Title
Essays on Expectation Driven Business Cycles and Monetary Policy

Author
Huang, Hanguo

Publication Date
2023

Peer reviewed|Thesis/dissertation

Essays on Expectation Driven Business Cycles and Monetary Policy

By

HANGUO HUANG

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

James Cloyne, Chair

Sanjay R. Singh

Alan M. Taylor

Committee in Charge

2023

To my loved ones.

Contents

List of Figures

List of Tables

ABSTRACT

Essays on Expectation Driven Business Cycles and Monetary Policy

This dissertation collects three essays on two hotly debated policy questions during and after the COVID-19 pandemic. Coupled with elevated uncertainty and falling confidence, the pessimism about future productivity and growth had dramatically discouraged business investment before the Great Lockdown. It has been widely accepted that heterogeneity matters for understanding the propagation of business cycles and how to target any policy intervention, but little is known about how firm heterogeneity matters in the propagation of expectation driven business cycles, i.e., cycles driven by news about future fundamentals.

In the first chapter, I use U.S. firm-level data from Compustat combined with identified news shocks from Structural VAR analysis to empirically study the heterogeneous responses of firms' capital accumulation following the sudden arrival of news about future technology. I find that non-dividend payers that are younger, smaller, with lower leverage, higher liquidity, or lower cash flow respond the most to news shocks over a five-year horizon. After taking into account the correlation among different characteristics, liquidity, age, and cash flow remain powerful in predicting the heterogeneity in firm investment. Besides, I also provide suggestive evidence on the transmission mechanism of technology news shocks to investment dynamics through the impact on firm finance. Contrary to the previous macro-level evidence, I find very limited and short-lived increases in firms' market values and share prices following good news. In contrast, firms' book values respond more persistently and prominently. The borrowing of long-term debts responds significantly more for younger, smaller, less indebted, and more liquid firms. The cash flow channel is less important in explaining the heterogeneity in investment response as the response in sales or earnings is short-lived and more homogeneous across firms.

Thanks to the massive policy measures put forward by both fiscal and monetary authorities in the last couple of years, the pandemic recession turned out to be short. But U.S. inflation soon hit a four-decade high record in June 2022 even though the Federal Reserves had taken increasingly bigger steps to raise interest rates. The effectiveness of monetary

policy to stabilize inflation greatly depends on the trade-off between inflation and economic slack. While numerous scholars have documented a weaker trade-off, known as the flattening of the Phillips curve, the precise reasons behind the change are still under debate. In the second chapter, I provide a new explanation and study the state dependence of the slope of the Phillips curve on the trend productivity growth. By merging two longitudinal databases, I present estimates of the "average" New Keynesian Phillips curve (NKPC) for 17 advanced economies across TFP growth regimes since 1890. Following the state-of-the-art method, I estimate the NKPC using trilemma monetary shocks as instruments and find that the Phillips curve is steeper (flatter) in high (low) growth regime. My empirical finding is consistent with the following mechanism: the structural changes that contribute to higher productivity growth could also result in a more competitive market, increasing the price elasticity of demand so that the pass-through of marginal costs from short-run demand changes is larger. This mechanism is qualitatively in tune with the recent secular trends of a flattening Phillips curve and productivity slowdown amid rising market concentration in major advanced economies. The policy implication is that structural reforms that can improve productivity and restore business dynamism help enhance the potency of monetary policy to stabilize inflation in the long run.

Despite decades of research, the consistent estimation of the Phillips curve remains one of the most challenging empirical tasks in macroeconomic studies due to pervasive endogeneity issues. In Chapter 2, I adapted the two-step approach proposed by Barnichon and Mesters (2020) in the estimation of the NKPC and imposed useful restrictions to improve point estimates. In the final chapter, I provide my justification by comparing the simulation performance of different estimators using monetary shocks as instruments. I show that the flexibility of the two-step approach to allow for extra controls in estimating the impulse responses gains its additional advantage over the other IV estimators and imposing range inequality constraints along with long-run restriction brings the point estimates closer to their true values when the rank condition fails. Nevertheless, the state-of-the-art method might be confronted with other pitfalls in the estimation of the Phillips curve, such as the many weak instruments problems and small sample bias, which deserve further research.

