$$l_t = \varphi_H(mc_{H,t} + p_{H,t}) - (1 - \lambda_H)\varphi\sigma c_{R,t}, \tag{116}$$

for $\varphi_H = \lambda_H \tilde{\varphi} + (1 - \lambda_H)\varphi$. This equation makes obvious that the presence of HtM households changes the effective labor supply elasticity of the local economy, a channel absent from a model with GHH preferences.

The determination of marginal costs imposing that $y = l$, yields:

$$mc_{H,t} = \frac{1}{\varphi_H} y_t + \frac{(1-\lambda_H)\varphi\sigma}{\varphi_H} c_{R,t} - p_{H,t} \tag{117}$$

Using the budget constraint for HtM consumers $c_{NR,t} = mc_{Ht} + p_{Ht} + l_{NR,t}$ plus the labor supply equation yields the following equation for consumption for HtM households:

$$c_{NR,t} = (1+\tilde{\varphi}) \left(\frac{1}{\varphi_H} y_t + \frac{(1-\lambda_H)\varphi\sigma}{\varphi_H} c_{R,t} \right) \tag{118}$$

Imposing these equations, we can characterize the model with separable preferences and hand-to-mouth consumers. The steps that follow are similar to the derivations before, but we include them for completeness.

$$\pi_{Ht} = \beta \mathbb{E}_t \pi_{H,t+1} + \kappa mc_{Ht} \tag{119}$$

$$\pi_{Ft} = \beta \mathbb{E}_t \pi_{F,t+1} + \kappa mc_{Ft} \tag{120}$$

$$c_{R,t} = \mathbb{E}_t c_{R,t+1} - \frac{1}{\sigma}(i_t - \mathbb{E}_t \pi_{t+1}) \tag{121}$$

$$c^*_{R,t} = \mathbb{E}_t c^*_{R,t+1} - \frac{1}{\sigma}(i_t - \mathbb{E}_t \pi^*_{t+1}) \tag{122}$$

$$mc_{H,t} = \frac{1}{\varphi_H} y_t + \frac{(1-\lambda_H)\varphi\sigma}{\varphi_H} c_{R,t} - p_{H,t} \tag{123}$$

$$mc_{F,t} = \frac{1}{\varphi_F} y^*_t + \frac{(1-\lambda_F)\varphi\sigma}{\varphi_F} c^*_{R,t} - p_{F,t} + q_t \tag{124}$$

$$c_{NR,t} = (1+\tilde{\varphi}) \left(\frac{1}{\varphi_H} y_t + \frac{(1-\lambda_H)\varphi\sigma}{\varphi_H} c_{R,t} \right) \tag{125}$$

$$c^*_{NR,t} = (1+\tilde{\varphi}) \left(\frac{1}{\varphi_F} y^*_t + \frac{(1-\lambda_F)\varphi\sigma}{\varphi_F} c^*_{R,t} \right) \tag{126}$$

