Essays on Nominal Rigidities, Bounded Rationality, and Macroeconomic Policy

by

Mikel Petri Castro

M.Sc., London School of Economics (2015)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2020

Author . . . .
Department of Economics
July 8, 2020

Certified by . . . .
Iván Werning
Robert M. Solow Professor of Economics
Thesis Supervisor

Certified by . . . .
Daron Acemoglu
MIT Institute Professor
Thesis Supervisor

Accepted by . . . .
Amy Finkelstein
John & Jennie S. MacDonald Professor of Economics
Chairman, Department Committee on Graduate Studies

Essays on Nominal Rigidities, Bounded Rationality, and Macroeconomic Policy

by

Mikel Petri Castro

Submitted to the Department of Economics
on July 8, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This thesis consists of three chapters about macroeconomic policy.

In the first chapter, I study the empirical relationship between nominal rigidities and the real effects of monetary policy. Nominal rigidities lie at the core of macroeconomics. The empirical evidence suggests that prices and wages adjust sluggishly to aggregate shocks, while theoretical models justify why and to what extent these rigidities imply monetary non-neutrality. However, direct evidence on nominal rigidities being the actual channel for the transmission of these shocks is relatively scarce. I construct a highly disaggregated measure of regional price stickiness for the U.S. and use it to provide evidence of this channel. My results are in line with sticky price models, indicating that employment in more rigid industries and commuting zones tend to have stronger reactions to monetary policy shocks.

In the second chapter, joint with Emmanuel Farhi and Iván Werning, we document the extreme sensitivity of New Keynesian models to fiscal policy announcements during a liquidity trap—a phenomenon we call the "fiscal multiplier puzzle". The response of current output to government spending grows exponentially in the horizon of the stimulus. Surprisingly, the introduction of rule-of-thumb hand-to-mouth agents, combined with deficit-financed stimulus, can easily generate negative multipliers that are equally explosive. This intuition translates to incomplete markets heterogeneous-agent New Keynesian models, leading to large negative multipliers when taxes are backloaded. We construct a belief-augmented New Keynesian framework to understand the role played by expectations in shaping the fiscal multiplier puzzle. The key element behind this result is the extreme coordination of the demand and supply blocks under rational expectations. Common knowledge between these two blocks induces an inflation-spending feedback loop. Government spending boosts aggregate demand and drives up inflation, which in turn leads to lower real rates and higher spending by households, increasing aggregate demand again. We break this strategic complementarity by introducing bounded rationality in the form of level-k thinking. In contrast to rational expectations, level-k multipliers are bounded and tend to zero over infinite horizons for all finite k. Moreover, level-k interacts strongly with incomplete markets in two different ways. First, the attenuation of the multipliers increases for any level of k on the degree of market incompleteness, especially in the future. Second, in contrast to complete markets, incomplete markets increase the magnitude of the multipliers for low levels of k when taxes are

backloaded, making deficits more effective at stimulating the economy.

In the third chapter, I explore the implications of downward nominal wage rigidities for fiscal policy and inflation in a liquidity trap. The standard Phillips Curve predicts big declines in economic activity should be accompanied by big deflation episodes. I study whether downward nominal wage rigidity can explain the missing deflation during the Great Recession. To do so, I introduce wage rigidity in a standard cash-in-advance liquidity trap model. My results show that nominal wage rigidities are consistent with mild deflationary episodes only when the trap is expected to be very short-lived. Away from this case, the model predicts large deflations and drops in output as in standard New Keynesian models. I also study the impact of fiscal policy in my setup, finding large multipliers that increase with the degree of wage rigidity. The main reason behind the effectiveness of government spending is its persistent effects on economic activity. Wage rigidity generates unemployment persistence due to pent-up wage deflation. Fiscal spending boosts aggregate demand and decreases deflationary pressures today. This increases output today and in the future by relaxing the downward wage rigidity constraint in all subsequent periods.

Keywords: nominal rigidities, price stickiness, monetary policy, regional, bounded rationality, incomplete markets, level-k, fiscal policy, downward nominal wage rigidity.
JEL Classification: E52, E62, E7.

Thesis Supervisor: Iván Werning
Title: Robert M. Solow Professor of Economics

Thesis Supervisor: Daron Acemoglu
Title: MIT Institute Professor

Acknowledgments

This thesis marks the end of a long road since I started my post-secondary education in Barcelona in 2007. It would have never been written without the love and continuous support of my family and friends. Being away from most of them was certainly the toughest part of my experience at MIT.

I am incredibly indebted to my advisors Iván Werning and Daron Acemoglu for their invaluable guidance and support. My admiration for them has only grown over time. I would have never finished this thesis without their help and support, particularly in the last steps of this process. I thank them for never pushing me into a particular path and supporting me in all the choices I made.

I am gratetful to all the teachers who directly or indirectly helped me to find my way to MIT. I thank Antón Aubanell for passing his passion for math and teaching on to me. I am grateful to Charo González for teaching me to be a critical thinker and develop a deep respect for humanities. I thank Judith Panadés for her passion in teaching economics. She planted the seed that eventually precipitated my transition to the field. I thank Xavier Sala-i-Martín for teaching me to think like an economist and inspiring me to become the best version of myself. Finally, I am grateful to Jordi Galí and Jaume Ventura for believing in me and paving my path to MIT.

MIT is the most intellectually rewarding place one could ever dream of. But life here would have been much less enjoyable without the company of good friends and colleagues. During this time, I met extraordinary people who influenced my thoughts and challenged my priors every time we talked. I will never forget the regular lunches with Ben Deaner, Ivan Badinski, Anton Popov, Pari Sastry, Layne Kirshon, Claire Lazar, Patricia Anghel, and many others. I am especially grateful to Mayara Silva and Masao Fukui for being the best possible office mates. Our converstaions about life, academia, and research shaped who I am today. In particular, I am grateful to Masao Fukui for all our lengthy discussions about this thesis. Special thanks go to Ari Bronsoler, Marc de la Barrera and Jaume Vives, whose friendship have made my social life at MIT worth living. They were constant sources of laughter, joy, and support, and I genuinely wish we had crossed paths earlier.

Finally, I am thankful for the endless suport of my family. The recurring conversations with my mum, Margarita Castro, and my brother, Manel Petri, have been crucial in maintaining my sanity. I am particularly grateful to my mum for always putting us and our education above everything else. Most of all, I am grateful to my wife, Xaidé Cáceres, for being there all the way through the Ph.D. I am extremely grateful to her for leaving everything behind and joining me in the pursuit of my dreams. She has been the bedrock that supported all the goods and bads of this long journey. Cheering me up in each and every setback, and sharing all my celebrations and successes. I would not have made it through graduate school without her constant love and support. She also taught me how to turn this hard experience into a shared dream by getting the most out of it. I keep learning from her every day and I look forward to the next chapter of our life in California. I cannot wait to discover the exciting dreams we manage to achieve together in the decades to come.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Regional and Sectoral Price Rigidity and Monetary Non-Neutrality: Evidence From Monetary Policy Shocks

1.1 Introduction

A central topic in the macroeconomics research revolves around studying the real effects of monetary policy. There is a broad consensus in the empirical literature establishing that monetary shocks do have sizable and persistent effects on the real economy—see Ramey (2016) for a survey on the literature. In light of these facts, macroeconomists have developed a series of theories that try to explain the causes behind this so-called monetary non-neutrality. Among these, nominal rigiditiy in the form of price or wage stickiness is arguably the leading mechanism present in most macroeconomic models, including the workhorse New Keynesian framework (Woodford, 2003; Galí, 2015). Researchers have found massive evidence backing up the idea that price changes are relatively infrequent.[1] Meanwhile, vast amounts of theoretical work have focused on understanding why and to what extent these micro rigidities imply aggregate monetary non-neutrality. However, direct evidence on nominal rigidities being the actual channel for the transmission of monetary shocks to the real economy is surprisingly relatively scarce.

In this paper, I investigate the relationship between price stickiness and the real effects of

[1] See for example Taylor (1999) for an early survey, Bils and Klenow (2004) and Nakamura and Steinsson (2008) for a more recent look into the U.S. microdata, and Dhyne et al. (2006) for a similar paper using European data.

monetary policy in the U.S. between 1995 and 2012 using regional variation in the degree of price rigidities. For this purpose, I construct a novel regional measure of nominal rigidities at the commuting zone level. According to macroeconomic theory, regions and sectors with higher degrees of price stickiness should have stronger responses to monetary policy shocks. Intuitively, the adjustment to an inflationary shock would immediately translate into higher prices in a neoclassical framework, leaving output unchanged. In the presence of nominal rigidities, however, prices cannot adjust immediately and output tends to increase in the short run. The size of the output response depends crucially on the capacity of prices to adjust to the shock, creating larger real effects when prices are more sticky. I test this hypothesis with data on monetary shocks from Romer and Romer (2004). They use a mix of a narrative and quantitative approach to identify their shocks, by isolating intended interest rate changes that cannot be explained by the forecasts about future economic conditions available to the Fed at the time of the FOMC meetings. I complement these results with monetary shocks from Nakamura and Steinsson (2018b), who construct their series using high-frequency changes in the futures markets happening in tight windows around the FOMC announcements.

The magnitude of the monetary non-neutrality in a wide class of models featuring nominal rigidities can be approximated by a sufficient statistic: the average duration of posted prices.[2] Consequently, I construct my regional stickiness index in two steps. First, I put together a measure of industry price duration by manually matching data on frequency of price adjustment for different CPI good categories from Nakamura and Steinsson (2008). Second, I define my regional duration index as a weighted average of the price durations of the industries located in each commuting zone. I find regional duration is fairly concentrated across commuting zones, with two thirds of my sample having durations between 15.5 and 20.2 months. This variation does not seem to be driven by stark differences in industry specialization, but rather different concentrations of sticky and flexible industries *within* 12 broad industry categories. Overall, regional price stickiness varies substantially more across commuting zones than across time. In geographic terms, average commuting zone duration and long-run duration changes are not particularly clustered over space, and do not seem correlated with other industry trends such as exposure to imports from China or automation.

Focusing on regional variation in price stickiness presents some advantages when studying the relationship between nominal rigidities and monetary non-neutrality. First, the United States forms a monetary union with common monetary policy across all its territories. Thus,

[2] Carvalho (2006) shows that this is the case in the context of a New Keynesian model with pricing à la Calvo (1983). In a more general result, Alvarez et al. (2016) prove that the degree of monetary non-neutrality in a wide class of state- and time-dependent pricing models can be approximated by the average duration and the kurtosis of the distribution of price changes.

we only need to focus on the *differential* response to a well-identified monetary shock, without worrying about comparability in the way monetary policy is conduced or the identification procedure of monetary shocks for different countries. Second, regions in the U.S. have lots of elements in common, such as a similar institutional setup and culture, reducing concerns about omitted variables bias and data comparability to some degree.

I use local projection methods to estimate the differential response of regions to monetary shocks induced by regional price duration. My findings suggests that sticky regions react more negatively to monetary shocks. Specifically, I find that one more month of price duration is associated with 3 to 10 basic points of relative employment reduction after an interest rate hike of 25 basic points. These results are robust to various outliers treatments, changes in the construction of my duration index, controlling for broad industry composition, and using different specifications and monetary shocks. I further show that these results hold at the industry level, with sticky industries having a more pronounced responses to monetary policy shocks. Finally, I crosscheck that my duration index captures the degree of nominal rigidity by exploring the reaction of inflation in these industries. These results are in line with the real estimates, with relatively flexible industries having stronger inflation reactions to monetary shocks. This suggests that my duration index is indeed capturing nominal rigidities as the transmission channel for monetary policy.

Related literature. Despite the prevalence of price stickiness as the main friction that generates monetary non-neutrality in business cycle models, there are only a few papers that directly assess the empirical relationship between them. One important exception is De Ridder and Pfajfar (2017) who, in a similar spirit to my paper, calculate the degree of downward nominal wage rigidity across U.S. states and finds that unemployment responses to different demand shocks tend to be more pronounced for states with relatively rigid wages. In contrast, I focus on the degree of price stickiness and I am able to obtain a much more disaggregated measure at the commuting zone level. This is, to the best of my knowledge, the first attempt in the literature to create a regional index of price rigidity. Another paper in this line is Olivei and Tenreyro (2007), who note that wage contracts tend to be renegotiated in the second half of the year, so we would expect monetary shocks to have different effects around the year because of varying levels of nominal rigidity. Indeed, they find suggestive evidence that output responses are smaller in the second half of the year from an estimated time-varying SVAR model. Pischke (2018) investigates the reaction of three real estate occupations with different wage arrangements to the housing cycle, finding that flexible occupations react less to cycle. Leaning towards the microeconomic side, Gorodnichenko and Weber (2016) study the effect of price stickiness at the firm level, finding that the conditional volatility of stock market returns after monetary policy announcements rises more for firms with stickier prices.

This paper also contributes to the literature on the regional impact of macroeconomic shocks. Carlino and DeFina (1998) uses a structural VAR model to examine the effect of monetary policy shocks on states and the eight BEA regions. They find that most aggregate regions have similar responses, while state-level evidence suggests the magnitude of the response is positively related to the share of manufacturing. Arnold (2001) replicates their analysis in the European Monetary Union, with similar results. Beraja et al. (2018) study the regional distribution of housing equity and its relationship with the mortgage refinancing channel of monetary policy. They find that the refinancing was weakest in the regions with the worst housing and labor market conditions. On the fiscal policy side, Nakamura and Steinsson (2014) exploit regional variation in military government spending and use the differential responses to estimate the aggregate effects of government spending.

Lastly, my work is related to the empirical literature on the sectoral effects of monetary shocks. Bils et al. (2003) use monetary shocks identified by a separately estimated structural VAR model and find that sectoral responsess of prices and output are inconsistent with a standard model of nominal rigidities. Boivin et al. (2009), on the other hand, estimate factor-augmenter VAR model including industry data. They point out that the strategy from Bils et al. (2003) gives raise to a substantial price puzzle, with inflation falling after an expansionary monetary shock, suggesting misspecification and monetary shocks that are not well identified. Their results support models with sticky prices, with flexible prices falling relatively more than sticky prices after a contractionary monetary policy shock. Bouakez et al. (2014) provide further evidence that sectors with more frequent price changes respond more rapidly to monetary policy by estimating a structural multi-sector model.

The rest of the paper is structured as follows. Section 1.2 introduces the different data sources used in the paper. Section 1.3 motivates and describes the construction of my regional index of price rigidity and presents its main characteristics. Section 1.4 describes the empirical methodology. Section 1.5 presents my results and various robustness checks. Section 1.6 concludes.

1.2 Data

In this section, I introduce the different data sources used in the paper.

1.2.1 Frequency of Price Changes

My main data consists of frequencies of price changes for different good categories included in the calculation of the CPI. These were estimated by Nakamura and Steinsson (2008), which I henceforth refer to as NS, using confidential microdata from the CPI Research Database at the BLS from 1998 to 2005.[3] The BLS collects samples of prices on thousands of goods and services every month and aggregates them into 264 categories to construct the CPI, called Entry Level Items (ELI).[4] NS use these data to construct two measures of the frequency of price adjustment for identical items at the ELI level. They note that a considerable share of price changes are temporary and go back to the previous price quite rapidly, corresponding to temporary or seasonal sales. These transient changes are generally considered to be unrelated to the actual desire of firms to change their prices, so they do not contribute to relaxing monetary non-neutrality (Kehoe and Midrigan, 2015; Nakamura and Steinsson, 2013). For that reason, they use several filters to identify these sales events and provide two types of frequency measures: a raw frequency including sales, and a filtered frequency excluding them. For my main analysis, I will focus on the frequency version that filters out sales unless otherwise stated.

Figure 1.1 plots the raw and filtered frequencies for every ELI code. Note the extreme heterogeneity in the frequency of price adjustment across different ELIs. The distribution is also skewed, so the (unweighted) mean frequency of adjustment is higher than the median. In addition, there seems to be an apparently large difference between both measures of price flexibility. Figure 1.2 plots the frequencies separately for services, durables, and non-durables. Notably, both measures of frequency are virtually identical for services, revealing that the difference between both measures is essentially driven by sales events in the goods sector. Still, both variables are very correlated, with a significant Pearson coefficient of 0.78.

This figure also suggests that all three expenditure categories seem to have a similar degree of stickiness, but this depends on the measure of frequency and whether expenditure weights are taken into account. Table 1.1 reports several weighted and unweighted descriptive statistics. Panel B shows that using expenditure weights, durables are the most flexible category according to both measures of frequency. But this is no longer true in

[3]The data is publicly available in Table 19 from the online supplement of the published paper.

[4]Altogether, these categories represent around 70% of household consumer expenditures. The remaining 30% corresponds to shelter expenditures, whose prices are excluded from the CPI Research Database.

the unweighted case. This is because, apart from the highly weighted categories of cars and motorcycles, a lot of durables such as electronic equipment and furniture are not that flexible, especially when sales are excluded. Finally, contrary to widespread beliefs, services and non-durables have similar levels of flexibility when sales are excluded. Non-durables are more flexible than services only when sales are considered.

1.2.2 Monetary Policy Shocks

My main source of monetary policy shocks is the work of Romer and Romer (2004), RR thereafter, who identify policy shocks from 1969 to 1996 using a combination of quantitative and narrative approaches. These shocks have been recently updated to later years by Wieland and Yang (2019), which is the source of my data. Monetary policy is known to be endogenous to macroeconomic conditions. Thus, RR construct their shocks by using the intended changes in interest rates conditional on the information available around the FOMC meetings. First, they analyze the narrative records around these meetings, such as reports and transcribed conversations, to infer the *intended* rate changes. These will be correlated with actual changes, but hopefully exclude some of the endogenous response of actual changes to the development of the economy. Then, they regress these intended policy changes on the Fed's *Greenbook* forecasts of inflation and real activity. The residuals of these regression are arguably free of endogenous and anticipatory effects, as this procedure purges policy actions taken in response to current and future economic conditions.

Furthermore, I complement the RR shocks with high-frequency shocks from Nakamura and Steinsson (2018b). These shocks are identified from changes in the futures market happening in a tight window around FOMC announcements. Due to the relatively recent availability of these markets, their shocks start at 1995. They are able to construct two kinds of shocks depending on the futures market they focus on. The FFR shock is based on Federal Funds Rate futures, while the Yield Curve shock is based on the largest Principal Component of changes in several futures markets at different horizons. These monetary shocks, together with the RR shocks, are plotted in Figure 1.3.

1.2.3 Regional and Industry Data

In my main analysis, I will construct an index of regional price rigidity and use it to test whether more sticky regions react relatively more to monetary policy shocks. The main data source for this task is the Quarterly Census of Employment and Wages (QCEW), which consists of a panel of employment and payroll by industry for the 3,195 counties in the U.S.

from 1995 to 2016.[5] The QCEW is maintained by the BLS and gathers data on employment and wages for establishments which report to the Unemployment Insurance (UI) programs. Employment covered by the QCEW represents roughly 97% of all wage and salary civilian employment in the United States. The industry-county data is disaggregated up to the 6-digit NAICS level, but I will focus on 4-digit industry classification because many observations are censored at such fine level for confidentiality reasons. For the rest of the paper, I will combine counties into the 724 commuting zones covering the U.S. continental territory, as defined by Tolbert and Sizer (1996).

1.2.4 Price Level Data

I complement my employment analysis with price level data from the BLS. They publish hundreds of price level series for different items at a monthly frequency starting as early as 1947 for the CPI and 1926 for the PPI, although only 30% of the series are available before 1969.

The CPI price-level data are published at the Item Stratum (IS) and Expenditure Class (EC) level, which are the aggregation levels immediately above the ELIs. The BLS establishes 70 ECs and 211 ISs, which are further broken down into the 264 ELIs. In total, they publish 129 IS series and 11 EC series. In contrast, the PPI price-level data are published using the NAICS industry codes. They publish 174 series at the 6-digit NAICS level.

1.3 Index of Regional Price Rigidity

In this section, I describe how to construct an index of regional price rigidity at the commuting zone level. First, I discuss how to construct an index that captures the effects of monetary non-neutrality. Second, I manually match every ELI code to NAICS industries. Third, I define the price rigidity of a commuting zone as a weighted average of the price duration of its industries.

1.3.1 Discussion: What Measure of Stickiness?

Average Duration of Price Changes

It is not immediately obvious how my data on frequency of price changes should be mapped to a regional index of price rigidity that captures the degree of monetary non-neutrality. A big part of the macroeconomic literature has focused on understanding when and why

[5] Employment is reported monthly, while payroll is available at a quarterly frequency.

price rigidities should matter in the aggregate to obtain non-neutrality. The most popular model for nominal rigidities is based on Calvo (1983), where only a fraction θ_i of randomly selected firms in industry i adjust their prices every period. It is precisely the popularity of the Calvo model what has arguably drawn most of the empirical literature in price setting to investigate the frequency of price adjustments in the real world.[6]

In the Calvo framework, the frequency at which firms adjust prices every period is the only variable capturing the degree of non-neutrality, which is often measured as the cumulated output response to a monetary shock. Carvalho (2006) shows that the degree of non-neutrality in the Calvo model depends on the average duration of the prices[7] in the economy

$$D = \int e_i D_i di, \tag{1.1}$$

where $D_i = -1/\ln(1-\theta_i)$ is the expected duration of prices in sector i, and e_i are expenditure or value added shares.

My measure of price rigidity will be based on the average duration defined at the commuting zone level. However, duration in equation (1.1) takes value added shares as weights for the different sectoral duration. Instead, I use payroll data as the closest proxy for value added shares at the commuting zone level. Then, my index of price rigidity for commuting zone c is constructed as the weighted average of the price durations of the different industries located in the commuting zone

$$D_{ct} = \sum_i p_{ict} D_i,$$

where D_i is the average duration of prices of industry i, and p_{ict} is the weight of industry i payroll in the total payroll of commuting zone c at time t. Note that my data for industry durations are based on the CPI data from Nakamura and Steinsson (2008), so I only have one data point per industry for the whole U.S. Hence, any variation in my index of price rigidity over regions and time will be ultimately driven by variation in the industrial composition. Going forward, I will refer to the index D_{ct} as the duration index for commuting zone c at time t, while I will refer to the *average* of the index across time, $\bar{D}_c$, as the *average* duration of commuting zone c.

[6]See for example Taylor (1999) for an early survey, Bils and Klenow (2004) and Nakamura and Steinsson (2008) for a more recent look into the U.S. microdata, and Dhyne et al. (2006) for a similar approach with European data.

[7]Price duration is just a measure of the average timespan that prices remain unchanged.

Caveats

Next, I discuss some potential caveats of using the average duration of price changes as a measure of monetary non-neutrality. Overall, I conclude that the average duration is a well-established statistic to measure the degree of non-neutrality across different types of models and setups.

Endogenous duration and menu cost models. Price duration and the frequency of price changes are exogenous in the Calvo model. This is not the case in more complicated pricing models, such as state dependent menu cost models, where duration is an endogenous microfunded object. In these models, the pricing decision arises as the result of an optimization problem where firms face a fixed cost of changing prices, while also incurring in some losses from deviating from their optimal price. Importantly, menu cost models can have completely opposite implications for the relation between frequency and monetary non-neutrality. Golosov and Lucas (2007) show that money can be roughly neutral in a quantitative menu cost model calibrated to target some moments of the empirical distribution of price changes, including the average duration. The intuition is that price change opportunities in the Calvo model come at random, so there are some firms which are arbitrarily far away from their optimal price but still cannot reset their prices. But changes are optimized in menu costs, implying that firms that choose to change their prices are the ones that are further away from their optimal price. Thus, the economywide price level increases much quicker in the Golosov-Lucas than in Calvo.

Nevertheless, the importance of this so-called 'selection effect' in state dependent models has been called into question in subsequent literature.[8] In an important contribution, Alvarez et al. (2016) use a sufficient statistic approach to show that the cumulative effect on output of a nominal shock in a large class of state and time dependent models, including Calvo and Golosov-Lucas models, is proportional to the product of two key moments of the distribution of price changes, the average duration of price changes and the average kurtosis. They conclude using estimated statistics from several datasets that the real effects of nominal shocks are just 30% below the Calvo benchmark. In other words, the evidence seems to point out that firms that choose to change prices are not as selected as one could potentially expect, so the Calvo model provides a good approximation to study monetary non-neutrality.

Finally, it could be argued using their framework that differences in sectoral kurtosis could be an omitted factor in our regressions if kurtosis is correlated with duration. Fortunately, they find that monetary non-neutrality only increases by 15% when they use a duration-

[8]Some leading examples include Gertler and Leahy (2008); Nakamura and Steinsson (2010); Midrigan (2011).

weighted kurtosis instead of the average kurtosis that corresponds to homogeneous sectors, suggesting a relatively mild positive correlation between duration and kurtosis in their data.

Input-output linkages. The result from Carvalho (2006) is derived without intermediate goods, but input-output linkages could potentially exacerbate the effects of monetary non-neutrality due to pricing decisions becoming strategic complements. This could lead to omitted variables bias in our analysis if the I/O amplification varies across commuting zones and is correlated with duration. Both Nakamura and Steinsson (2010) and Pasten et al. (2019) find a significant amplification of monetary shocks when I/O linkages are introduced in a menu cost and a Calvo model, respectively. However, they both find that *heterogeneity* in the I/O links and the fraction of intermediates across sectors have a very small impact on their aggregate measures of non-neutrality. Nonetheless, this problem could be a more important issue when evaluating the differential response to monetary shocks across sectors rather than commuting zones. Even then, Pasten et al. (2019) show in their Figure 3 that the ranking of the sectoral output responses to monetary shocks in their estimated model is mainly driven by differences in duration across sectors, and introducing heterogeneous expenditure shares or I/O linkages does not dramatically change the overall ranking. In short, these papers suggest that I/O effects on non-neutrality can be approximated by a constant scaling factor across regions and industries.

1.3.2 Construction of the Regional Duration Index

In this section, I describe how I use my data to construct a regional index of price duration. First, I match my data on frequency of price adjustments for different goods and services to 6-digit NAICS industries. Then, I use my payroll data from the QCEW to construct a measure of regional stickiness as a weighted average of its industries price durations.

Matching ELI codes to NAICS. I manually match the 264 ELIs to the NAICS industry codes by looking at common keywords in their titles and descriptions in the Census website. This resulted in 470 matches out of a total of 1,175 6-digit NAICS codes.[9] Despite matching less than half of all the possible industries, matched sectors represent 86% and 76% of total U.S. employment and payroll between 1995 and 2016, respectively.

Figure 1.4 plots the matched frequencies across 12 broad industry categories defined in Table 1.2.[10] Red columns represent the goods-producing sectors, while blue columns are

[9]Bils and Klenow (2004) also matched a subset of an older version of the ELI codes to manufacturing and wholesale NAICS industries. I asked the authors for their data and it mostly coincides with mine in our overlapping observations.

[10]I could not find any match from ELIs to the Construction sector, so I omit it.

the service-providing sectors. We can immediately observe that the level of granularity is much higher for the goods and retail sectors. As we argued before, services tend to be more sticky than the goods producing sectors, although some services such as rentals or traveler accommodation are particularly flexible. Further details can be found in Table 1.3, which presents the most sticky and flexible industries within each broad industry category. Figure 1.4 and Table 1.3 are important because they capture the kind of variation that I will be exploiting when controlling for broad industry categories.

Censoring and missing industries in the QCEW. One important issue in the construction of the index relates to censored data in the QCEW, as well as my lack of data on price duration for some industries. Because of confidentiality reasons, employment and payroll figures at the county-industry level cannot be disclosed when the underlying observations could be easily identified. In order to ameliorate this problem, I aggregate my industry duration variable D_i to 4-digit NAICS industries as in Table 1.3, using the expenditure shares from the CPI. This coarser level of granularity gives lower variation to my index, but makes the censoring less likely. Still, the remaining censoring and the lack of duration for some industries significantly limits my ability to match industry duration with payroll data at the county-industry level. Overall, I am able to match a 40.3% of total U.S. payroll. Moreover, Figure 1.5 shows there is a lot of heterogeneity in the time-average share of matched payroll across commuting zones, ranging from 0 to 90 percent. Consequently, I drop from my dataset the commuting zones with a matched-payroll share below the 20th percentile, which is around 7.3%. This leaves us with 537 commuting zones.

Another additional concern about censoring is that it might affect industries and commuting zones differentially. This could potentially alter the duration ranking depending on the relationship between duration and share of matched payroll. For example, relatively flexible industries such as agriculture could be systematically censored in smaller commuting zones, which would create artificial differences in my duration measure depending on the size of the commuting zone. I explore this issue in Panel A of Figure 1.6, which plots the fraction of matched payroll and the average duration for all commuting zones. We can observe there is no systematic relationship between censoring and price duration—the correlation is 0.03. As one could actually expect, Panel B actually shows that the main determinant of the matched-payroll share across commuting zones is its size. However, the is no evidence that smaller regions tend to censor relatively more flexible or sticky industries, with size being essentially uncorrelated with my duration index—the correlation is 0.02.

1.3.3 Descriptive Statistics of the Regional Duration Index

The distribution of the average duration index by commuting zone, $\bar{D}_c$, is presented in Figure 1.7. The distribution is quite symmetric and bell-shaped, with an average and median durations of 17.6 and 18.1 months, respectively. Except for some outliers, duration is fairly concentrated, with two thirds of my sample having durations between 15.5 and 20.2 months. These numbers are somewhat higher than the usual estimates obtained in the literature, which are around 11.7 and 13 months when excluding sales and substitutions (Nakamura and Steinsson, 2008). The discrepancy is due to the U.S. being generally concentrated in services, and the NAICS classification system being relatively more detailed for the goods sector. This makes the goods sector relatively more censored than the service sector in the QCEW dataset, increasing the average duration for all commuting zones.[11]

Cross-sectional and time variation of the duration index. Differences in industrial composition across commuting zones is the only driver behind the variation of my duration measure. This could lead to interpretation problems in my empirical analysis, as duration could just be reflecting how different industries react to monetary shocks for other reasons unrelated to nominal rigidities. Figure 1.8 illustrates the different patterns of specialization in broad industry categories for commuting zones in different quintiles of the average duration distribution. Surprisingly, there is no clear pattern of broad industry specialization across different quintiles of my duration index, with the exception of the first quintile. Commuting zones with the lowest duration values are relatively more concentrated in the primary sector and relatively flexible services, such as utilities; wholesale; accommodation and leisure; and other services. But the overall industrial composition is broadly similar for the rest of commuting zones. Therefore, we can conclude that most of the variation in the duration index is coming from differences *within* broad industry categories. This variation is highlighted in Figure 1.1 and Table 1.3. In other words, when comparing sticky and flexible commuting zones, we are not comparing areas that are very focused on manufacturing versus areas that are concentrated in health, education, or information services. Instead, we are mostly comparing areas that have similar shares of manufacturing and services, but specialize in more flexible or sticky industries within these categories. This reduces some of the concerns and let us be more confident on nominal rigidities driving the differential responses for different levels of duration.

We can also infer that the regional duration outliers in Figure 1.7 are probably driven by censoring, and hence the 'true' distribution of commuting zone duration is probably more

[11] As I discussed in Section 1.3.2, there is no evidence that this goods sector censorship is affecting commuting zones differentially.

concentrated than my estimated measure. Panel B in Figure 1.6 reports higher censoring for smaller commuting zones. Even though they are not systematically specialized in flexible or sticky industries, their small size makes their largest sectors more likely to survive non-disclosure rules when finer industry classifications are considered. This makes these small commuting zones look more extreme, overweighting the duration of sectors where they are specialized in. As we discussed earlier, there is no evidence that this affects the commuting zone duration ranking in a systematic way, but it could lead to attenuation of my estimates.[12]

The variation of regional price rigidity across time is substantially smaller than in the cross section. Panel A in Figure 1.9 provides insight into the evolution of price rigidity over time. The red line shows that the average cross-sectional duration remains considerably constant throughout the sample. This same trend holds for the entire cross-sectional distribution over time. Panel B shows that the distribution of long-run changes in duration between 1995 and 2016 is narrowly concentrated around 0, with two thirds of the commuting zones lying between -0.74 and 1.97. The average and median long-run changes are 0.57 and 0.71, respectively, indicating a mild upward trend.

Geographic distribution of duration. Panel A in Figure 1.10 depicts the geographic distribution of the different quintiles of the average duration, $\bar{D}_c$. We can see that my duration index is not particularly clustered over space. The top and bottom quintiles are slightly more concentrated in the Southeast and the West, respectively, but overall spatial distribution is fairly spread out for the rest of the quintiles. The exact same conclusions can be derived from long-run differences in duration, shown in Panel B. Most importantly, neither the cross-sectional nor the time variation seem to be correlated with familiar secular industrial trends such as imports from China and Mexico, offshoring, or automation; which are substantially clustered around the Midwest and the Southeast.[13] My duration measure is not capturing either the large manufacturing concentration in the Rust Belt. These reassures us that the variation in duration is not reflecting typical reasons why monetary policy could be more effective, such as high concentration in capital-intensive industries or durable goods manufacturing.

[12] Although not exactly the same, this has a similar flavor to classical measurement error that is correlated with size but uncorrelated with my duration measure. It blows up the variance of duration for small commuting zones, probably leading to attenuation bias of my estimates.

[13] See Figure 3 in Acemoglu and Restrepo (2017) for details on the geographic distribution of these secular trends.

1.4 Empirical Methodology

My main empirical analysis focuses on studying whether regions in the U.S. respond differently to nominal shocks based on their degree of price rigidity. To do so, I will estimate the *relative* impulse response functions (IRF) to monetary policy shocks using the local projection methods introduced by Jordà (2005). Over the last decade, local projections have become a very popular method to estimate impulse responses over the classic vector autoregression (VAR) alternative. In contrast to local projections, VAR imposes much more structure on the data generating process and estimates the IRFs by iterating the autoregressive system forward. Consequently, it is widely believed that local projections estimates are more robust to misspecification and non-linearities (Ramey, 2016; Nakamura and Steinsson, 2018a), although this view has recently been called into question.[14]

The estimation of *relative* IRFs consists of estimating equations of the form

$$\Delta_{h+1} y_{ct+h} = \alpha_c^h + \alpha_t^h + \beta^h \cdot D_{ct} \cdot R_t + \Gamma^h \cdot W_{ct} + \epsilon_{ct+h}^h \tag{1.2}$$

at different horizons $h = 0, 1, 2 \ldots$, where subscripts refer to commuting zone c at time t; Δ_h is the long difference of h periods so $\Delta_{h+1} y_{ct+h} = y_{ct+h} - y_{ct-1}$; y is the percentage change in employment $100 \times \log(\text{employment})$; R is the monetary policy shock measured in percentage points; W is a vector of control variables; α denotes time and commuting zone fixed effects; and ϵ is the error term.

The *relative* impulse response is traced out by the different coefficients β^h coming out from every regression. Note that these coefficients do not represent the typical impulse response in levels, which is absorbed by the time fixed effect. This is rather a continuous difference-in-difference design with multiple treatments. They must be interpreted as the average *differential* response of employment to a monetary shock, in percentage points, induced by an additional month of regional duration. In light of the discussion in Section 1.3.1, we would expect to find negative estimates for β^h. In other words, regions with longer price durations—that is a high degree of nominal rigidities—should react more negatively to interest rates hikes.

Shift-share structure and identification. Note that my specification in equation (1.2) has a shift-share structure with a single national shifter, given by the monetary policy shock. In this view, the commuting zone duration D_{ct} captures the exposure of region c to the national shifter at time t. Identification and inference in shift-share designs has recently

[14] Plagborg-Møller and Wolf (2020) show that local projections and VARs are equivalent in their population estimands when the lag structure is unrestricted, although they might differ in finite samples.

received a lot of attention, stressing that identification can come from exogeneity of the shares (Goldsmith-Pinkham et al., 2018), or the shocks (Borusyak et al., 2018; Adão et al., 2019). Using the latter approach, the identifying assumption for the coefficients β^h is given by

$$E[R_t \mu_{th}] = 0,$$

where $\mu_{th} = E[D_{ct}\epsilon^h_{ct+h}]$ is a cross-regional average of the product of the regional durations D_{ct} and the unobserved error components ϵ^h_{ct+h}. Intuitively, this condition would not hold if regions with higher duration tend to have higher growth for unmodeled reasons (i.e., $\mu_{th} > 0$) in periods with expansionary monetary shocks. In other words, it highlights that any impact of monetary policy resulting from channels other than duration, and not captured by the controls in α and W, must affect regions uniformly. Thus, it is *not* required that the regional duration is randomly assigned, but rather that high and low duration commuting zones are not affected *differentially* by other channels that are correlated with duration.

This also stresses why it is preferable to use a series of plausibly exogenous monetary shocks. Monetary policy decisions depend on a lot of economic factors, and therefore encapsulate many aggregate variables that could affect high- and low-duration regions heterogeneosuly through many different mechanisms. Removing part of the endogenous component of monetary policy facilitates the requirement of interest rate changes not inducing differential employment effects that are correlated with duration but are caused by other reasons than nominal rigidities.

Baseline specification. The baseline specification estimates equation (1.2) with monthly data at the commuting zone level. Commuting zone fixed effects are in fact allowing for differential trends, as the specification is in differences. Control variables W will typically include industry shares and their interactions with the monetary shock at one of the aggregation levels defined in Table 1.2, that is, broad industry categories and 2-digits NAICS. This allows for heterogeneous effects of monetary shocks across industries, and identifies the effect of duration *exclusively* from differences in duration *within* the controlled industries. Additional controls include the baseline duration index, a lag of my variable of interest, $D_{ct-1}R_{t-1}$, and the first difference of the dependent variable to control for any persistence and further purge any past dependence of the monetary shock.[15] The popular recursivity assumption in SVAR models claiming that real variables do not react immediately to monetary policy can be implemented by including the contemporaneous value of the dependent

[15]Despite the RR shocks are supposed to be exogenous, they have a mild first lag autocorrelation of 0.17.

variable on the right-hand side.[16] Unless otherwise noted, the regressions are weighted by commuting zone total employment. Finally, I cluster my standard errors both by commuting zone and time. Commuting zone clustering accounts for any residual autocorrelation in the monetary shocks and the dependent variable. Clustering by time allows for correlation across commuting zones due to other aggregate shocks. Time clustering is also the correct way of clustering my shift-share specification according to Adão et al. (2019) when, using their language, every region-time observation is 'specialized' in one sector-time shock.[17]

Industry specification. I complement my regional analysis by studying differential effects across industries using aggregate U.S. data. Specifically, I estimate

$$\Delta_{h+1} y_{it+h} = \alpha_i^h + \alpha_t^h + \gamma^h \cdot D_i \cdot R_t + \Gamma^h \cdot W_{it} + \epsilon_{it+h}^h \tag{1.3}$$

where subscript i denotes industry and y denotes aggregate employment or prices. Note that my duration index is not time-varying in these regressions, so the relative IRFs are identified *exclusively* from cross-industry variation in duration and time variation of the monetary shock. In this case we would also expect employment in more sticky industries to react relatively more to monetary policy shocks, so $\gamma^h < 0$, as they are less capable of adjusting prices. On the other hand, we would expect $\gamma^h > 0$ in the case for prices, as sticky industries tend to decrease their prices relatively less. Control variables will be the same as in the baseline specification, changing industry shares for industry dummies. Because of the same reasons highlighted above, robust standard errors are two-way clustered by industry and time.

1.5 Results

In this section, I report the main findings of my analysis. I find that sticky regions react relatively more to monetary shocks. The results are robust to controlling for broad industry shares, using employment weights for the duration index, leaving extreme observations out, and using high-frequency monetary shocks. In the aggregate, I also find that sticky industries tend to reduce employment relatively more in response to a monetary tightening. Finally, these results are consistent with monetary non-neutrality induced by nominal rigidities, as

[16] This makes all coefficients equal to 0 at $h = 0$ since the right-hand side contains the dependent variable. See Ramey (2016) for a discussion of this assumption.

[17] In my specification, every region-time observation $\Delta_{h+1} y_{ct+h}$ is 'exposed' to the monetary shock h periods before $D_{ct} \cdot R_t$. Another example can be found in Chodorow-Reich et al. (2019), who use a similar specification and argument.

the data suggests that flexible industries and products are able to decrease their prices relatively more after interest rate hikes.

1.5.1 The Effects of Monetary Shocks Across Commuting Zones

The baseline results consists of estimating equation (1.2) using my whole sample, the duration index as defined in Section 1.3, and the monetary shocks from Romer and Romer (2004). Figure 1.11 presents the impulse responses of employment coming from the coefficients β^h with different industry controls. It shows that the effects of monetary policy are significantly different depending on the degree of price rigidity. On average, employment decreases by 0.1 to 0.25 percentage points more in the peak of the response for every additional month in price duration.