ACKNOWLEDGMENTS

First and foremost, words cannot express my gratitude to my main advisor and chair of my committee, James Cloyne, for his invaluable patience, motivation, inspiration, and genuine feedback. He's the best instructor, mentor, and advisor for me, I could not have completed this journey without his constant guidance since the summer before the oral exam. Being his RA after advancing to candidacy got me exposed to immense practical knowledge in conducting empirical research. I am extremely grateful to my advisor Sanjay Singh, who has always been ready to share his experience and expertise to guide me through the toughest stage of my research. I will definitely miss the numerous thought-provoking meetings that have shaped my dissertation. I also would like to express my sincere appreciation to my advisor Alan Taylor, who kindly provided a detailed clarification of the empirical analysis in his joint work with Prof. Òscar Jordà on trilemma monetary shocks, which greatly expanded the depth and breadth of my second chapter. I wish I had asked for his advice and suggestions earlier and more often over the years.

I would also like to thank Professor Òscar Jordà, Andrés Carvajal, Nicolas Caramp, Athanasios Geromichalos, and Colin Cameron. I benefited a lot from their excellent lectures in the first two years and insightful comments on my research projects. I would be remiss in not mentioning the rest of the Economics faculty members at UC Davis, who always made me feel warm and supported. Thanks should also go to my peers and fellow apprentices for their continuous care and moral support. The list includes but is not limited to Haopeng Shen, Mingzhi Xu, Ece Ozge Emeksiz, Yumeng Gu, Ninghui Li, Yijing Wang, Matthew (Kepler) Illich, Shotaro Nakamura, Manho Kang, Jaeyoung Yoo, and Xiangtao Meng.

Above all, I'm deeply indebted to my family. My parents spared no effort in providing the best possible education and the best role models for me, not only supporting me financially but also emotionally. My elder brother, who has been the backbone of the family, relieved my worries and helped me focus on pursuing my dreams abroad. None of this would have been possible without their devotion, which I will never forget!

Chapter 1

Technology News Shocks, Firm Heterogeneity and Investment Dynamics

Beyond the current steep economic contraction, the pandemic is likely to leave lasting scars on the global economy by undermining consumer and investor confidence, human capital, and global value chains.

Global Economic Prospects (June, 2020)

1.1 Introduction

In the Global Economic Prospects about the impacts of the pandemic after its outbreak in early 2020, it was widely believed that COVID-19 had triggered the deepest global recession since the Second World War. "Beyond the staggering economic impacts, the pandemic will also have severe and long-lasting socio-economic impacts that may well weaken long-term growth prospects." Coupled with elevated uncertainty and falling confidence, the pessimism about future productivity and growth dramatically discouraged consumption and investment *before* the Great Lockdown.

Such dramatic changes in economic activities remind me of the discussion back after the Global Financial Crisis. One strand of the macroeconomic literature seems to have shifted the focus from modern business cycle models to the "old" Keynesian model, where pessimism among consumers and investors about the future economy will simultaneously lower aggregate economic activities through an increase in unemployment and lower capital utilization.[1] Although the main driving forces of each crisis might be different, the actions agents take in response to the news about future productivity are probably similar in certain ways.

Originally advanced by Pigou (1927) and recently re-emerged in Beaudry and Portier (2004, 2006), the news-driven business cycles hypothesis posits that business cycles might arise on the basis of expectations of future fundamentals and absent any actual change in fundamentals themselves yet. A collection of macro-level evidence using structural VAR (SVAR) analysis has shown that positive (negative) news about future productivity leads to persistent and prolonged growth (decline) in macro aggregates, although the on-impact responses of investment and hours worked might be slightly the opposite in some cases, see Barsky and Sims (2011). However, little is known about how technology news is propagated across different types of firms in their investment decisions, or the specific mechanisms involved. Besides, understanding how different firms respond to business cycles is also important for understanding who might be most responsive to macro policy interventions. To shed light on this, I use U.S. firm-level microdata from Compustat combined with identified news shocks from macro-level SVAR analysis to empirically study the heterogeneous responses of firms' capital accumulation following the sudden arrival of news about future technology.