$$c_t = \lambda_H c_{NR,t} + (1-\lambda_H) c_{R,t} \tag{127}$$

$$c^*_t = \lambda_F c^*_{NR,t} + (1-\lambda_F) c^*_{R,t} \tag{128}$$

$$\pi_t = \phi \pi_{Ht} + (1-\phi)\pi_{Ft} \tag{129}$$

$$\pi^*_t = \phi \pi_{Ft} + (1-\phi)\pi_{Ht} \tag{130}$$

$$i_t = \frac{\psi}{2}(\pi_t + \pi^*_t) + \epsilon_t \tag{131}$$

$$y_t = \phi c_{Ht} + (1-\phi) c^*_{Ht} \tag{132}$$

$$y^*_t = (1-\phi) c_{Ft} + \phi c^*_{Ft} \tag{133}$$

$$c_t = \phi c_{Ht} + (1-\phi) c_{Ft} \tag{134}$$

$$c^*_t = \phi c^*_{Ft} + (1-\phi) c^*_{Ht} \tag{135}$$

$$c_{Ht} = c_t - \eta(p_{Ht}) \tag{136}$$

$$c_{Ft} = c_t - \eta(p_{Ft}) \tag{137}$$

$$c^*_{Ht} = c^*_t - \eta(p_{Ht} - q_t) \tag{138}$$

$$c^*_{Ft} = c^*_t - \eta(p_{Ft} - q_t) \tag{139}$$

$$p_{Ht} = \log P_{Ht} - \log P_t \tag{140}$$

$$p_{Ft} = \log P_{Ft} - \log P_t \tag{141}$$

$$q_t = p^*_t - p_t. \tag{142}$$

We obtain

$$\pi_{Ht} = \beta \mathbb{E}_t \pi_{H,t+1} + \kappa \left(\frac{1}{\varphi_H} y_t + \frac{(1-\lambda_H)\varphi\sigma}{\varphi_H} c_{R,t} - p_{H,t} \right) \tag{143}$$

$$\pi_{Ft} = \beta \mathbb{E}_t \pi_{F,t+1} + \kappa \left(\frac{1}{\varphi_F} y_t^* + \frac{(1-\lambda_F)\varphi\sigma}{\varphi_F} c_{R,t} + \frac{\phi\varphi_F - (2\phi-1)\lambda_F\varphi}{(1-\phi)\varphi_F} p_{H,t} \right) \tag{144}$$

$$c_{R,t} = \mathbb{E}_t c_{R,t+1} - \frac{1}{\sigma}\left(\left(\frac{\psi}{2}(\pi_{Ht} + \pi_{Ft}) + \epsilon_t \right) - \mathbb{E}_t \left(\phi \pi_{H,t+1} + (1-\phi)\pi_{F,t+1} \right) \right) \tag{145}$$

$$c_t = \lambda_H \frac{(1+\tilde{\varphi})}{\varphi_H} y_t + (1-\lambda_H) \left(1 + \frac{\lambda_H(1+\tilde{\varphi}\varphi\sigma)}{\varphi_H} \right) c_{R,t} \tag{146}$$

$$c_t^* = \frac{\lambda_F(1+\tilde{\varphi})}{\varphi_F} y_t^* + \frac{(1-\lambda_F)}{\varphi_F} \left(\lambda_F (1+\tilde{\varphi})\varphi\sigma + \varphi_F \right) c_{R,t} + \frac{(1-\lambda_F)(2\phi-1)}{\sigma(1-\phi)\varphi_F} \left(1 + \lambda_F(1+\tilde{\varphi})\varphi\sigma \right) p_{H,t} \tag{147}$$

$$y_t = \phi c_t + (1-\phi) c_t^* - \frac{\eta\phi}{1-\phi} p_{Ht} \tag{148}$$

$$y_t^* = (1-\phi) c_t + \phi c_t^* + \eta p_{Ht} \tag{149}$$

$$p_{Ht} - p_{H,t-1} = \pi_{Ht} - \pi_t \tag{150}$$

With the model in this form, we can compute partial derivatives of the equilibrium impulse response functions in the model with respect to a parameter of interest. With this machinery we can compute the effect of model parameters on local and national impulse responses according to equations 64 and 65

Share of HtM consumers In the first exercise we set $\lambda_H = \lambda + \delta$, and $\lambda_F = \lambda - \delta$, so δ has the interpretation of a mean-preserving spread on the share of HtM consumers across regions keeping constant the national share, or an increase in the dispersion of HtM consumers. The rest of the parameters of the model are kept the same across regions.