Comparing Panel A and B, we can see that adding broad industry controls reduces the precision. These controls partial out any influence in my duration index driven by differences in broad industry composition, so they only use the heterogeneity in duration *within* broad industry categories.[18] The relative response becomes marginally significant at the 90% level only in year 2, while it was significant for the entire first two years when no controls are used. Still, broad industry composition does not alter the main features of the relative response. This is a logical conclusion given that Figure 1.8 showed that broad industry shares were not that different for regions in different parts of the duration distribution. The differential impact of monetary policy peaks around 2 years and fades away after roughly 4 years, which is consistent with the usual shape of estimated unconditional responses in the literature.[19] There seems to be a marginally significant effect in the opposite direction after 5 years, but it disappears when controlling for broad industry composition.

Finally, Panel C reports the same specification increasing the number of industry controls from 12 broad industry categories to the 23 NAICS 2-digit sector shares. In this case the results are similar, with sticky regions reacting more to interest rate hikes. Quantitatively, the peak point estimate for the relative IRF of 0.12 percentage points after 3 years, even though the difference is estimated less precisely and no longer significant. This is reasonable as doubling the number of industry controls removes some of the variation in duration and, despite not affecting the point estimates much, it increases the standard errors to some degree.

Overall, my baseline results suggests that sticky regions react more strongly to monetary policy. Next, I do some robustness checks to see whether these results hold are robust to

[18] I also tried interacting the industry shares with the time fixed effects, which allows for differential impact of aggregate shocks depending on industry shares. This results in very similar results.

[19] See Ramey (2016) for a survey on the effects of macroeconomic shocks and their propagation.

changes in the sample, the duration index, the shocks and the specification.

Excluding outliers. First, I explore the influence of outliers in my estimates. The manual matching of CPI codes to NAICS industries together with the censoring in the QCEW data leads to a rather noisy measure of regional duration. Figure 1.7 showed that the distribution in duration is somewhat skewed to the left, with some outliers particularly concentrated below 10 months of duration. I analyze the sensitivity of my estimates to these observations in Panels A and B of Figure 1.12, which estimate equation (1.2) excluding commuting zones in the first quintile of the average duration distribution. The results have a similar shape and are slightly larger when these outliers are omitted, with stickier regions decreasing employment between 0.15 and 0.375 percentage points more at the peak for every additional month of price duration, depending on the broad industry controls.

I additionally inspect if dropping outliers from both ends of the distribution affects my results. In this case, only observations with duration values within one standard deviation of the median—between 12 and 23.84 months—are included in the analysis. This effectively drops 17% of the most extreme observations in my sample. Results are plotted in Panel C and D of Figure 1.12. The relative impulse responses are again very similar to my baseline results and the previous outlier analysis, with a peak relative impact on employment of 0.14 to 0.30 percentage points for every additional month of price duration.

Different measures of duration. The regional duration index was constructed from industry durations based on the filtered measure of frequency, and using payroll weights for each commuting zone. I crosscheck that my results are not particularly sensitive to these decisions by estimating equation (1.2) with alternative duration measures. Panel E and F of Figure 1.12 uses employment weighed inflation as the independent variable. The employment-weighted duration yields slightly odd results without industry controls. The relative response is negative and non-significant for the first two years, but it turns significantly positive after year 4. This late fact mostly goes away when controlling for broad industry shares. The peak relative responses around 2 years are still of roughly 0.2 percentage points of additional employment reaction to contractionary monetary shocks. In the end, if value added or expenditure shares are the right weights for duration according to the literature reviewed in Section 1.3.1, it is reasonable that payroll will offer a better proxy for those than employment counts.

On the other hand, the results look surprisingly better when the raw frequency of price changes is considered. Panel G and H in Figure 1.12 show these specifications with and without broad industry controls. The hump after year 4 is mostly gone in both specifications, and the significance of the results are quite strong even with broad industry controls. This could be due to Nakamura and Steinsson (2008) actually introducing noise in their frequency measures when they apply their sales filters. This duration measure based on the raw frequencies of price adjustment reveals a similar magnitude for the relative impulse responses, with employment dropping by 0.2 to 0.3 percentage points more after a monetary tightening.

High-frequency monetary shocks. Next, I test whether my results are robust to alternative monetary policy shocks based on high-frequency identification. Panels A-D of Figure 1.13 shows my estimations for the two kinds of shocks available from Nakamura and Steinsson (2018b), a shock to the federal funds rate and another one to the yield curve—capturing forward guidance. Interestingly, the estimated relative impulse responses are larger and more precise when high-frequency shocks are considered. The differential response peaks around 3 years for both cases and is significant all the periods from 6 months to roughly five years after the shock. The estimates in Panel A for the yield curve shock implies average negative employment responses between 1 and 1.5 percentage points larger for every additional month of price duration. Given that this shock affects the whole yield curve, it is reasonable that it induces larger responses on real variables.[20] The shocks to the federal funds rate in Panels C and D of Figure 1.13 still induce larger responses than the RR shocks, with peak relative impulse responses between 0.55 and 0.7 percentage points in employment for one more month in regional duration.

Alternative Specifications. Overall, these results are surprising given that I was able to match less than 20% of the total payroll for most commuting zones. This would suggest a low signal-to-noise ratio from our duration index, compromising the precision of our estimates and probably attenuating the results. Nevertheless, this problem is in part weakened by the employment weights in the regression. Panel B of Figure 1.6 indicated that the fraction of payroll that I am able to match is very related to the employment size of the commuting zone. Therefore, by weighting observations by employment I am effectively increasing the influence of the less noisy observations in my dataset, and vice versa. In order to see this, I estimate the unweighted version of equation (1.2) in Panels E and F of Figure 1.13. Indeed, the unweighted regression is a lot noisier than the baseline results. The estimates are significant and of a similar magnitude without industry controls, but become essentially indistinguishable from 0 when broad industry shares are included. All in all, this supports the view that employment weights are partially solving the measurement error from my matching procedure.

[20]The shock significantly affects the Treasury yields up to 5 years into the future. See Table 1 in Nakamura and Steinsson (2018b) for further details on the effects of this shock on the yield curve.

Furthermore, local projections are usually run with the dependent variables in levels, rather than differences. This should affect the interpretation of the relative impulse response coefficients, but it matters for the actual specification of the panel data model. As it is, the commuting zone fixed effects allows for differential trends, and the time fixed effect controls for shocks to the aggregate growth rate of employment. Also, the lag of the dependent variable is set in differences rather than in levels. This would be the correct way of modeling this dataset if employment is an $I(1)$ process with differential trends for every region and with common aggregate shocks. In contrast, when the specification is run on levels, it assumes a common trend across regions that is absorbed by the time fixed effect. Overall, the specification in difference is much more demanding as it is a lot more flexible in fitting different time series models for the different regions. Still, I report the results in levels in Panel G and H of Figure 1.13, which are virtually the same as in the baseline, if not stronger in terms of statistical significance.

Other robustness checks. I run some additional checks that are omitted for brevity. Including the contemporaneous value of the left-hand side variable, the recursivity assumption from Ramey (2016), keeps all the results virtually unchanged. All the results are robust to using the first lag of the duration variable, obtaining very similar responses in terms of shape and magnitude. Including a larger number of lags of the dependent variable removes a lot of the variation in the baseline specification leading to noisy results. This is reasonable given how demanding the specification is, as I already differenced the dependent variable and allowed for differential trends across regions, which are probably enough to control for persistence. However, increasing the number of lags does not affect the results much in the specification in levels, which is the standard in the literature. Finally, in section 1.3 I mentioned that I dropped the 20% of the commuting zones with the lowest values of the matched-payroll share. The results are robust to moving that threshold between 0 and 40%, with the results varying very little. This threshold only affects the results in the unweighted regression, as one could expect, leading to relatively stable results for thresholds between 15% and 40%.[21]

[21] The baseline regression is weighted by employment, which is the main determinant of the matched-payroll share. Thus, dropping the commuting zones with the lowest values of matched-payroll share just excludes the observations with the lowest weights, thereby leaving the results mostly unchanged. This is not the case for the unweighted regression, though.

1.5.2 The Effects of Monetary Shocks Across Industries

The last section showed that employment in more sticky commuting zones tend to react more to monetary policy shocks. This can be explained by many models featuring nominal rigidities, which link the average price duration of the economy to the degree of monetary non-neutrality. In this section, I study the same question using aggregate industry data. I find that industries with larger degrees of price stickiness exhibit smaller responses in prices, but larger drops in employment after a monetary shock. Again, these results are mostly robust to removing outliers, constructing the duration measure with the raw frequencies of price changes, using high-frequency monetary shocks, using weights in the estimation, and running the specification in levels. Compared to the regional results, the industry response relative responses tend to peak slightly later, around 3-4 years, and feature higher persistence.

Employment Responses

The baseline results are presented in Figure 1.14, which consist of estimating equation (1.3) at the 6-digit NAICS level with different sets of industry dummies. Note that in this case the duration measure is not varying across time, so all the variation identifying of the relative impulse responses is cross-sectional. Overall, the figures indicates that an additional month of duration makes industry employment fall by 0.09 to 0.2 percentage points more after an interest rate hike of 1 percentage point. The estimated relative impulse response in Panel A is only significantly different from 0 between 8 months and 3 years after the shock, but adding industry controls actually improves the precision and indicates longer-lasting effects. The peak of the relative responses is reached slightly later than in my regional estimates, ranging from 2.5 to 4.5 years.

On the whole, results are similar to those from the regional analysis, with stickier industries presenting stronger employment reactions to monetary policy shocks. Next, I run the same robustness check as in the previous section and show that the results are are robust to changes in the sample, the duration index, the shocks and the specification.

Robustness. I analyze the sensitivity of my estimates extreme observations by keeping industries with duration values within one standard deviation of the mean—between 3.5 and 24.7 months. This effectively drops 23% of the most extreme observations in my sample. Results are shown in Panels A and B of Figure 1.15 The relative impulse response is dampened and less precise without industry controls, but recovers the baseline shape and magnitude in Panel B after including broad industry dummies.

Raw frequency estimates are presented in Panels C and D of figure 1.15. The relative response peaks slightly earlier, between 2 and 3 years without controls, setting these estimates closer to the regional results. Quantitatively, the estimates are fairly close to the baseline, with employment falling by 0.09 to 0.12 percentage points more in the peak for every additional month of duration.

Panels E and F result from estimating equation (1.3) without weighting industries by employment. In this case the results are somewhat puzzling, with the relative response gets delayed with more sticky industries only starting to respond more after 2 years, and fading away in the fifth year after the shock. Nonetheless, the data keep pointing out that sticky industries experience responses to monetary shocks that are more pronounced. Additionally, running the specification in levels, plotted in Panels E and F of Figure 1.16, yields virtually the same estimates as in the baseline analysis.

Finally, the estimates using high-frequency shocks are shown in Panels A-D of figure 1.16. Both shocks produce very similar estimates, which are significantly different from 0 for the whole horizon. This is not what one would expect from typical nominal rigidities models where the effects of monetary policy are just temporary. The relative responses just seem to revert after the fifth year. We can also see that, in line with the regional results, high-frequency shocks tend to give slightly higher estimates for the effect of duration.

Price Responses

Next, I use PPI price-level data from the BLS to assess whether my matching procedure from ELI to NAICS industries is consistent with price of more flexible industries reacting more to monetary policy shocks. This would provide evidence that my duration measure is indeed capturing the effect of nominal rigidities and therefore induces different industry and regional responses through this mechanism.

The baseline results estimate equation (1.3) using my whole PPI sample and the filtered frequency measure.[22] Results are presented in Figure 1.17, which plots the negative of the coefficients γ^h from estimating equation (1.3). Plotting $-\gamma^h$ leads to a more natural intepretation, as we would expect more *flexible* industries (i.e., lower duration) to have a more pronounced reaction and decrease prices relatively more, which translates to $-\gamma^h < 0$. Prices of flexible and sticky industries evolve in the same way for the first two years, precisely the peak moment of most of my regional estimates. We know from Romer and Romer (2004) and Ramey (2016) that RR shocks lead to no effect on aggregate prices for the first two years, so my estimates suggest that the flexible industries drop prices relatively more when they start to change after 2 years. The relative response of prices peaks at roughly 4.5 years, and then reverts back and I cannot reject that the gap is closed around 6 years. The results

[22]The regressions are not weighted as I do not have data on employment before 1995.

remain strikingly unaltered even when controlling for broad industry category or 2-digit NAICS dummies. Also, using CPI rather than PPI data leads to the same conclusions and the results are omitted for conciseness. Overall, Figure 1.17 suggests that industries with one more percentage point of frequency of price adjustment tend to drop prices by an additional 0.04 percentage points after a 1 percent hike in interest rates.

Robustness. The same battery of robustness checks that I used in previous sections are used to assess the robustness of my results. These are shown in Figure 1.18 and Figure 1.19. The price results are very similar across all figures, with the exception of high-frequency shocks. Surprisingly, high-frequency shocks yield very noisy results throughout the horizon, and even produce the opposite sign around 1.5 years after the shock, suggesting that flexible industries react less to monetary shocks. This is completely at odds with the regional and industry-employment results that we obtained in earlier sections, and is the main reason why I focus on RR shocks instead despite their somewhat weaker results in the other parts. This could signal that prices have become less sensitive to monetary policy shocks, or that high-frequency shocks are not well-suited to study aggregate effects of monetary policy because of their tiny size and limited scope, as suggested by Ramey (2016).

1.6 Conclusion

Nominal rigidities are the main friction introduced in most business cycle models to generate real effects of monetary policy. However, despite its predominance, there is very little empirical evidence that directly links the extent of nominal rigidities to monetary non-neutrality. In this paper, I have constructed a novel measure of regional price stickiness and used it to study the relationship between nominal rigidities and monetary non-neutrality. My results suggest larger reactions of stickier regions and industries to monetary shocks. Furthermore, in line with predictions of sticky price models, prices in more flexible industries tend to fall relatively more after an interest rate hike, suggesting that the channel captured by my measure is indeed related to price rigidity. Overall, this paper provides useful evidence in support for one of the most prevalent assumptions in modern macroeconomic models.

I believe this paper opens the door to many new research venues related to my regional duration index. Some possibilities include applying a similar analysis to other demand and supply shocks, and extending the dataset to earlier periods and other countries. Furthermore, leveraging new price data based on web scrapping could add useful time variation to the stickiness measure by industry. An important point to be addressed in future research is to understand the determinants of the differences in price stickiness across industries, and how they relate to other factors such as downward nominal wage rigidities, balance sheet composition, or capital intensity. Understanding these fundamentals could be helpful in finding instruments for price stickiness that allows for a more robust causal interpretation of my estimates. A structural approach could also be used to assess the robustness of my results by estimating a multi-region multi-sector New Keynesian model. All things considered, I see all these possibilities as a potentially fruitful and very promising research agenda.

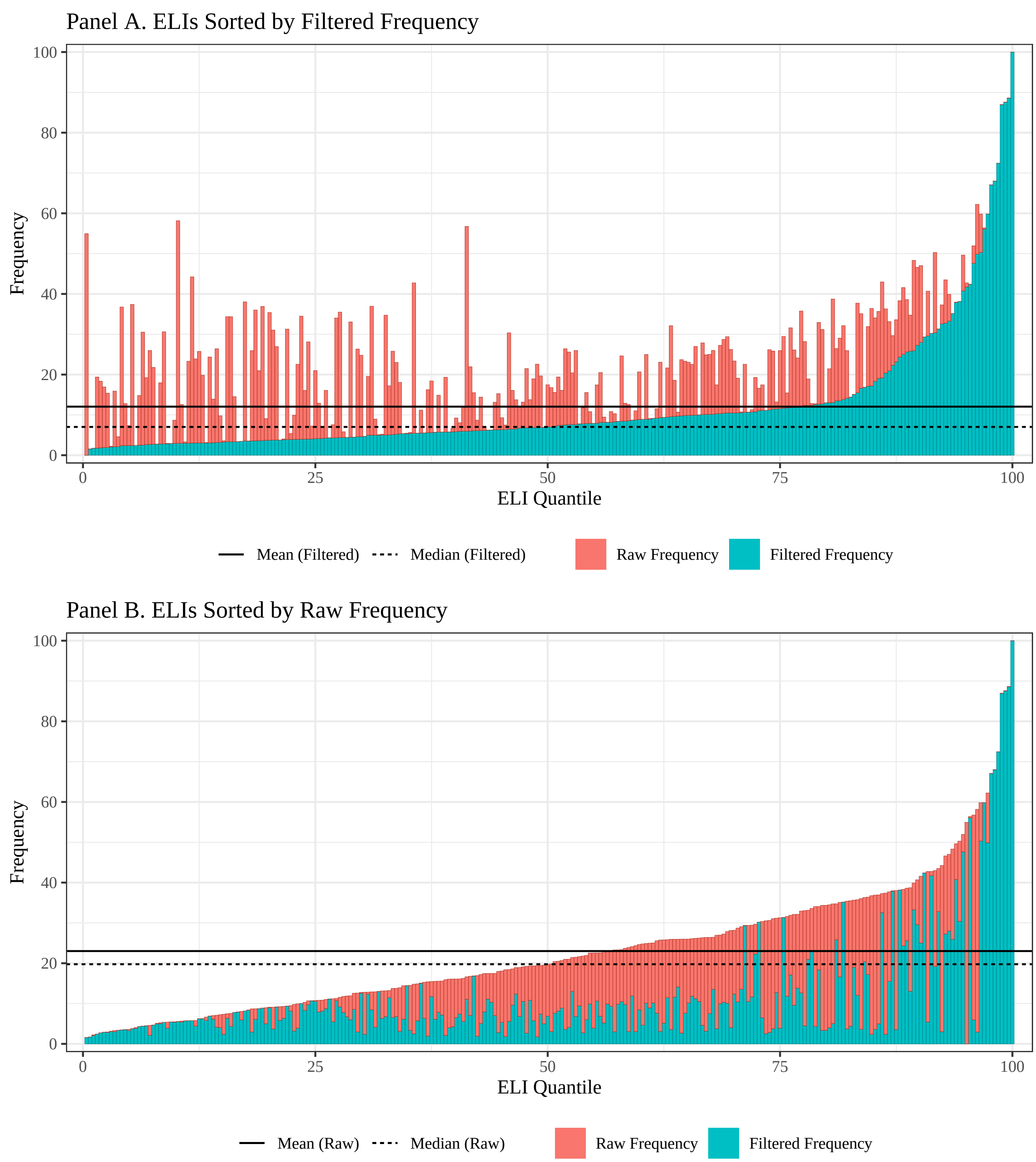


Figure 1.1: Frequency Measures by ELI Code.

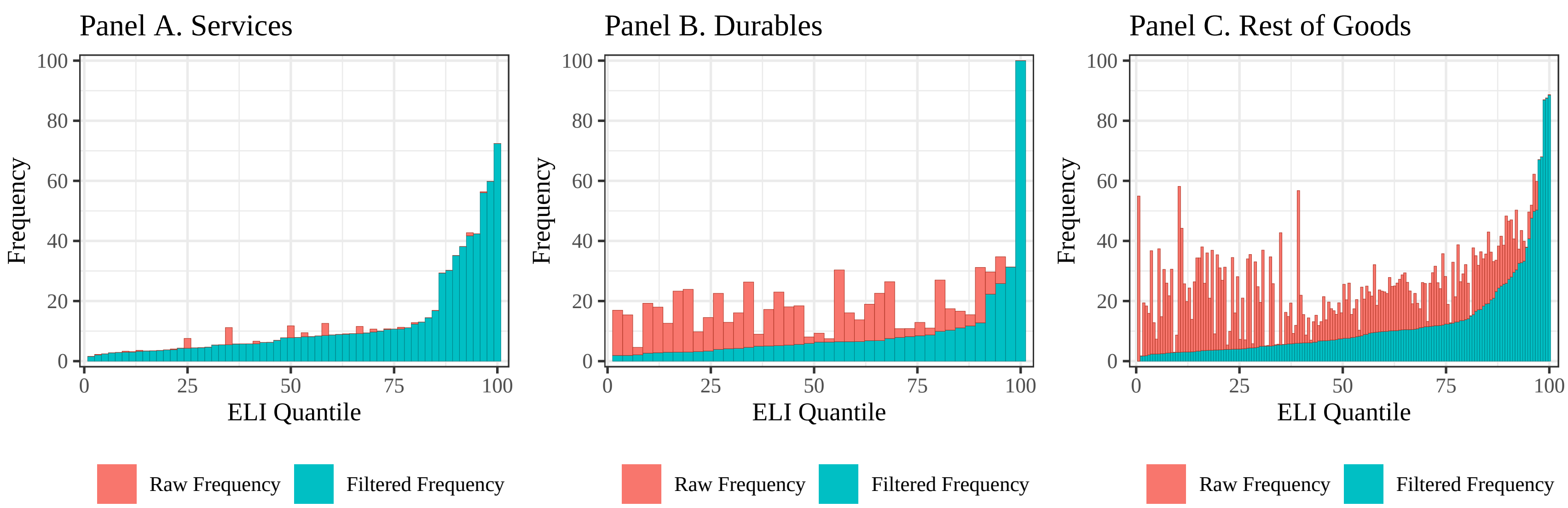


Figure 1.2: Frequency Measures for Goods and Services.

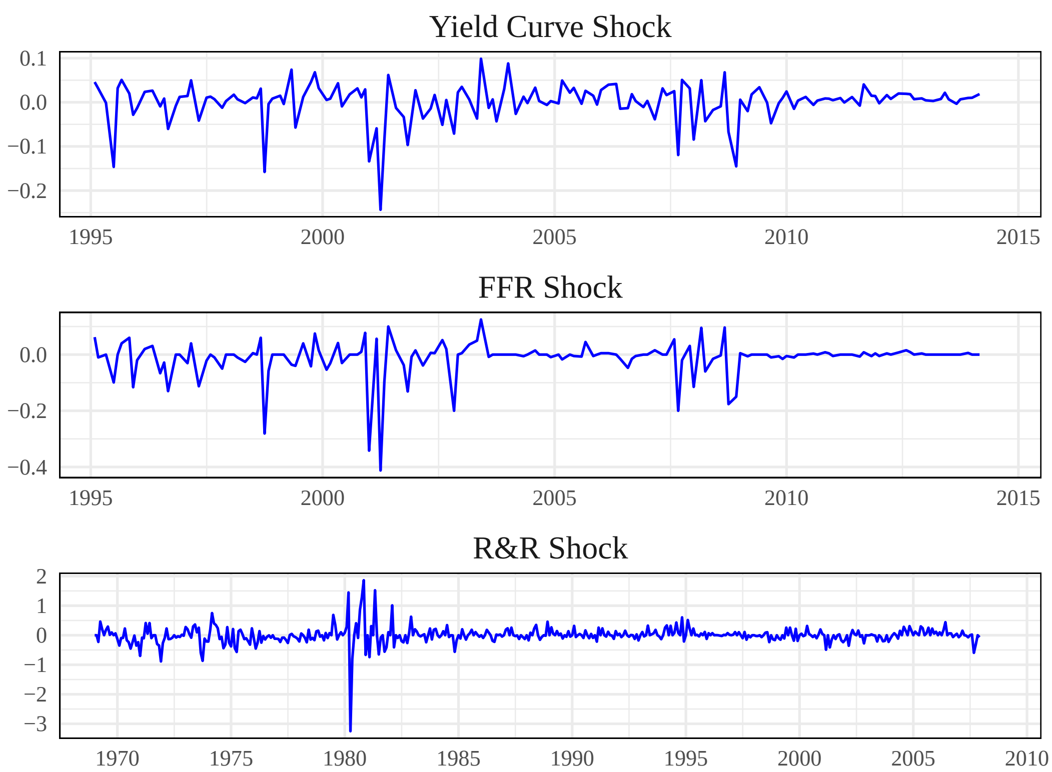


Figure 1.3: Monetary Policy Shocks.

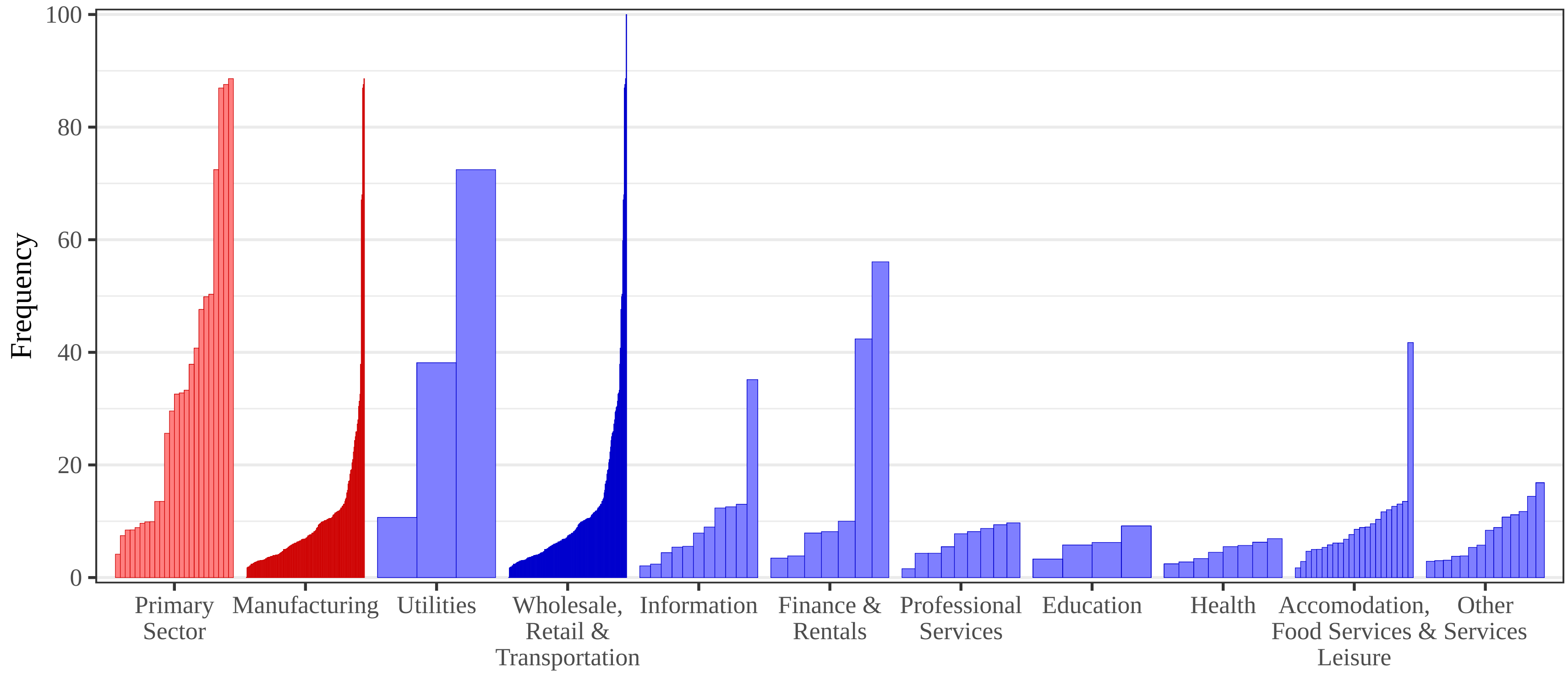


Figure 1.4: Frequency Measures of NAICS Industries.

Note: The y-axis represents the filtered measure.

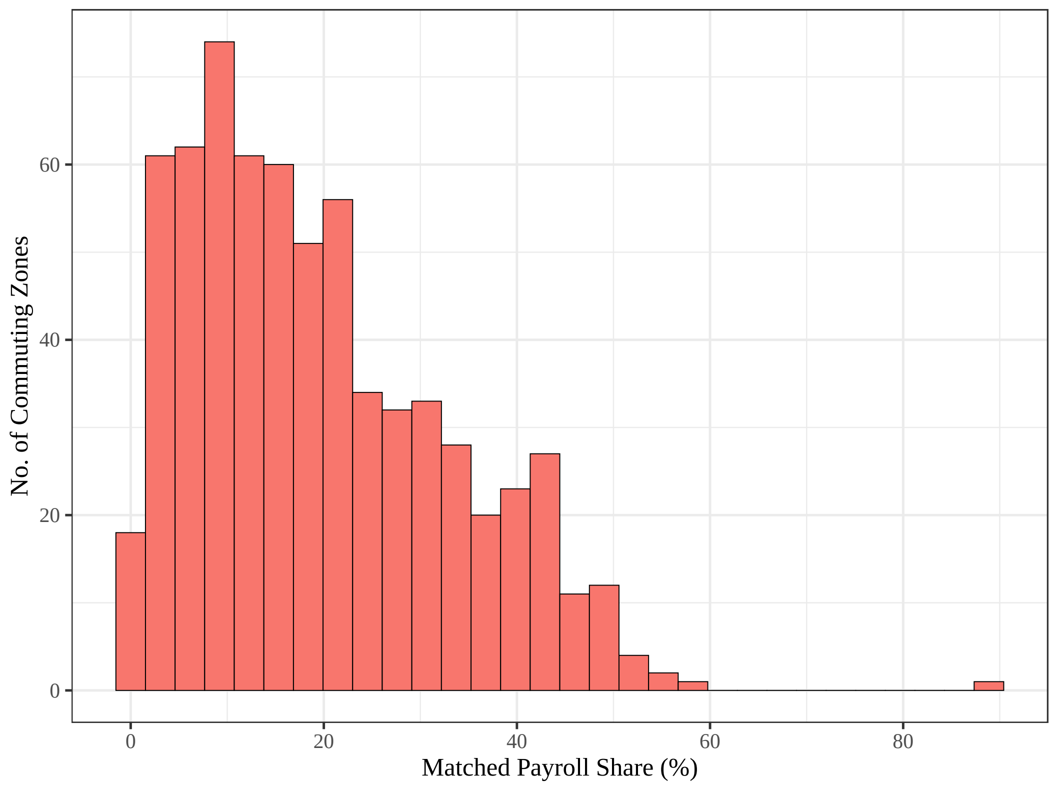


Figure 1.5: Histogram of Matched-Payroll Share.

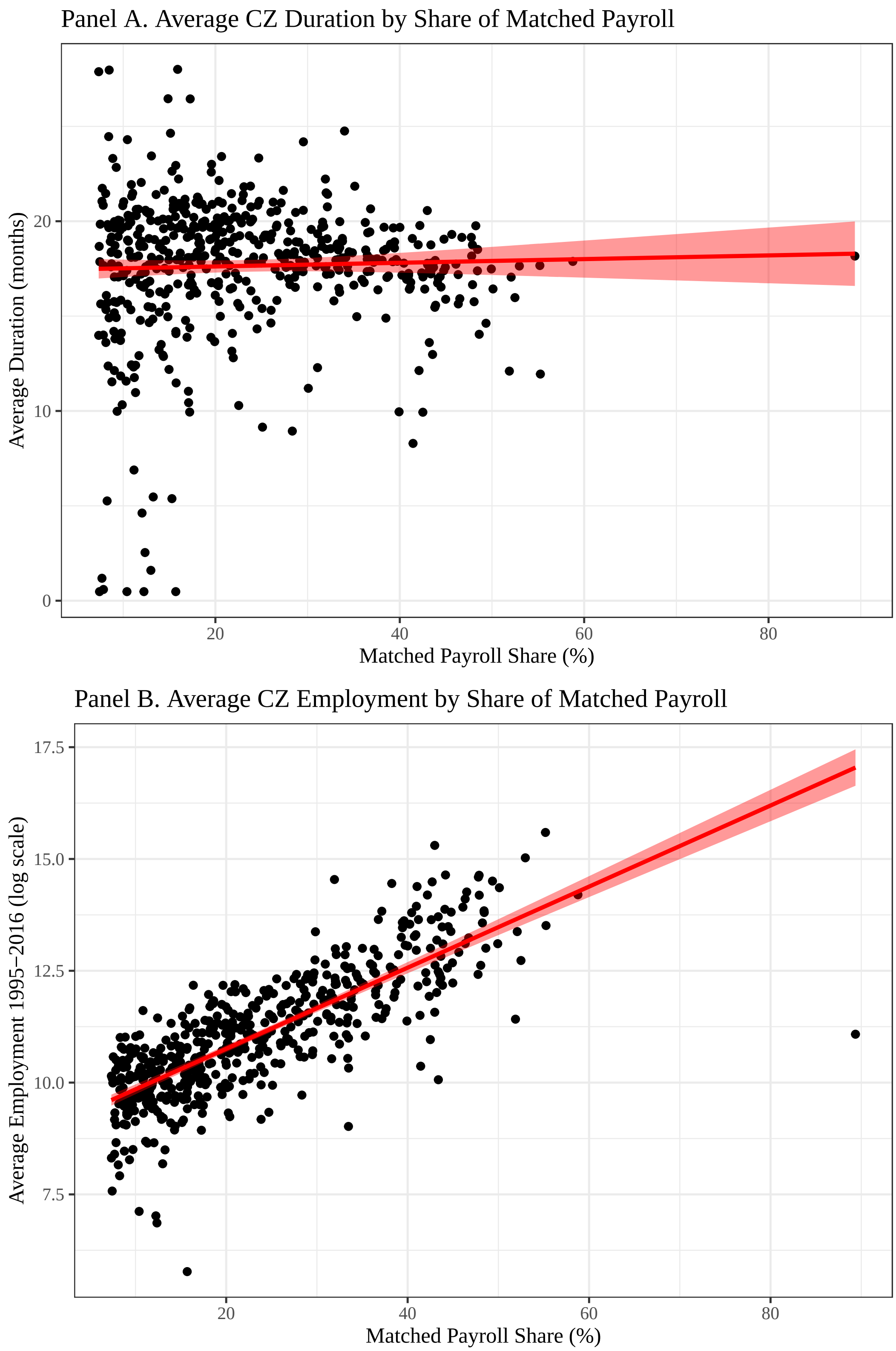


Figure 1.6: Average CZ Duration by Share of Matched Payroll.

Notes: The red lines are the OLS regression lines with their corresponding 95% confidence intervals. The non-significant OLS coefficient in Panel A is robust to removing the right-most outlier.

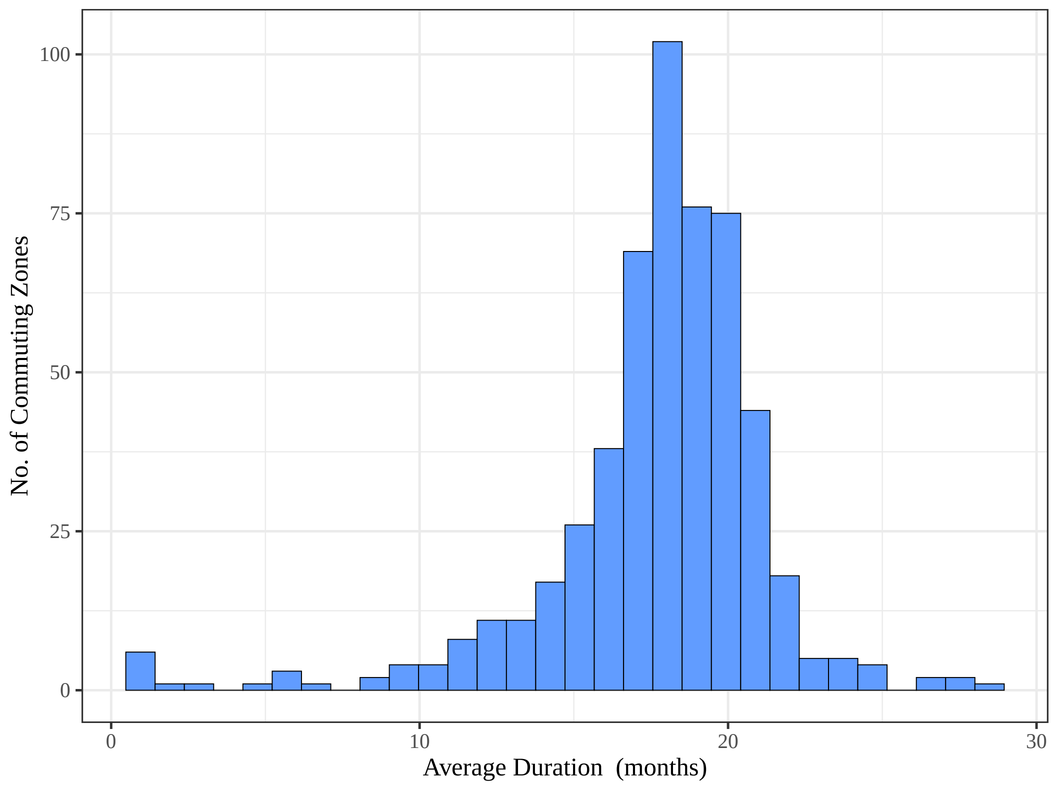


Figure 1.7: Histogram of CZ Average Duration.

Note: The duration index is averaged over time for every commuting zone.

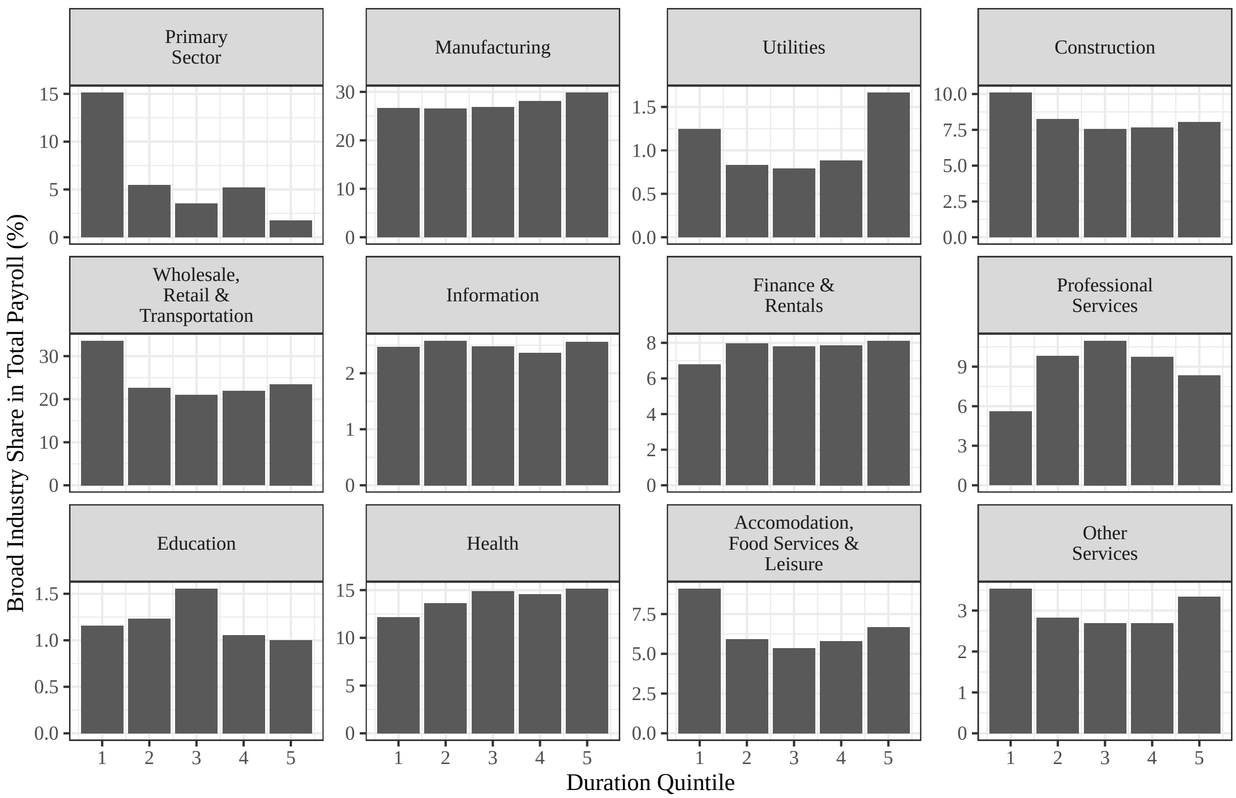


Figure 1.8: CZ Average Broad Industry Shares by Duration Quintile.