Speaking of investment dynamics, it is a commonly held view that financial frictions play a central role in firms' investment decisions in response to macroeconomic shocks, especially in the study of monetary policies. In the news shock literature, financial frictions play an important role in matching theoretical predictions with empirical findings.[2] However, to the best of my knowledge, empirical evidence on the role of financial frictions in the propagation of news shocks to firm investment remains limited. Without taking a stand on

[1] See Krusell and McKay (2010) for more discussions.

[2] On the theoretical side, several papers try to incorporate the impact of news shocks in firm's problem through collateral constraints (e.g., Chen and Song, 2013; Kobayashi et al., 2012; Walentin, 2014, etc) or enforcement constraints on working capital loans à la Jermann and Quadrini (2012) (e.g., Kamber et al., 2017). On the empirical side, several studies have focused on the supply side of the credit market and monetary policy reactions to TFP news shocks, see e.g., Görtz et al. (2019); Görtz et al. (2022).

how different firm characteristics are related to financial constraints, which largely depends on the specific model, I first document empirical evidence on the heterogeneous sensitivity of firm investment to technology news shocks conditional on a range of commonly discussed firm characteristics. Then I assess the bitterness and sweetness of a financial accelerator story in explaining the observed heterogeneity in firms' investment dynamics.

To do so, I employ local projections in the spirit of Jordà (2005*a*) and estimate the heterogeneous sensitivity of firms' physical capital accumulation to the identified news shocks conditional on firm characteristics. In terms of the firm characteristics, I first consider two financial positions that have drawn a lot of attention in the literature, i.e., leverage and liquid asset ratio. I find that over a five-year horizon following good news about future technology, firms with *lower leverage* or *higher liquid asset holdings* prior to the shock exhibit relatively stronger capital accumulation. Quantitatively, in response to a 1 percentage point (pp) positive news shock, which predicts that the aggregate total factor productivity (TFP) will rise by 0.3 pp on average in the next 5 years, a 10 pp lower leverage ratio or a 10 pp higher ratio of liquid assets to total assets predict about 0.13 and 0.3 pp faster capital growth during the two and a half years after the shock. When simultaneously controlling for liquid asset ratio, however, the relevance of leverage becomes less prominent in explaining the differential investment responses. In contrast, the estimates for the liquid asset ratio barely change when conditioning on leverage.

Next, I consider firm age and size as their predictions of the existence of financial frictions are relatively clear in the corporate finance literature. It is widely accepted that young and small firms are more financially constrained than more established large firms. I find that *young* and *small* firms tend to invest more in physical capital over a five-year horizon after the sudden arrival of good news. However, When simultaneously controlling for age, the relevance of firm size almost disappears. Furthermore, cash flow and dividend status are also useful to predict the investment dynamics across firms. Firms with *lower* cash flow to asset ratio are more responsive to technology news shocks. Among the more responsive firm groups, the *non-dividend payers* are responsible for the observed heterogeneity. These findings are largely robust to various robustness checks.

As small and young non-dividend payers with low cash flow tend to be more financially constrained, the main findings are largely in line with the financial accelerator mechanism. The conventional wisdom suggests that financially constrained firms will significantly increase capital investment in response to expansionary shocks. The results regarding the relevance of leverage and liquid asset holdings require careful scrutiny. If we view firms with high liquidity and low leverage as constrained, then the findings that less indebted and more liquid firms respond more to news shock are in tune with the financial accelerator story. If leverage is viewed as a measure of default risk as in Ottonello and Winberry (2020), then my finding is consistent with their argument that expansionary shocks will stimulate the less risky firms to invest. But if firms can't borrow against future improvements in productivity in the presence of financial frictions, why would they respond more today? I argue that the constraint itself is loosened following good news because the firm if then is perceived as more profitable, raising the pledgeability of its future cash flow and asset values to borrow externally.