The following plots will show the derivative of local and foreign impulse responses to δ evaluated at $\delta = 0$. We calibrate the model as in the main text with the exception of setting $\lambda = 0.1$ so that we can introduce a mean-preserving spread around it. Formally the plots are showing $\vartheta_{x,\epsilon}^h$, the partial derivative with respect to δ of the impulse response of variable x with respect to a monetary policy shock ϵ at horizon h of the impulse response function. Formally,

$$\frac{d\vartheta_{x,i}^h}{d\delta} = \frac{d}{d\delta}\frac{dx_{t+h}}{d\epsilon_t}. \tag{151}$$

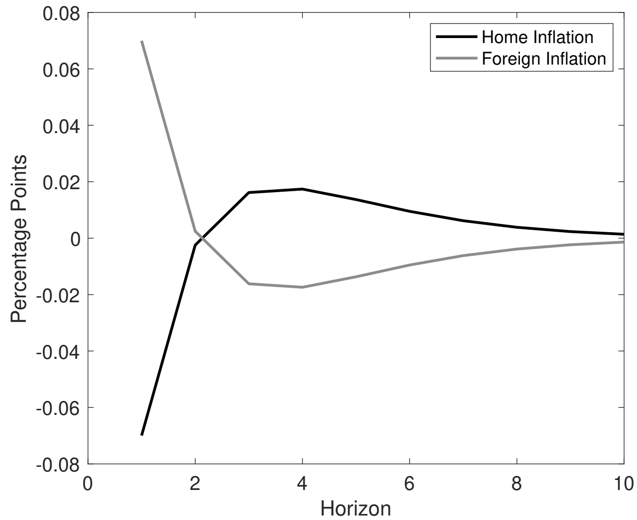


Figure A.16: Derivatives of local inflation with respect to the share of hand-to-mouth consumers

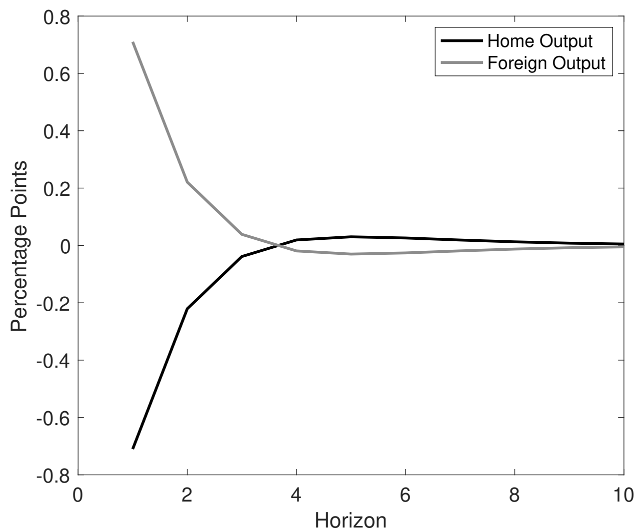


Figure A.17: Derivatives of local output with respect to the share of hand-to-mouth consumers

Figure A.16 shows that after a monetary policy tightening, the IRF of home inflation gets amplified when it has a higher share of HtM consumers, and the IRF of Foreign inflation becomes dampened when it has a lower share.

Figure A.17 shows similar results. Local output falls by more, and Foreign output falls by less after a monetary tightening when there is a reallocation of the mass of hand-to-mouth consumers across space.

We next explore the aggregate implications of this heterogeneity. We compute the

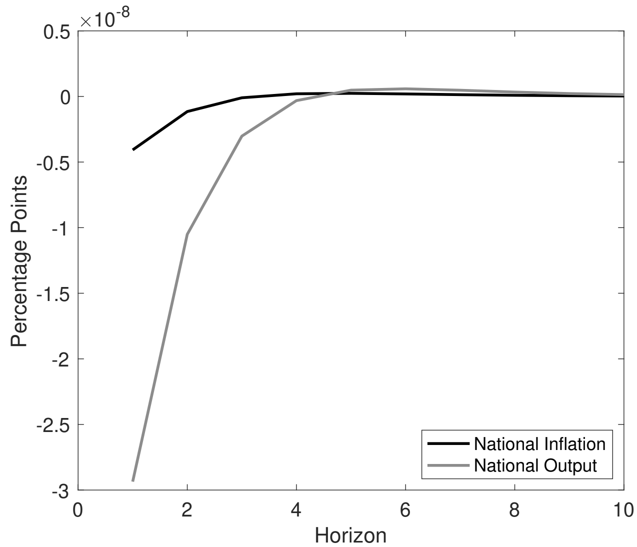


Figure A.18: Derivatives of national inflation and output with respect to the share of hand-to-mouth consumers