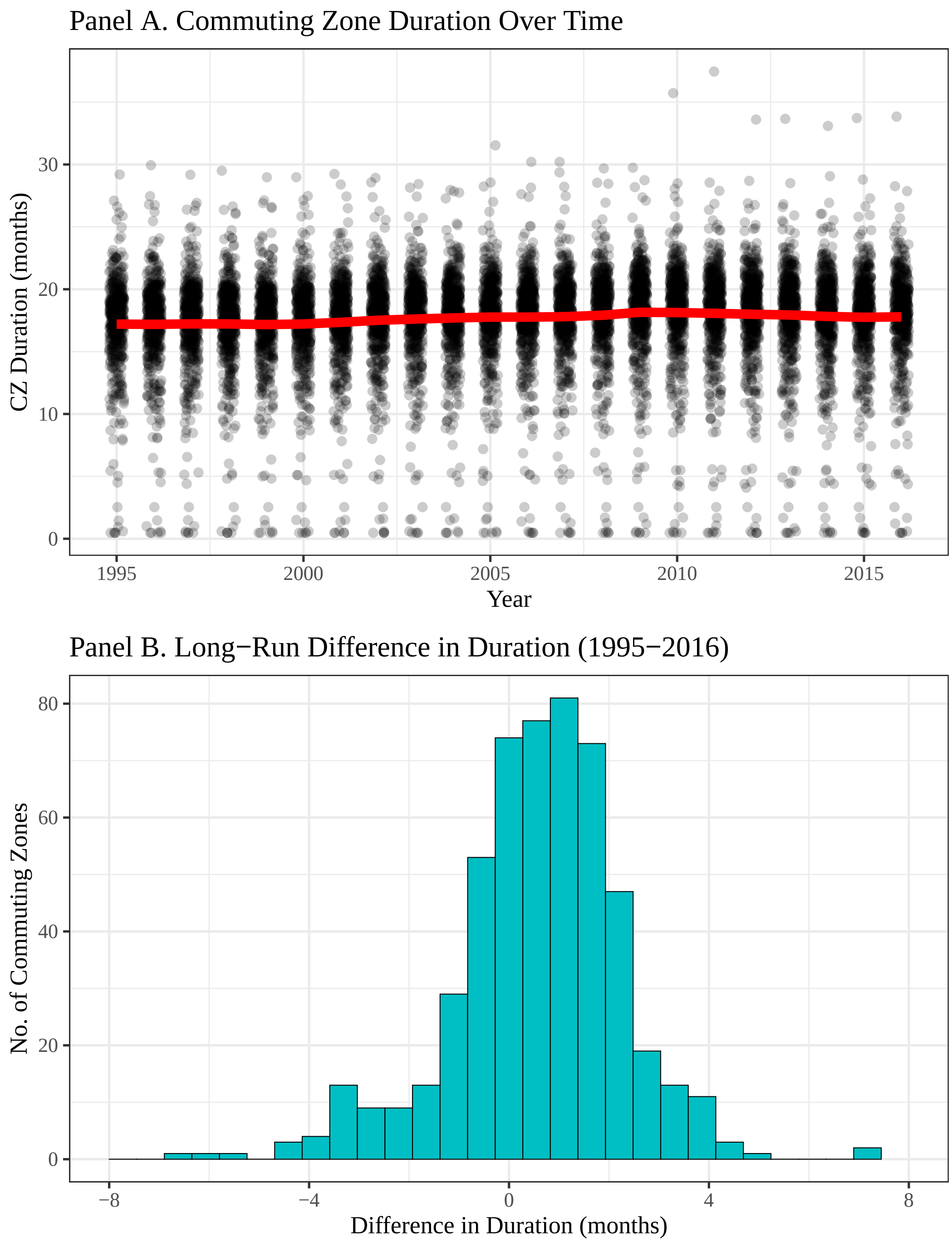
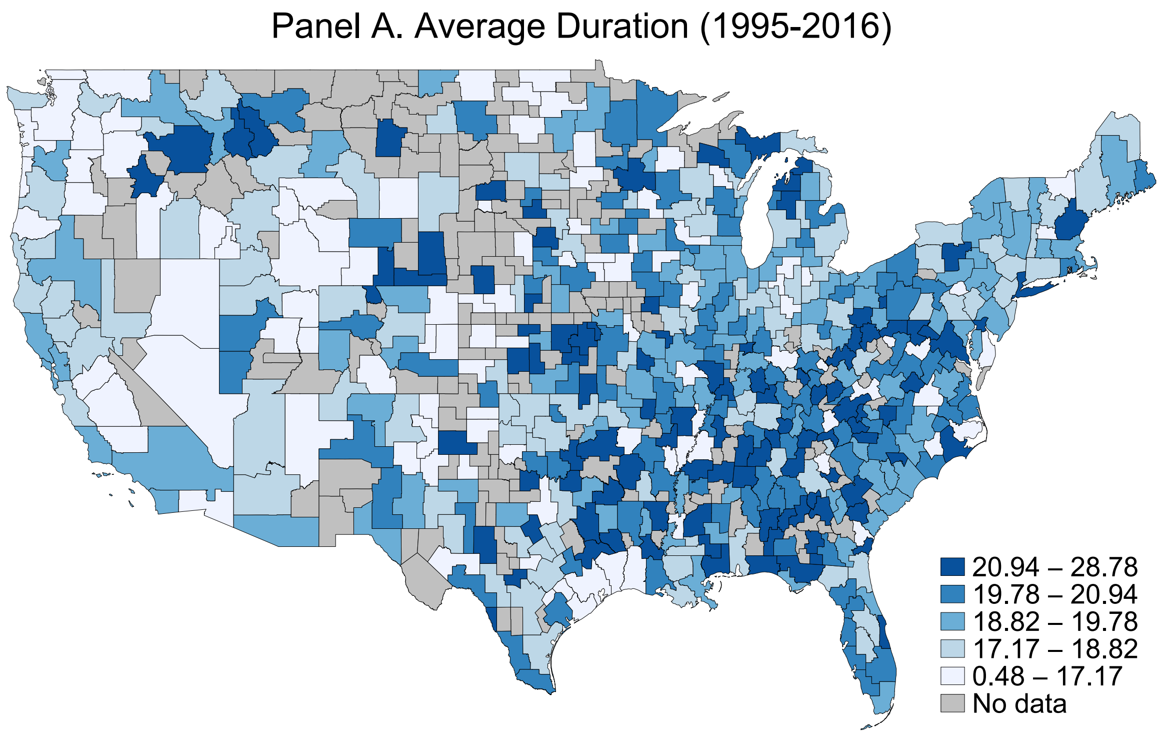


Figure 1.9: Variation of Regional Duration Across Time.



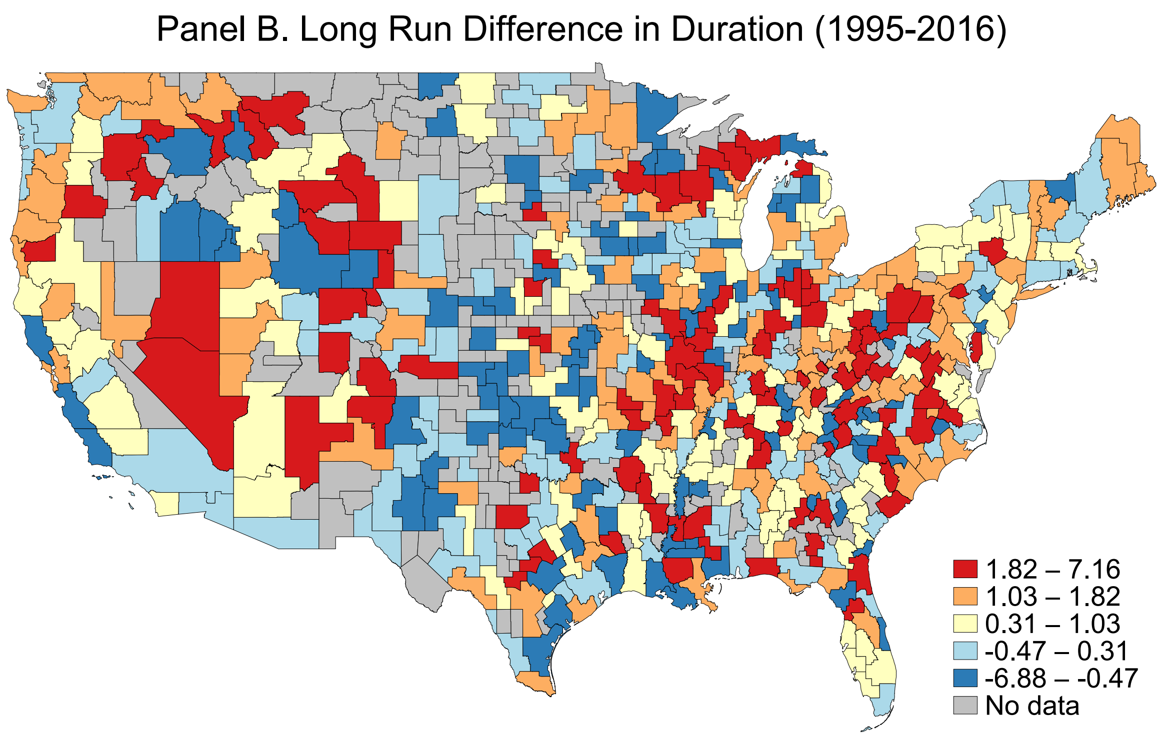


Figure 1.10: Geographic Distribution of Average Duration.

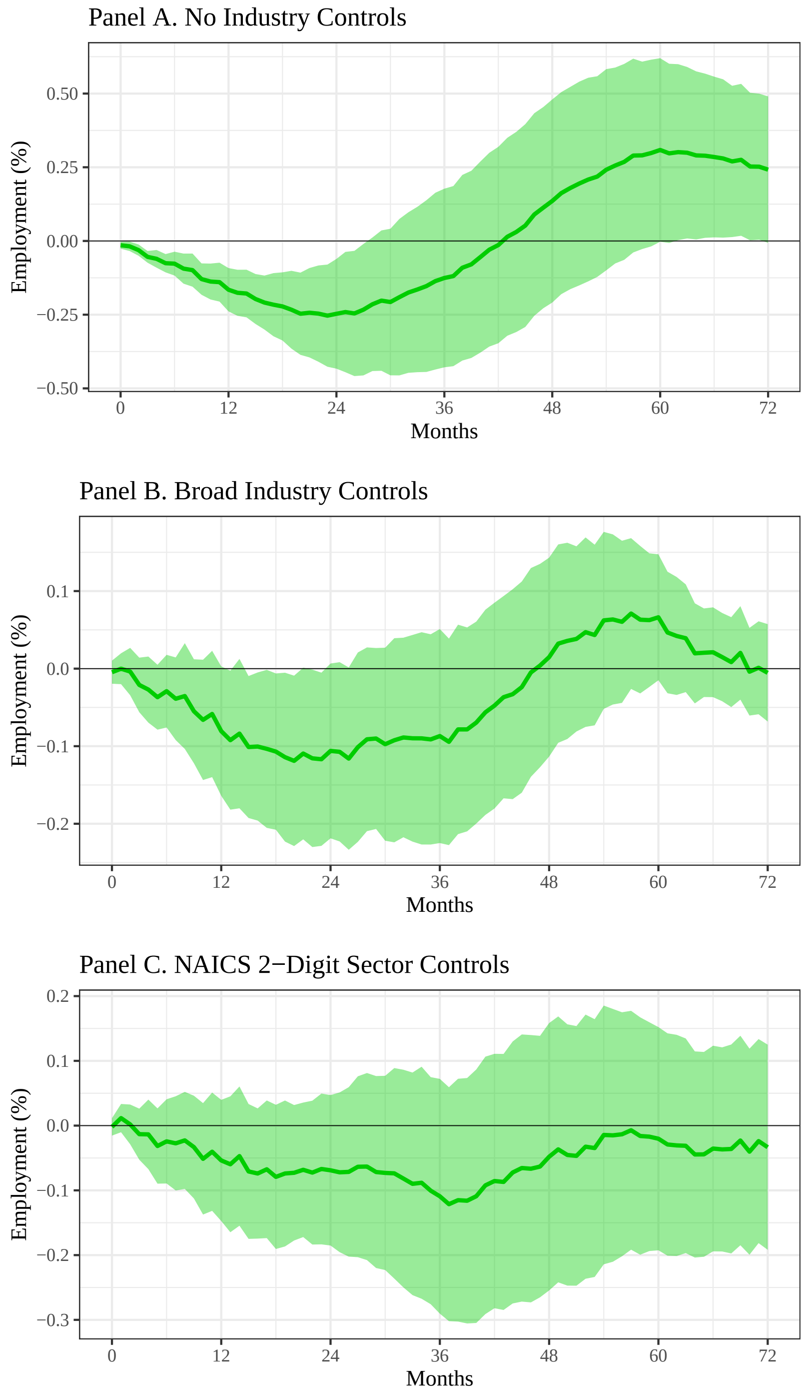


Figure 1.11: Regional Relative Impulse Responses to Monetary Shocks.

Notes: The figures plot the coefficients β^h from estimating equation (1.2) with different controls. The horizontal axis represents the horizon h in months. Panel B controls for the industrial composition of the CZs by using 12 broad industry shares and their interactions with the monetary shock. Panel C uses 23 NAICS 2-Digit Sector shares and their interactions with the monetary shock. 90% confidence intervals are shown. Robust standard errors are clustered by commuting zone and time.

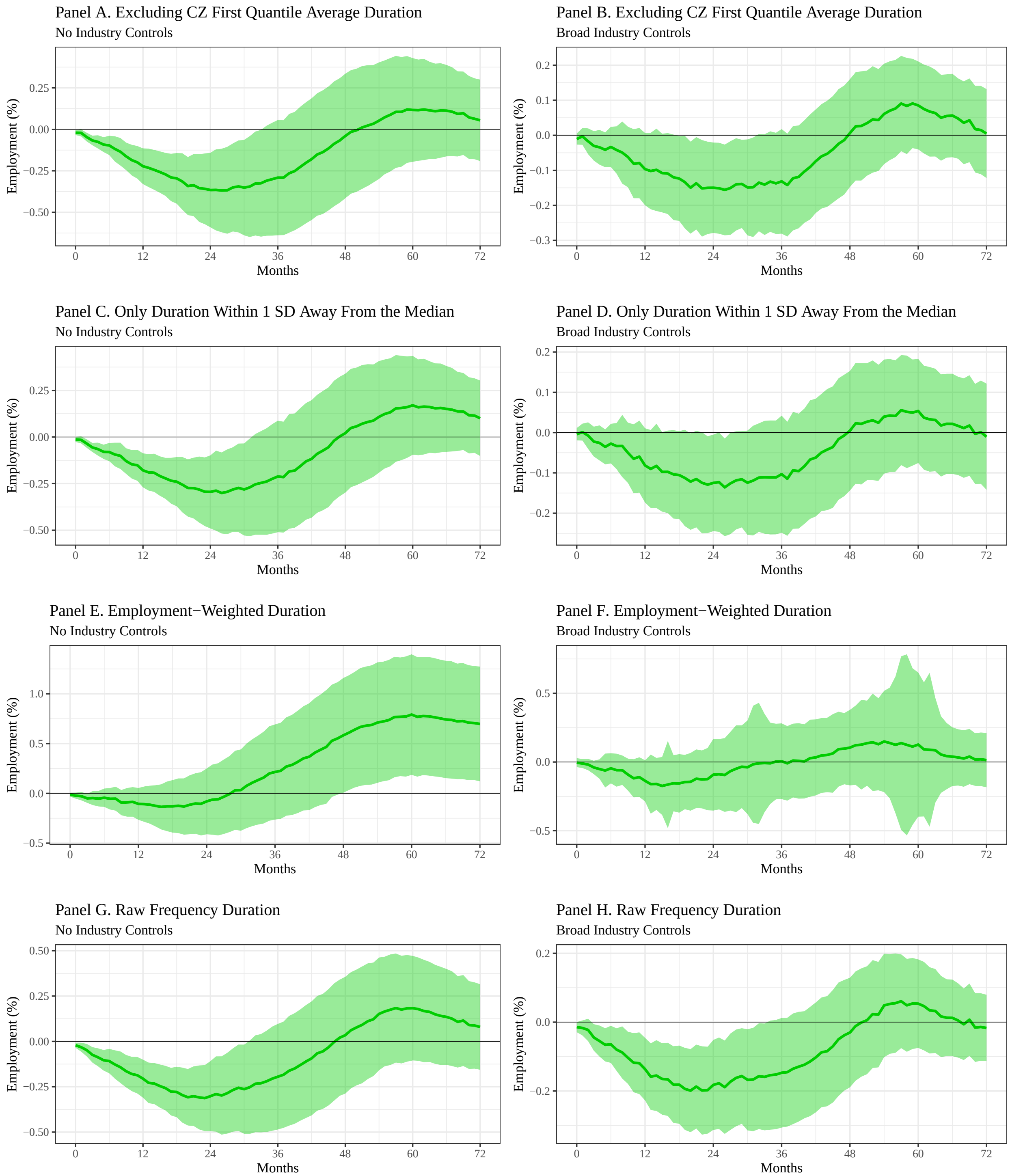


Figure 1.12: Regional Results Robustness.

Notes: The figures plot the coefficients β^h from estimating equation (1.2) with different controls. The horizontal axis represents the horizon h in months. Controls for the industrial composition of the CZs include 12 broad industry shares and their interactions with the monetary shock. 90% confidence intervals are shown. Robust standard errors are clustered by commuting zone and time.

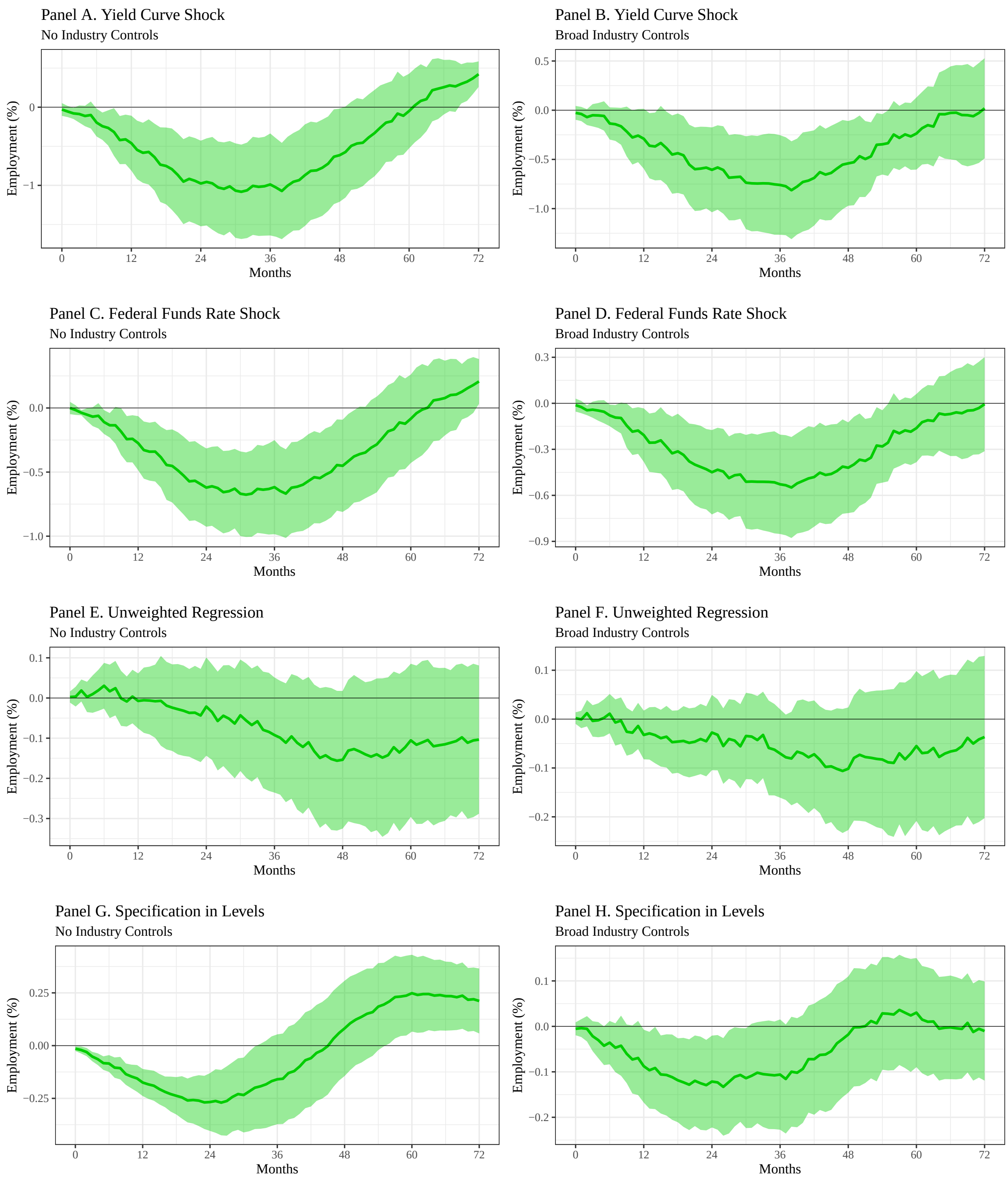


Figure 1.13: REGIONAL RESULTS ADDITIONAL ROBUSTNESS.

Notes: The figures plot the coefficients β^h from estimating equation (1.2) with different controls. The horizontal axis represents the horizon h in months. Controls for the industrial composition of the CZs include 12 broad industry shares and their interactions with the monetary shock. 90% confidence intervals are shown. Robust standard errors are clustered by commuting zone and time.

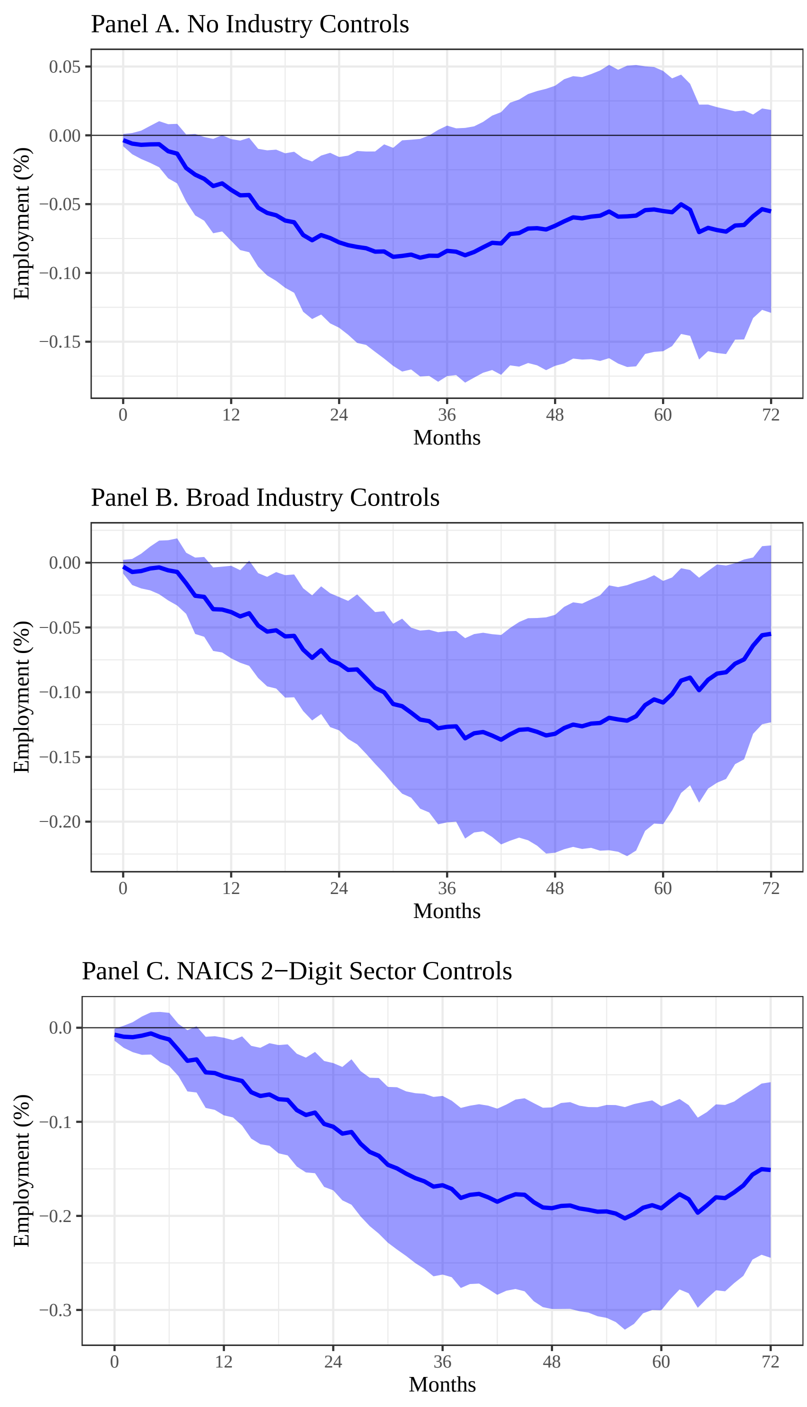


Figure 1.14: Industry Relative Impulse Responses to Monetary Shocks.

Notes: The figures plot the coefficients γ^h from estimating equation (1.3) using industry employment with different controls. Panel B includes 12 broad industry dummies and their interactions with the monetary shock. Panel C uses 23 NAICS 2-Digit Sector dummies and their interactions with the monetary shock. 90% confidence intervals are shown. Robust standard errors are clustered by 6-digit NAICS industry and time.

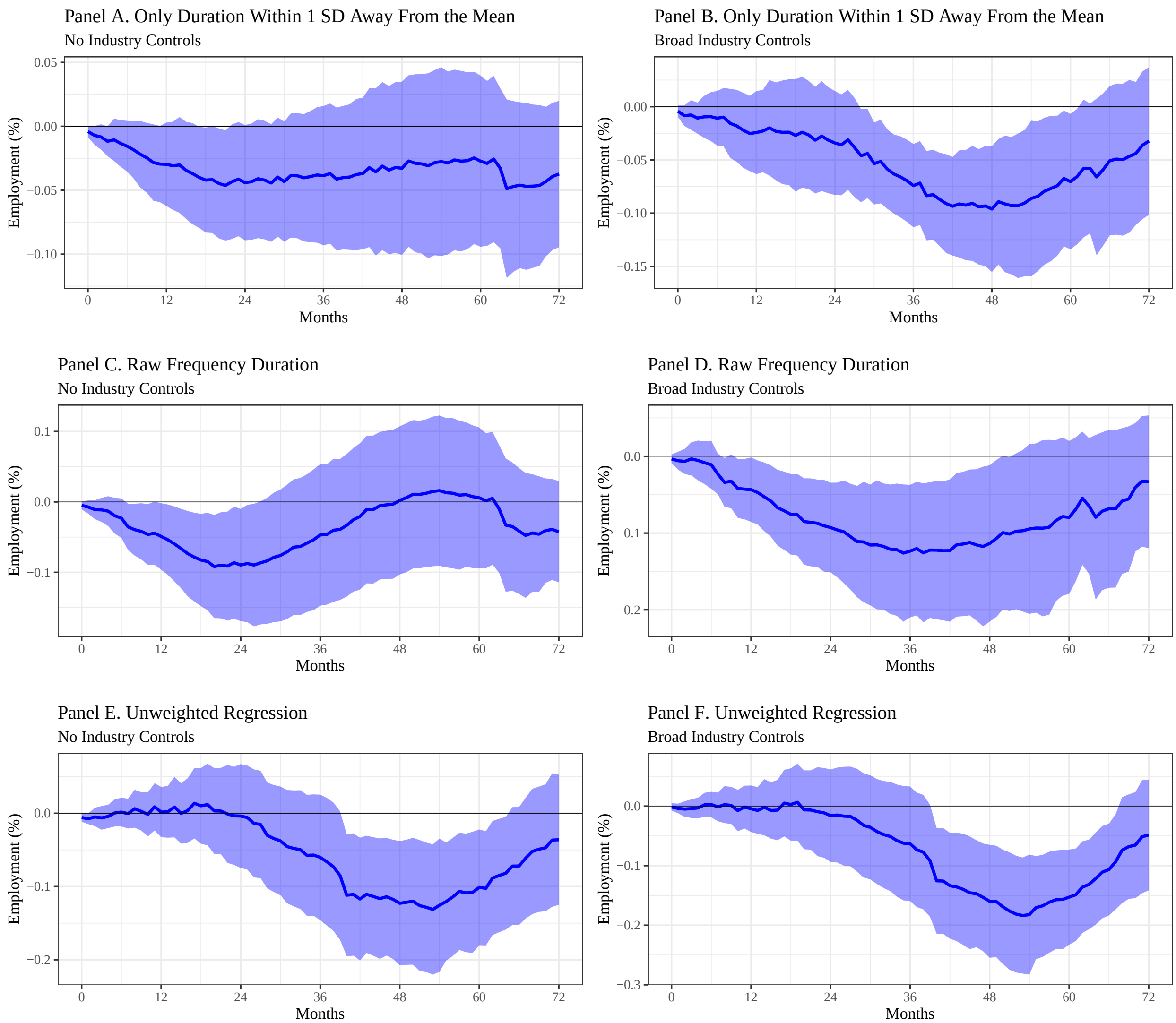


Figure 1.15: Industry Results Robustness.

Notes: The figures plot the coefficients γ^h from estimating equation (1.3) using industry employment with different controls. Panel B includes 12 broad industry dummies and their interactions with the monetary shock. Panel C uses 23 NAICS 2-Digit Sector dummies and their interactions with the monetary shock. 90% confidence intervals are shown. Robust standard errors are clustered by 6-digit NAICS industry and time.

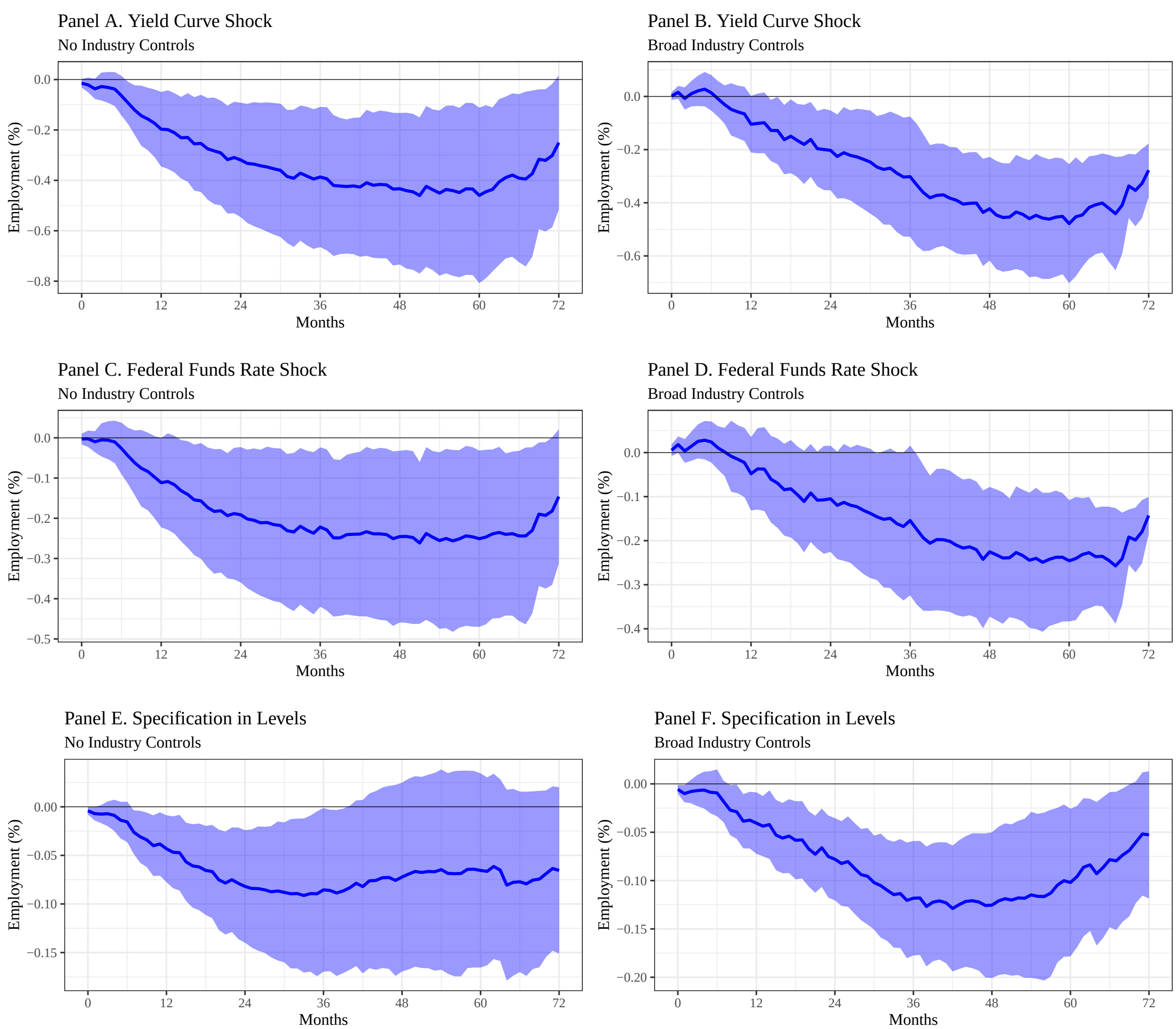


Figure 1.16: Industry Results Additional Robustness.

Notes: The figures plot the coefficients γ^h from estimating equation (1.3) using industry employment with different controls. Panel B includes 12 broad industry dummies and their interactions with the monetary shock. Panel C uses 23 NAICS 2-Digit Sector dummies and their interactions with the monetary shock. 90% confidence intervals are shown. Robust standard errors are clustered by 6-digit NAICS industry and time.

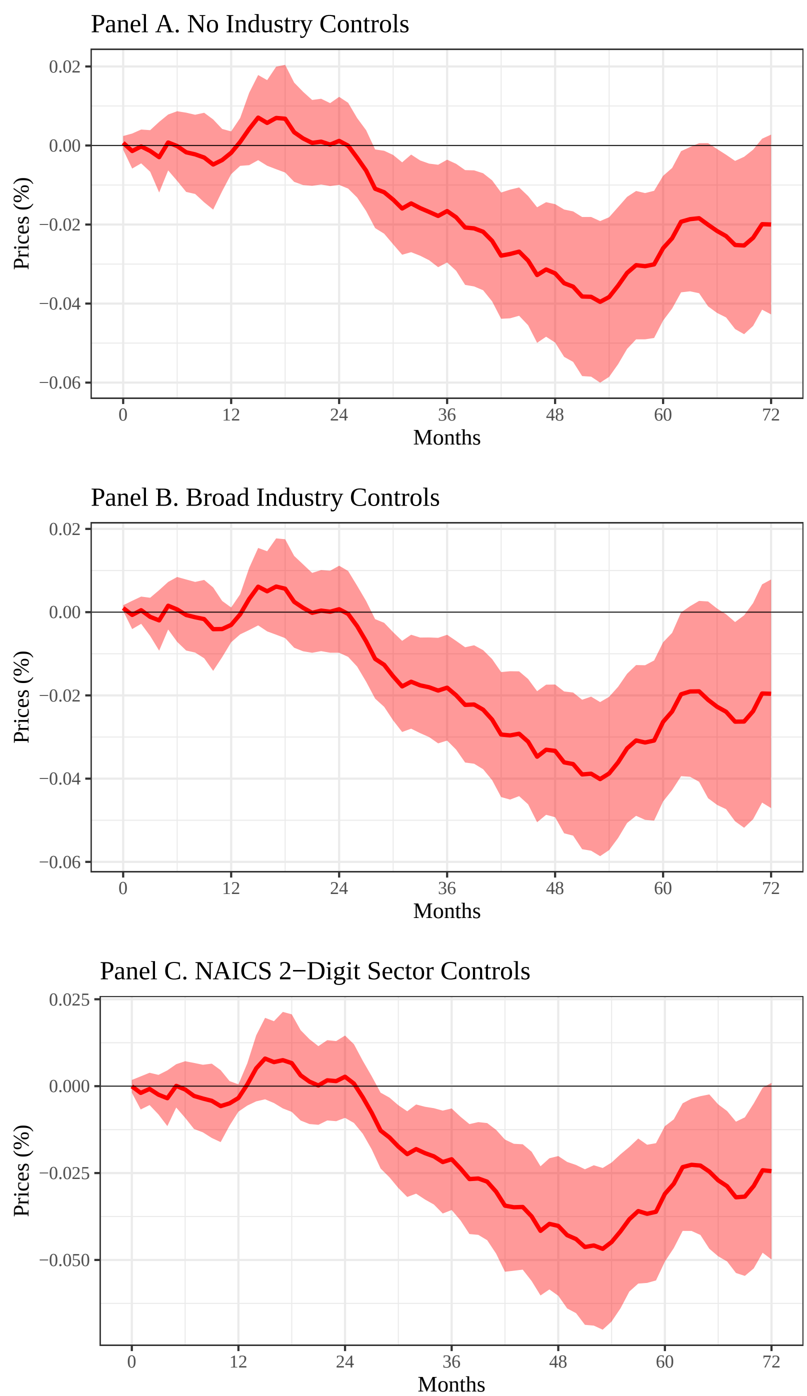


Figure 1.17: RELATIVE PRICE IMPULSE RESPONSES TO MONETARY SHOCKS.

Notes: The figures plot the coefficients $-\gamma^h$ from estimating equation (1.3) using industry prices with different controls. Panel B includes 12 broad industry dummies and their interactions with the monetary shock. Panel C uses 23 NAICS 2-Digit Sector dummies and their interactions with the monetary shock. 90% confidence intervals are shown. Robust standard errors are clustered by 6-digit NAICS industry and time.

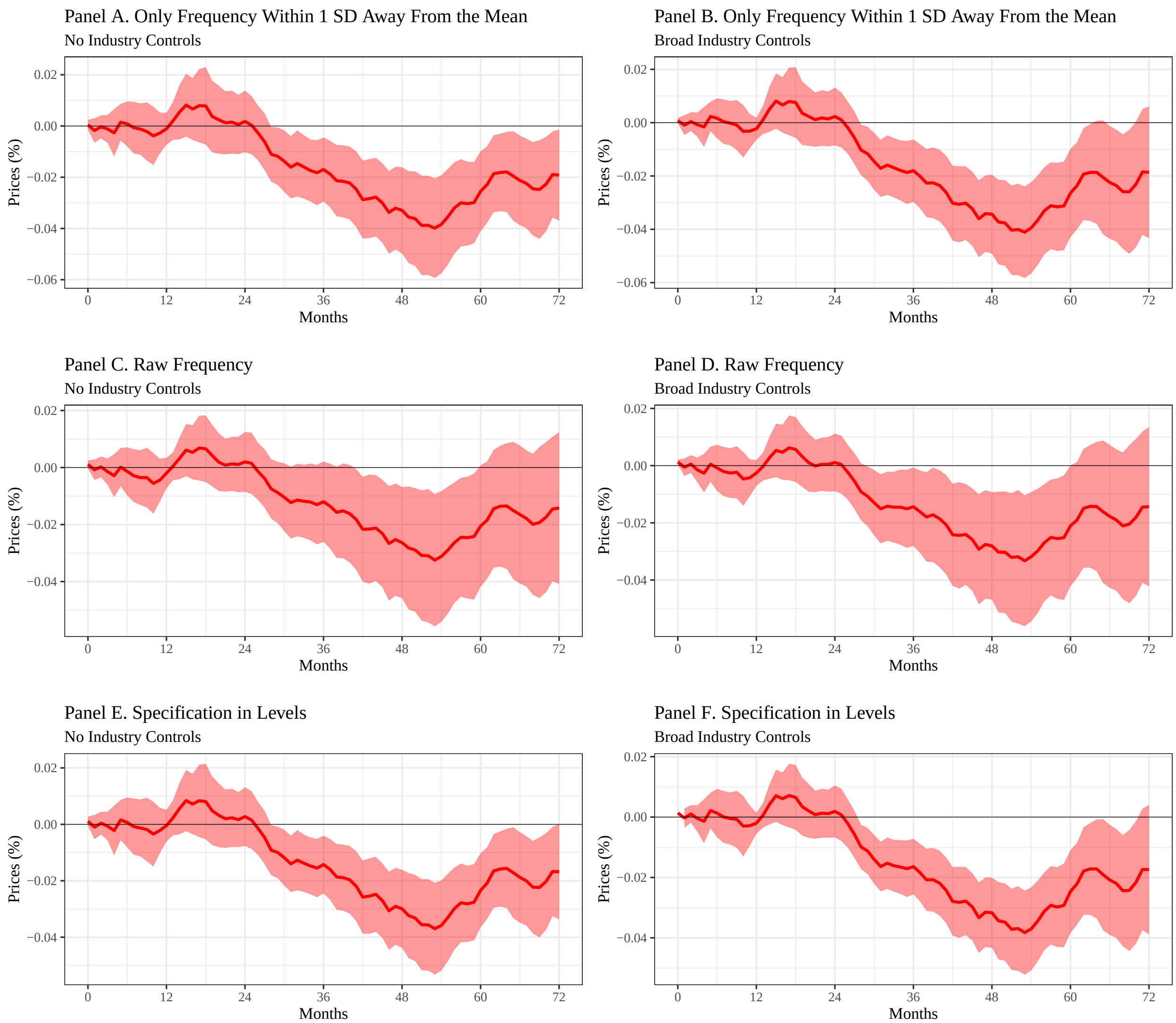


Figure 1.18: PRICE RESULTS ROBUSTNESS.