To understand the transmission mechanism of technology news shocks to firms' investment dynamics, I further explore the heterogeneous responses in different sources of firm finance. I find that in response to a favorable news shock, firms' market values (and share prices) jump up on impact, but they quickly recover and, if anything, even fall thereafter. The short-lived increase in share prices is consistent with the macro impulse response of the stock price index to a patent-based news shock, but greatly contradicts the response to a TFP news shock. This calls into question the stance of the TFP news shock identified based on strict short-run and/or long-run assumptions and casts doubt on the justification of the asset pricing channel based on the macro impulse responses of stock prices.

In contrast, the firm's book equity experiences a significant and prolonged increase over five years following a positive news shock. It turns out that firms that are more responsive in capital adjustment also show a more pronounced increase in net worth. This enables them to have more pledgeability over their investment projects to finance with external funding as they face less severe asymmetric information problems, thus less severe financial frictions. Indeed, I find that the borrowing of long-term debts increases significantly, especially for younger, smaller, less indebted, and more liquid firms. The borrowing of short-term debts increases more uniformly, which combined with the rises in sales and earnings, implies that good news about future technology fosters a consumption boom, sending a signal to firms to expand production. To meet the higher demand, firms borrow more short-term debts to pay for wages and other input costs. But the cash flow channel is less important in explaining the heterogeneity in investment response as the rise in sales or earnings is short-lived and more homogeneous across firms.

Related Literature This paper contributes to several strands of the literature. First, there is a fast-growing empirical literature of studies on firm heterogeneity, financial frictions, and their relevance in firms' investment dynamics in the face of macroeconomic shocks. Monetary policy shocks are often used as a source of aggregate variation to study the heterogeneity in firms' responses as an indication of the presence of financial frictions. Several earlier papers, such as Gertler and Gilchrist (1994), Oliner and Rudebusch (1996) and Bernanke et al. (1996), use firm size as a proxy for the financial constraints they might be facing, and find that small firms are relatively more responsive to contractionary monetary policy changes. Three recent papers closest to this one in methodology are by Ottonello and Winberry (2020), Cloyne et al. (2018), and Jeenas (2019). Ottonello and Winberry (2020) study Compustat firms' capital accumulation responses to monetary policy shocks conditional on three proxies for default risk, i.e., leverage, credit ratings, and distance to default. They find that firms with low default risk – those with low leverage, high credit ratings, and large distance to default – are the most responsive to monetary policy shocks. Cloyne et al. (2018) take a non-parametric approach instead and emphasize the explanatory power of firm age. They find that younger firms paying no dividends exhibit the largest and most significant change in capital expenditure and drive the responses of aggregate investment. Jeenas (2019) evaluates the role of firms' balance sheet liquidity in the transmission of monetary policy to investment. He finds that both firms with higher leverage or lower liquid asset ratio reduce investment relative to others, but controlling for liquid assets, leverage loses its explanatory power in explaining such heterogeneity while liquid asset holdings remain important when conditioning on leverage. This paper adds to this strand of literature by studying the impact of a quite different type of shock – news about future productivity – as a source of aggregate fluctuation.

Compared to the conventional monetary policy shock and other macroeconomic shocks, such as government spending shock, tax shock, and traditional technology surprise shock, technology news shock distinguishes itself in several ways. First, due to the slow diffusion of new technology, it takes time for the actual production efficiency to fully materialize, but households and firms respond actively upon the arrival of the news about future technology. Therefore, most of the short-run effect of news shock is due to anticipation. Second, news shock has been found to have a persistent macroeconomic impact that may last for about a decade, while most of the shocks mentioned above are transitory and may not necessarily have long-run effects. Third, unlike monetary policy shock, news shock does not directly affect firms' borrowing costs, which may lead to a new and interesting role of financial frictions in accounting for the heterogeneity in investment dynamics. Thanks to these unique features, technology news shock provides us with a new environment to study the role firm heterogeneity played in the investment response to shocks.