derivatives of the national output and inflation impulse responses, which we present in Figure A.18. We can see that increased dispersion in the share of hand-to-mouth consumers increases the aggregate response of national inflation and output in a very standard New Keynesian model. Naturally, since increased heterogeneity creates a reallocation of production and increases in prices across space, the national responses are dampened with respect to the local derivatives.

These results confirm the results of the simulation in in Section A.7.1

A.8 Estimation of Model Parameters using a Simulated Method of Moments

In this Appendix we describe our approach for inferring the model parameters that replicate the cross-sectional slope between employment and price effects.

We conduct a Simulated Method of Moments (SMM) on model-simulated data, and minimize the distance between three data moments and their data counterparts. These three moments are the slope between employment and price responses shown in Figure 3 in the body of the paper, and the cross-sectional dispersion of cumulative prices and employment effects of a monetary policy shock. Notice than in a model without heterogeneity, the dispersion in impulse responses in nominal and real variables would be zero and the slope between price and employment responses would not be well-defined.

We divide the parameter vector of the model, which we will call Θ into two subsets, $\Theta = \Theta_1 \cup \Theta_2$, where $\Theta_2 = \{\theta, \lambda\}$, and Θ_1 is the set of all other structural parameters in the model.

Before proceeding, it is worth highlighting two considerations that complicate the link between our empirical results and the Impulse Response Functions (IRFs) implied by standard New Keynesian models. The New Keynesian model has the feature that IRF of national quantities will tend towards zero as the horizon of the IRF increases. The IRF of inflation will be given by discounted sum of future expected output gaps, which given the shape of the output IRF also implies that the IRF of inflation will decay towards zero. All these patterns are rejected by the data. To complicate things further, a New Keynesian model of a monetary union implies that price differences induced by a monetary policy shock must disappear in the long run. The reason is that PPP deviations across locations are a relative (real) prices that affect allocations even in models with flexible prices.

These two considerations make the mapping between the empirics and the model complicated, since the local projections we estimate exhibit hump-shapes, and, our estimates imply that, as far as our impulse responses go, relative price differences across places do not close down, which implies very persistent effects of nominal disturbances across space.

Given these two considerations, we make the decision of setting Θ_2 in order to minimize the distance between our "on impact effect" and the cumulative effects in the data. The rational is that the model and the data imply different timing patterns of when the relative effects reach their highest value. Since the data implies that relative effects are the highest at the end of the horizon, and the model implies that the relative effects are highest on impact, we target these effects at different horizons.

Now formally, we set $\{\theta, \lambda\}$ in order to

$$\min_{\theta,\lambda} S' W \tilde{S}(\Theta_1), \tag{152}$$

and we use $W = I$, the identity matrix.

Notice that, as in our baseline model, we set the share of hand-to-mouth consumers to zero, so the coefficient λ should be interpreted as the difference in the share of hand-to-mouth consumers across space. As in our benchmark calibration, one period in the model is meant to represent one quarter.

The result of the SMM model are that $\{\hat{\lambda}, \hat{\theta}\} = \{0.249, 0.635\}$. Notice that these two parameters are calibrated using cross-sectional moments across regions, not properties of the national impulse responses. Our estimate for θ implies that at the monthly frequency 86% of prices remain unchanged. A monthly frequency of price changes of 14%, inside the range of average mean and median frequency of price changes reported by Nakamura and Steinsson (2008). Our estimate of the difference in the share of hand-to-mouth consumers across regions of roughly 25% is also qualitatively in line with our inferred dispersion of the share of hand-to-mouth consumers coming from the P10-P90 difference across regions using data from the CPS and the estimates from Patterson (2019) that we presented in Figure A.10.