Notes: The figures plot the coefficients $-\gamma^h$ from estimating equation (1.3) using industry prices with different controls. Panel B includes 12 broad industry dummies and their interactions with the monetary shock. Panel C uses 23 NAICS 2-Digit Sector dummies and their interactions with the monetary shock. 90% confidence intervals are shown. Robust standard errors are clustered by 6-digit NAICS industry and time.

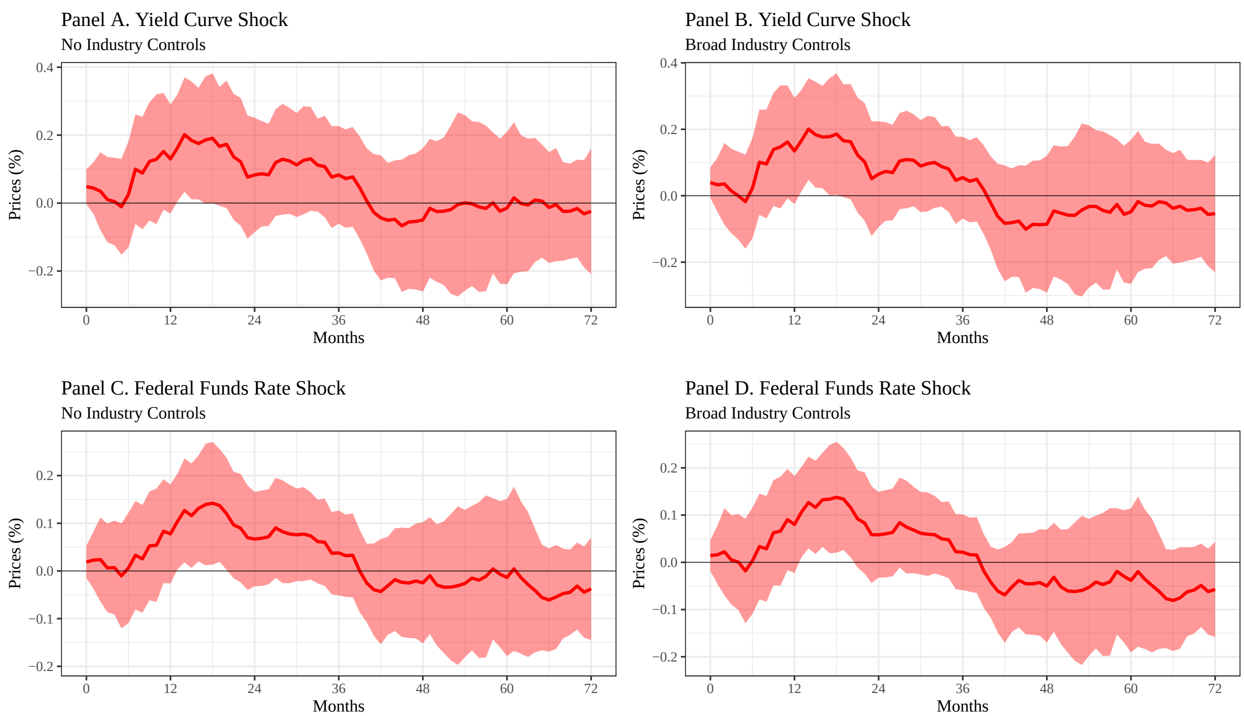


Figure 1.19: Relative Price Responses to High-Frequency Shocks.

Notes: The figures plot the coefficients $-\gamma^h$ from estimating equation (1.3) using industry prices with different controls. Panel B includes 12 broad industry dummies and their interactions with the monetary shock. Panel C uses 23 NAICS 2-Digit Sector dummies and their interactions with the monetary shock.90% confidence intervals are shown. Robust standard errors are clustered by 6-digit NAICS industry and time.

Table 1.1: CPI Data Descriptive Statistics.

	Filtered Frequency							Raw Frequency						
Category	Mean	SD	Min	Q1	Median	Q3	Max	Mean	SD	Min	Q1	Median	Q3	Max
	Panel A. Unweighted statistics													
All	12.05	15.28	0.00	4.13	7.02	11.59	100.00	23.02	16.41	1.57	10.79	19.75	31.29	100.00
Durables	9.71	15.71	1.90	3.92	6.35	8.46	100.00	20.09	14.77	4.59	12.88	17.20	23.30	100.00
Other Goods	12.44	15.28	0.00	4.07	7.62	12.35	88.61	27.43	15.56	1.72	17.10	25.79	35.25	88.63
Services	12.59	15.10	1.57	4.39	7.85	10.69	72.44	13.02	15.03	1.57	4.46	8.28	11.59	72.44
	Panel B. Statistics weighted by expenditure shares													
All	21.26	25.53	0.00	5.42	8.88	31.33	100.00	26.70	24.61	1.57	7.48	20.98	36.28	100.00
Durables	29.75	35.89	1.90	6.36	16.68	31.33	100.00	34.17	33.48	4.59	12.31	29.17	31.33	100.00
Other Goods	19.76	26.65	0.00	4.99	8.99	21.08	88.61	30.20	24.78	1.72	13.50	25.96	36.74	88.63
Services	19.94	19.45	1.57	6.17	8.95	38.14	72.44	20.23	19.43	1.57	6.26	10.32	38.14	72.44

Table 1.2: Industry Categories Hierarchy.

Broad Industry Category	NAICS Sector Code and Title
Primary Sector	11 Agriculture, Forestry, Fishing, and Hunting 21 Mining
Construction	23 Construction
Manufacturing	31 Food and Textile Manufacturing 32 Wood, Paper, Plastics, and Chemical Manufacturing 33 Metal, Machinery, Electronics, Transportation, and Furniture Manufacturing
Utilities	22 Utilities
Wholesale, Retail & Transportation	42 Wholesale Trade 44 Major Retail (Vehicles, Furniture, Electronics, Groceries, Clothing...) 45 Specialized Retail and Department Stores 48 Transportation 49 Couriers and Storage
Information	51 Information
Finance & Rentals	52 Finance and Insurance 53 Real Estate and Rental and Leasing
Professional Services	54 Professional, Scientific, and Technical Services 55 Management of Companies and Enterprises 56 Administrative and Waste Services
Education	61 Educational Services
Health	62 Health Care and Social Assistance
Accomodation, Food Services & Leisure	71 Arts, Entertainment, and Recreation 72 Accommodation and Food Services
Other Services	81 Other Services

Table 1.3: Matched NAICS Frequencies From CPI ELIs.

Primary Sector			Manufacturing		
NAICS	Title	Freq.	NAICS	Title	Freq.
1129	Other Animal Production	4.1	3159	Apparel Accessories and Other Apparel Manufacturing	1.7
2123	Nonmetallic Mineral Mining and Quarrying	7.4	3345	Navigational, Measuring, Electromedical, and Control Instruments Manufacturing	2.2
1151	Support Activities for Crop Production	9.6	3332	Industrial Machinery Manufacturing	2.4
1111	Oilseed and Grain Farming	9.9	3272	Glass and Glass Product Manufacturing	2.9
	...			...	
1113	Fruit and Tree Nut Farming	38.5	3251	Basic Chemical Manufacturing	25.2
1112	Vegetable and Melon Farming	39.5	3341	Computer and Peripheral Equipment Manufacturing	25.8
1123	Poultry and Egg Production	47.6	3361	Motor Vehicle Manufacturing	30.9
2111	Oil and Gas Extraction	80.9	3241	Petroleum and Coal Products Manufacturing	82.8

Wholesale, Retail & Transportation			Information		
NAICS	Title	Freq.	NAICS	Title	Freq.
4542	Vending Machine Operators	1.7	5182	Data Processing, Hosting, and Related Services	2.1
4483	Jewelry, Luggage, and Leather Goods Stores	3.3	5191	Other Information Services	3.9
4511	Sporting Goods, Hobby, and Musical Instrument Stores	3.4	5111	Newspaper, Periodical, Book, and Directory Publishers	5.9
4482	Shoe Stores	3.5	5112	Software Publishers	7.9
	...		5121	Motion Picture and Video Industries	9.0
4811	Scheduled Air Transportation	59.8	5172	Wireless Telecommunications Carriers (except Satellite)	10.0
4411	Automobile Dealers	61.2	5152	Cable and Other Subscription Programming	12.4
4247	Petroleum and Petroleum Products Merchant Wholesalers	85.9	5179	Other Telecommunications	34.4
4471	Gasoline Stations	87.6	5171	Wired Telecommunications Carriers	34.9

continued on next page...

Matched NAICS Frequencies From CPI ELIs. (Continued)

Utilities			Education		
NAICS	Title	Freq.	NAICS	Title	Freq.
2213	Water, Sewage and Other Systems	10.7	6112	Junior Colleges	5.8
2211	Electric Power Generation, Transmission and Distribution	38.1	6113	Colleges, Universities, and Professional Schools	5.8
2212	Natural Gas Distribution	72.4	6111	Elementary and Secondary Schools	6.2
2213	Water, Sewage and Other Systems	10.7	6114	Business Schools and Computer and Management Training	9.2
2211	Electric Power Generation, Transmission and Distribution	38.1	6115	Technical and Trade Schools	9.2
2212	Natural Gas Distribution	72.4			

Finance & Rentals			Professional Services		
NAICS	Title	Freq.	NAICS	Title	Freq.
5242	Agencies, Brokerages, and Other Insurance Related Activities	7.9	5411	Legal Services	1.6
5241	Insurance Carriers	8.1	5615	Travel Arrangement and Reservation Services	4.3
5322	Consumer Goods Rental	8.5	5412	Accounting, Tax Preparation, Bookkeeping, and Payroll Services	5.5
5242	Agencies, Brokerages, and Other Insurance Related Activities	7.9	5617	Services to Buildings and Dwellings	5.9
5241	Insurance Carriers	8.1	5419	Other Professional, Scientific, and Technical Services	9.0
5322	Consumer Goods Rental	8.5	5621	Waste Collection	9.4
5222	Nondepository Credit Intermediation	35.6			
5321	Automotive Equipment Rental and Leasing	44.7			

continued on next page...

Matched NAICS Frequencies From CPI ELIs. (Continued)

Health			Other Services		
NAICS	Title	Freq.	NAICS	Title	Freq.
6241	Individual and Family Services	2.8	8123	Drycleaning and Laundry Services	3.0
6213	Offices of Other Health Practitioners	3.1	8121	Personal Care Services	3.1
6211	Offices of Physicians	3.2	8129	Other Personal Services	4.1
6216	Home Health Care Services	3.7	8114	Personal and Household Goods Repair and Maintenance	6.7
6212	Offices of Dentists	4.5	8129	Other Personal Services	4.1
6221	General Medical and Surgical Hospitals	6.3	8114	Personal and Household Goods Repair and Maintenance	6.7
6222	Psychiatric and Substance Abuse Hospitals	6.3	8122	Death Care Services	8.9
6223	Specialty (except Psychiatric and Substance Abuse) Hospitals	6.3	8112	Electronic and Precision Equipment Repair and Maintenance	11.1
6244	Child Day Care Services	6.9	8111	Automotive Repair and Maintenance	14.9

Accomodation, Food Services & Leisure		
NAICS	Title	Freq.
7213	Rooming and Boarding Houses	4.7
7221	Full-Service Restaurants	5.0
7223	Special Food Services	5.9
7222	Limited-Service Eating Places	7.6
7113	Promoters of Performing Arts, Sports, and Similar Events	8.1
7139	Other Amusement and Recreation Industries	8.6
7224	Drinking Places (Alcoholic Beverages)	9.8
7211	Traveler Accommodation	41.7

Note: The frequency measure is the filtered frequency. Industry frequencies are aggregated at the 4-digits NAICS level using expenditure shares from the CPI.

Chapter 2

The Fiscal Multiplier Puzzle: Liquidity Traps, Bounded Rationality, and Incomplete Markets

2.1 Introduction

The Great Recession had profound implications for macroeconomic stabilization policy. Most developed economies quickly hit the zero lower bound, removing the key policy instrument that central banks have to stimulate the economy. Macroeconomists and policy makers quickly emphasized fiscal policy as an alternative stabilization tool in the absence of conventional monetary policy. This spurred a large body of research focused on studying the impact of government spending when monetary policy is constrained by the zero lower bound. Christiano et al. (2011a), Eggertsson (2011) and Woodford (2011) are some examples that use the workhorse New Keynesian model to study the effects of fiscal policy in a liquidity trap. They all find sizable fiscal multipliers much larger than one. In fact, the extreme sensitivity of the New Keynesian model to fiscal policy in a liquidity trap is arguably too large to be credible. Farhi and Werning (2016) show that multipliers grow explosively when government spending is backloaded further into the future. Far from that, they also show that fiscal policy can have a completely opposite effect and become recessionary when hand-to-mouth household are included and fiscal deficits are incurred. Finally, they find that these extreme results become worse when prices are relatively more flexible, echoing similar results from Werning (2012) about the depth of liquidity trap recessions.

In this paper, we study the reasons behind the extreme sensitivity of the New Keynesian model to fiscal policy announcements—a phenomenon we call the "fiscal multiplier puzzle"—

and provide a solution to it. We begin by deviating from the standard assumption of rational expectations and allow for arbitrary belief formation at the time of the policy announcement. This allow us to see the explicit role played by expectations in characterizing the response of the economy. We find that the fiscal multiplier puzzle depends crucially on the inflation-feedback loop that results from common knowledge between the demand and supply blocks of the economy. Next, we implement a particular form of bounded rationality based on level-k thinking. This is a recursive belief formation process that nests rational expectations as a limit case when $k \to \infty$. In contrast to rational expectations, level-k fiscal multipliers are bounded and tend to zero as fiscal spending is realized further into the future. We also explore the consequences for fiscal policy of interacting bounded rationality and incomplete markets. Incomplete markets models turn out to be plagued by the same multiplier puzzles featured in standard complete markets models with hand-to-mouth households. Compared to complete markets, bounded rationality dampens the multipliers in incomplete markets to a larger extent, with the mitigation increasing in the degree of market incompleteness. Moreover, backloading taxes has recessionary effects in complete markets both under rational expectations and bounded rationality. In contrast, running deficits helps stimulating the economy when both bounded rationality and incomplete markets are considered.

Our exploration on the role of expectations in producing the fiscal multiplier puzzles relies on the concept of temporary equilibrium.[23] This concept defines equilibrium outcomes conditional on exogenous sets of beliefs held by the agents in the economy. We develop a belied-augmented New Keynesian model that allow us to decompose the impact of fiscal policy in two parts: the naive multipliers and the belief multipliers. The naive multipliers characterize the response of the economy when expectations about future variables do not react to the policy announcement. On the other hand, the belief multipliers consist of the reaction induced by changes in the path of beliefs of future economic conditions. We show that the explosive response to fiscal policy in rational expectations must come from the belief multipliers. Then, we use this framework to isolate what element of rational expectations is key in inducing the puzzle. In particular, massive multipliers arise when we allow for common knowledge between household and firms in their pricing and spending decisions. Government spending boosts aggregate demand and inflation, which decreases the real interest rate and encourages higher spending by households due to intertemporal substitution—the inflation-spending feedback loop.

Given the critical role played by rational expectations, we study the effects of fiscal policy under bounded rationality, in particular, level-k thinking. This is a recursive belief

[23]See Evans (2001) for a discussion on the origins of temporary equilibria in economics and the learning literature.

formation process previously introduced in macroeconomics by Farhi and Werning (2019), similar to the concept of "reflective equilibrium" in Garcia-Schmidt and Woodford (2019). Starting from a baseline rational expectations equilibrium, agents form expectations about future macroeconomic variables based on a finite deductive process about others' behavior, involving k iterations. This belief formation process has been proposed in the literature on game theory as a solution to counterfactual predictions of Nash equilibria that require tremendous amounts of coordination among lots of agents and are very sensitive to the common knowledge of rationality assumption.[24] We believe these characteristics make level-k particularly well-suited for the kind of announcements that we are interested in. The zero lower bound is a totally new environment that has never been experienced before in the last century. Similarly, fiscal policy was mostly abandoned as a countercyclical tool in favor of monetary policy a long time ago. These features make other belief processes such as learning appear implausible.

Level-k is very effective at reducing the pervasive effect of the inflation-spending feedback loop. Fiscal multipliers under this form of bounded rationality are bounded, and tend to zero as the policy action occurs further into the future. Intuitively, level-k breaks the feedback loop by only allowing k iterations of the loop. Government spending raises income and aggregate demand, but this is a surprise to households and firms in the level-zero equilibrium as they belief all aggregate variables will remain in the steady state. They turn out to be wrong though, so household get higher income than expected and firms end up paying larger wages than expected. Then, level-one agents belief everybody else is a level-zero agent and thus think that aggregate variables will be consistent with the level-zero equilibrium. Because of the nature of this coordination setup and the actual discounting of future events, events that happen further away into the future require further rounds of the loop to have an effect at $t = 0$. Spending multipliers in the level-five equilibrium are roughly 50 percent lower than the rational expectations benchmark for shocks happening three years into the future, and 90 percent lower for shocks in a five year horizon. As the level of rationality increases, they will incorporate further rounds of the reactions of others and the beliefs will eventually converge to rational expectations when $k \to \infty$. However, given the finite nature of the loop for any finite k and the effective discounting of future events, unbounded multipliers only arise in the limit as $k \to \infty$.

Farhi and Werning (2019) show the multipliers can be infinitely recessionary when hand-

[24] For evidence supporting level-k thinking in games of full information, see Stahl and Wilson (1994), Stahl and Wilson (1995), Nagel (1995a), Bosch-Domènech et al. (2002), Camerer et al. (2004), Mauersberger and Nagel (2018). They all find evidence for low levels of reasoning (between 0 and 3) when individuals are confronted with a new game and have the rules explained to them, despite them knowing that everybody else knows the rules as well.

to-mouth agents are introduced in the New Keynesian model, the so-call Two-Agent New Keynesian (TANK) model. We show the same result holds under incomplete markets à la Bewley-Hugget Aiyagari, where households are pushed towards their borrowing constraint endogenously. Level-k effectively mitigates this recessionary effect in the presence of hand-to-mouth both with complete and incomplete markets. Interestingly though, bounded rationality and incomplete markets have a strong interaction and exhibit much larger attenuation of the fiscal multipliers when k is low and market incompleteness is large.[25] Increasing the fraction of hand-to-mouth households from 20 to 80 percent increases the attenuation of the incomplete markets response in the level-two equilibrium from 30 to 70 percent with respect to the complete markets benchmark, for a shock 4 years into the future. Moreover, the interaction between bounded rationality and incomplete markets makes running deficits expansionary, at least for low levels of reasoning. Deficits tend to be recessionary in the future when taxes are raised. This induces future deflation and offsets the expansionary effects of early spending. Under incomplete markets though, issuing debt has an expansionary effect by providing liquidity services. Altogether, the liquidity services and the higher attenuation of future events with level-k and incomplete markets makes deficit-financed stimuli relatively more expansionary than equivalent budget-balanced stimuli.

Related literature. This paper is related to the growing literature that deviates from rational expectations and introduces bounded rationality and informational frictions in macroeconomic models. The closest paper to ours is Farhi and Werning (2019), who study the impact of introducing bounded rationality and incomplete markets for the forward guidance puzzle, that is, the extreme sensitivity of the New Keyenesian model to monetary policy. They find that both channels together can effectively mitigate the effect of forward guidance. Similarly, Angeletos and Lian (2018) show that introducing informational frictions and lack of common knowledge about the monetary policy announcement can also rationalize the forward guidance puzzle. Gabaix (2017) also rationalizes the forward guidance puzzle implementing a microfunded myopia towards distant atypical events. Garcia-Schmidt and Woodford (2019) use a similar equilibrium concept to ours with a continuous index of reflection and use it to challenge the neo-Fisherian predictions that higher interest rates lead to higher inflation rates under rational expectations. Molavi (2019) explores an environment were agents learn about the economy with misspecified low-factor models.

We also contribute to the literature analyzing the impact of fiscal policy in incomplete market models. Hagedorn et al. (2019) explores quantitatively the impact of fiscal policy in

[25] This same strong interaction between bounded rationality and incomplete markets is reported by Farhi and Werning (2019) when evaluating the effects of monetary policy announcements.

a large incomplete markets model with nominal rigidities. In contrast to our paper, they do not find the fiscal multiplier puzzle we try to solve. The main reason is that their analysis is based on a different nonstandard equilibrium selection criteria based on a long-run nominal debt anchor that forces the price level to go back to its previous steady state value in the long run. This prevents the large inflationary and deflationary episodes that we emphasize here and makes their multipliers more comparable to the currency-union multipliers in Farhi and Werning (2016). Auclert et al. (2018) also study the fiscal multipliers in an incomplete market framework with rigid prices and derive a set of sufficient statistics that characterize the response of the economy, the intertemporal marginal propensities to consume. They find that the fiscal response is determined by the interaction of their sufficient statistics and the path of public deficits. We also find a similar decomposition but have additional channels coming from the reaction of inflation.

Finally, the concept of level-k thinking originated from the literature on game theory as a solution to counterfactual predictions of Nash equilibria. Apart from theoretical developments, the literature has focused on measuring the level of rationality on laboratory experiments in Stahl and Wilson (1994), Stahl and Wilson (1995), Nagel (1995a), Bosch-Domènech et al. (2002), Camerer et al. (2004), and Mauersberger and Nagel (2018). They all find very low levels of rationality bellow 3 in relatively simple games.

The remainder of the paper is organized as follows. Section 2.2 describes the fiscal multiplier puzzle using a standard rational expectations New Keynesian framework. It also shows the prevalence of the puzzle in incomplete market models. Section 2.3 develops a belief-augmented New Keynesian framework based on the concepts temporary equilibrium, and use it to show that the fiscal multiplier puzzle is caused by the inflation-spending feedback loop that results from common knowledge between the demand the supply blocks of the economy. Section 2.4 incorporates level-k thinking to the model from the previous sections and derives the fiscal multipliers under this form of bounded rationality. It also studies the interaction of incomplete markets and bounder rationality. Section 2.5 concludes.

2.2 The Fiscal Multiplier Puzzle

In this section, we present the extreme sensitivity of different variants of the rational expectations New Keynesian (NK) model to fiscal policy announcements, particularly in the future, when monetary policy is constrained by the Zero Lower Bound—a phenomenon we call the “fiscal multiplier puzzle.”

2.2.1 Fiscal Multipliers With a Representative Agent

Consider a standard textbook Representative Agent New Keynesian (RANK) model with fiscal policy, as in Woodford (2011). The log-linear approximation of the model can be characterized by the (aggregate) Euler equation

$$c_t = \mathbb{E}c_{t+1} - \bar{\sigma}(i_t - \mathbb{E}_t \pi_{t+1}), \tag{2.1}$$

which aggregates spending decisions by households, and the Phillips Curve

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa \left[c_t + (1-\xi) g_t\right], \tag{2.2}$$

which aggregates the pricing decisions by firms. Lowercase variables represent percentage deviations from their zero-inflation steady-state level. In particular, $c_t = (C_t - C)/Y$ denotes the consumption gap normalized as percentage points of steady-state GDP; π_t denotes inflation, $g_t = (G_t - G)/Y$ denotes government spending gap normalized by steady-state GDP; i_t is the nominal interest rate gap; $\bar{\sigma}$ is the aggregate elasticity of intertemporal substitution; β denotes the discount factor; κ reflects the degree of price rigidities; and ξ denotes the fiscal multiplier under flexible prices, which is positive but less than 1 and is driven entirely by a wealth effect on labor supply.

Our objective is to understand how temporary fiscal shocks g_t affect the economy when monetary policy is constrained (e.g. the Zero Lower Bound), so interest rates do not respond to changes in the economic conditions. For simplicity, suppose $i_t = 0$ for all periods. Also, assume the fiscal policy announcement takes the form of an MIT shock, so the economy receives no other shocks after the announcement. Then, rational expectations implies perfect foresight and $\mathbb{E}_t x_{t+h} = x_{t+h}$ for all periods t and horizons h. Iterating forward the Euler equation and the Phillips Curve gives[26]

$$c_t = \bar{\sigma} \sum_{h=0}^{\infty} \pi_{t+1+h}, \tag{2.3}$$

$$\pi_t = \kappa \sum_{h=0}^{\infty} \beta^h \left[c_{t+h} + (1-\xi) g_{t+h}\right]. \tag{2.4}$$

These equations let us analyze some of the mechanisms driving the response of the economy to changes in fiscal policy. First, equation (2.3) establishes that *all* the effects of fiscal policy on consumption must come exclusively through *future inflation* in the RANK model. Future

[26]We assume the economy goes back to its steady state in the long run, when the fiscal expansion is over. Thus, $\lim_{h\to\infty} c_{t+h} = \lim_{h\to\infty} \pi_{t+h} = 0$. One way to do this could be to enforce a Taylor rule with the appropriate coefficients after some period T far into the future.

inflation affects real interest rates, motivating households to change their spending path due to intertemporal substitution. Combining this with the goods market clearing condition $y_t = c_t + g_t$, one can conclude that fiscal multipliers must be equal to 1 when prices are *completely rigid* ($\kappa = 0$) and there is no inflation whatsoever. This can be explained through the lens of Ricardian equivalence. Consumption remains unchanged as higher g_t involves a one-to-one increase in income with rigid prices, but also involves an equivalent increase in the present value of future taxes. Second, equation (2.4) indicates that government spending stimulates inflation by boosting aggregate demand directly and indirectly, through its effect on consumption. Firms facing higher demand require more labor and bid up wages, so they tend to adjust prices to reflect this path of higher marginal costs.

The linkage between consumption and inflation constitutes the inflation-spending feedback loop; higher *expected* inflation leads to higher consumption, and higher *expected* consumption leads to higher inflation. This mechanism amplifies the initial direct response of inflation via its indirect effect on consumption. As a matter of fact, this loop is extremely powerful under rational expectations. To see this, we solve the system of difference equations formed by the Euler equation (2.1) and the Phillips Curve (2.2) in Appendix A. Following Farhi and Werning (2016), define the spending multipliers as

$$\alpha_h^g = \frac{\bar{\sigma}\kappa(1-\xi)}{v_H - v_L}(v_H^h - v_L^h),$$

where $0 < v_L < 1 < v_H$ are the eigenvalues of the system. Then, equilibrium output under rational expectations is given by

$$y_t^{RANK} = g_t + c_t = g_t + \sum_{h=1}^{\infty} \alpha_h^g g_{t+h}. \tag{2.5}$$

The second term in the equation characterizes the response of current *consumption* due to the inflation-spending feedback loop. The spending multipliers α_h^g are shown in Figure 2.1.[27] They capture the change in aggregate demand due to an increase in government spending h periods from now. We can immediately notice that the response of current output starts at zero for current spending and increases exponentially as spending is backloaded into the future. The reason behind this explosive behavior is the inflation-spending feedback loop. Intuitively, fiscal expansion creates a boom, putting upward pressure on prices and boosting inflation. Higher inflation decreases real interest rates and encourages higher consumption in earlier periods, which feeds back into higher prices and further rounds of higher inflation and

[27] Given the exponential growth of the spending multipliers, we use a logarithmic scale and add one to α_s^g to be able to represent a zero effect.

consumption. Rational expectations makes this interaction particularly strong, as everybody has common knowledge of this mechanism and take it into account when forming beliefs about everyone else's actions. Moreover, the spending multipliers increase as prices become more flexible due to larger reactions of inflation in the feedback loop.

We assess the quantitative magnitude of the multipliers by running a series of fiscal experiments using an arguably standard calibration described in Table 2.1. The experiments consist of fiscal stimuli starting today that last for several periods. Figure 2.2 shows the response of current output depending on the length of the stimulus, where darker lines correspond to a larger degree of price stickiness. Not only fiscal multipliers explode when spending happens infinitely far into the future, but they also grow incredibly fast for finite horizons. Away from the extreme case of rigid prices, our calibrations yield a minimum value for the dominating eigenvalue v_H of 1.5. Thus, the response of *consumption* increases by at least 50% for every additional quarter of government spending. Even with an average price duration of 6 quarters, it only takes a fiscal expansion of roughly 7 quarters to make the impact multiplier larger than 3, which is arguably the upper bound of most empirical estimates in the literature (Ramey, 2016). Finally, we can also observe that impact multipliers become larger and steeper when prices are revised more often, reflecting the importance of the inflation-spending feedback loop. Highly flexible prices lead to larger fluctuations in inflation for a given change in aggregate demand. This generates a more pronounced response of consumption, thereby reinforcing the feedback loop.

2.2.2 Fiscal Multipliers With Hand-to-Mouth Households

We next consider a Two Agent New Keynesian (TANK) model similar to Gali et al. (2007), by adding a constant fraction of "constrained" households to the RANK model from previous section. In contrast to regular "Ricardian" households, constrained agents are prevented from access to financial markets, thereby behaving as hand-to-mouth (HtM) consumers. Still, they optimize their labor supply decision in the same way as Ricardian households. The log-linear approximation to the model around the symmetric steady-state can be characterized by a slightly updated version of the previous Euler equation,

$$c_t = \mathbb{E}c_{t+1} - \tilde{\sigma}(i_t - \mathbb{E}\pi_{t+1}) - (\Omega_g \mathbb{E}\Delta g_{t+1} - \Omega_\tau \mathbb{E}\Delta t_{ht+1}), \tag{2.6}$$

and the same Phillips Curve (2.2) as in the RANK model. Here, t_{ht} denotes lump-sum taxes collected from HtM agents at time t; Δ is the first difference operator $\Delta x_t = x_t - x_{t-1}$; $\tilde{\sigma}$ is proportional to the elasticity of intertemporal substitution; and Ω_g and Ω_τ are positive parameters that measure the impact of g_t and t_{ht} on aggregate consumption through its effect on HtM income.[28]

Two crucial differences stand out in comparison to the RANK model. First, the path of government spending plays a direct role in determining aggregate consumption. This is because government purchases increase income for HtM households and they immediately decide to spend it. Second, Ricardian equivalence no longer holds, so the precise path for HtM taxes also has an impact on aggregate consumption allocation. To be precise, both the timing and share of taxes paid by each type of household affects the response of consumption, and hence the fiscal multipliers.

However, before moving on with the discussion, it is crucial to acknowledge that fiscal policy induces redistribution away from the representative agent case. Here, the split of taxes and the cyclicality of income for each household type, which is determined by the split of profits, fully characterize this redistribution channel. This channel plays an important role in shaping the fiscal multipliers in a TANK economy. In order to isolate some of the effects of these additional features that come with HtM agents, we initially consider a "neutral" case that shuts down the redistribution channel. This can be achieved by distributing profits uniformly across households, so that individual income is proportional to aggregate income, and raising lump-sum taxes proportionally to individual earnings.[29] Under these assumptions, we show in the Appendix A that $\tilde{\sigma} = \bar{\sigma}$, $\Omega_g = \Omega_\tau = \Omega$ and $t_{ht} = t_t$, where t_t denotes the aggregate amount of lump-sum taxes levied at time t. Then, the Euler equation (2.6) becomes

$$c_t = \mathbb{E}c_{t+1} - \bar{\sigma}(i_t - \mathbb{E}\pi_{t+1}) - \Omega\mathbb{E}\Delta(g_{t+1} - t_{t+1}). \tag{2.7}$$

Again, we are interested in understanding how a TANK economy responds to temporary fiscal policy shocks when monetary policy is constrained. We iterate equation (2.7) forward assuming, for simplicity, perfect foresight and $i_t = 0$ for all t. As shown in the Appendix A, this gives

$$c_t = \Omega(g_t - t_t) + \bar{\sigma}\sum_{h=0}^{\infty}\pi_{t+1+h}. \tag{2.8}$$

At first glance, the effects of government spending and taxes in this model can be analyzed independently from each other. First, the impact of government spending *alone* in the TANK model is necessarily larger thank in the RANK model. This is a consequence of the additional term Ωg_t in (2.8). Fixing future inflation, fiscal stimulus directly increases consumption due to the HtM receiving more income, the typical Keynesian cross channel emphasized in undergraduate macroeconomics. Then, equation (2.4) reveals that the initial reaction

[28] These parameters depend on the share of constrained agents and the way profits are split between both types of households.

[29] Similar assumptions are made in Werning (2015) to obtain neutrality with respect to monetary policy.

of inflation to government spending will be amplified due to the increase in consumption from the HtM. This larger increase in inflation in the initial round gets exacerbated by the inflation-spending feedback loop, suggesting larger effects of fiscal policy in TANK.

On the other hand, the effect of levying taxes is precisely the opposite of government spending. Taxes have a direct *negative* impact on consumption by reducing the resources of HtM households. Moreover, this gets transmitted to inflation and back into consumption through the familiar feedback loop, thereby magnifying its initial impact.

All in all, fiscal multipliers in the TANK model could be either higher or lower than in the RANK benchmark. Spending and taxes are constrained to satisfy the intertemporal budget constraint of the government, which will help in determining what channel is more important. Interestingly, these countervailing forces cancel each other out when the fiscal expansion is budget-balanced, that is, $g_t = t_t$ for all t. Note that equations (2.7) and (2.8) reduce to its RANK counterparts (2.1) and (2.3) when the budget is balanced. Thus, fiscal multipliers in the RANK and the "neutral" TANK models must be exactly the same in this particular case.[30]

Generally speaking, the inflation-feedback loop is incredibly powerful here as well. We follow the same steps as before and solve the system of difference equations given by (2.7) and (2.2). Define the deficit multipliers as

$$\alpha_h^{g-t} = \frac{\bar{\sigma}\kappa}{v_H - v_L}\left[v_H^{h-1}(v_H + 1) - v_L^{h-1}(v_L + 1)\right],$$

where $0 < v_L < 1 < v_H$ are the eigenvalues of the system. As we show in the Appendix A, the response of output can be decomposed according to

$$y_t = y_t^{RANK} + \Omega(g_t - t_t) + \sum_{h=1}^{\infty} \alpha_h^{g-t}\Omega(g_{t+h} - t_{t+h}), \tag{2.9}$$

where y_t^{RANK} corresponds to the level of output that would prevail in a RANK version of this economy featuring no HtM households, given by equation (2.5). We can easily observe that including HtM consumers adds additional effects, on top of the RANK response, that depends on the path of (primary) deficits. The first difference from the RANK multipliers comes from the term $\Omega(g_t - t_t)$, implying that deficits have a contemporaneous impact on output. This effect comes from the Keynesian cross mechanism we mentioned earlier. Besides that, the deficit multipliers α_h^{g-t}, which are shown in Figure 2.1, capture the additional impact in

[30]This result holds more generally, away from the neutral case. The countervailing forces of government spending and taxes on HtM consumption cancel each other out whenever $\Omega_g g_t = \Omega_\tau t_{ht}$ for all periods. Then, TANK multipliers are equivalent to those in a RANK with an elasticity of intertemporal substitution equal to $\tilde{\sigma}$. A similar result is obtained by Auclert et al. (2018) in a setup with rigid prices and constant real rate.

current output of a fiscal deficit happening h periods from now. At first glance, we can see that this additional channel also explodes when deficits are backloaded further away into the future. The exponential growth is also caused by the familiar inflation-spending feedback loop. Yet, a crucial difference compared to the RANK channel is that it can induce *negative* responses when agents expect the government to run fiscal surpluses in the future, as in a fiscal consolidation. Intuitively, taxes will decrease HtM spending and put downward pressure in prices. This depresses spending early on through higher real interest rates, creating a deflation-spending feedback loop. A surprising consequence is that fiscal deficits today can easily be recessionary, in that they will induce fiscal surpluses in the future. This could seem counterintuitive, as the presence of HtM households is often used as an argument for fiscal deficits. These agents are unable to optimize intertemporally so, in contrast to Ricardian households, they ignore the raise in future taxes implied by current deficits and further stimulate the economy by spending all their additional income. But this argument backfires under rational expectations; current deficits imply future surpluses, and future shocks have a larger impact on current output due to the feedback loop.[31]

We conclude by running a series of experiments with fiscal deficits, to demonstrate quantitatively how easily they lead to recessions in TANK. We use the same calibration as in 2.2.1 with 6 quarters of average price duration, adding a somewhat conservative 20 percent share of constrained households.[32] The experiments consist of announcements to increase government spending and taxes for a number of periods, where spending starts today but taxes are raised with some delay. To be precise, taxes are levied according to $t_{t+d} = R^d g_t$, where d is the delay and $R = \beta^{-1}$ is the steady-state gross real interest rate. Panel B in Figure 2.2 shows the impact multiplier of current output as a function of the length of the stimulus, where lighter lines correspond to larger time delays in raising taxes. First, notice that the budget-balanced response is exactly the same as the 6-quarters line in Panel A. As we argued above, RANK and TANK multipliers are identical when the budget is balanced. However, as taxes are raised further away into the future, multipliers become more negative. This highlights the interaction of deficits and the inflation-spending feedback loop. In fact, running a deficit-financed expansion for just 6 quarters and immediately raising taxes for

[31] A similar argument determines that frontloading taxes and backloading spending leads to larger booms in TANK versus RANK. We focus on the deficit case because of the central role it plays in policy discussions to stimulate the economy.

[32] Two potential targets to calibrate the share of HtM agents in the model are the actual share of HtM agents in the data, or the average marginal propensity to consume (MPC). Kaplan, Violante, and Weidner (2014) estimate that between 25 and 40 percent of U.S. households are HtM, while an extensive literature measuring MPCs out of tax rebates conclude that households spend roughly 15-25 percent of the rebate in the same quarter (see, e.g., Johnson, Parker, and Souleles 2006; Parker et al. 2013; Misra and Surico 2014; Broda and Parker 2014).

the next 6 quarters afterwards yields a multiplier of -5. It is worth emphasizing that the degree price of stickiness and the share of HtM households assumed in this experiment are relatively conservative compared to empirical estimates, so the negative multipliers would be even worse under more plausible calibrations.

So far we have only considered the neutral case, where redistributional effects of fiscal policy are shut down. Away from this case, income cyclicality and the amount of taxes paid by constrained households could potentially dampen or magnify the effects of government spending. Income cyclicality is determined in TANK by the share of profits distributed to each type of household. The basic New Keynesian framework features sticky prices and flexible wages, leading to countercyclical markups and procyclical real wages. This makes profits relative to aggregate income countercyclical. Consequently, whoever receives relatively less profits will have a larger share of its earnings coming from wages, and hence relatively more procyclical income. In general, demand shocks will be amplified when income from HtM households is relatively more procyclical, as in the earnings heterogeneity channel described in Auclert (2019). The reason is that booms (recessions) redistribute income towards (away from) the HtM in relative terms. Constrained households have larger MPCs, and hence magnify the initial shock. This is the reason why the aggregate EIS in TANK, $\tilde{\sigma}$, is generally different than the EIS in RANK, $\bar{\sigma}$, depending on the distribution of profits. Similarly, collecting lower taxes from HtM households (e.g. progressive taxes) increases the multipliers by redistributing income away from Ricardian low-MPC households towards HtM.

The effects of redistribution can be massive as both the Keynesian cross and the inflation-spending feedback loop are affected, and they interact strongly with each other. Take for example countercyclical inequality and progressive taxes, in line with the empirical literature. On the one hand, for a given EIS, a stronger Keynesian cross channel magnifies the effect of government spending on consumption and leads to larger inflation and real interest rate responses. On the other hand, for a given change in future inflation, the initial response by Ricardian agents also gets amplified. This suggests that the "neutral" multipliers we showed in Figure 2.2 are arguably a lower bound of the true multipliers when distributional concerns are realistically addressed.