Second, to study the dynamic effects of the technology news shock on firm investment, we need a series of identified news shocks. This allows us to evaluate the micro-level performances of different identification strategies proposed in this literature. Earlier studies have almost exclusively relied on exploiting the movements in utilization-adjusted TFP by Fernald (2014*a*). In the original paper of Beaudry and Portier (2006), they present two orthogonalized moving average representations for TFP and stock prices: one based on an impact restriction and one based on a long-run restriction and find that they are highly correlated and constitute one of the main drivers of business cycles. Beaudry and Lucke (2010) suggest using a combination of short-run and long-run restrictions to identify news shock and find that news shock dominates technology surprise shock, monetary policy shock, and preference shock in explaining economic fluctuations.[3]

[3]The news shock is identified to be orthogonal to measures of TFP and the relative price of investment on impact but unrestricted in the long run. They also work with an alternative identification scheme that imposes fewer short-run restrictions and relies more on long-run restrictions.

More recently, Barsky and Sims (2011) proposed an alternative identification strategy closely related to the maximum forecast error variance approach by Francis et al. (2007). They define the technology news shock as the shock that does not affect TFP in the short run but drives most of its variations *over* some long horizons. Their partial identification scheme can be conducted on a system with any number of variables without imposing auxiliary assumptions about other shocks. By applying their identification strategy, they find that a good realization of news shock is associated with an increase in consumption but declines in output, hours worked, and investment on impact. After impact, aggregate variables largely track predicted movements in measured TFP. These are exactly what the standard business cycle models would predict. However, several recent papers apply their identification scheme but yield quantitatively and qualitatively different results using different vintages of utilization-adjusted TFP series updated by Fernald (2014*a*).[4] Kurmann and Sims (2021) give a profound discussion about the measurement issues with the TFP series and propose an alternative and arguably more robust identification method based on Francis et al. (2007). They define the news shock as the shock that maximizes the FEV of TFP *at* some long horizon without the zero-impact restriction.

Despite the popularity of using macro-level information sets to identify technology news shock, there are also attempts to refine identification by incorporating information from micro-level data. Miranda-Agrippino et al. (2019) use data on monthly patent applications to construct an external instrument for the identification of technology news shocks in an information-rich VAR. The shock elicits a slow, large, and positive response of quantities, and a sluggish contraction of prices followed by an endogenous easing of the monetary stance. Cascaldi-Garcia and Vukotić (2022) exploit firm-level data on patent grants and subsequent reactions of their stocks to identify technology news shocks. In particular, they make use of the patent-based innovation index constructed by Kogan et al. (2017) and apply Cholesky recursive formulation in a VAR with the innovation index ordered first. The patent-based identification stands out in several aspects against the TFP-based news shocks. For example, unlike TFP news shocks, the patent-based news shocks generate a pos-

[4] Among others, Chen et al. (2018) apply their approach and find that output and hours worked will *increase* in response to positive TFP shock using 2015 vintage of TFP series.

itive response in inflation and the federal funds rate, which is consistent with the standard New Keynesian model without additional features.[5] As a result, I study firms' investment responses to the patent-based news shocks in the baseline. Results are qualitatively robust if we switch to TFP news shocks. As far as I know, this is the first paper to look at how firm-level financial variables respond to the news shocks identified using different approaches based on a similar macro information set. The micro-level evidence can shed light on the plausibility of different identification strategies as public firms account for more than 50% of aggregate business investment.

Third, this paper also speaks to the relevance of different proxies for the financial constraints that firms are facing in their investment decisions when the economy is hit by aggregate shocks. The existing literature has proposed a number of factors to measure financial friction, but the findings about their predictive powers are mixed. Hadlock and Pierce (2010) cast doubt on the validity of KZ index (Kaplan and Zingales, 1997) as a measure of financial constraints, and offer mixed evidence on the validity of other common measures in predicting constraints. One of their key results that catches my attention immediately is that firms with more cash are actually more likely to be constrained.[6] They also find that firm size and age are particularly useful predictors of financial constraint levels, and propose a measure based solely on them called the SA index.