In conclusion, HtM agents introduce additional effects of fiscal policy through the breakdown of Ricardian equivalence, the Keynesian cross channel, and their interactions with the inflation-spending feedback loop. Deficits tend to generate negative multipliers that grow unboundedly as taxes are further backloaded into the future. Moreover, despite redistribution being important, it can only affect the *levels* and/or *signs* of the multipliers and cannot prevent the explosive forward-looking behavior of the feedback loop. If anything, realistic assumptions such as countercyclical income risk and progressive taxes are likely to exacerbate the amplification of the feedback loop, the Keynesian cross, and the magnitude of the multipliers.

2.2.3 Fiscal Multipliers With Incomplete Markets

In this section, we study the rational expectations fiscal multipliers in a standard Heterogeneous Agent New Keynesian (HANK) model. The model deviates from RANK by adding idiosyncratic income risk, incomplete markets and borrowing constraints à la Bewley-Hugget-Aiyagari. This leads to an endogenous joint distribution of wealth and marginal propensities to consume, where low-wealth households endogenously exhibit hand-to-mouth behavior.

The insights built from RANK and TANK can be helpful to analyze the underlying mechanisms operating in HANK. A priori, there are reasons to suspect that HANK and TANK multipliers could be related, as these two models have some important elements in common. A key similarity is that both incorporate HtM households, hence breaking Ricardian equivalence. Also, redistributional consequences of fiscal policy matter in determining the response of the economy, in particular the incidence of taxes and the correlation between MPCs and the elasticity of individual earnings to aggregate income.

Despite all similarities, a key difference sets both models apart. Specifically, Ricardian households in TANK are fundamentally different from non-HtM agents in HANK. The former have infinite horizon and are always on their optimal spending path. However, households in HANK face some probability of hitting their borrowing constraint at some point in the future, thus becoming HtM temporarily. That limits their ability to smooth consumption by borrowing against their future income and hinders their forward-looking behavior. In principle, this could imply a dampening of the inflation-spending feedback loop, as it crucially depends on the extreme forward-looking behavior of Ricardian households and their unlimited ability to smooth consumption.

We run the same deficit-financed experiments as in Section 2.2.2 to see how fiscal multipliers in HANK compare to those in RANK and TANK. It is well known that HANK models cannot be solved analytically, so we rely on simulations. The model is calibrated to be comparable to the TANK model from previous section, with an average duration of prices of 6 quarters, a 19 percent share of HtM households and and average MPC of 22 percent. The details of the model and the calibration are described in Appendix B. Again, we first first focus on a "neutral" case that shuts down redistributional effects. Neutrality in HANK can be achieved by using the same assumptions we discussed in Section 2.2.2, together with a unitary EIS and endogenous asset revaluation proportional to aggregate after-tax income.[33]

[33] Assets revalution ensures that the whole budget constraint scales with aggregate after-tax income, which leads to neutrality with respect to real interest rate changes (Werning, 2015). This can be obtained with

Panel C in Figure 2.2 displays the response of current output depending on the length of the stimulus. Lighter lines represent longer delays in raising taxes. Surprisingly, HANK multipliers under neutrality are the same as in RANK and TANK when the budget is balanced. Budget-balanced spending leaves the budget constraints unchanged, while the additional assumptions from Werning (2015) guarantee a neutral response to changes in the real rate driven by inflation. As a whole, the multipliers in HANK bear a stark resemblance to those in TANK. They are unbounded and easily become negative as taxes are backloaded into the future. Waiting two years to raise taxes is enough to generate negative multipliers as low as -6 when the stimulus lasts for three years. These results emphasize how the lessons from RANK and TANK can help us disentangle the effects of fiscal policy in more complex models, like HANK.

While HANK seems to slightly weaken the inflation-spending feedback loop when spending is deficit-financed, the multipliers still preserve all the extreme properties we highlighted in Section 2.2.2. This tiny attenuation in the negative effects of deficits could be caused by a liquidity channel going in the opposite direction. Higher debt increases liquidity and facilitates consumption smoothing. This reduces the precautionary savings motive and stimulates aggregate demand. We will talk further about these 'debt' multipliers in Section 2.4.2, when we explore the interaction of level-k thinking with incomplete markets.

Redistributional effects also play a similar role as in TANK. Government spending is more effective when high-MPC low-wealth individuals have more procyclical income or pay fewer taxes compared to the neutral case. Similarly, this countercyclical inequality also amplifies the effects of the inflation-spending feedback loop, as stressed in Werning (2015) and Auclert (2019). As discussed above, empirical evidence supports countercyclical income risk and progressive taxes, so the benchmark multipliers in Figure 2.2 are likely to be a lower bound of the true multipliers generated by this model under more realistic assumptions.

In conclusion, the rough behavior of rational expectations fiscal multipliers in HANK is well approximated by a simpler TANK model, with multipliers turning negative when government spending is deficit financed. Even though the amplification of backloading through the inflation-spending feedback loop is slightly reduced versus TANK, the multipliers are still massively negative for even reasonable horizons. Furthermore, realistic assumptions about distributional effects would make this problem even worse. Given the tight relationship between the inflation-feedback loop and rational expectations, in the next section we present a NK framework that allows for arbitrary belief formation.

real assets whose dividend scales with aggregate after-tax income, or even with nominal bonds if combined with the appropriate tax policy. Further details can be found in Appendix B.

2.3 Belief-Augmented New Keynesian Framework

Last section showed the exaggerate sensitivity of the rational expectations New Keynesian model to fiscal policy shocks when monetary policy is constrained and unresponsive. In this section, we relax the assumption of rational expectations and allow for arbitrary belief formation. We describe our main equilibrium concept, the Temporary Equilibrium, and use it to show that the unreasonable results we reported in the previous section crucially depend on the extreme level of coordination achieved under rational expectations.

2.3.1 Basic Framework

We begin by introducing a baseline New Keynesian model with arbitrary beliefs about the future. There is no aggregate uncertainty. We adopt continuous time because it can be analyzed parsimoniously, but no conclusion relies specifically on this assumption. As in the previous section, the economy is populated by hand-to-mouth and Ricardian households. All variables thereafter are in real terms unless explicitly noted.

Hand-to-mouth households. The economy is populated by a share χ of constrained households. They do not have access to financial markets to save or borrow, so they consume all their income. They receive income from their labor services,$W_t N_{ht}$, and a share of uncapitalized profits from the private sector, Φ_{ht}, which are taken as given.[34] They pay lump-sum taxes T_{ht} to the government, which are also taken as given. Even though the HtM cannot smooth consumption intertemporally, they still optimize their labor supply decision. In sum, their behavior is characterized by

$$C_{ht} = W_t N_{ht} + \Phi_{ht} - T_{ht}, \tag{2.10}$$

$$W_t = N_{ht}^{\gamma} C_{ht}^{1/\sigma}, \tag{2.11}$$

Ricardian households. The remaining $1-\chi$ share of households are Ricardian. At every period t, these households choose paths for real consumption $\{C^e_{t+s}\}$ and labor supply $\{N^e_{t+s}\}$ that maximize their *expected* lifetime utility

$$\int_0^{\infty} e^{-\rho s} \left[\frac{\left(C^e_{rt+s}\right)^{1-1/\sigma}}{1-1/\sigma} - \frac{\left(N^e_{rt+s}\right)^{1+\gamma}}{1+\gamma} \right] ds,$$

[34]These profits can be thought of as a mix of rent sharing by firms with their workers and transfers by the government.

subject to their budget constraints, where ρ denotes the discount factor, σ denotes the denotes elasticity of intertemporal substitution and γ denotes the inverse Frisch elasticity. They receive income from their labor services, $W_t N_{rt}$, and a share of uncapitalized profits from the private sector, Φ_{rt}, which are taken as given. They pay lump-sum taxes T_{rt} to the government, which are also taken as given. In contrast to HtM households, Ricardian agents can trade real and nominal risk-free assets. Specifically, there are Lucas trees X_{rt} with price Q_t that pay dividend D_t, and nominal government bonds $P_t B_{rt}$ which pay interest i_t. In sum, agents believe assets evolve according to the *expected* budget constraints

$$Q^e_{t+s}\frac{d}{ds}X^e_{rt+s}+\frac{d}{ds}B^e_{rt+s}=(i^e_{t+s}-\pi^e_{t+s})B^e_{rt+s}+D^e_{t+s}X^e_{rt+s}+W^e_{t+s}N^e_{rt+s}+\Phi^e_{rt+s}-C^e_{it+s}-T^e_{rt+s},$$

for $s \geq 0$ with the initial conditions $B^e_{rt} = B_{rt}$ and $X^e_{rt} = X_{rt}$, where the e superscript denotes beliefs about future variables. However, in reality assets evolve following the *actual* budget constraints

$$Q_t\frac{d}{dt}X_{rt}+\frac{d}{dt}B_{rt}=(i_t-\pi_t)B_{rt}+D_tX_{rt}+W_tN_{rt}+\Phi_{rt}-C_{rt}-T_{rt},$$

where $C_t = C^e_t$ and $N_t = N^e_t$at time t for $s = 0$.[35] This distinction between *expected* and *actual* budget constraints is crucial. Ricardian households optimize according to their beliefs about the future, but *ex post* savings will generally differ from *ex ante* optimal savings when their beliefs turn out to be wrong.

The first order conditions that characterize the optimal decisions by Ricardian households are given by

$$\frac{d}{ds}\frac{C^e_{rt+s}}{C^e_{rt+s}}=\sigma\left(i^e_{t+s}-\pi^e_{t+s}-\rho\right), \tag{2.12}$$

$$W^e_{t+s}=\left(N^e_{rt+s}\right)^{\gamma}\left(C^e_{rt+s}\right)^{1/\sigma}, \tag{2.13}$$

the transversality condition

$$\lim_{s\to\infty} e^{-\int_0^s i^e_{t+z}-\pi^e_{t+z}dz}\left(Q_sX_{rs}+B_{rs}\right)=0,$$

and the no-arbitrage condition

$$\frac{\frac{d}{ds}Q^e_{t+s}+D^e_{t+s}}{Q^e_{t+s}}=i^e_{t+s}-\pi^e_{t+s}.$$

[35]Formally, the notation should be $C^e(t,s)$ and $N^e(t,s)$, denoting the optimal plan for consumption and labor supply held at time t for period $t+s$. Thus, $C_t = C^e(t,0)$ and $N_t = N^e(t,0)$. We save on notation by adopting the shortcut $C^e_{t+s} = C^e(t,s)$ and $N^e_{t+s} = N^e(t,s)$.

This last condition arises because the Lucas trees and the government bonds both have the same *ex ante* return.[36] Then, the price of the Lucas Tree must satisfy

$$Q_t = \int_0^\infty e^{-\int_0^s i^e_{t+z} - \pi^e_{t+z} dz} D^e_{t+s} ds, \tag{2.14}$$

and the budget constraints can be expressed in terms of total financial wealth, $A_{rt} = Q_t X_{rt} + B_{rt}$, so

$$\frac{d}{ds} A^e_{rt+s} = (i^e_{t+s} - \pi^e_{t+s}) A^e_{rt+s} + W^e_{t+s} N^e_{rt+s} + \Phi^e_{rt+s} - C^e_{rt+s} - T^e_{rt+s}, \tag{2.15a}$$

$$\frac{d}{dt} A_{rt} = \left[\left(i^e_t - \pi^e_t + \frac{D_t - D^e_t}{Q_t}\right)\lambda_t + (i_t - \pi_t)(1 - \lambda_t)\right] A_{rt} + W_t N_{rt} + \Phi_{rt} - C_{rt} - T_{rt}, \tag{2.15b}$$

with initial condition $A^e_{rt} = A_{rt}$, where λ_t is the weight of Lucas trees on total wealth $Q_t X_{rt} / A_{rt}$.[37] Most importantly, the actual *ex post* return on total wealth depends on the realized interest rate and inflation, which affects nominal bonds; the realized dividends, which affects the Lucas trees; and the portfolio composition between these two assets. Similarly, the transversality condition becomes

$$\lim_{s \to \infty} e^{-\int_0^s i^e_{t+z} - \pi^e_{t+z} dz} A_{rs} = 0. \tag{2.16}$$

Final good producers. The final good is produced by a perfectly competitive sector using intermediate good varieties $Y_t(j)$ with the CES production function

$$Y_t = \left(\int_0^1 Y_t(j)\, dj\right)^{\frac{\theta}{\theta-1}}.$$

The demand for variety j from the final good sector is given by

$$Y_t(j) = \left(\frac{P_t(j)}{P_t}\right)^{-\theta} Y_t,$$

where $P_t(j)$ is the price of the variety and P_t is the price for the final good

$$P_t = \left(\int_0^1 P_t(j)^{1-\theta} dj\right)^{\frac{1}{1-\theta}}. \tag{2.17}$$

[36] In general, the return of the two assets will differ *ex post* depending on the actual realization of the dividends, the interest rate, and inflation.

[37] Actual portfolio composition for households is undetermined when both assets have the same *ex ante* return, so it will be determined in equilibrium by relative asset supplies.

Intermediate good producers. Varieties of the intermediate good are produced by monopolistically competitive firms using a Cobb-Douglas technology

$$Y_t(j) = N_t(j)^{1-\delta}.$$

Prices are set *à la* Calvo, that is, firms can only change prices when they are hit by a Poisson shock with rate ρ_c. Those firms that are hit by the shock choose their reset price P_t^* to maximize the present discounted value of *expected* lifetime profits

$$\max_{P_t^*} \int_0^\infty e^{-\rho_c s - \int_0^s i_{t+z}^e dz} \left[P_t^* Y_{t+s|t}^e(j) - P_{t+s}^e C_s \left(Y_{t+s|t}^e(j) \right) \right] ds,$$

taking beliefs W_t^e, Y_t^e and P_t^e as given, where $Y_{t+s|t}^e(j) = \left(P_t^* / P_{t+s}^e \right)^{-\theta} Y_{t+s}^e$ is the demand for their variety j, and $C_s\left(Y_s(j)\right) = W_s Y_s(j)^{1/(1-\delta)}$ is the real cost function for intermediates at time s. The real reset price is

$$\frac{P_t^*}{P_t} = \left[\frac{\theta}{\theta - 1} \frac{\int_0^\infty e^{-\rho_c s - \int_0^s i_{t+z}^e - (\theta + \Theta^{-1}) \pi_{t+z}^e dz} Y_{t+s}^e MC_{t+s}^e}{\int_0^\infty e^{-\rho_c s - \int_0^s i_{t+z}^e - \theta \pi_{t+z}^e dz} Y_{t+s}^e ds} \right]^\Theta, \tag{2.18}$$

where $\Theta = \frac{1-\delta}{1-\delta+\theta\delta}$, and $MC_{t+s}^e = (1-\delta)^{-1} W_{t+s}^e \left(Y_{t+s}^e \right)^{\delta/(1-\delta)}$ is proportional to the *expected* average real marginal cost at time $t+s$. Then, equation (2.17) implies that inflation of the final good is

$$\pi_t = \frac{\rho_c}{1-\theta} \left[\left(\frac{P_t^*}{P_t} \right)^{1-\theta} - 1 \right]. \tag{2.19}$$

Lucas trees and profits. There is a unit net supply of Lucas trees. These trees capitalize a share of the profits of the intermediate good sector[38] and their dividend is

$$D_t = \delta Y_t. \tag{2.20}$$

The remaining uncapitalized profits

$$\Phi_t = Y_t - D_t - W_t N_t, \tag{2.21}$$

[38] The profits capitalized by the Lucas trees could be interpreted as the rents of a fixed factor used in the intermediate sector. Then, the Lucas trees would represent the shares of a competitive mutual fund that owns the entire supply of the fixed factor in the economy.

are lump-sum transferred to both types of households. The split of profits depends on a parameter o so

$$\Phi_{ht} = (1-o)\left(Y_t - W_t N_t\right), \tag{2.22}$$

where $o \in \left[-\frac{1-\chi}{\chi}, 1\right]$. Note that total income of HtM households is proportional to aggregate output when $o = 0$, corresponding to the neutral case mentioned in Section 2.2. In fact, this parameter captures the income cyclicality of the HtM. The value of o determines the amount of profits accruing to the HtM, affecting the share of their total earnings coming from labor income. Thus, HtM earnings will be relatively more procyclical or countercyclical when o is positive or negative, respectively.[39]

Government. The government finances its expenditures by raising (real) taxes T_t and issuing nominal government bonds $P_t B_t$

$$T_t = \chi T_{ht} + (1-\chi) T_{rt}, \tag{2.23}$$
$$T_t + \frac{d}{dt} B_t = G_t + (i_t - \pi_t) B_t,$$

where G_t is real government spending and $(i_t - \pi_t)B_t$ is the real cost of serving the debt. Fiscal policy announcements $\{G_{t+s}, T_{t+s}\}$ must satisfy the intertemporal government budget constraint

$$B_t = \int_t^\infty e^{-\int_t^s i_z - \pi_z dz}\left(T_s - G_s\right) ds, \tag{2.24}$$

so the government must raise enough taxes over time to finance its spending path and the current level of public debt.

Active vs. passive fiscal policy. The announcement consists of a path for government spending and taxes $\{G_t, T_{ht}, T_{rt}\}$ that satisfies the intertemporal government budget constraint.[40] The literature usually considers two ways of approaching fiscal policy: active or passive (see Leeper, 1991; Sims, 1994). In a nutshell, an active fiscal policy announces a path for fiscal policy that might not satisfy equation (2.24), but forces the price level to adjust to restore back the equality through its impact on the real stock of debt B_0—this is known as the Fiscal Theory of the Price Level. On the other hand, the arguably more standard passive fiscal policy takes B_0 as given, so the announced paths are restricted to satisfy (2.24)

[39] As we argued in Section 2.2, labor income is procyclical and profits tend to be countercyclical in typical New Keynesian models with sticky prices and flexible wages.

[40] Note that equation (2.24) depends on *actual* inflation and interest rates, so fiscal policy might not satisfy the budget constraint according to agents beliefs.

in equilibrium. For our purposes, an active fiscal policy could be seen as a particular case of a passive policy where the government adjust taxes at $t = 0$ proportionally to debt holdings by every agent, guaranteeing that equation (2.24) is satisfied in equilibrium. Therefore, the particular fiscal policy approach is not crucial for most of our paper, and we will not take a stance unless explicitly noted.

Monetary policy. Monetary policy is constrained, so interest rates do not react to changes in the economy. Specifically, we assume that $i_t = \rho$ for all $t < T$, and follows a standard Taylor rule satisfying the Taylor principle for all $t > T$. This guarantees that inflation will be pinned down and solves undeterminacy problems under an interest rate peg.

Aggregation and market clearing. Aggregate variables are obtained by aggregating individual quantities over all households

$$C_t = \chi C_{ht} + (1 - \chi)C_{rt}, \tag{2.25a}$$

$$N_t = \chi N_{ht} + (1 - \chi)N_{rt}, \tag{2.25b}$$

$$A_t = (1 - \chi)A_{rt}, \tag{2.25c}$$

$$\Phi_t = \chi\Phi_{ht} + (1 - \chi)\Phi_{rt} \tag{2.25d}$$

and the goods, labor and assets market clear whenever

$$Y_t = C_t + G_t, \tag{2.26a}$$

$$N_t = Y_t^{\frac{1}{1-\delta}}\Delta_t, \tag{2.26b}$$

$$Q_t = \lambda_t A_t, \tag{2.26c}$$

$$B_t = (1 - \lambda_t)A_t, \tag{2.26d}$$

where $\Delta_t = \int_0^1 (P_t(j)/P_t)^{-\theta/(1-\delta)} dj$ is an index of price dispersion.

Belief formation. We assume beliefs are static (i.e., do not get updated over time) for all variables that agents take as given in their optimization problems. This implicitly rules out learning or any kind of revelation of information. There is only one sequence of beliefs, common across all agents. Beliefs can only possibly adjust when a new fiscal policy is announced, reflecting how agents belief other agents will react to the announcement. This does not apply to variables that agents optimize over, though. At every period t, agents form optimal plans for their current and future decisions. The beliefs for these future decision variables could potentially be adjusted over time if their beliefs turned out to be wrong in the past. Intuitively, they might need to reoptimize their plan after unforeseen circumstances arise.

Temporary equilibrium. We want to deviate from rational expectations by allowing agents to have more general beliefs about the future. We build on previous work from Grandmont (1977, 1988) and use the concept of temporary equilibrium. Intuitively, a temporary equilibrium describes the outcome of an economy where agents optimize with respect to some arbitrary beliefs about the future and all markets clear at all times.

Definition 1 (Temporary equilibrium)**.** *A temporary equilibrium (TE) of this economy, where $t \in \mathbb{R}^+$, consists of paths for aggregates*

$$\{C_t, N_t, Y_t, W_t, Q_t, A_t, \lambda_t, i_t, \pi_t, T_t, B_t\},$$

and household quantities

$$\{C_{rt}, C_{ht}, N_{rt}, N_{ht}, A_{rt}\},$$

such that, given a fiscal policy path announced at $t = 0$

$$\{G_t, T_{rt}, T_{ht}\},$$

and arbitrary paths of beliefs

$$\{W^e_t, \Phi^e_{rt}, i^e_t, \pi^e_t, D^e_t, T^e_{rt}, Y^e_t\},$$

households decisions are optimal and satisfy the Euler equation (2.12), the labor supplies (2.13) and (2.11), the transversality condition (2.16) and the budget constraints (2.15) and (2.10) together with (2.20) and (2.22), intermediate and final good producers set prices optimally according to (2.18) and (2.19), aggregate profits are given by (2.21), fiscal policy satisfies the government budget constraint (2.24), aggregation conditions (2.23) and (2.25) hold, and market clearing conditions (2.14) and (2.26) are satisfied.

The TE defines an operator $\mathcal{T}$ mapping paths of beliefs to paths of actual outcomes. Using this definition, we can easily see that a rational expectations equilibrium is a particular case of a TE. In particular, a rational expectations equilibrium is just a fixed point of the mapping $\mathcal{T}$.

Definition 2 (Rational expectations equilibrium)**.** *A rational expectations equilibrium (REE) of this economy is a temporary equilibrium where beliefs satisfy perfect foresight, that is, be-*

liefs about future variables coincide with their actual outcomes

$$\{W^e_t, \Phi^e_{rt}, i^e_t, \pi^e_t, D^e_t, T^e_{rt}, Y^e_t\} = \{W_t, \Phi_{rt}, i_t, \pi_t, D_t, T_{rt}, Y_t\},$$

for all t.

2.3.2 Temporary Equilibrium Euler Equation and Philips Curve

In this section we obtain a parsimonious two-equation log-linear representation of the temporary equilibria in the previous model. Interestingly, deviating from rational expectations leads to discounted versions of the standard Euler equation and Phillips Curve, as in Gabaix (2017) and McKay et al. (2016). We show that fiscal multipliers depend crucially on the reaction of expectations to the change in policy.

First we log-linearize the equilibrium conditions around the symmetric zero-inflation steady state with constant government spending and debt.

Demand side. The aggregate behavior of the demand side of the economy can be summarized by the aggregate Euler equation. Starting from the aggregation condition (2.25a), together with the equilibrium conditions (2.10), (2.11), (2.20), and (2.21) we obtain the aggregate consumption function

$$c_t = \Omega_c c_{rt} + \Omega_g g_t - \Omega_\tau t_{ht}.$$

The coefficients Ω_g, Ω_τ and Ω_c can be found in the Appendix A and measure the impact of these variables on aggregate demand through the usual Keynesian cross channel, explained in Section 2.2. The aggregate Euler equation follows immediately

$$\dot{c}_t = \Omega_c \dot{c}_{rt} + \Omega_g \dot{g}_t - \Omega_\tau \dot{t}_{ht}, \tag{2.27}$$

where we use the standard notation $\dot{c}_{rt}$ to denote the derivative with respect to time $\frac{d}{dt} c_{rt}$.

It might be tempting to conclude that the term $\dot{c}_{rt}$ corresponds to the standard Euler equation (2.12). However, equation (2.12) only holds *ex ante* under the path described by the expected budget constraints (2.15a). In other words, it only reflects the path of expected consumption if everything goes as planned.[41] In reality, the actual *ex post* Euler equation

[41] We use a slight abuse of notation to keep it simple. You should think about consumption by Ricardian households as a function of both t and s. Then, $c_r(t, s)$ denotes beliefs held at time t of consumption at time $t + s$. This implies that $\frac{d}{dt} c_r(t, s)$ and $\frac{d}{ds} c_r(t, s)$ will be different in general. Realized consumption $c_{rt} = c^e_{rt}$ is given by $c_r(t, 0)$. The term $\dot{c}_{rt}$ in the text is equivalent to $\frac{d}{dt} c_r(t, 0)$, while $\frac{d}{ds} c^e_{rt+s}$ represents $\frac{d}{ds} c_r(t, s)$.

will be generally different from (2.12) if households beliefs are incorrect. Ultimately, the consumption plan continuously readjusts when facts differ from beliefs. We can see this by considering the consumption function by Ricardian households

$$c_{rt} = \bar{\rho}\left(a_{rt} + \int_t^{\infty} e^{-\rho(s-t)}\left[\left(\frac{\mathcal{A}}{1-\chi} - \frac{\bar{\sigma}}{\bar{\rho}}\right)(i_s^e - \pi_s^e) + \mathcal{L}(1+\gamma^{-1})w_s^e + \phi_{rs}^e - t_{rs}^e\right] ds\right), \tag{2.28}$$

which comes from the expected budget constraints (2.15a) and the equilibrium conditions (2.12), (2.16), and (2.13). Here, $\bar{\sigma} = \sigma C/Y$, $\bar{\rho} = \rho/\left[1 + \mathcal{L}(\bar{\sigma}\gamma)^{-1}\right]$ is the marginal propensity to consume (MPC) corrected for the wealth effect on labor supply, $\mathcal{A} = A/Y$ is the steady-state wealth-to-GDP ratio, and $\mathcal{L} = WN/Y$ is the steady-state labor share. Ricardians decide their consumption purely based on expectations about their income path. Actual income plays no role in determining consumption because of the continuous time assumption. Therefore, incorrect beliefs fully manifest in the actual evolution of savings, which differs from the agent *ex ante* plan. Taking the derivative of the consumption function together with the actual budget constraint (2.15b) gives the actual *ex post* Euler equation

$$\dot{c}_{rt} = \bar{\sigma}(i_s^e - \pi_s^e) + \bar{\rho}\hat{\mathcal{E}}_{rt}, \tag{2.29}$$

where

$$\hat{\mathcal{E}}_{rt} = \mathcal{B}_r(\hat{i}_t - \hat{\pi}_t) + (1-\chi)^{-1}\hat{d}_t + \mathcal{L}(1+\gamma^{-1})\hat{w}_t + \hat{\phi}_{rt} - \hat{t}_{rt}.$$

Here, $\mathcal{B}_r = \mathcal{B}/(1-\chi)$, $\mathcal{B} = B/Y$ is the government debt-to-GDP ratio in the steady state, and hat variables $\hat{x}_t = x_t - x_t^e$ denote surprises, that is, deviations of actual realizations from their expected value. Thus, the variable $\hat{\mathcal{E}}_{rt}$ represents surprises in earnings of Ricardian households, coming from the *ex post* real returns on nominal bonds and the Lucas trees, wages, profits or taxes. This *ex post* Euler equation can be very easily interpreted. Suppose there is a positive surprise and profits turn out to be higher than expected. This leads to higher savings than previously expected, which in turn induces a raise in consumption modulated by the corrected MPC $\bar{\rho}$. Unsurprisingly, $\dot{c}_{rt}$ coincides with the standard Euler equation (2.12) when there are no surprises and beliefs coincide with reality.

The aggregate Euler equation can be obtained by plugging (2.29) into (2.27) together with the equilibrium conditions to get

$$\dot{c}_t = \tilde{\sigma}(i_t^e - \pi_t^e) + \Omega_g \dot{g}_t - \Omega_\tau \dot{t}_{ht} + \bar{\rho}\Omega_c\hat{\mathcal{E}}_{rt}, \tag{2.30}$$

where $\tilde{\sigma} = \Omega_c\bar{\sigma}$ is the aggregate EIS. Some key differences arise from comparing (2.31) to its

REE counterpart, characterized by perfect foresight so $\hat{\mathcal{E}}_{rt} = 0$, $i_t^e = i_t$ and $\pi_t^e = \pi_t$.[42] First, the real rate driving consumption growth due to intertemporal substitution is the expected rate $i_t^e - \pi_t^e$, in contrast to the actual one $i_t - \pi_t$. This rate is precisely the *ex ante* return on savings that consumers perceive when they optimize. Second, the additional term $\bar{\rho}\Omega_c\hat{\mathcal{E}}_{rt}$ captures the aggregate effect of earnings surprises for Ricardian households. Finally, we can use the equilibrium conditions to express wages, dividends and profits in $\mathcal{E}_{rt}$ in terms of aggregate quantities

$$\mathcal{E}_{rt} = \mathcal{B}_r(i_t - \pi_t) + \rho(\Omega_c\bar{\rho})^{-1}c_t + \Omega_{rg}g_t - t_{rt},$$

and plug it into (2.30) to get

$$\dot{c}_t = \rho c_t + \Omega_g\dot{g}_t - \Omega_\tau\dot{t}_{ht} + \bar{\rho}\Omega_c\left[\mathcal{B}_r(i_t - \pi_t) + \Omega_{rg}g_t - t_{rt}\right] - \mathcal{D}_t^e, \tag{2.31}$$

where $\mathcal{D}_t^e = -\tilde{\sigma}(i_t^e - \pi_t^e) + \bar{\rho}\Omega_c\mathcal{E}_{rt}^e$ is a composite variable representing the relevant beliefs for the demand side, and Ω_{rg} is a parameter governing the effect of government spending on the income of Ricardian households. Note that the term ρc_t implies that the Euler equation in a temporary equilibrium is discounted at rate ρ, that is, the discount factor of Ricardian households.

Supply side. The aggregate behavior of the supply side is encapsulated in the Phillips Curve. Starting from the optimal pricing rule (2.18) and the definition of inflation (2.19) we get

$$\pi_t = \rho_c\lambda\int_t^\infty e^{-(\rho+\rho_c)(s-t)}\left(mc_s^e + \Theta^{-1}\int_t^s \pi_z^e dz\right)ds,$$

where $\lambda = \Theta(\rho + \rho_c)$ and $mc_s^e = w_s^e + \frac{\delta}{1-\delta}y_s^e$. Inflation depends on the optimal reset price, which in turn depends positively on the expected paths for real marginal costs and inflation. Taking the derivative with respect to time leads to the TE Phillips Curve

$$\dot{\pi}_t = (\rho + \rho_c)\pi_t - \rho_c\mathcal{S}_t^e, \tag{2.32}$$

where $\mathcal{S}_t^e = \pi_t^e + \lambda mc_s^e$ is a composite variable representing the relevant beliefs for the supply side. In contrast to the REE Phillips Curve, the TE version is discounted at a higher rate $\rho + \rho_c > \rho$. This is the relevant discount factor that firms use to discount future marginal costs and inflation in their pricing decisions.

[42] This is just the continuous time analog of equation (2.6) in Section 2.2.

2.3.3 Fiscal Multipliers in Temporary Equilibria

The new TE Euler equation (2.31) and Phillips Curve (2.32) explicitly reveal the role played by expectations in the economy. This is hard to think about in the standard rational expectations framework, given that beliefs and actual realizations are exactly the same. This will allow us to break down the fiscal multipliers into two components, namely, the naive and the belief multipliers. Finally, we explore the strategic interactions uncovered by this formulation and show how the inflation-spending feedback loop is the main driver of the explosive behavior of the rational expectations multipliers.

The aggregate response of output to fiscal policy announcements can be obtained by solving the system formed by the Euler equation (2.31) and the Phillips Curve (2.32). Assuming the economy goes back to its steady state in the long run, we derive the following proposition.

Proposition 1 (Temporary equilibrium fiscal multipliers)**.** *The TE response of output to a fiscal policy announcement at* $t = 0$ *is given by*

$$y_t = \underbrace{(1+\Omega_g)g_t - \Omega_\tau t_{ht} + \bar{\rho}\Omega_c(1-\chi)^{-1}b_t}_{\text{Naive Multipliers}} + \underbrace{\int_t^\infty e^{-\rho(s-t)}\mathcal{D}^e_s ds}_{\text{Belief Multipliers}}. \tag{2.33}$$

Proposition 1 highlights that for any temporary equilibrium, including rational expectations, we can decompose the multipliers into two components that operate independently. Specifically, fiscal policy affects the economy directly through its effects on aggregate demand and household disposable income—we call this the "naive multipliers". Besides that, it also has an indirect effect through a shift in expectations and the resulting reaction by decision-makers today—the belief multipliers.

This result holds regardless of the monetary policy set by the central bank and the actual path for inflation, as they do not show up in equation (2.33). Nevertheless, this does not imply that monetary policy or inflation are innocuous. Far from that, central bank decisions affect interest rates, which together with actual inflation determine the *ex post* real rate of government bonds. Thus, monetary policy affects the budget constraint of the government, shaping the evolution of debt and taxes. Depending on how beliefs are formed, it could also affect how the beliefs multipliers respond to fiscal policy, as it does with rational expectations. In short, this proposition just clarifies that the effects of monetary policy and actual inflation must be transmitted through the elements in equation (2.33).

Naive multipliers. These multipliers summarize the response of the economy under *naive* expectations, that is, when expectations remain unaltered after the shock. In other words, they exclusively capture the response of output induced by the contemporaneous impact of fiscal policy on the allocation of resources. In particular, government spending and taxes affect HtM disposable income and their level of consumption. This translates into aggregate demand through the Keynesian cross channel discussed in Section 2.2.2. Similarly, the impact on earnings of Ricardian households is reflected on their savings, which must parallel the evolution of debt for the bonds market to clear. Current wealth is one of the key factors determining consumption by Ricardian households, implying that the stock of government debt affects output directly. This is a direct consequence of Ricardian equivalence failing under *naive* expectations, as agents do not revise their relevant beliefs about income, inflation or marginal costs. In other words, they fail to see that debt is not net wealth, as it comes with higher taxes in the future.[43] In the end, they increase consumption proportionally to their MPC out of wealth $\bar{\rho}$, leading to an additional Keynesian multiplier effect $\Omega_c(1-\chi)^{-1}$ adjusted for redistributional effects.

Belief multipliers. The remaining impact of fiscal policy arises from a shift in expectations. In essence, agents must form an opinion about how the fiscal policy announcement will affect their planning problem. The gold standard in macroeconomics is to assume rational expectations, where agents perfectly observe the shocks, fully understand the economic setting and have common knowledge about all this. In that case, $\mathcal{D}_s^e = \mathcal{D}_s = \tilde{\sigma}(i_t - \pi_t) + \bar{\rho}\Omega_c\mathcal{E}_{rt}$. But many alternatives have been proposed in the literature for this belief formation process, including learning, rational inattention, sparsity, level-k thinking, and uncertainty in high-order beliefs.[44]

The belief multipliers are the ones driving the explosive behavior of output in the REE. But a relevant feature of Proposition 1 is that beliefs are explicitly discounted. Therefore, models that are known to discount the future more heavily, like HANK, could potentially give less weight to this channel of fiscal policy. This intuition does not hold for rational expectations because the actual evolution of the economy together with the belief formation process make it look like as if there was no discounting in the Euler equation. However, we will show that incomplete markets certainly dampen the beliefs multipliers away from the knife-edge case of rational expectations. Before going there though, we first try to understand what drives the explosive multipliers in the REE by breaking it down in two different assumptions about the extent of common knowledge between agents in the economy.

[43]One could argue that this is not a relevant benchmark since households should at least understand that they will pay higher taxes, as this variable enters directly in their decision problem. However, the concept of TE does not require beliefs to be rational or satisfy any consistency whatsoever. We believe this benchmark is indeed useful to clearly isolate the role of expectations in shaping the response to fiscal policy.

[44]Note that in our framework, belief updates happen only once at $t = 0$, so dynamic updating processes such as learning cannot be represented by Proposition 1.

This allow us to isolate a number of strategic interactions arising in the REE and pin down the main culprit behind the fiscal multiplier puzzle, the inflation-spending feedback loop.

2.3.4 Partial Common Knowledge and Strategic Interactions in the REE

Rational expectations imposes a huge degree of coordination across agents in the economy. But what exactly drives the extreme growth of the belief multipliers in a liquidity trap? We explore this question by breaking down the common knowledge in the REE. The household belief variable $\mathcal{D}_t^e$ depends on expected inflation π_t^e, and expected income for Ricardian households $\mathcal{E}_t^e$. One interpretation is that households are trying to guess the behavior of firms through π_t^e, and the behavior of other households through the different components forming expected income $\mathcal{E}_t^e$. Similarly, the belief variable for firms $\mathcal{S}_t^e$ depends on on expected inflation π_t^e as well, and expected marginal costs mc_t^e. Again, this could be interpreted as if firms were trying to predict the behavior of other firms through expected inflation π_t^e, and the behavior of households through the marginal costs mc_t^e.[45] This interpretation allows us to study the fiscal multipliers in a liquidity trap[46] under different assumptions about the strategic interaction and the degree of common knowledge between the two blocks of the economy—the demand and supply sides.

Within-block common knowledge. Assume that households and firms guess the reaction *within* their own type correctly and hold naive expectations *between* each other, so $\mathcal{E}_t^e = \mathcal{E}_t$ and $\pi_t^e = 0$ in the Euler equation (2.31); and $\pi_t^e = \pi_t$ and $mc_t^e = 0$ in the Phillips curve (2.32).[47] Then, these equations become

$$\begin{aligned}\dot{c}_t &= \Omega_g \dot{g}_t - \Omega_\tau \dot{t}_{ht},\\ \dot{\pi}_t &= \rho \pi_t.\end{aligned}$$

Note that the Euler equation is no longer discounted and the additional discounting ρ_c in the Phillips curve vanished. This can be easily understood in terms of common knowledge and

[45] This is just a heuristic interpretation which is not totally accurate. Household income and marginal costs are jointly determined in equilibrium by both demand and supply sides, so there is no notion of household or firm behavior determining these in isolation.

[46] We follow Section 2.2 and set $i_t = i_t^e = 0$ for all $t < T$. Afterwards, monetary policy follows a Taylor rule that guarantees $c_T = \pi_T = 0$ for all $t > T$. We take the limit as $T \to \infty$ to characterize a liquidity trap of an arbitrary length.

[47] Technically, inflation determines the *ex post* real rate on bonds. So getting $\mathcal{E}_t^e$ right requires predicting inflation correctly as well. We opt to ignore the effect of inflation on nominal debt momentarily, as it gets us cleaner expressions without altering our main message.

the strategic interactions induced by it. Suppose households receive news that prices will go up in the future. Everything else equal, that encourages households to spend more before prices increase and less afterwards. The initial increase in spending will be smaller when the raise in prices happens further into the future due to discounting. However, if households have common knowledge about the shock, they understand that all households will raise their spending before the inflation shock. Higher income in all periods before the shock induces another round of higher spending, feeding back into a spending-spending feedback loop. Spending decisions by households are strategic complements, and this feedback loop effectively cancels out the discounting.

The response of output under within-block common is straightforward and given by

$$y_t^W = (1+\Omega_g)g_t - \Omega_\tau t_{ht}.$$

We can see that this common knowledge assumption cancels out the non-ricardian effects of debt on output from the naive multipliers. Ricardian households guess their income correctly and understand that government debt involves either a raise in future taxes, or future drops in income—as taxes raised to the HtM depress aggregate demand. Note that he missing discounting in the Euler equation could amplify any additional effect, such as changes in expected inflation. In general though, the spending-spending feedback loop is not strong enough to generate the explosive behavior documented in Section 2.2 by itself.

Between-block common knowledge. Alternatively, assume that households and firms are able to correctly predict the reaction *between* them while holding naive expectations *within* their own block, so $\pi_t^e = \pi_t$ and $\mathcal{E}_t^e = 0$ in the Euler equation (2.31); and $mc_t^e = mc_t$ and $\pi_t^e = 0$ in the Phillips curve (2.32). Then, these equations become

$$\dot{c}_t = \rho c_t - \tilde{\sigma}\pi_t + \Omega_g \dot{g}_t - \Omega_\tau \dot{t}_{ht} + \Omega_c \bar{\rho}\left[-\mathcal{B}_r \pi_t + \Omega_{rg} g_t - t_{rt}\right], \tag{2.34a}$$

$$\dot{\pi}_t = (\rho+\rho_c)\pi_t - \kappa\left[c_t + (1-\xi)g_t\right], \tag{2.34b}$$

where $\kappa = \frac{\rho_c \lambda}{\bar{\sigma}\xi}$ and $\xi = \frac{1-\delta}{1-\delta+\bar{\sigma}(\gamma+\delta)}$. The key implication of between-block common knowledge is the introduction of the familiar inflation-spending feedback loop that we discussed in Section 2.2. Pricing and spending decisions are strategic complements, in that higher aggregate demand boosts marginal costs and inflation, and inflation in turn encourages spending by decreasing the real rate. Nevertheless, these equations still feature the additional discounting missing in the REE.

The response of output with this common knowledge assumption is given by

$$y_t^B = (1+\Omega_g)g_t - \Omega_\tau t_{ht} + \frac{\bar{\rho}\Omega_c}{1-\chi}b_t + \kappa\tilde{\sigma}\int_0^\infty \alpha_s^B \left[(1-\xi+\Omega_g)g_{t+s} - \Omega_\tau t_{ht+s} + \frac{\bar{\rho}\Omega_c}{1-\chi}b_{t+s}\right]ds,$$

where

$$\alpha_s^B = \frac{e^{-v_L s} - e^{-v_H s}}{v_H - v_L},$$

and $v_H > v_L$ are the eigenvalues of the system (2.34). Now, whether the response of output explodes with the horizon of the fiscal shock depends crucially on α_s^B, which in turn depends on the signs of the eigenvalues.

Proposition 2 (Bounded multipliers with between-block common knowledge)**.** *Assume between-block common knowledge. Then, the belief multipliers α_s^B are bounded if and only if*

$$\rho \geq \frac{\rho_c \Theta \Omega_c}{\xi},$$

where the bound is given by $\alpha^B(s) \leq (\rho_c^2 + 4\kappa\tilde{\sigma})^{-1/2}$.

This result echoes the discussion we had in the previous section about discounted beliefs. Whenever the discount factor is big enough to cancel out the inflation-spending feedback loop, the multipliers will be bounded. The strength of the feedback loop is determined by the curvature of the production function Θ, the redistributional effect of aggregate demand Ω_c, the wealth effect on labor supply ξ, and most importantly, the degree of price flexibility ρ_c.

Rational expectations is obtained by simply combining both assumptions. Intuitively, we could think as the inflation-spending feedback loop carrying most of the action for multipliers, while the spending-spending loop effectively dropping the discounting.

2.4 Level-k Fiscal Multipliers

The last section introduced an equilibrium notion that allows for different belief formation processes. Making the role of expectations explicit highlights the role of belief discounting in determining the response of output to fiscal policy. In this section, we allow for a particular belief formation process based on level-k thinking (Nagel, 1995b; Camerer et al., 2004). This recursive belief formation process lies in between naive expectations, when $k = 0$, and rational expectations, which arises in the limit as $k \to \infty$. We characterize the liquidity trap fiscal multipliers under level-k and show that, for any finite k, they are always bounded and converge to zero as the fiscal shock occurs far away into the future. Moreover, level-k interacts strongly with incomplete markets because of the additional discounting resulting from hitting the borrowing constraint. The spending multipliers for HANK are significantly smaller and converge faster to zero than its counterparts in RANK and TANK. Finally, incomplete markets and level-k is the only way deficit-financed fiscal policy can be more effective at stimulating the economy than running a balanced budget, in line with common intuition in the policy-making world.

2.4.1 Level-k Fiscal Multipliers in RANK and TANK

We next use the general framework developed in Section 2.3 to study the effects of fiscal policy when expectations are updated using level-k thinking. The level-k framework is a recursive belief formation process where agents try to guess relevant outcomes based on a notion of depth of reasoning about how others will act.

Recursive belief formation. The level-zero equilibrium corresponds to a temporary equilibrium where beliefs are given by the original REE, the steady-state, so agents do not update their beliefs after seeing the announcement. This means they hold naive expectations as discussed in the previous section, and hence the response of output will be given by the naive multipliers. This could be interpreted as if agents did not understand the announcement or did not pay attention to it, so they default to act as if everything stayed the same.[48] This equilibrium acts as an initial condition for the recursive process of expectation formation.

Starting the recursion, the level-one equilibrium is a temporary equilibrium with beliefs given by the outcomes that would arise in the level zero equilibrium. In other words, all agents think everybody else is a level-zero agent who will not react to the announcement, and their knowledge of the environment makes them capable of predicting the implications of such behavior on aggregate variables and their own planning problem.

Moving forward, the level-k equilibrium can be defined as a temporary equilibrium with beliefs given by the outcomes that would arise in the level-$k-1$ equilibrium. It is important to stress that the belief formation process happens only once at $t = 0$, and agents do not update their beliefs further even as they turn out to be wrong over time. These equilibria could be seen as if agents were engaged in a cognitive process where they try to guess how

[48]It is standard in the level-k literature to assume that level zero players act in a non-strategic way, defaulting to some random action drawn from a (usually uniform) probability distribution (Nagel, 1995b; Camerer et al., 2004). In our setup, we think it is natural to assume for the most naive agents to behave as if nothing really happened after the announcement.

others will behave using a notion of depth of reasoning. Then, using their knowledge about the environment, they try to deduce the consequences to their own optimization problem. Specifically, for any level-k equilibrium with $k > 0$, they all believe that everybody else is one step behind in the recursive belief formation process; so they expect wages, marginal costs, profits and any other outcome to to have their level-$k-1$ equilibrium values.

Strong temporary equilirbrium. A useful feature of any level-k equilibrium is that their beliefs always constitute a temporary equilibrium. This is guaranteed by the recursive structure of level-k; the beliefs of agents are the outcomes that would arise in a level-k–1 equilibrium, which is itself a temporary equilibrium. Also, given our current decision to start the belief recursion with the REE, the level-0 equilibrium also has beliefs that are a temporary equilibrium. This allow us to use the equilibrium conditions on beliefs to characterize them in terms of the aggregate variables and the fiscal shocks, $\{c_t^e, \pi_t^e, g_t^e, t_{ht}^e, t_{rt}^e\}$, and take advantage of the recursive structure of level-k thinking. The Euler equation and the Phillips curve can be restated as

$$\begin{aligned}
\dot{c}_t &= \rho\hat{c}_t - \tilde{\sigma}\pi_t^e + \Omega_g\dot{g}_t - \Omega_\tau\dot{t}_{ht} + \bar{\rho}\Omega_c\left[\Omega_{rg}\hat{g}_t - \mathcal{B}_r\hat{\pi}_t - \hat{t}_{rt}\right], \\
\dot{\pi}_t &= \rho\pi_t + \rho_c\hat{\pi}_t - \kappa\left[c_t^e + (1-\xi)g_t^e\right],
\end{aligned}$$

where hat variables $\hat{x}_t = x_t - x_t^e$ denoted surprises, that is, deviations of actual realizations from their expected value.

Definition 3 (Strong temporary equilibrium)**.** *A strong temporary equilibrium (STE) is a temporary equilibrium such that the paths for beliefs $\{\mathcal{D}_t^e, \mathcal{S}_t^e\}$ are themselves (part of) a temporary equilibrium.*

Neutral case. Following Section 2.2, we focus on a neutral case that shuts down redistribution induced by fiscal policy. First, the split of profits is such that income for both agents is proportional to aggregate output, so $o = 0$. Second, we assume there is not debt in the steady state, so $\mathcal{B} = 0$, to abstract momentarily from the effects of surprise inflation on nominal debt. Third, taxes are equal for both types of agents so $t_{ht} = t_{rt} = t_t$. This leads to the following definition.

Definition 4 (Level-k equilibrium under neutrality)**.** *A level-k equilibrium under neutrality is a strong temporary equilibrium consisting of a set of paths $\{c_t^k, \pi_t^k\}$ indexed by $k \geq 0$ such that, given a fiscal policy announcement $\{g_t, t_t\}$ at $t = 0$, solve the recursive system of* *differential equations*

$$\dot{c}_t^{k+1} = \rho \hat{c}_t^{k+1} - \bar{\sigma}\pi_t^{k-1} + \Omega(\dot{g}_t - \dot{t}_t),$$
$$\dot{\pi}_t^{k+1} = \rho\pi_t^{k+1} + \rho_c \hat{\pi}_t^{k+1} - \kappa\left[c_t^k + (1-\xi)g_t\right],$$

where $\hat{x}_t^{k+1} = x_t^{k+1} - x_t^k$ *for any variable* x_t*, with initial conditions*

$$c_t^0 = \Omega(g_t - t_t) + \bar{\rho}(1-\chi)^{-1}b_t,$$
$$\pi_t^0 = 0,$$

corresponding to the STE with expectations given by the steady-state REE, $\{c_t^e, \pi_t^e, g_t^e, t_t^e\} = \{0,0,0,0\}$.

Level-k multipliers. The reaction of the economy to the fiscal policy announcement varies greatly with the degree of reasoning k. Lower values for k means that agents are not very sophisticated at understanding the consequences of fiscal shocks and the strategic interactions they promote.[49] Thus, the multipliers will be close to the naive multipliers, with contemporaneous Keynesian cross and the non-ricardian effects of debt driving most of the response. On the other hand, as k grows agents are more refined and believe others to be so as well, so they anticipate larger strategic interactions through the feedback loops that we emphasized for rational expectations.

Proposition 3 (Level-k fiscal multipliers)**.** *The response of the economy to a fiscal policy announcement* $\{g_t, t_t\}$ *at* $t = 0$ *in a level-k equilibrium is given by*

$$c_t^k = \Omega(g_t - t_t) + \int_0^\infty \left[\alpha_{cg}^k(s) g_{t+s} - \alpha_{c\tau}^k(s) t_{t+s}\right] ds,$$
$$\pi_t^k = \int_0^\infty \left[\alpha_{\pi g}^k(s) g_{t+s} - \alpha_{\pi\tau}^k(s) t_{t+s}\right] ds.$$

The multipliers

$$\Lambda_s^k = \begin{pmatrix} \alpha_{cg}^k(s) & \alpha_{c\tau}^k(s) \\ \alpha_{\pi g}^k(s) & \alpha_{\pi\tau}^k(s) \end{pmatrix}$$

are bounded and tend to zero as $s \to \infty$ *for any finite level of* k*. They can be calculated by using the recursion*

$$\Lambda_s^{k+1} = e^{-As}\left(C + \int_0^s e^{Au} B \Lambda_u^k \, du\right),$$

[49] Or alternatively, they presume others are not sophisticated enough to understand these mechanisms.

with initial condition $\Lambda_s^0 = e^{-As}C^0$, *where*

$$A = \begin{pmatrix} \rho & 0 \\ 0 & \rho + \rho_c \end{pmatrix}, \quad B = \begin{pmatrix} \rho & \tilde{\sigma} \\ \kappa & \rho_c \end{pmatrix},$$
$$C = \kappa \begin{pmatrix} 0 & 0 \\ 1 - \xi + \Omega & \Omega \end{pmatrix}, \quad C^0 = -\frac{\bar{\rho}\Omega_c}{1-\chi} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

The coefficients Λ_s^k summarize the response of output and inflation to fiscal policy in a level-k equilibrium. In contrast to rational expectations, the multipliers are bounded and tend to zero as shocks happen further away into the future. In order to analyze them thoroughly, we decompose the multipliers into different channels as in the rational expectations case. In particular, the RANK model introduces the expansionary effect of government spending by rising inflation, while TANK adds an additional effect of deficits that can can create inflation or deflation and be either expansionary or contractionary. Moreover, there is one more channel in level-k absent in rational expectations, coming from the choice of initial conditions at $k = 0$. To be precise, the terms in the integral in Proposition 3 can be expressed as

$$\alpha_{cg}^k(s)g_{t+s} - \alpha_{c\tau}^k(s)t_{t+s} = \alpha_s^{k,g}g_{t+s} + \alpha_s^{k,g-t}(g_{t+s} - t_{t+s}) + \alpha_s^{k,b}(t_{t+s} - g_{t+s}),$$

where $\alpha_s^{k,g}$ represent the spending multipliers characteristic of RANK, $\alpha_s^{k,g-t}$ denote the deficit multipliers introduced in TANK, and $\alpha_s^{k,b}$ corresponds to the debt multipliers from our particular choice of equilibrium at $k = 0$.

Spending multipliers (RANK). Government spending drives up aggregate demand, which increases wages and marginal costs. Firms decide to rise prices, and the generated inflation stimulates consumption by lowering real interest rates—the inflation-spending feedback loop. This channel drives the full explosive response of the economy in the rational expectations RANK model, and likewise is the only channel present in level-k when there are no HtM and $\chi = 0$.

In contrast, the strength of the feedback loop in level-k depends crucially on the depth of reasoning k. Panel A in Figure 2.3 shows the magnitude of this RANK effect in the REE and level-k equilibria.[50] At first glance, we can see that the spending multipliers are markedly attenuated in level-k. The level-one effect is zero, coinciding with the rigid price multipliers highlighted in Section 2.2.1. There is no inflation in the level-zero equilibrium, as firms

[50]Given the exponential growth of the rational expectations multipliers, we use a logarithmic scale and add one to $\alpha_s^{k,g}$ to be able to represent a zero effect.

expect everything to stay the same after the announcement. Thus, level-one households expect zero inflation and do not engage in intertemporal substitution. But even higher levels of reasoning lead to considerable dampening. Spending multipliers in the level-five equilibrium are roughly 50 percent lower than the REE benchmark for shocks happening three years into the future, and 90 percent lower for shocks in a five year horizon.

The reason behind the attenuation of this channel in level-k is the limited scope of the inflation-spending feedback loop. The number of interactions where households react to inflation by increasing consumption, and firms react to consumption by raising prices, is limited by the rounds of reasoning they do when forming their expectations. Interestingly, because of the discounting of future decisions by both households and firms, shocks that are further away into the future require higher rounds of reasoning k to make an impact in the present. Discounting by each block is compounded so that faraway shocks have very small direct effects on current inflation and spending, so they require a lot of rounds of strategic interaction for them to gain some traction. This induces a hump shape to the multipliers. For a given finite level of k, shocks after the peak in the spending multipliers are in need of more rounds of the feedback loop for them to make a dent in the response today. In our simulations, the multipliers end up decaying at rate ρ, illustrating the intuition above.

Deficit multipliers (TANK). Deficits have no additional effect in the rational expectations RANK model because of Ricardian equivalence. This changes when HtM agents are introduced though, as rising taxes to the HtM reduces their consumption and depresses aggregate demand. Then, future taxes depress consumption today in anticipation of future deflation and higher real rates, which creates further deflation due to lower marginal costs—a deflation-spending feedback loop. A striking implication of this deficit channel is that deficit-financed fiscal policy can lead to negative multipliers because of the negative effect that future fiscal surpluses have on current output.

Again, the strength of this channel in a level-k equilibrium depends critically on the level of reasoning k. We can see in Panel B in Figure 2.3 that the attenuation of the deficit multipliers in level-k is similar to that of the spending multipliers. The level-one multipliers are zero given that no inflation occurs in the level-zero equilibrium. As k rises, agents understand the positive (negative) impact of deficits (surpluses) in the future and hence adjust their current spending and pricing decisions accordingly. Even for $k = 5$, the attenuation is around 50 percent for a horizon of three years, and 90 percent for five years.

The reasons behind the attenuation of the deficit multipliers are the same as for the inflation multipliers. Level-k limits the number of rounds in the feedback loop between deflation and spending due to future surpluses, leading to bounded and hump-shaped deficit multipliers for any finite k. Consequently, a flatter profile for the deficit multipliers substantially mitigates the negative effect of deficit-financed fiscal policy.

Debt multipliers (Initial conditions). The initial conditions chosen at $k = 0$ have an impact in the successive level-k equilibria. In our case, we chose the REE as the beliefs for the naive level-zero agents. This leads to non-ricardian effects of debt, where agents perceive government bonds as net wealth because they cannot foresee the future raise in taxes.

We use the intertemporal government budget constraint to express the debt multipliers $\alpha_s^{k,b}$ in terms of the impact of future surpluses, to facilitate the comparison with the deficit multipliers. In contrast to the other channels, the "fundamental" non-ricardian effect of debt is not present in level-k equilibria other than $k = 0$. Level-zero agents belief debt is net wealth because they do not appreciate the raise in future taxes, but this is no longer true for $k \geq 1$. In the level-one equilibrium, households no longer believe that debt is making them richer. However, they belief others will increase spending when debt is issued, so they anticipate larger aggregate and individual income and increase their spending accordingly. The successive level-k households already understand that increases in income caused by government spending are offset by raises in taxes. They just get wrong the spending and inflation responses by others, which are in turn caused by their wrong beliefs about others. The only reason why debt has an effect when $k \geq 1$ is because of high-order beliefs. Agents belief that others will increase spending in the presence of debt, or belief that others belief that others will increase spending, etc. This differentiates the debt multipliers from the spending and deficit multipliers, in that agents are somehow reacting to a sort of "sunspot" instead of a "real" shock.[51]

We find that these debt multipliers have the same properties as the other multipliers: they are bounded, hump-shaped, and tend to zero as $s \to \infty$. We can see in Panel C in Figure 2.3 that the debt multipliers are orders of magnitude smaller than the other multipliers. The effect tends to fade away for short horizons as k increases, converging to the rational expectations response at zero. On the other hand, the multipliers grow for long horizons as k increases, illustrating the power of the inflation-feedback loop even for non-fundamental shocks.[52] As a matter of fact, despite the very weak impact for short horizons shown in Panel C, they tend to overgrow the deficit multipliers for large enough horizons of decades. Given that debt and deficit multipliers have opposite signs, this suggests that very long-lasting

[51] It is not a sunspot in the usual sense of a coordination device for multiple equilibria in rational expectations, but rather a belief that is not driven by fundamentals and just comes from high-order beliefs.

[52] This fact can be consistent with the debt multipliers converging to the rational expectations zero effect by using a formal notion of convergence. Debt multipliers seem to converge pointwise to zero for all s as $k \to \infty$, while it is clear that they do not converge uniformly given their growth for large s as k increases.

debt could be expansionary under level-k if the zero lower bound remains binding for long enough.

Alternative initial conditions at $k = 0$. So far we have only considered the REE as the initial set of beliefs for level-zero equilibrium, leading to a temporary equilibrium where debt has non-ricardian effects. This provides a neat interpretation and demands very little from agents in the economy, but other options could be considered.

One possibility would be that agents understand the effect of announced taxes on their planning problem. This is a reasonable assumption considering taxes enter directly into their budget constraints. In RANK, agents forecast lower lifetime income due to higher taxes, without perceiving the benefits of government spending. Thus, the level-zero economy would start at a depressed state. As the surprise income shock from government spending occurs, the economy would go back to the steady state. In this case, the same debt multipliers $\alpha_s^{k,b}$ would apply to just the path of g_{t+s}, rather than the path for surpluses $t_{t+s} - g_{t+s}$. Things would more complicated in TANK, as they see the impact of taxes on their budget constraints but ignore the general equilibrium effects of the Keynesian cross from HtM households. In the case of deficit-financed fiscal policy, the economy would start depressed as in RANK, then overshoot the steady state as government spending is realized, to finally go down again as taxes turn out to reduce their earnings by more than expected through lower labor income.

From all the possible alternatives for initial conditions, there is one in particular that deserves special attention. Suppose that level-zero households hold beliefs that take into account the effects of fiscal policy on their budget constraints, except for the inflation that it could cause. Specifically, households understand the impact of fiscal policy on wages and profits, including the impact of the Keynesian cross if HtM are present. In other words, households understand Ricardian equivalence and do not perceive debt as net wealth. This beliefs would precisely correspond to a REE with rigid prices. Then, the level-zero equilibrium coincides exactly with this rigid price REE, and the spending and deficit multipliers would be the only ones featured in any subsequent level-k equilibrium. This assumption makes the debt multipliers to be zero for all k, and corresponds to the zero initial condition for the fiscal multipliers $\Lambda_s^0 = 0$ in Proposition 3. Despite this being a useful benchmark, we thought our choice of the steady-state REE as the initial beliefs for level-zero agents was more plausible in a bounded rationality setting.

Surprise inflation and nominal debt (Fisherian channel). An additional channel arises in level-k thinking compared to rational expectations because of the possibility of surprise inflation together with nominal debt. So far we opted to simplify the discussion by conveniently shutting down nominal debt in the steady state (i.e., setting $\mathcal{B} = 0$). When government debt is allowed though, unexpected inflation decreases the real value of nominal debt, redistributing from creditors to debtors. This Fisherian channel is well known in the literature, at least since Fisher (1933). The most immediate implication of this channel is that fiscal policy needs to be adjusted across different level-k equilibria for the government budget constraint (2.24) to hold under varying levels of inflation. In this case, the choice of active versus passive fiscal policy does matter in calculating the level-k multipliers.

Consider that fiscal policy is active, so the paths $\{g_t, t_t\}$ can stay the same across level-k equilibria. Then, all the adjustment required by the different inflation rates is done through b_0. For example, a fiscal policy path that tends to create more inflation that deflation will put upward pressure on b_0 as the government would be able to repay larger levels of debt for the same fiscal surpluses. Given that future inflation will eventually catch agents by surprise, they currently believe that they are richer than they actually are and increase spending accordingly. Thus, the effects of fiscal policy will be magnified when the Fisherian channel is present and fiscal policy is active.

On the other hand, passive fiscal policy complicates the analysis significantly. Assuming that the path of government spending stays the same, taxes must vary across different level-k equilibria to satisfy the intertemporal budget constraint of the government. Thus, taxes become endogenous objects subject to surprises not only in the level-zero equilibrium, but in every level-k equilibrium. In principle, there can be a wide range of implications depending on how and when taxes adjust in the presence of inflation, but a neutrality result can still be obtained. Consider that taxes of the HtM remain unchanged across equilibria and all the adjustment comes from taxes to the Ricardian agents, tracking one-to-one any surprise in inflation across equilibria. Then, the Fisher channel is completely shut down as Ricardians find themselves with fewer savings but lower taxes whenever there is surprise inflation, canceling each other out. In a way, this is like the government guaranteeing that the *ex post* return on debt is equal to *ex ante* return by using lump-sum taxes.

2.4.2 Level-k Fiscal Multipliers in HANK

The previous section showed that level-k is very powerful at reducing the effect of future fiscal policy announcements on current output—the fiscal multiplier puzzle—when financial markets are complete. Next, we study the interaction of level-k thinking with incomplete markets in a standard heterogeneous agent New Keynesian (HANK) model à la Bewley-Hugget-Aiyagari. We find that incomplete markets interact strongly with level-k in two different ways. First, the expansionary effects of a balanced-budget fiscal stimulus are further dampened under level-k when incomplete markets are introduced, particularly in the future. Second, level-k fiscal multipliers in HANK tend to be larger when the fiscal stimulus is financed through debt rather than current taxes. This result cannot be obtained under complete markets or rational expectations, restoring the common intuition in policy-making circles that running deficits are better at stimulating the economy.

We begin by sketching the main differences from adapting the framework in Section 2.3 to a discrete-time HANK model.

Households. The economy is populated by a continuum of households subject to idiosyncratic income risk and borrowing constraints. Incomplete markets limits their ability to smooth consumption, so they engage in precautionary savings and occasionally hit their borrowing constraints, becoming hand-to-mouth temporarily. Households maximize their expected lifetime utility subject to the actual and expected budget constraints

$$C^e_{it+s} + A^e_{it+s} = \frac{R^e_{t+s-1}}{\Pi^e_{t+s}} A^e_{it+s-1} + W^e_{t+s} Z_{it+s} N^e_{it+s} + \Phi^e_{it+s} - T^e_{it+s},$$

$$C_{it} + A_{it} = \left[\frac{R_{t-1}}{\Pi_t}(1-\lambda_{it-1}) + \left(\frac{R^e_{t-1}}{\Pi^e_t} + \frac{D_t - D^e_t}{Q_{t-1}}\right)\lambda_{it-1}\right] A_{it-1} + W_t Z_{it} N_{it} + \Phi_{it} - T_{it},$$

where $Z_{it}N_{it}$ is the effective labor supply, and z_{it} is an idiosyncratic productivity shock following the stochastic process $\ln Z_{it} = \rho_z \ln Z_{it-1} + \varepsilon_{it}$, with $\varepsilon_{it} \sim N\left(\mu_\varepsilon, \sigma^2_\varepsilon\right)$ being i.i.d. over time and across agents.[53]. We assume that the portfolio composition is the same for all households, so $\lambda_{it} = \lambda_t = Q_t/A_t$. Then, their optimality conditions are given by

$$C_{it}^{-1/\sigma} \geq \beta \frac{R_t}{\Pi^e_{t+1}} C_{it+1}^{-1/\sigma}$$

$$Z_{it} W_t = C_t^{1/\sigma} N_{it}^{\gamma}.$$

Aggregation conditions. Aggregate variables are obtained by aggregating individual quantities over all households

$$C_t = \int C_{it}\, di,$$

$$N_t = \int Z_{it} N_{it}\, di,$$

$$A_t = \int A_{it}\, di, ,$$

$$\Phi_t = \int \Phi_{it}\, di.$$

[53]The mean of ε is chosen so that $\mathbb{E}_i z_{it} = 1$. In particular, $\mu_\varepsilon = -\sigma^2_\varepsilon (1-\rho_z)^{-1}/2$.

Lucas trees and profits. Dividends of the Lucas trees are taxed, so

$$D_t = \delta(Y_t - T_{dt}).$$

Households receive profits proportionally to their labor income

$$\Phi_{it} = \frac{(Z_{it}N_{it})^{1+o}}{\int (Z_{it}N_{it})^{1+o}\, di}\Phi_t,$$

where the coefficient o measures the convexity of earned profits with labor income. Importantly, households take profits as given, so they do not internalize the effect of their labor supply on earned profits. Similar to equation (2.22), higher values for o imply that low-income households receive relatively less profits, so the income of high-MPC agents will tend to fluctuate more than aggregate income. The neutral case arises with $o = 0$, making individual earnings proportional to aggregate income.

Government. Total taxes raised by the government include the taxes from households and the tax on dividends, so

$$T_t = \int T_{it} di + \delta T_{dt}.$$

Neutral case. As always, we base our results in a neutral case that shuts down redistributive effects of fiscal and monetary policy. First, individual profits are proportional to labor income, so $o = 0$. Second, $\sigma = 1$ so we have log utility. Third, the effect of surprise inflation on the value of nominal debt is canceled out by issuing a tax rebate to households proportional to their debt holdings. This shuts down the redistribution from the Fisherian channel in level-k equilibria. Fourth, individual taxes are raised proportionally to labor earnings. Thus, individual taxes are given by

$$T_{it} = \frac{(Z_{it}N_{it})^{1+o}}{\int (Z_{it}N_{it})^{1+o}\, di}(1-\delta)T_{wt} + \left(\frac{R_{t-1}}{\Pi_t} - \frac{R^e_{t-1}}{\Pi^e_t}\right)(1-\lambda_{t-1})A_{it-1},$$

where T_{wt} is part of the fiscal policy announcement. Note that agents take taxes as given, and they do not internalize the impact of their decisions on the final taxes they will pay. Finally, dividends and labor income are taxed equally so $T_{wt} = T_{dt} = T_t - (R_{t-1}/\Pi_t - R)\, B_{t-1}$.

Level-k spending multipliers in HANK. First, we analyze the inflationary effects of government spending emphasized by the spending multipliers $\alpha_s^{k,g}$. We showed in Section 2.2 that HANK, TANK and RANK yield the same rational expectations multipliers when the government runs a balanced budget and neutrality assumptions shut down redistribution. Therefore, we use budget-balanced stimuli to compare the spending multipliers under level-k with and without incomplete markets. Panel A in Figure 2.4 plots the ratio of the spending multipliers with and without complete markets, that is, the response of current output to a balanced-budget shock s quarters into the future. We find that the level-k response to government spending is dampened when incomplete markets are considered. For example, the reaction of output in a level-two equilibrium is 30 percent smaller than in the RANK benchmark for a fiscal shock 4 years out. Moreover, Panel B and C show that the interaction between incomplete markets and level-k gets amplified when liquidity becomes scarce. Increasing the fraction of hand-to-mouth households from 20 to 80 percent attenuates the incomplete markets response by more than 70 percent with respect to the level-two RANK benchmark for a shock 4 years into the future. Even when $k = 5$, the response is almost 40 percent smaller than in complete markets for this latter shock. These results are in line with Farhi and Werning (2019), who finds a similar interaction between level-k and incomplete markets for monetary policy announcements.

Deficit-financed stimulus in level-k HANK. HANK and TANK have different rational expectations multipliers, complicating any comparison between them. As we saw in Figure 2.2, HANK negative multipliers are smaller, suggesting that the impact of future surpluses and the resulting deflation is attenuated in HANK. Interestingly, this difference seems to be exacerbated in level-k, to the point where deficits tend to reduce the level-k multipliers in TANK, while increasing them in HANK. Figure 2.5 shows the reaction of output at $t = 0$ for the same experiment we run in Section 2.2, with both models having roughly 20 percent of their agents being HtM. This consisted of a fiscal stimulus of varying length, where taxes were raised with some delay. Starting with the left column, we can see that TANK multipliers tend to decrease for all k, as taxes are delayed further into the future. This suggests that deficits reduce the effects of fiscal policy not only in the REE, but also in level-k. Nonetheless, the negative impact of backloading taxes is certainly reduced in level-k, where multipliers remain positive for reasonable horizons. Naturally, the negative impact of deficits increases with k, as higher levels of reasoning helps agents to better anticipate the deflation feedback loop. On the other hand, the right column clearly shows how backloading taxes actually magnifies the impact multipliers in HANK for a wide range of level-k equilibria. Thus, deficits intensify the expansionary effects of fiscal spending in the level-k HANK. This restores the usual intuition in policy-making circles that combining fiscal stimulus with debt issuing tends to be more expansionary than keeping the budget balanced.

There might be several reasons behind the expansionary effects of deficits in HANK under level-k thinking. First, the comparison between the spending multipliers in HANK and RANK from Figure 2.4 shows that level-k multipliers tend to be lower when markets are incomplete, particularly in the future. Therefore, the additional future discounting from occasionally hitting the borrowing constraints in HANK could be downsizing the impact of the deflation loop from future taxes. Nonetheless, the quantitative impact of this interaction does not seem to be that large when only 20 percent of the agents are constrained. An alternative explanation could be related to the expansionary effect of debt when markets are incomplete. More liquidity facilitates consumption smoothing in HANK, reduces the precautionary savings motive and increases aggregate demand. Then, because the expansionary effect of higher liquidity is more frontloaded than the contractionary effect of future taxes and debt reduction, it gets featured more prominently in low-reasoning level-k equilibria. We believe this liquidity channel to be the main driver behind the amplification of deficit-financed fiscal stimuli when level-k and incomplete markets are paired together.

2.5 Conclusion

In this paper, we have reported that the New Keynesian model exhibits strange behavior in a liquidity trap beyond the usual focus of the literature on monetary policy and forward guidance. The fiscal multipliers grow unboundedly as spending is backloaded into the future, but can easily yield the opposite effect and be extremely recessionary when hand-to-mouth agents are introduced and the stimulus is deficit-financed. Surprisingly, the incomplete markets heterogeneous-agent New Keynesian model inherits this same property and produces negative multipliers when taxes are backloaded. We have built a belief-augmented New Keynesian framework that makes explicit the role played by expectations in shaping this results. Adopting bounded rationality in the form of level-k thinking provides a powerful and tractable tool to reduce the extreme sensitivity to fiscal policy embedded in rational expectations. Interestingly, it interacts strongly with incomplete markets by further mitigating the multipliers and rationalizing the common belief that deficits are more effective at stimulating the economy.

We belief that an important direction for future research is to generalize level-k thinking to a dynamic stochastic environment. Here, we focused on a temporary and unanticipated shock and did not allow for belied revision or any kind of learning. However, it might be natural to assume in our context that forecasting errors could induce revisions about the level of rationality of other agents. This is an ambitious idea given one needs to specify what contingencies agents are able to predict, whether different kind of shocks or policies deserve different levels of rationality, how the level-k gets updated in presence of new information, and the initial condition for the recursion. Furthermore, in reality new agents are continuously entering the economy which could lead to interesting dynamics for the perceived level-k depending on how people believe these new agent will act. It is our opinion that exploring the implications of these different specifications and mustering empirical evidence to help sort through the different theoretical possibilities is a demanding but promising agenda.

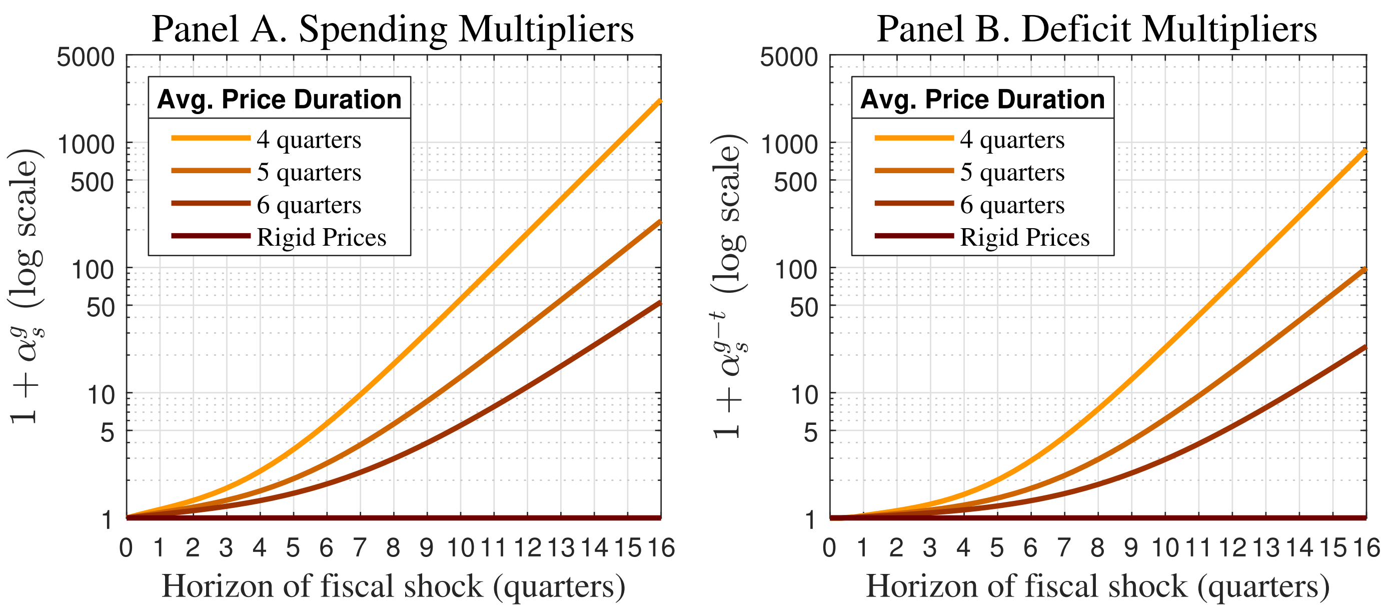


Figure 2.1: Spending and Deficit Multipliers from RANK and TANK.

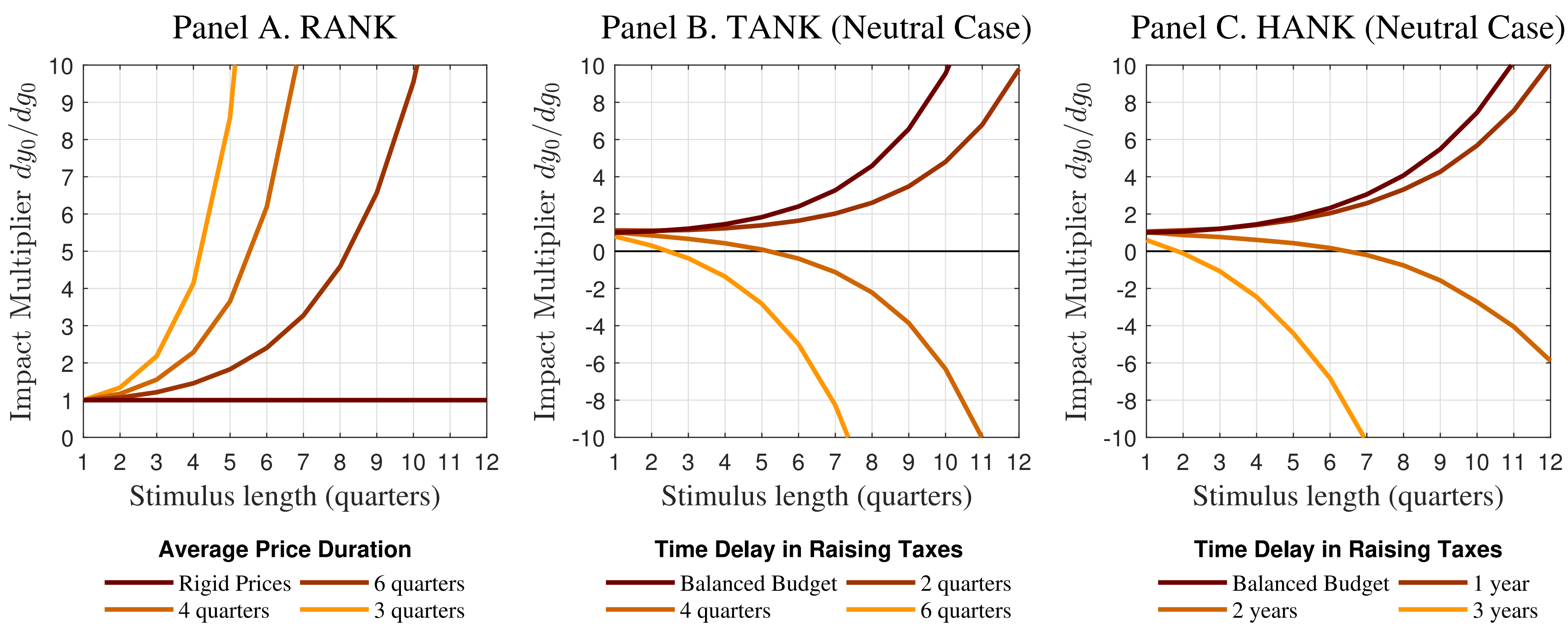


Figure 2.2: Output Response to a Persistent Fiscal Stimulus in NK Models.

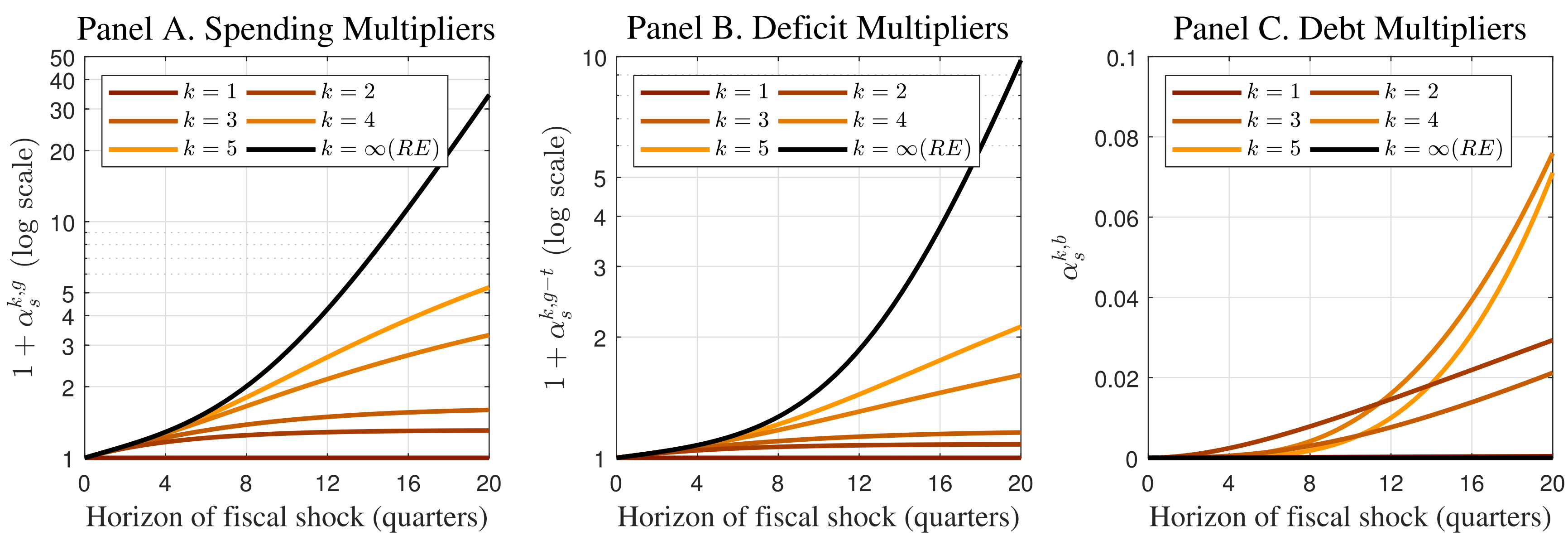


Figure 2.3: Decomposition of Level-k Fiscal Multipliers.

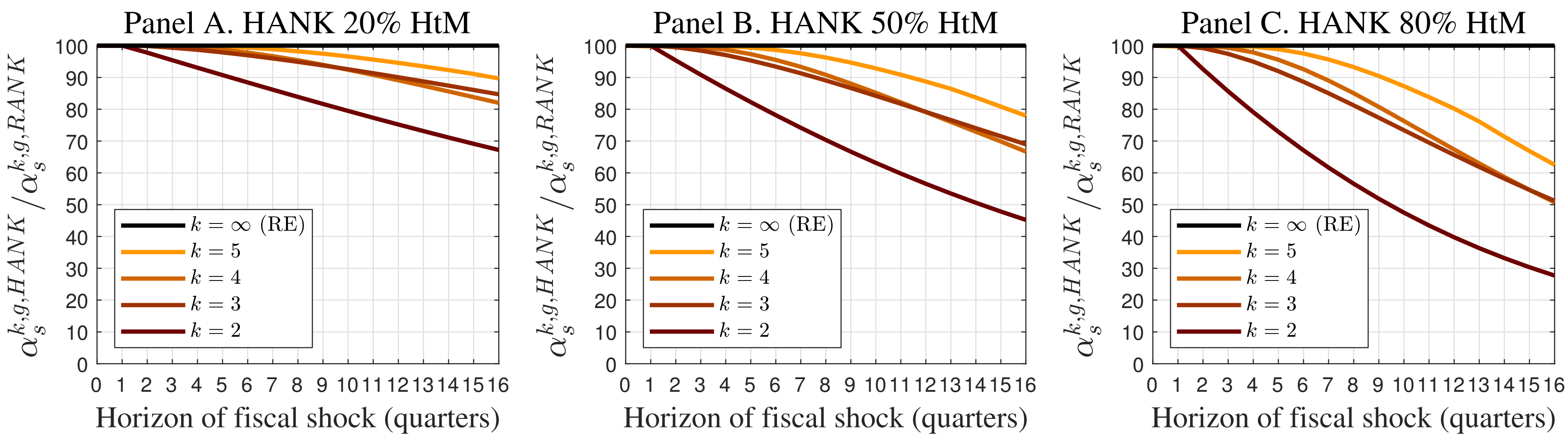


Figure 2.4: Comparison of Level-k Spending Multipliers in HANK and RANK.

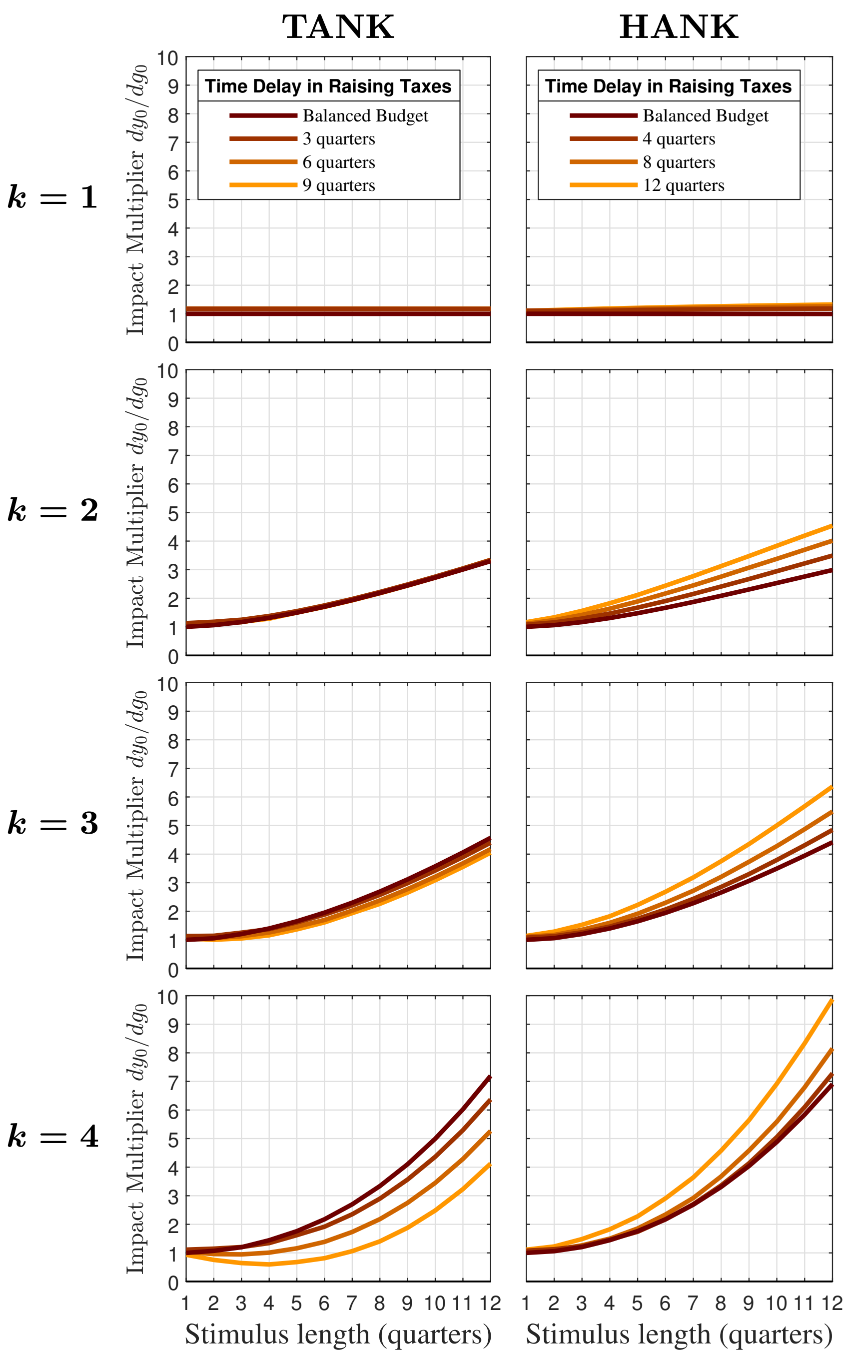


Figure 2.5: Comparison of Deficit-Financed Stimulus in TANK and HANK.

Table 2.1: CALIBRATION OF SIMULATIONS.

Parameter	Description	Values	Target/Source
β	Discount factor	$1.02^{-1/4}$	Two percent annual interest rate.
R	Steady-state interest rate	$1.02^{1/4}$	Endogenous in HANK.
σ	Elasticity of intertemporal substitution	1	Neutrality in Werning (2015).
$1/\gamma$	Frisch elasticity	1/2	Chetty et al. (2013).
δ	Capital share	0.0286	144% Liquid Assets to GDP (McKay et al., 2016).
$1-\lambda$	Fraction of firms adjusting prices	15%	Christiano et al. (2011b).
θ	Elasticity of demand	6	20% markup (Christiano et al., 2011b).
ρ_z	Income process: persistence	0.966	Floden and Lindé (2001)
σ_ε^2	Income process: shock variance	0.017	Floden and Lindé (2001)
χ	Share of hand-to-mouth households	20%	Consistent with HANK outcome.

Appendices

Here, we describe the main setup of the rational expectations New Keynesian models featured in Section 2.2 and derive the fiscal multipliers.

A Representative-Agent and Two-Agent Models

We begin by describing the environment in RANK & TANK that allow us to get a closed form solution for the fiscal multipliers.

Hand-to-mouth households. The economy is populated by a share χ of constrained households. They do not have access to financial markets to save or borrow, so they consume all their income. They receive income from their labor services,$W_t N_{ht}$, and a share of uncapitalized profits from the private sector, Φ_{ht}, which are taken as given.[54] They pay lump-sum taxes T_{ht} to the government, which are also taken as given. Even though the HtM cannot smooth consumption intertemporally, they still optimize their labor supply decision. In sum, their behavior is characterized by

$$C_{ht} = W_t N_{ht} + \Phi_{ht} - T_{ht}, \tag{37}$$

$$W_t = N_{ht}^{\gamma} C_{ht}^{1/\sigma}, \tag{38}$$

Ricardian households. The remaining $1-\chi$ share of households are Ricardian. At every period t, these households choose paths for real consumption $\{C_{t+s}\}$ and labor supply $\{N_{t+s}\}$ that maximize their lifetime utility

$$\sum_{s=0}^{\infty} \beta^s \left[\frac{\left(C^e_{rt+s}\right)^{1-1/\sigma}}{1-1/\sigma} - \frac{\left(N^e_{rt+s}\right)^{1+\gamma}}{1+\gamma} \right] ds,$$

subject to their budget constraints, where β denotes the discount factor, σ denotes the denotes elasticity of intertemporal substitution, and γ denotes the inverse Frisch elasticity. They receive income from their labor services, $W_t N_{rt}$, and a share of uncapitalized profits from the private sector, Φ_{rt}, which are taken as given. They pay lump-sum taxes T_{rt} to the government, which are also taken as given. In contrast to HtM households, Ricardian agents can trade real and nominal risk-free assets. Specifically, there are Lucas trees X_{rt} with price Q_t that pay dividend D_t, and nominal government bonds $P_t B_{rt}$ which pay interest R_t. Thus,

[54] These profits can be thought of as a mix of rent sharing by firms with their workers and transfers by the government.

assets evolve according to the budget constraints

$$C_{rt} + Q_t X_{rt+1} + B_{rt+1} = \frac{R_{t-1}}{\Pi_t} B_{rt} + (D_t + Q_t) X_{rt} + W_t N_{rt} + \Phi_{rt} - T_{rt},$$

for all t.

The first order conditions that characterize the optimal decisions by Ricardian households are given by

$$C_{rt}^{-1/\sigma} = \beta \frac{R_t}{\Pi_{t+1}} C_{rt+1}^{-1/\sigma}, \tag{39}$$

$$W_t = N_{rt}^{\gamma} C_{rt+s}^{1/\sigma}, \tag{40}$$

the transversality condition

$$\lim_{s \to \infty} \frac{Q_s X_{rs} + B_{rs}}{\prod_{k=0}^{s-1} R_{t+k}/\Pi_{t+k+1}} = 0,$$

and the no-arbitrage condition

$$Q_{t+1} + D_{t+1} = Q_t \frac{R_t}{\Pi_{t+1}}.$$

This last condition arises because the Lucas trees and the government bonds both have the same *ex ante* return. Then, the price of the Lucas Tree must satisfy

$$Q_t = \sum_{s=0}^{\infty} \frac{D_{t+s+1}}{\prod_{k=0}^{s} R_{t+k}/\Pi_{t+k+1}}, \tag{41}$$

and the budget constraints can be expressed in terms of total financial wealth, $A_{rt} = Q_t X_{rt} + B_{rt}$, so

$$C_{rt} + A_{rt+1} = \frac{R_{t-1}}{\Pi_t} A_{rt} + W_t N_{rt} + \Phi_{rt} - T_{rt}, \tag{42}$$

Similarly, the transversality condition becomes

$$\lim_{s \to \infty} \frac{A_{rs}}{\prod_{k=0}^{s-1} R_{t+k}/\Pi_{t+k+1}} = 0. \tag{43}$$

Final good producers. The final good is produced by a perfectly competitive sector using intermediate good varieties $Y_t(j)$ with the CES production function

$$Y_t = \left(\int_0^1 Y_t(j)\, dj \right)^{\frac{\theta}{\theta-1}}.$$

The demand for variety j from the final good sector is given by

$$Y_t(j) = \left(\frac{P_t(j)}{P_t} \right)^{-\theta} Y_t,$$

where $P_t(j)$ is the price of the variety and P_t is the price for the final good

$$P_t = \left(\int_0^1 P_t(j)^{1-\theta} dj \right)^{\frac{1}{1-\theta}}. \tag{44}$$

Intermediate good producers. Varieties of the intermediate good are produced by monopolistically competitive firms using a Cobb-Douglas technology

$$Y_t(j) = N_t(j)^{1-\delta}.$$

Prices are set *à la* Calvo, that is, firms can only change prices with a constant probability $1-\lambda$. Those firms that can change prices choose their reset price P_t^* to maximize the present discounted value of lifetime profits

$$\max_{P_t^*} \sum_{s=0}^{\infty} \frac{P_t^* Y_{t+s|t}(j) - P_{t+s} C_s \left(Y_{t+s|t}(j) \right)}{\prod_{k=0}^{s-1} R_{t+k} / \Pi_{t+k+1}},$$

where $Y_{t+s|t}(j) = \left(P_t^* / P_{t+s} \right)^{-\theta} Y_{t+s}$ is the demand for their variety j, and $C_s \left(Y_s(j) \right) = W_s Y_s(j)^{1/(1-\delta)}$ is the real cost function for intermediates at time s, and aggregate variables W_t, Y_t and P_t are taken as given. Then, the real reset price is

$$\frac{P_t^*}{P_t} = \left[\frac{\theta}{\theta - 1} \frac{\sum_{s=0}^{\infty} \frac{\lambda_c^s}{\prod_{u=0}^{s-1} R_{t+u} / \Pi_{t+u+1}^{\theta+1/\Theta}} Y_{t+s} MC_{t+s}}{\sum_{s=0}^{\infty} \frac{\lambda_c^s}{\prod_{u=0}^{s-1} R_{t+u} / \Pi_{t+u+1}^{\theta}} Y_{t+s} ds} \right]^{\Theta}, \tag{45}$$

where $\Theta = \frac{1-\delta}{1-\delta+\theta\delta}$, and $MC_{t+s} = (1-\delta)^{-1} W_{t+s} \left(Y_{t+s} \right)^{\delta/(1-\delta)}$ is proportional to the average real marginal cost at time $t+s$. Then, equation (44) implies that inflation of the final good is

$$\Pi_t = \left[\frac{1 - (1-\lambda) \left(P_t^* / P_t \right)^{1-\theta}}{\lambda} \right]^{1/(\theta-1)}. \tag{46}$$

Lucas trees and profits. There is a unit net supply of Lucas trees. These trees capitalize a share of the profits of the intermediate good sector[55] and their dividend is

$$D_t = \delta Y_t. \tag{47}$$

The remaining uncapitalized profits

$$\Phi_t = Y_t - D_t - W_t N_t, \tag{48}$$

are lump-sum transferred to both types of households. The split of profits depends on a parameter o so

$$\Phi_{ht} = (1-o)\left(Y_t - W_t N_t\right), \tag{49}$$

where $o \in \left[-\frac{1-\chi}{\chi}, 1\right]$. Note that total income of HtM households is proportional to aggregate output when $o = 0$, corresponding to the neutral case mentioned in Section 2.2. In fact, this parameter captures the income cyclicality of the HtM. The value of o determines the amount of profits accruing to the HtM, affecting the share of their total earnings coming from labor income. Thus, HtM earnings will be relatively more procyclical or countercyclical when o is positive or negative, respectively.

Government. The government finances its expenditures by raising (real) taxes T_t and issuing nominal government bonds $P_t B_t$

$$\begin{aligned} T_t &= \chi T_{ht} + (1-\chi) T_{rt}, \\ T_t + \frac{d}{dt} B_t &= G_t + (i_t - \pi_t) B_t, \end{aligned} \tag{50}$$

where G_t is real government spending and $(i_t - \pi_t)B_t$ is the real cost of serving the debt. Fiscal policy announcements $\{G_{t+s}, T_{t+s}\}$ must satisfy the intertemporal government budget constraint

$$B_t = \sum_{s=0}^{\infty} \frac{T_{t+s} - G_{t+s}}{\prod_{k=0}^{s-1} R_{t+k}/\Pi_{t+k+1}}, \tag{51}$$

so the government must raise enough taxes over time to finance its spending path and the current level of public debt.

[55] The profits capitalized by the Lucas trees could be interpreted as the rents of a fixed factor used in the intermediate sector. Then, the Lucas trees would represent the shares of a competitive mutual fund that owns the entire supply of the fixed factor in the economy.

Monetary policy. Monetary policy is constrained, so interest rates do not react to changes in the economy. Specifically, we assume that $R_t = \beta$ for all $t < T$, and follows a standard Taylor rule satisfying the Taylor principle for all $t > T$. This guarantees that inflation will be pinned down and solves undeterminacy problems under an interest rate peg.

Aggregation and market clearing. Aggregate variables are obtained by aggregating individual quantities over all households

$$C_t = \chi C_{ht} + (1-\chi) C_{rt}, \tag{52a}$$
$$N_t = \chi N_{ht} + (1-\chi) N_{rt}, \tag{52b}$$
$$A_t = (1-\chi) A_{rt}, \tag{52c}$$
$$\Phi_t = \chi \Phi_{ht} + (1-\chi) \Phi_{rt} \tag{52d}$$

and the goods, labor and assets market clear whenever

$$Y_t = C_t + G_t, \tag{53a}$$
$$N_t = Y_t^{\frac{1}{1-\delta}} \Delta_t, \tag{53b}$$
$$Q_t = \lambda_t A_t, \tag{53c}$$
$$B_t = (1-\lambda_t) A_t, \tag{53d}$$

where $\Delta_t = \int_0^1 (P_t(j)/P_t)^{-\theta/(1-\delta)} dj$ is an index of price dispersion.

Log-linearization. We log-linearize the equilibrium conditions around the zero inflation steady-state with no inflation and no government debt. As in Farhi and Werning (2016), the equilibrium is characterized by the Euler equation and the Phillips curve

$$\begin{aligned} c_t &= c_{t+1} - \tilde{\sigma}(i_t - \pi_{t+1}) - \left(\Omega_g \Delta g_{t+1} - \Omega_\tau \Delta t_{ht+1}\right), \\ \pi_t &= \beta \pi_{t+1} + \kappa \left[c_t + (1-\xi) g_t\right], \end{aligned}$$

where

$$\tilde{\sigma} = \frac{\sigma}{1 - G/Y} \frac{(1-\chi)(\mu\sigma\gamma + 1 - \delta)}{\sigma\gamma\left[\mu(1-\chi(1-o)) - \chi o(1+\gamma)\right] + (1-\delta)\left[1 - \chi(1+o\gamma)\right]}$$

is the aggregate intertemporal elasticity of substitution and

$$\Omega_g = \frac{\chi\sigma\gamma\left[\mu + o(1+\gamma-\mu)\right]}{\sigma\gamma\left[\mu(1-\chi(1-o)) - \chi o(1+\gamma)\right] + (1-\delta)\left[1-\chi(1+o\gamma)\right]},$$

$$\Omega_t = \frac{\chi\mu\sigma\gamma}{\sigma\gamma\left[\mu(1-\chi(1-o)) - \chi o(1+\gamma)\right] + (1-\delta)\left[1-\chi(1+o\gamma)\right]}.$$

Fiscal Multipliers. Putting the system in matrix form we get

$$\begin{pmatrix} c_t \\ \pi_t \end{pmatrix} = \begin{pmatrix} 1 & -\tilde{\sigma} \\ \kappa & \beta \end{pmatrix} \begin{pmatrix} c_{t+1} \\ \pi_{t+1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \kappa & 0 \end{pmatrix} \begin{pmatrix} c_t \\ \pi_t \end{pmatrix} + \begin{pmatrix} \Omega_g \Delta g_{t+1} - \Omega_\tau \Delta t_{ht+1} \\ \kappa(1-\xi) g_t \end{pmatrix},$$

or

$$X_t = AX_{t+1} + MX_t + C_t.$$

Then,

$$X_t = (I-M)^{-1}\left[AX_{t+1} + C_t\right],$$

and the only non-explosive solution is given by

$$X_t = \sum_{h=0}^{\infty} B^h (I-M)^{-1} C_t,$$

where $B = (I-M)^{-1}A$. Then the reaction of consumption to fiscal policy in the RANK model ($\chi = 0$) is given by

$$\begin{aligned} c_t^{RANK} &= \sum_{h=0}^{\infty} B^h (I-M)^{-1} E_2 \kappa (1-\xi) g_{t+h} = \\ &= \frac{\tilde{\sigma}\kappa(1-\xi)}{v_H - v_L}(v_H^h - v_L^h) g_{t+h}, \end{aligned}$$

where

$$v_H = \frac{1 + \beta + \kappa\tilde{\sigma} + \sqrt{(1-\beta+\kappa\tilde{\sigma})^2 + 4\kappa\tilde{\sigma}(1+\beta)}}{2},$$

$$v_L = \frac{1 + \beta + \kappa\tilde{\sigma} - \sqrt{(1-\beta+\kappa\tilde{\sigma})^2 + 4\kappa\tilde{\sigma}(1+\beta)}}{2},$$

are the eigenvalues of B and $E_2 = (0,1)'$ is a unit vector.

For TANK, we have the same response plus the first element of C_t, so

$$
\begin{aligned}
c_t^{TANK} &= \sum_{i=0}^{\infty} B^i (I-M)^{-1} \left[E_2 \kappa (1-\xi) g_{t+i} + E_1 \left(\Omega_g \Delta g_{t+i} - \Omega_\tau \Delta t_{ht+i}\right)\right] = \\
&= c_t^{RANK} + \Omega_g g_t - \Omega_\tau t_{ht} + \sum_{i=1}^{\infty} B^{i-1}(I-B)(I-M)^{-1}\left(\Omega_g g_{t+i} - \Omega_\tau t_{ht+i}\right) = \\
&= c_t^{RANK} + \Omega_g g_t - \Omega_\tau t_{ht} + \sum_{i=1}^{\infty} \alpha_i^{g-t} \left(\Omega_g g_{t+i} - \Omega_\tau t_{h+i}\right),
\end{aligned}
$$

where $E_1 = (1,0)'$ and

$$
\alpha_i^{g-t} = \frac{\tilde{\sigma}\kappa \left[v_H^{h-1}(v_H+1) - v_L^{h-1}(v_L+1)\right]}{v_H - v_L}.
$$

This expressions coincide with the ones in the main text under the neutrality assumptions of $o = 0$ and $t_{ht} = t_{rt} = t_t$, as it implies $\tilde{\sigma} = \bar{\sigma} = \sigma/(1-G/Y)$ and $\Omega_g = \Omega_t = \Omega = \frac{\chi\sigma\gamma\mu}{(1-\chi)(\chi\sigma\gamma\mu+1-\delta)}$.

B Incomplete Markets Model

The setup is exactly the same as in A, except for the following differences.

Households. The economy is populated by a continuum of households subject to idiosyncratic income risk and borrowing constraints. Incomplete markets limits their ability to smooth consumption, so they engage in precautionary savings and occasionally hit their borrowing constraints, becoming hand-to-mouth temporarily. Households maximize their expected lifetime utility subject to the budget constraints

$$
C_{it} + A_{it} = \frac{R_{t-1}}{\Pi_t} A_{it-1} + W_t Z_{it} N_{it} + \Phi_{it} - T_{it},
$$

where $Z_{it} N_{it}$ is the effective labor supply, and z_{it} is an idiosyncratic productivity shock following the stochastic process $\ln Z_{it} = \rho_z \ln Z_{it-1} + \varepsilon_{it}$, with $\varepsilon_{it} \sim N\left(\mu_\varepsilon, \sigma_\varepsilon^2\right)$ being i.i.d. over time and across agents.[56] For simplicity, we assume that there are no government bonds in the initial period so $B_{i,-1} = 0$ for all households and consequently they only hold Lucas

[56]The mean of ε is chosen so that $\mathbb{E}_i z_{it} = 1$. In particular, $\mu_\varepsilon = -\sigma_\varepsilon^2 (1-\rho_z)^{-1}/2$.

trees in their portfolios $A_{i,-1} = Q_t X_{i,-1}$. Their optimality conditions are given by

$$C_{it}^{-1/\sigma} \geq \beta \frac{R_t}{\Pi_{t+1}} C_{it+1}^{-1/\sigma}$$
$$Z_{it} W_t = C_t^{1/\sigma} N_{it}^{\gamma}.$$

Aggregation conditions. Aggregate variables are obtained by aggregating individual quantities over all households

$$C_t = \int C_{it}\, di,$$
$$N_t = \int Z_{it} N_{it}\, di,$$
$$A_t = \int A_{it}\, di, ,$$
$$\Phi_t = \int \Phi_{it}\, di.$$

Lucas trees and profits. Dividends of the Lucas trees are taxed, so

$$D_t = \delta(Y_t - T_{dt}).$$

Households receive profits proportionally to their labor income

$$\Phi_{it} = \frac{(Z_{it} N_{it})^{1+o}}{\int (Z_{it} N_{it})^{1+o}\, di} \Phi_t,$$

where the coefficient o measures the convexity of earned profits with labor income. Importantly, households take profits as given, so they do not internalize the effect of their labor supply on earned profits. Similar to equation (2.22), higher values for o imply that low-income households receive relatively less profits, so the income of high-MPC agents will tend to fluctuate more than aggregate income. The neutral case arises with $o = 0$, making individual earnings proportional to aggregate income.

Government. Total taxes raised by the government include the taxes from households and the tax on dividends, so

$$T_t = \int T_{it} di + \delta T_{dt}.$$

Neutral case. Neutrality is achieved by shutting down the redistributive effects of fiscal and monetary policy. First, individual profits are proportional to labor income, so $o = 0$. Second, $\sigma = 1$ so we have log utility. Third, individual taxes are raised proportionally to labor earnings. Thus, individual taxes are given by

$$T_{it} = \frac{(Z_{it}N_{it})^{1+o}}{\int (Z_{it}N_{it})^{1+o}\, di}(1-\delta)T_{wt},$$

where T_{wt} is part of the fiscal policy announcement. Note that agents take taxes as given, and they do not internalize the impact of their decisions on the final taxes they will pay. Finally, dividends and labor income are taxed equally so $T_{wt} = T_{dt} = T_t - (R_{t-1}/\Pi_t - R)\, B_{t-1}$.

Under these assumptions, it is trivial to check that the budget constraints remain unchanged when fiscal policy maintains a balanced budget, $G_t = T_t$. Thus, the only effect of fiscal policy must come through changes in the real rate induced by inflation. However, as proven in Werning (2015), an incomplete markets economy reacts in exactly the same way as if it had a representative agent under our assumptions. Therefore, the response of fiscal policy in HANK when the budget is balanced is exactly the same as in the RANK and TANK models.

Calibration. The calibration for the simulations in all three RANK, TANK and HANK models is given in Table 2.1.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Downward Nominal Wage Rigidity: Fiscal Policy and the Missing Deflation

3.1 Introduction

The Great Recession has raised a lot of interest in the macroeconomic consequences of the zero lower bound. A huge literature has been developed trying to find policy remedies that can alleviate the inefficacy of monetary policy in a liquidity trap. However, standard work-horse macroeconomic models have had a hard time in predicting the evolution of the crisis. As Cochrane (2013) notes, standard new-Keynesian models display extreme behavior in a liquidity trap scenario. Authors such as Eggertsson (2010), Christiano et al. (2011a), and Werning (2012) are some modern examples of these atypical predictions such as deep deflation and recessions, completely unusual policy recommendations or high public spending multipliers. On the other hand, there are other empirical facts such as jobless recoveries that are hard to generate in a standard model.

There are several empirical puzzles still awaiting to be resolved. In this paper I focus on the puzzle of the "missing deflation", that is, the absence of deflation during the Great Recession predicted by the Phillips Curve and the New Keynesian model. It is commonly argued in policy circles and informal disccussions that downward nominal wage rigidity might be the reason behind this lack of deflation. Empirically, there is vast evidence in favour of downward wage rigidities (IMF, 2013; Elsby and Solon, 2019). In light of this evidence, it is argued that firms are unable to decrease prices dramatically if one of the main components of their costs, labor compensation, is rigid downwards.

I build on the idea that nominal wage rigidity could explain the missing deflation. The novelty of my approach is to introduce this feature into a standard cash-in-advance model based on the seminal work by Krugman (1998). I follow the wage rigidities scheme used in Schmitt-Grohé and Uribe (2012) due to its simplicity, but focusing on its link with inflation dynamics. This consists of adding a constraint that nominal wages cannot fall by a certain amount that depends on the slack of the labor market. My findings suggests that it is hard to reconcile the idea that the missing deflation during a liquidity trap can be explained by downward nominal wage rigidities. The model predicts mild inflation episodes only when the liquidity trap is very short-lived. Otherwise, it displays sharp unrealistic deflation and recessions in line with the predictions of standard New Keynesian models. I show these conclusions holds both under deterministic and stochastic liquidity traps.

I also use my steup to study the relationship between downward nominal wage rigidity and fiscal policy. Christiano et al. (2011a), Eggertsson (2011), and Woodford (2011) all find large fiscal multipliers in New Keynesian models under the zero lower bound. I find similar large multipliers in my model, but they arise due to a different mechanism. The aforementioned papers find large multipliers due to the government causing an inflation spiral that induces large changes in the real rate and the economic activity through intertemporal substitution. On the other hand, the large multipliers in my model arise due to the persistence of unemployment in the future. High unemployment today means high unemployment in the future because of pent-up wage deflation. Government expenditures boosts aggregate demand and decreases deflationary pressures. This relaxes the wage rigidity constraint for all subsequent periods, inducing higher economic activity in the present and the future. In other words, government spending "greases the wheels" of the labor market.

This paper is related to several strands of the literature. In a similar spirit to my paper, Daly and Hobijn (2014) show in a more complicated model that adding downward nominal wage rigidity à la Calvo in a DSGE model bends the wage inflation Phillips curve, but they do not study the implications of fiscal policy and do not report different results depending on the duration of the trap. Amano and Gnocchi (2017) introduce downward wage rigidities in a standard New Keynesian models and study the optimal inflation target. Lindé and Trabandt (2019) show that the non-linear solution to a model feautirng price and wage rigidities with strategic complementarities can solve the missing deflation puzzle. Finally, Rendahl (2016) demonstrates that fiscal policy in a search and mathcing model with completely rigid wages can induce large multipliers for the similiar reasons emphasized in my model.

The rest of the paper is organized as follows. Section 2 provides the theoretical framework I use in the rest of the paper. Section 3 describes the response of the economy to different kind of negative natural rate shocks that pushes the economy into a liquidity trap. Section

4 presents the main findings related to fiscal policy. Section 5 concludes.

3.2 The Model

In this section, I develop a cash-in-advance model based on Krugman (1998). The main difference of my model lies in the way I introduce nominal frictions; while Krugman assumes that prices are predetermined, I rather introduce monetary non-neutrality due to downward nominal wage rigidity as in Schmitt-Grohé and Uribe (2012). Thus, involuntary unemployment may arise as a response to a shock that induces a reduction in the growth rate of the full-employment nominal wage.

3.2.1 Households

The economy is inhabited by a large number of infinitely-lived households with preferences given by

$$E_0 \sum_{t=0}^{\infty} \left(\prod_{s=0}^{t} \beta_{s-1} \right) U(c_t), \tag{3.1}$$

where c_t denotes real consumption, $\beta_t > 0$ is the random discount factor which can take values $\beta_L < 1$ or $\beta_H > 1$, and $E_t(\cdot)$ is the expectation operator conditional on information available in period t. Period utility $U(\cdot)$ is assumed to take the functional form

$$U(c) = \frac{c^{1-\sigma} - 1}{1 - \sigma},$$

where $\sigma > 0$.

There are two markets, an asset market and a simultaneous goods/labor market. At the beginning of period t, households observe the realization of the exogenous shocks, receive a monetary injection V_t from the central bank, and pay a lump-sum tax T_t to the government. Second, they trade money balances M_t and nominal bonds B_t in the asset market. Bonds are nominally risk-less, offering a gross nominal interest rate of R_t. Finally, the goods and labor market open. For expositional purposes, following Lucas (1982), every period the representative household splits into a 'worker' and a 'buyer', and they cannot communicate until the market closes. The buyer buys consumption goods in the market using money holdings, subject to the cash-in-advance constraint

$$p_t c_t \leq M_t, \tag{3.2}$$

where p_t is the price level in period t and M_t denotes the money holdings acquired previously in the asset market. The worker supply his labor services inelastically at level $\bar{l}$, normalized to one. However, due to nominal wage rigidities, he is only able to sell the amount demanded by firms, earning a wage $W_t l_t$.

Accordingly, the budget constraint of the household at time t is given by

$$M_t + B_t = W_{t-1}l_{t-1} + (M_{t-1} - p_{t-1}c_{t-1}) + R_{t-1}B_{t-1} + V_t - T_t + \Phi_{t-1}, \tag{3.3}$$

where Φ_t are the nominal dividends from the ownership of the firms, which are taken as exogenously given by the household.

The optimization problem of the household consists in choosing C_t, M_t and B_t to maximise equation (3.1) subject to the budget constraint equation (3.3), the cash-in-advance constraint equation (3.2), and the no-Ponzi game condition

$$\lim_{j\to\infty} E_t \left[\left(\prod_{s=t}^{t+j-1} \beta_s \right) U'(c_{t+j}) \frac{M_{t+j} + B_{t+j}}{p_{t+j}} \right] \geq 0 \tag{3.4}$$

for all $t = 0, 1, 2, ...$ The corresponding optimality conditions are given by the budget equation (3.3) and the cash-in-advance equation (3.2) constraints, the no-Ponzi game constraint equation (3.4) holding with equality, and

$$\begin{aligned}
\left(\prod_{s=0}^{t} \beta_{s-1} \right) U'(c_t) =& p_t(\mu_t + \lambda_{t+1}) \\
\lambda_t =& \mu_t + \lambda_{t+1} \\
\lambda_t =& R_t \lambda_{t+1} \\
\mu_t(p_t c_t - M_t) =& 0 \\
\mu_t \geq& 0,
\end{aligned}$$

where λ_t and μ_t are the Lagrange multipliers associated with the budget and the cash-in-advance constraints respectively.

These conditions lead to three important features of this model. First, the zero lower bound arises naturally in this economy as an implication of the non-negativity constraint for μ_t, given that money holdings always earn a nominal return equal to one. Second, combining the three first conditions, we obtain a typical consumption Euler equation of the form

$$U'(c_t) = \beta_t R_t E_t \left(\frac{U'(c_{t+1})}{\pi_{t+1}} \right) \tag{3.5}$$

where, as I argued before, $R_t \geq 1$ and $\pi_{t+1} \equiv p_{t+1}/p_t$ denotes the gross rate of inflation between period t and $t+1$. Finally, the fourth condition together with $\mu_t > 0$ implies that

$$p_t c_t = M_t \tag{3.6}$$

when the economy is not in a liquidity trap, that is, $R_t > 1$. As we will see shortly, this will allow the central bank to control the price level and the inflation rate when the economy is outside the trap.

3.2.2 Firms

The single consumption good is produced by a large number of identical perfectly competitive firms using a production function of the form

$$y_t = f(l_t) = l_t^{\alpha},$$

where y_t denotes output and $\alpha \in (0, 1)$.

Firms participate in perfectly competitive product and labor markets. They choose l_t to maximise profits after observing prices p_t and nominal wages W_t. Profits take the form of

$$\Phi_t = p_t f(l_t) - W_t l_t.$$

The optimality condition for the firm's problem is

$$w_t = f'(l_t) = \alpha l_t^{\alpha-1}, \tag{3.7}$$

where $w_t \equiv W_t/p_t$ is the real wage. This first-order condition constitutes the demand for labor of the firm. It guarantees that firms are always on their labor demand curve. Thus, equation equation (3.7) will help to determine the unemployment rate of the economy.

3.2.3 Downward Nominal Wage Rigidity

I introduce downward nominal wage rigidity following the wage-setting specification used by Schmitt-Grohé and Uribe (2012). For any given period, nominal wages are allowed to fall at most by a factor $\varphi(l_t) \geq 0$. That is,

$$W_t \geq \varphi(l_t) W_{t-1}.$$

Specifically, I assume that the factor $\varphi(l_t)$ takes the functional form

$$\varphi(l) = \pi l^{\gamma}, \tag{3.8}$$

where π is the inflation rate in the steady state. This setup allows for wage rigidity to be less constraining when unemployment $u_t \equiv 1 - l_t$ increases, which seems reasonable. In addition, it states that in a steady state with full employment, nominal wages are indexed at least to $\varphi(1) = \pi$, that is, the steady-state inflation rate.

The presence of this nominal friction generally prevents the labor market to clear when inflation dynamics requires the real wage w_t to be greater than its full-employment level α. Finally, we assume that the following slackness condition on nominal wages and employment

$$(1 - l_t)\left(W_t - \varphi(l_t)W_{t-1}\right) = 0$$

holds for all t. This requires that whenever $l_t < 1$ and involuntary unemployment arises, the lower bound on nominal wages must be binding. Correspondingly, whenever the lower bound on nominal wages is not binding, the economy must operate at full employment.

3.2.4 Fiscal and Monetary Authorities

The government's period budget constraint is given by

$$G_t + R_t B^g_{t-1} = T_t + B^g_t,$$

where G_t and T_t denote government spending and lump-sum taxes, and B^g_t is the stock of one-period nominally riskless government bonds issued in period t. For the moment, we assume that fiscal policy consists on $G_t = T_t = B^g_t = 0$. I will relax this assumption in Section 4, where I analyze the impact of fiscal policy.

Monetary policy follows an inflation target $\pi^* > 1$. The central bank tries to implement the target through monetary injections V_t to the households that increase the money supply

$$M^s_t = M^s_{t-1} + V_t.$$

It chooses M^s_t according to the following rule

$$M^s_t = c_t p_{t-1}\pi^*.$$

This rule guarantees that $\pi_t = \pi^* = \pi$ whenever the economy is outside the liquidity trap.

This is a straightforward implication given that equation (3.2) holds with equality and $M_t = M_t^s$ in equilibrium. On the other hand, when the economy hits the zero lower bound, the cash-in-advance constraint equation (3.2) is slack and households will hoard cash. This prevents the monetary authority from reaching the inflation target for any M_t^s, as I will show in Section 3.3.

3.2.5 Equilibrium

In equilibrium, the goods and asset markets must clear, and money demand must equal money supply so

$$\begin{aligned} y_t =& c_t \\ B_t =& B_t^g = 0 \\ M_t =& M_t^s. \end{aligned}$$

A competitive equilibrium is therefore defined as a set of processes $\{y_t, l_t, w_t, \pi_t, R_t, M_t\}_{t=0}^{\infty}$ satisfying

$$y_t^{-\sigma} = \beta_t R_t E_t \left(\frac{y_{t+1}^{-\sigma}}{\pi_{t+1}} \right) \tag{3.9}$$

$$R_t \geq 1 \tag{3.10}$$

$$y_t = l_t^{\alpha} \tag{3.11}$$

$$w_t = \alpha l_t^{\alpha-1} \tag{3.12}$$

$$l_t \leq 1 \tag{3.13}$$

$$w_t \geq \frac{\pi}{\pi_t} l_t^{\gamma} w_{t-1} \tag{3.14}$$

$$(1 - l_t) \left(w_t - \frac{\pi}{\pi_t} l_t^{\gamma} w_{t-1} \right) = 0 \tag{3.15}$$

$$p_t c_t \leq M_t \tag{3.16}$$

$$(R - 1)(p_t c_t - M_t) = 0 \tag{3.17}$$

$$M_t = c_t p_{t-1} \pi^*, \tag{3.18}$$

given the exogenous process β_t and the initial condition w_{t-1}, which I assume to be equal to the full-employment real wage α.

The following proposition establishes the existence and uniqueness of the deterministic steady state.

Proposition 4 (Existence and Uniqueness of the Deterministic Steady State)**.** *Suppose $\beta = \beta_L$. Then, there exists a unique non-stochastic steady state. Moreover, at the steady state there is full employment and inflation equals the inflation target π^*.*

Proof. First, condition equation (3.15) implies that $l = 1$. Combining it with equation (3.11) and equation (3.12) we have $y = 1$ and $w = \alpha$. From equation (3.9) we have that $R = \pi/\beta_L$. Combining with equation (3.17) we have two possibilities. If $R = 1$ then $\pi = \beta_L$. But this is impossible since it would violate the no-Ponzi game condition holding with equality since $\lim_{j\to\infty} E_t \left[\beta_L^j U'(1) \frac{\bar{M}^s}{p_t \beta_L^j}\right] > 0$. Then we must have that equation (3.16) holds with equality. Combining with equation (3.18) we obtain that $\pi = \pi^*$ so $R = \pi^*/\beta_L > 1$, which solve conditions equation (3.9) and equation (3.17). $\square$

3.3 The Liquidity Trap

In this section I study the response of the model to a positive discount factor shock. The economy enters in a liquidity trap when the downward pressure on the interest rate R_t makes it hit the zero lower bound $R_t = 1$. First, I study a perfect foresight equilibrium where the future evolution of the shocks is known in period t as in Schmitt-Grohé and Uribe (2012). Next, I contrast my results with those obtained by these authors replacing the strict inflation target for an interest rate feedback rule. Finally, I allow for stochastic duration of the liquidity trap as in Eggertsson and Woodford (2003) to see the implications of uncertainty in response to the same shock used before.

3.3.1 Perfect Foresight

The economy starts at the steady state characterized by Proposition 4. The natural real interest rate is defined as the real rate of interest that would prevail in the absence of nominal frictions, given by β_t^{-1}. Hence, a positive discount factor shock is equivalent to a negative natural rate shock.

In order to study the effects of a temporary liquidity trap, households realize in period t that $\beta_t = \beta_H$ and $\beta_s = \beta_L$ for all $s \geq t$. This positive shock to β_t can be accommodated by the market decreasing R_t. However, the Euler equation equation (3.9) implies that $R_t = 1$ when $\beta_H > \pi^*$. Furthermore, households know that next period $\beta_{t+1} = \beta_L$ and the economy will be out of the liquidity trap, which implies that inflation will be at the target π^*. Combining the equilibrium conditions equation (3.9) and equation (3.11) with $R_t = 1$ and $E_t(\pi_{t+1}) = \pi^*$ we obtain

$$l_t^{-\alpha\sigma} = \beta_H \frac{1}{\pi^*} l_{t+1}^{-\alpha\sigma}, \tag{3.19}$$

where there is no need for the expectations operator since we are working in a perfect foresight environment.

The intertemporal relation equation (3.19) points out to one of the main implications of this model. That is, when the economy enters in a liquidity trap, the adjustment must come through a reduction in output y_t and involuntary unemployment emerges. Moreover, unemployment in period t implies that the real wage increases and $\pi_t < \pi^*$ from equation (3.15) and equation (3.14). Thus, the inflation rate will be bellow the target in spite of downward nominal wage rigidity. This in turn generates persistent unemployment in the future given that the real wage will be above its steady-state level and will converge monotonically from above towards it following the law of motion

$$w_t = l_t^{\gamma} w_{t-1}, \tag{3.20}$$

where equation (3.12) and equation (3.13) imply that $l_t^{\gamma} \to 1$ as $w_t \to \alpha$.

To illustrate the dynamics of the economy set in motion by a temporary time discount factor shock, I simulate a calibrated version of the model. This calibration, summarized in Table 3.1, assumes that a time period is one quarter. I set $\sigma = 2$ as it is a standard value in the literature. I take the standard labor share in the growth literature of $\alpha = 2/3$. I assign values to β_L and β_H to match a real interest rate of 2% per year in the steady state and -4% per year when the economy is in a liquidity trap. That is, I set $\beta_L = 1.02^{-1/4}$ and $\beta_H = 0.96^{-1/4}$. I use the standard inflation target of the Federal Reserve and the European Central Bank of 2% per year to set $\pi^* = 1.02^{1/4}$. Finally, I follow Schmitt-Grohé and Uribe (2012) in the calibration of γ imposing the restriction that at an unemployment rate of 5%, nominal wages can fall at most by 2% every year. The implied value for γ is 0.195.[57]

Figure 3.1 shows the evolution of the model triggered by a temporary rise in the discount factor of just one period. I solve for the exact dynamics of the non-linear model. The model shows high persistence of unemployment to the shock that lasts substantially beyond the end of the trap. The shock generates a slight deflation in period zero, and inflation back to the target afterwards. We can see that the interest rate immediately jumps above its steady state value after the shock disappears. Given that the central bank can reach the target again in period one, the interest rates remains above its steady state value of 4% while the economy keeps growing towards full employment.

In order to check whether these dynamics are specific to my setup, I add to Figure 3.1 the response of the economy when the monetary policy follows a standard Taylor rule of the

[57] This corresponds to the value of γ that satisfies $0.98^{1/4} = \varphi(0.95)$.

form

$$R_t = \max\left\{1, \frac{\pi^*}{\beta_L} + \phi_\pi(\pi_t - \pi^*) + \phi_y \ln y_t\right\}, \tag{3.21}$$

where the parameters ϕ_π and ϕ_y take standard values used in the literature, shown in Table 3.1. The main difference lies in the evolution of inflation. The reason is that the interest rate rule establishes a trade-off between inflation and unemployment. Specifically, the central bank is willing to tolerate some inflation above the target when the output is bellow potential. Thus, households anticipate that inflation will be higher in the future when the shock vanishes and unemployment is positive. This relaxes the downward wage constraint in the future and alleviates the situation already in period zero. Consequently, the same shock that causes a liquidity trap in my baseline model is not big enough to push the interest rate against the zero lower bound when the Taylor rule is in motion.

Thus far, I have characterized the dynamics set off by a short-lived temporary shock to the natural real rate of interest. Next I analyze the consequences of a persistent shock to the discount factor. In particular, I calculate the exact response of the non-linear model to a positive shock on β_t that lasts for 10 quarters as in Schmitt-Grohé and Uribe (2012). Formally, we have that $\beta_t = \beta_H$ for $t = 0, ..., 9$ and $\beta_t = \beta_L$ for $t \geq 10$. The results, using the calibration in Table 3.1, are shown in Figure 3.2. In contrast to the short-lived trap of Figure 3.1, a persistent fall in the natural rate produces a big disinflation and output loss on impact. Subsequent inflation is fairly mild, until it reaches the objective in period 10. The evolution is qualitatively similar under both monetary policy rules. Both the baseline and the Taylor rule models experience sharp deflation in period zero that leads to an increase in real wages and massive unemployment. The Tayor rule generates a slightly worse recession during the trap, but accelerates the recovery after the shock is gone. As in the previous exercise, the interest rate rule allows for inflation to be slightly above the target around period 10, when the natural rate reverts to its steady state value. This allows the economy to recover faster after the trap as the additional inflation helps reducing pent-up wage deflation. The pace of the other variables are quite similar except for the interest rate, that leaves the zero lower bound already in period 6 in the Taylor rule model, right when inflation starts being positive.

In this section I have shown the implications of introducing downward nominal wage rigidity in a cash-in-advance model with a deterministic liquidity trap. The results show that downward rigidity for wages can only induce mild responses of inflation when the trap is very short-lived. Otherwise, the liquidity trap features sharp declines in inflation and employment, in line with standard New Keynesian models. The main qualitative findings for inflation and employmentremain remain unchanged when monetary policy follows a standard interest rate feedback rule.

3.3.2 Stochastic Liquidity Trap

In this section I analyze an environment where the duration of the liquidity trap follows a stochastic process. Specifically, suppose that in period t there is an unexpected negative shock to the natural rate from β_L^{-1} to β_H^{-1} and it reverts back to the steady-state level with a fixed probability q every period. Then, the economy is expected to stay in the liquidity trap for $1/q$ periods. This setup assumes that once the natural rate gets back to its steady-state level, it will remain there forever. Therefore, after the trap, the economy converges to the steady state eventually as the central bank is able to reach the inflation target when $\beta_t = \beta_L$.

I start the analysis adapting the Euler equation (3.9) to the setup I just described. The Euler equation in period t combined with equation (3.11) becomes

$$l_t^{-\alpha\sigma} = \beta_H R_t \left(\frac{q}{\pi^*} \bar{l}_{t+1}^{-\alpha\sigma} + \frac{1-q}{\pi_{t+1}} l_{t+1}^{-\alpha\sigma} \right), \tag{3.22}$$

where $\bar{l}$ and l denote the level of employment when the natural rate reverts and does not revert, respectively. This implies that the economy will fall into a liquidity trap when $\beta_H > \pi^*$, exactly the same condition that I found in the previous section. To see this, assume that that the economy is not in a liquidity trap neither today nor tomorrow, so $l_t = l_{t+1} = 1$ and $\pi_{t+1} = \pi^*$. Then, equation (3.22) collapses to

$$R_t = \frac{\pi^*}{\beta_H}.$$

However, since $R \geq 1$, the economy cannot be at full employment when $\beta_H > \pi^*$.

Next I analyze the implications of having a shock $\beta_H > \pi^*$. First, I substitute away to leave equation (3.22) only in terms of l_t and l_{t+1}. From equation (3.14) and equation (3.15) we obtain that employment tomorrow when the shock reverts and inflation is at the target is given by

$$\bar{l}_{t+1} = l_t^{\frac{1-\alpha}{1-\alpha+\gamma}}. \tag{3.23}$$

The level of employment after the trap depends on the level before the trap due to the pent-up wage deflation we saw in Section 3.3.1. Combining again equation (3.14) and equation (3.15), we get that inflation tomorrow, when the shock does not revert, is given by

$$\pi_{t+1} = \pi^* l_{t+1}^{1-\alpha+\gamma} l_t^{\alpha-1}. \tag{3.24}$$

Thus, inflation during the trap is below the target as there will be positive unemployment.

Substituting (3.23) and (3.24) back into equation (3.22) we obtain

$$l_t^{-\alpha\sigma} = \frac{\beta_H}{\pi^*}\left(q l_t^{\frac{\alpha\sigma(\alpha-1)}{1-\alpha+\gamma}} + (1-q) l_{t+1}^{\alpha(1-\sigma)+\gamma-1} l_t^{1-\alpha}\right). \tag{3.25}$$

This complex expression implicitly defines l_{t+1} as a function of l_t. I solve numerically equation equation (3.25) using the calibration in Table 3.1 for several values of q. The results are displayed in Figure 3.3, where the thickest black line corresponds to the 45-degree degree line. In blue, it shows the implicit laws of motion for reversion probability $q = 0.1, 0.2, 0.3, ..., 0.9$.

There are two important conclusions to extract from the graph. First, for low value of the probability of reverting, namely less than 1/3, there is no fixed point in the law of motion. Second, when there is a fixed point, the slope of the law of motion is above one. This means any level of employment away from the fixed point will tend to explote, so employment must immediately jump to the fixed point and stay there while the discount factor shock is still active. This, combined with the first point, implies there is no rational expectations equilibrium for low values of q, since employment would eventually hit its ceiling. Thus, while the economy remains in the liquidity trap, inflation, employment and all relevant variables stay constant. Table 3.2 reports the annualized values for inflation, output and employment during the liquidity trap for every q in Figure 3.3.

This table clearly shows that for low probabilities of reverting to the positive natural rate, corresponding to short trap durations, we obtain small inflation values bellow the target. However, inflation and output decreases sharply when longer trap durations are considered, consistent with my results from the dterministic trap.

3.4 Fiscal Policy Implications

So far I have studied the implications for inflation of introducing downward nominal wage rigidity in a cash-in-advance model subject to a negative natural rate shock. In the following section, I analyze the implications of wage rigidities for fiscal policy in a liquidity trap.

We have already seen from Section 3.3 that a positive discount factor shock may induce a liquidity trap when it is big enough. In such a scenario, even with a purely temporary deterministic shock, unemployment shows persistence for some periods even beyond the trap. This means that boosts in aggregate demand improve the situation today, but also in the future due to the persistence of unemployment. This opens the door for the government to intervene and get high fiscal multipliers.

Consider, for simplicity, a one period discount factor shock. Equation (3.20) together with equation (3.12) implies that

$$l_{t+1} = l_t^{\frac{1-\alpha}{1-\alpha+\gamma}}. \tag{3.26}$$

This relation nests perfectly with the idea of a inertial labor markets in Rendahl (2016), given that $\frac{1-\alpha}{1-\alpha+\gamma} \in (0,1)$. This paper illustrates large multipliers using this assumption, microfunding it through a matching model. However, it arises naturally in my setup due to nominal wage rigidities in a much simpler model.

Now government can spend money satisfying the equilibrium condition

$$y_t = c_t + g_t,$$

where g_t denotes real public expenditure. Combining the Euler equation and equation (3.26) and assuming $g_{t+1} = 0$ for simiplicity we obtain

$$(y_t - g_t)^{-\sigma} = \beta_H \frac{1}{\pi} y_t^{\frac{\sigma(\alpha-1)}{1-\alpha+\gamma}}. \tag{3.27}$$

Applying the implicit function theorem to equation (3.27) around $g_t = 0$ give us the fiscal spending multiplier

$$\frac{dy_t}{dg_t} = \frac{1}{1 - \frac{1-\alpha}{1-\alpha+\gamma}} > 1.$$

Notably, the fiscal multiplier will be higher when the wage rigidities are stronger. This corresponds to low values of the elasticity of the wage constraint to the level of employment, γ . Intuitively, lower elasticity γ implies that the ability of wages to decrease is less sensitive to the level of employment, making the wage rigidity constraint more stringent. This makes unemployment more persistent due to the inability of real wages to go down quickly, thereby opening the door to large multipliers by affecting future output. The effect on future output can be seen by deriving an expression for the present value of the fiscal multipliers, the so-called cumulative multipliers. Taking equation (3.26) and putting it in terms of output, we find that the multiplier effect for $j > 0$ periods ahead takes the form

$$\frac{dy_{t+j}}{dg_t} = \kappa^j y_t^{\kappa^j - 1} \frac{dy_t}{dg_t} = \kappa^j \frac{y_{t+j}}{y_t} \frac{dy_t}{dg_t} = \frac{\kappa^j \beta_H \beta_L^{j-1} \prod_{k=0}^{j-1} R_{t+k}}{\pi^j} \frac{dy_t}{dg_t},$$

where $\kappa = \frac{1-\alpha}{1-\alpha+\gamma}$. Thus, the cumulative multiplier is given by

$$\sum_{j=0}^{\infty} \frac{\pi^j}{\prod_{k=0}^{j-1} R_{t+k}} \frac{dy_{t+j}}{dg_t} = \left[1 + \frac{\kappa \beta_H}{1 - \kappa \beta_L}\right] \frac{\partial y_t}{\partial g_t} = \frac{1 + \kappa(\beta_H - \beta_L)}{(1-\kappa)(1-\kappa\beta_L)}.$$

Thus, the cumulative multiplier is large, well above one, and exceeding the square of the current multiplier.

It is worth pointing out that the mechanism behind these large multipliers does not rely on the government boosting future inflation, as in Christiano et al. (2011a) and Eggertsson (2011). Inflation throughout the previous experiment always stays constant. The current mechanism works through better employment expectations in the future. Public spending boosts aggregate demand and improves future economic perspectives due to employment persistence. It is useful to remark though, that if the government tried to push the economy above its potential of $l_t = 1$, interest rate would increase and totally crowd out government spending.

3.5 Final Remarks

Downward nominal wage rigidity is arguably a good candidate to complement or even substitute the prevalence of sticky price models, and there is vast evidence in favour of it. In this paper I have studied the implications for inflation of introducing downward nominal wage rigidities in a standard cash-in-advance liquidity trap model. I find that it is hard to reconcile the idea that the missing deflation puzzle might be due to downward nominal wage rigidities.

Overall, the model generates large deflation episodes unless the duration of the trap is extremely short. This result holds under deterministic and stochastic liquidity traps, and different monetary policy rules. Despite the differences in the models, this predictions are in line with standard New Keynesian models. Further research is needed in this promising setup, for example studying the full stochastic model rather than using an absorbing state that simplifies calculations.

I also study the impact of fiscal policy under downward nominal wage rigidities. I find huge fiscal multipliers, considerably larger than one, that increase with the degree of wage rigidity. This result works through a different channel than those usually seen in the literature. Specifically, large multipliers arise due to the persistence of current unemployment in the future. High unemployment today means high unemployment in the future. Because of this channel, government spending pushes up current aggregate demand and improves employment today and in subsequent periods.

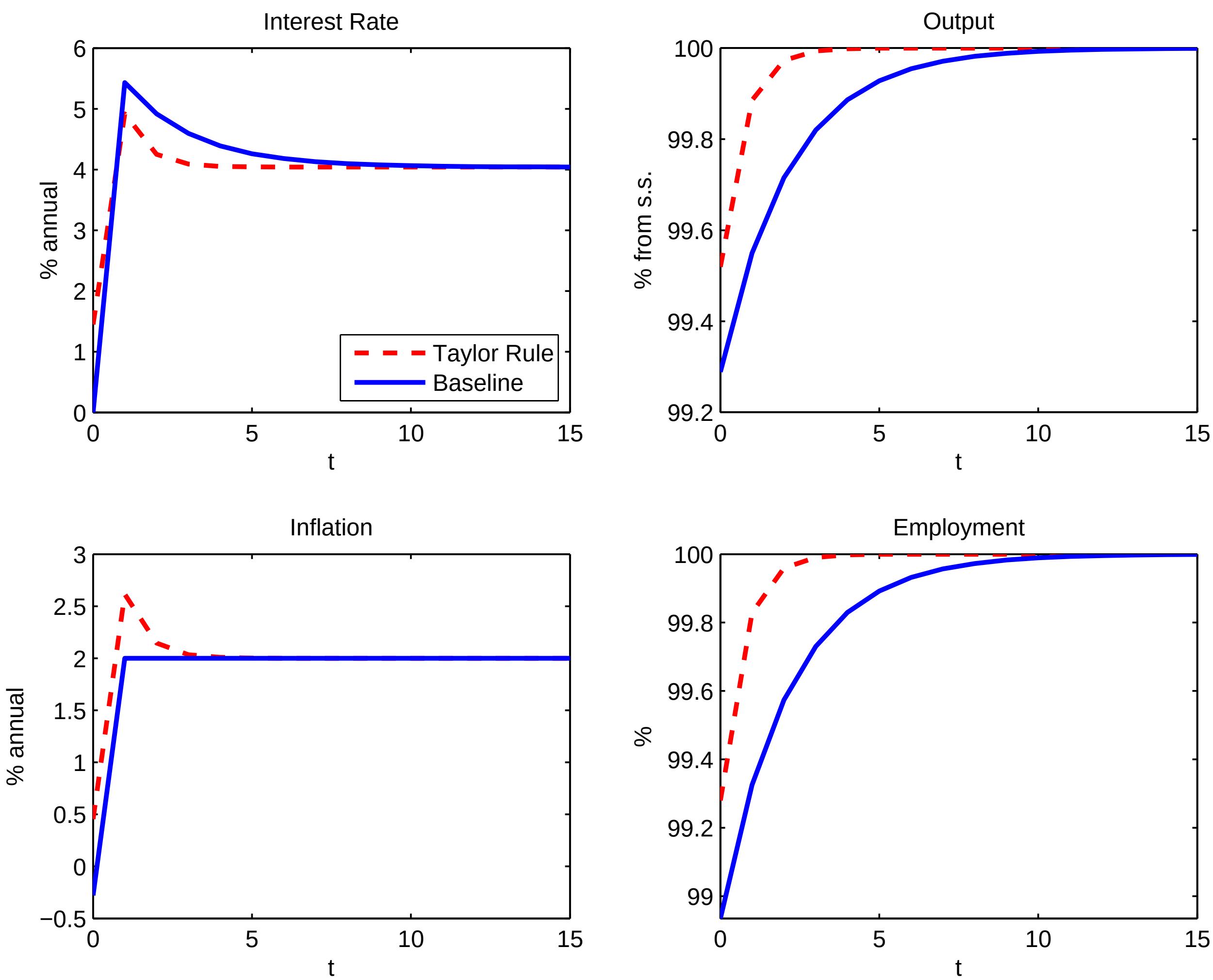


Figure 3.1: Response to a Short-lived Fall in the Natural Real Rate.

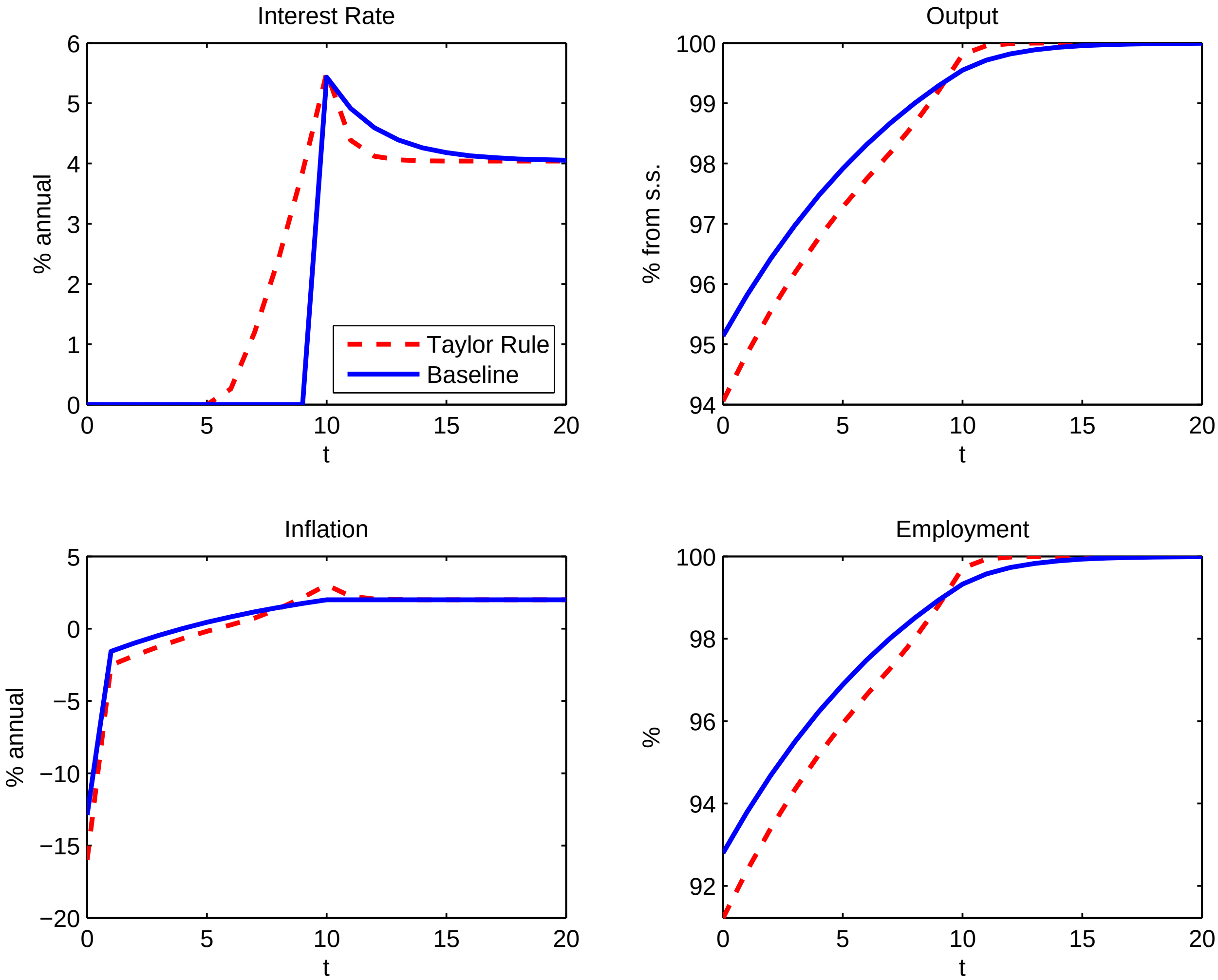


Figure 3.2: Response to a Persistent Fall in the Natural Real Rate.

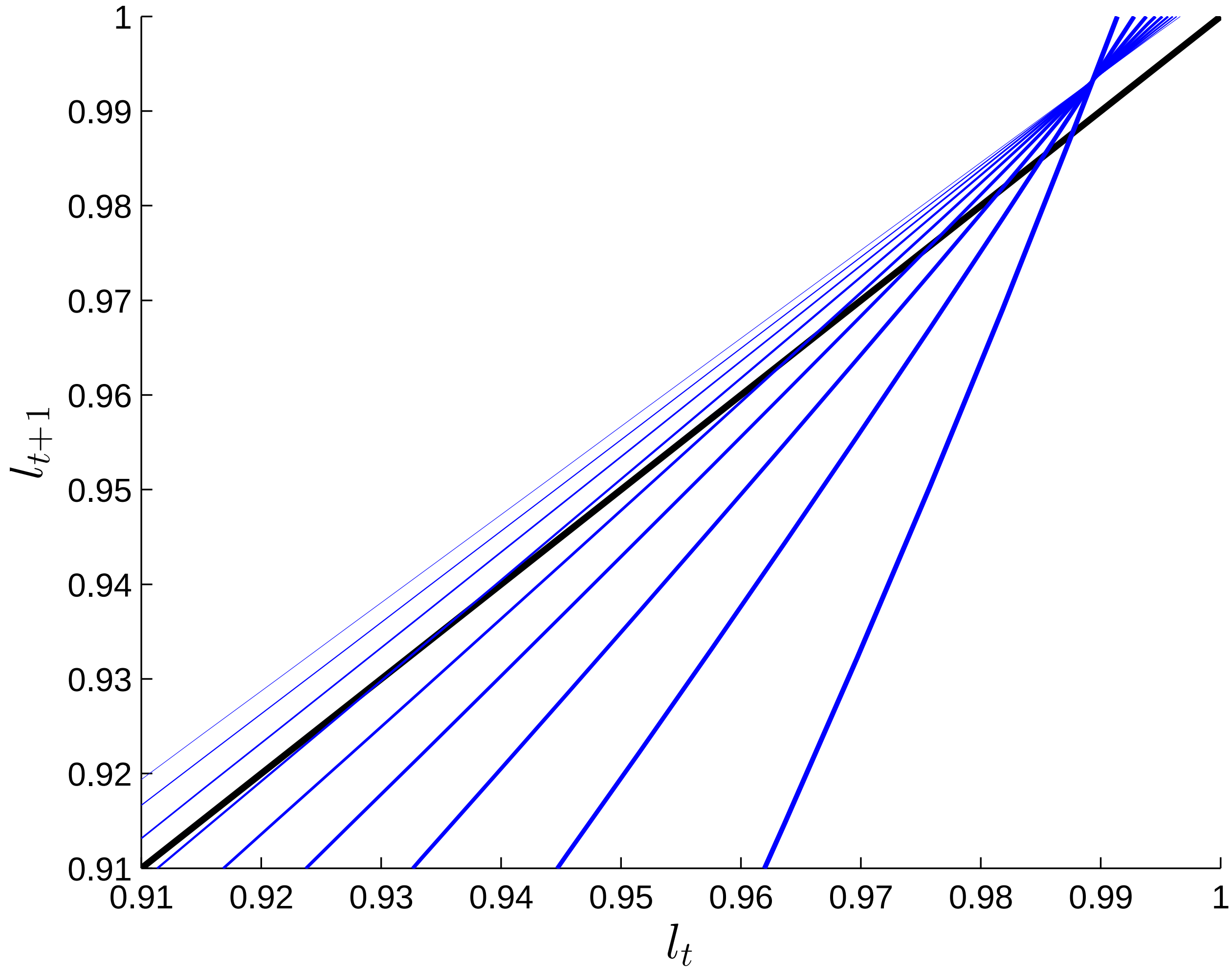


Figure 3.3: Law of Motion for Employment Conditional in a Liquidity Trap.

Note: Thiner lines represent the law of motion for lower probabilities q of leaving the trap.

Table 3.1: Baseline Calibration for Figures.

Parameter	Description	Value
Baseline Model		
α	Labor share	2/3
σ	Coefficient of relative risk aversion	2
β_L	Steady-state discount factor	0.9951
β_H	Discount factor shock	1.0103
π^*	Inflation target	1.005
γ	Elasticity of wage rigidity with respect to employment	0.195
Taylor Rule Model		
ϕ_π	Taylor Rule inflation coefficient	1.5
ϕ_y	Taylor Rule output coefficient	0.125

Table 3.2: Liquidity Trap Variable Values (Annualized %).

q	π_{LT}	y_{LT}	l_{LT}	Expected Duration (quarters)
0.4	-3.36	95.49	93.31	2.5
0.5	-0.82	97.63	96.47	2
0.6	0.08	98.39	97.59	1.67
0.7	0.54	98.78	98.17	1.42
0.8	0.82	99.01	98.52	1.25
0.9	1.01	99.17	98.76	1.11

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

Acemoglu, Daron and Pascual Restrepo (2017) "Robots and Jobs: Evidence from US Labor Markets," Working Paper No. 23285, National Bureau of Economic Research.

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales (2019) "Shift-Share Designs: Theory and Inference," *The Quarterly Journal of Economics*, Vol. 134, pp. 1949–2010.

Alvarez, Fernando, Hervé Le Bihan, and Francesco Lippi (2016) "The Real Effects of Monetary Shocks in Sticky Price Models: A Sufficient Statistic Approach," *American Economic Review*, Vol. 106, pp. 2817–2851.

Amano, Robert A and Stefano Gnocchi (2017) "Downward nominal wage rigidity meets the zero lower bound,"Technical report, Bank of Canada Staff Working Paper.

Angeletos, George-Marios and Chen Lian (2018) "Forward guidance without common knowledge," *American Economic Review*, Vol. 108, pp. 2477–2512.

Arnold, Ivo J.M. (2001) "The Regional Effects of Monetary Policy in Europe," *Journal of Economic Integration*, Vol. 16, pp. 399–420.

Auclert, Adrien (2019) "Monetary Policy and the Redistribution Channel," *American Economic Review*, Vol. 109, pp. 2333–2367.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub (2018) "The Intertemporal Keynesian Cross," Working Paper No. 25020, National Bureau of Economic Research.

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra (2018) "Regional Heterogeneity and the Refinancing Channel of Monetary Policy," *The Quarterly Journal of Economics*, Vol. 134, pp. 109–183.

Bils, Mark and Peter J. Klenow (2004) "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, Vol. 112, pp. 947–985.

Bils, Mark, Peter J. Klenow, and Oleksiy Kryvtsov (2003) "Sticky Prices and Monetary Policy Shocks," *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol. 27, pp. 2–9.

Boivin, Jean, Marc P Giannoni, and Ilian Mihov (2009) "Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data," *American Economic Review*, Vol. 99, pp. 350–384.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel (2018) "Quasi-Experimental Shift-Share Research Designs," Working Paper No. 24997, National Bureau of Economic Research.

Bosch-Domènech, Antoni, José G Montalvo, Rosemarie Nagel, and Albert Satorra (2002) "One, Two, (Three), Infinity, . . . : Newspaper and Lab Beauty-Contest Experiments," *American Economic Review*, Vol. 92, pp. 1687–1701.

Bouakez, Hafedh, Emanuela Cardia, and Francisco Ruge-Murcia (2014) "Sectoral price rigidity and aggregate dynamics," *European Economic Review*, Vol. 65, pp. 1–22.

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

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong (2004) "A Cognitive Hierarchy Model of Games," *The Quarterly Journal of Economics*, Vol. 119, pp. 861–898.

Camerer, Colin, Teck-Hua Hong, and Juin-Kuan Chong (2004) "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, Vol. 119, pp. 861–898.

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

Carvalho, Carlos (2006) "Heterogeneity in Price Stickiness and the Real Effects of Monetary Shocks," *Frontiers in Macroeconomics*, Vol. 6.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber (2013) "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities," *NBER Macroeconomics Annual*, Vol. 27, pp. 1–56.

Chodorow-Reich, Gabriel, Plamen Nenov, and Alp Simsek (2019) "Stock Market Wealth and the Real Economy: A Local Labor Market Approach," Working Paper No. 25959, National Bureau of Economic Research.

Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo (2011a) "When Is the Government Spending Multiplier Large?," *Journal of Political Economy*, Vol. 119, pp. 78 – 121.

——— **(2011b)** "When Is the Government Spending Multiplier Large?," *Journal of Political Economy*, Vol. 119, pp. 78–121.

Cochrane, John H (2013) "The New-Keynesian Liquidity Trap,"Technical report, National Bureau of Economic Research.

Daly, Mary C. and Bart Hobijn (2014) "Downward Nominal Wage Rigidities Bend the Phillips Curve," *Journal of Money, Credit and Banking*, Vol. 46, pp. 51–93.

De Ridder, M. and D. Pfajfar (2017) "Policy Shocks and Wage Rigidities: Empirical Evidence from Regional Effects of National Shocks,"Technical report, Cambridge INET Working Paper Series.

Dhyne, Emmanuel, Luis J Álvarez, Hervé Le Bihan, Giovanni Veronese, Daniel Dias, Johannes Hoffmann, Nicole Jonker, Patrick Lünnemann, Fabio Rumler, and Jouko Vilmunen (2006) "Price Changes in the Euro Area and the United States: Some Facts from Individual Consumer Price Data," *Journal of Economic Perspectives*, Vol. 20, pp. 171–192.

Eggertsson, Gauti B (2010) "The paradox of toil," *FRB of New York Staff Report*.

Eggertsson, Gauti B. (2011) "What Fiscal Policy Is Effective at Zero Interest Rates?" *NBER Macroeconomics Annual*, Vol. 25, pp. 59–112.

Eggertsson, Gauti B. and Michael Woodford (2003) "Optimal monetary policy in a liquidity trap."

Elsby, Michael W. L. and Gary Solon (2019) "How Prevalent Is Downward Rigidity in Nominal Wages? International Evidence from Payroll Records and Pay Slips," *Journal of Economic Perspectives*, Vol. 33, pp. 185–201.

Evans, George (2001) *Learning and expectations in macroeconomics*, Princeton, N.J: Princeton University Press.

Farhi, E. and I. Werning (2016) "Fiscal Multipliers: Liquidity Traps and Currency Unions," in *Handbook of Macroeconomics*: Elsevier, pp. 2417–2492.

Farhi, Emmanuel and Iván Werning (2019) "Monetary Policy, Bounded Rationality, and Incomplete Markets," *American Economic Review*, Vol. 109, pp. 3887–3928.

Fisher, Irving (1933) "The Debt-Deflation Theory of Great Depressions," *Econometrica*, Vol. 1, p. 337.

Floden, Martin and Jesper Lindé (2001) "Idiosyncratic Risk in the United States and Sweden: Is There a Role for Government Insurance?" *Review of Economic Dynamics*, Vol. 4, pp. 406–437.

Gabaix, Xavier (2017) "A Behavioral New Keynesian Model," harvard working paper, Harvard.

Galí, Jordi (2015) *Monetary Policy, Inflation, and the Business Cycle*: Princeton University Press.

Gali, Jordi, J. David Lopez-Salido, and Javier Valles (2007) "Understanding the Effects of Government Spending on Consumption," *Journal of the European Economic Association*, Vol. 5, pp. 227–270.

Garcia-Schmidt, Mariana and Michael Woodford (2019) "Are low interest rates deflationary? A paradox of perfect-foresight analysis," *American Economic Review*, Vol. 109, pp. 86–120.

Gertler, Mark and John Leahy (2008) "A Phillips Curve with an Ss Foundation," *Journal of Political Economy*, Vol. 116, pp. 533–572.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift (2018) "Bartik Instruments: What, When, Why, and How," Working Paper No. 24408, National Bureau of Economic Research.

Golosov, Mikhail and Robert E. Lucas (2007) "Menu Costs and Phillips Curves," *Journal of Political Economy*, Vol. 115, pp. 171–199.

Gorodnichenko, Yuriy and Michael Weber (2016) "Are Sticky Prices Costly? Evidence from the Stock Market," *American Economic Review*, Vol. 106, pp. 165–199.

Grandmont, Jean-Michel (1977) "Temporary General Equilibrium Theory," *Econometrica*, Vol. 45, pp. 535–72.

——— **(1988)** *Temporary Equilibrium: Selected Readings*, Boston: Academic Press.

Hagedorn, Marcus, Iourii Manovskii, and Kurt Mitman (2019) "The Fiscal Multiplier," Working Paper No. 25571, National Bureau of Economic Research.

IMF (2013) "World Economic Outlook," April.

Jordà, Òscar (2005) "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, Vol. 95, pp. 161–182.

Kehoe, Patrick and Virgiliu Midrigan (2015) "Prices are sticky after all," *Journal of Monetary Economics*, Vol. 75, pp. 35–53.

Krugman, Paul R. (1998) "It's Baaack: Japan's Slump and the Return of the Liquidity Trap," *Brookings Papers on Economic Activity*, Vol. 29, pp. 137–206.

Leeper, Eric M. (1991) "Equilibria under 'active' and 'passive' monetary and fiscal policies," *Journal of Monetary Economics*, Vol. 27, pp. 129–147.

Lindé, Jesper and Mathias Trabandt (2019) "Resolving the missing deflation puzzle."

Lucas, Robert E. (1982) "Interest rates and currency prices in a two-country world," *Journal of Monetary Economics*, Vol. 10, pp. 335–359.

Mauersberger, Felix and Rosemarie Nagel (2018) "Levels of Reasoning in Keynesian Beauty Contests: A Generative Framework," *Handbook of Computational Economics*, Vol. 4, Article in press.

McKay, Alisdair, Emi Nakamura, and Jón Steinsson (2016) "The Power of Forward Guidance Revisited," *American Economic Review*, Vol. 106, pp. 3133–3158.

Midrigan, Virgiliu (2011) "Menu Costs, Multiproduct Firms, and Aggregate Fluctuations," *Econometrica*, Vol. 79, pp. 1139–1180.

Molavi, Pooya (2019) "Macroeconomics with learning and misspecification: A general theory and applications," *Unpublished manuscript*.

Nagel, Rosemarie (1995a) "Unraveling in Guessing Games: An Experimental Study," *The American Economic Review*, Vol. 85, pp. 1313–1326.

——— **(1995b)** "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, Vol. 85, pp. 1313–26.

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

——— **(2010)** "Monetary Non-Neutrality in a Multisector Menu Cost Model," *Quarterly Journal of Economics*, Vol. 125, pp. 961–1013.

——— **(2013)** "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications," *Annual Review of Economics*, Vol. 5, pp. 133–163.

——— **(2014)** "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, Vol. 104, pp. 753–792.

——— **(2018a)** "Identification in Macroeconomics," *Journal of Economic Perspectives*, Vol. 32, pp. 59–86.

——— **(2018b)** "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect," *The Quarterly Journal of Economics*, Vol. 133, pp. 1283–1330.

Olivei, Giovanni and Silvana Tenreyro (2007) "The Timing of Monetary Policy Shocks," *American Economic Review*, Vol. 97, pp. 636–663.

Pasten, Ernesto, Raphael Schoenle, and Michael Weber (2019) "The propagation of monetary policy shocks in a heterogeneous production economy," *Journal of Monetary Economics*.

Pischke, Jörn-Steffen (2018) "Wage Flexibility and Employment Fluctuations: Evidence from the Housing Sector," *Economica*, Vol. 85, pp. 407–427.

Plagborg-Møller, Mikkel and Christian K. Wolf (2020) "Local Projections and VARs Estimate the Same Impulse Responses," Working Paper.

Ramey, V.A. (2016) "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*: Elsevier, pp. 71–162.

Rendahl, Pontus (2016) "Fiscal Policy in an Unemployment Crisis," *The Review of Economic Studies*, Vol. 83, pp. 1189–1224.

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

Schmitt-Grohé, Stephanie and Martín Uribe (2012) "The Making of a Great Contraction with a Liquidity Trap and a Jobless Recovery."

Sims, Christopher A. (1994) "A simple model for study of the determination of the price level and the interaction of monetary and fiscal policy," *Economic Theory*, Vol. 4, pp. 381–399.

Stahl, Dale II and Paul W. Wilson (1994) "Experimental evidence on players' models of other players," *Journal of Economic Behavior & Organization*, Vol. 25, pp. 309–327.

——— **(1995)** "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, Vol. 10, pp. 218–254.

Taylor, John B. (1999) "Chapter 15 Staggered price and wage setting in macroeconomics," in *Handbook of Macroeconomics*: Elsevier, pp. 1009–1050.

Tolbert, Charles M. and Molly Sizer (1996) "U.S. Commuting Zones and Labor Market Areas: A 1990 Update," Staff Reports 278812, United States Department of Agriculture, Economic Research Service.

Werning, Ivan (2012) "Managing a liquidity trap: Monetary and fiscal policy."

Werning, Iván (2015) "Incomplete Markets and Aggregate Demand," Working Paper 21448, National Bureau of Economic Research.

Wieland, Johannes F. and Mu-Jeung Yang (2019) "Financial Dampening," *Journal of Money, Credit and Banking*, Vol. 52, pp. 79–113.

Woodford, Michael (2003) *Interest and Prices*: Princeton University Press.

——— **(2011)** "Simple Analytics of the Government Expenditure Multiplier," *American Economic Journal: Macroeconomics*, Vol. 3, pp. 1–35.