This paper will *not* take a stand on this ongoing debate, but I borrow the insights that many of these proxies are correlated, and this paper attempts to find relatively stable predictors in characterizing the heterogeneous investment response across firms. Most related to this paper is Chen and Song (2013). They analytically show that variations in financial frictions in response to news shock can trigger aggregate TFP fluctuations before the actual technology change is materialized. Using the Compustat dataset, they empirically find a significant countercyclical pattern for capital misallocation, measured by the relative capital productiv-

[5] For TFP news shock to general business cycle comovement, several papers have attempted to modify standard RBC models to match empirical impulse responses, e.g., Beaudry and Portier (2004, 2006); Christiano et al. (2008); Jaimovich and Rebelo (2009); Kobayashi et al. (2012) and Jermann and Quadrini (2012), etc.

[6] Their findings on cash holdings is consistent with the notion that constrained firms store cash for precautionary reasons, a hypothesis that has prior support in the literature, see the reference therein.

ity of financially constrained to unconstrained firms.[7] This paper takes a different approach and directly studies the heterogeneous effects of technology news shocks on firms' capital accumulation by different proxies for financial frictions.

The rest of the paper is organized as follows. Section 1.2 reviews the identification strategy of patent-based news shocks, followed by a discussion of the macro-level impulse responses and a comparison with TFP news shocks. In Section 1.3, I present the data, the construction of firm-level capital stock and firm characteristics, the empirical specifications employed, and the average effect of news shock on firm-level investment. Section 1.4 presents the estimates of the heterogeneity in investment response to a patent-based news shock by firm characteristics. In Section 1.5, I conduct an extensive set of robustness checks. In Section 1.6, I investigate how technology news shock affects firm finance that leads to the observed heterogeneity in investment dynamics across firms. Section 1.7 concludes.

1.2 Identification of Patent-Based News Shocks

The news shock literature generally defines technological news shocks as advanced information about technological progress that will materialize in the future with some delay due to technological diffusion. In this section, I briefly review the identification strategy of patent-based news shocks proposed by Cascaldi-Garcia and Vukotić (2022), followed by a discussion of the macro-level impulse responses. In Appendix 1.8.1, I illustrate two representative identification strategies of TFP-based news shocks. The section ends with a comparison between patent-based and TFP-based news shocks.

1.2.1 Identification Strategy

Unlike the earlier identifications that rely on exploiting the movements in the utilization-adjusted TFP by Fernald (2014*a*), the patent-based news shocks are identified using a direct aggregate measure of technological innovation – patent-based innovation index by Kogan et al. (2017) (KPSS, henceforth). The KPSS innovation index is constructed by combining rich firm-level data on patents issued to US firms during the 1926-2010 period with firm stock price movements. They estimate the private economic value of each patent by exploit-

[7] Chen and Song (2013) classify firms into constrained and unconstrained groups using the firm size and SA index.

ing movements in stock prices following the days that patents are issued to the firm. They document larger trading activity and more volatile returns in the stock of patent-granted firms on patent grant days, suggesting that valuable information is released to the market. They then filter the component of firm return that is related to the value of the patent from noise by making several distributional assumptions. The aggregate innovation index is calculated as the sum of the private economic value of patents granted each year to the firms, scaled by aggregate output.

The original KPSS innovation index is in an annual frequency, Cascaldi-Garcia and Vukotić (2022) are able to construct a quarterly measure and then identify the patent-based news shocks from a structural VAR analysis. The information set of the VAR system consists of real macroeconomic variables along with forward-looking variables in levels, which is quite standard in the news shock literature. The patent-based news shock is then identified under a conventional lower triangle Cholesky decomposition, with the patent-based innovation index ordered first. Besides, the benchmark VAR system also includes utilization-adjusted TFP, real output, consumption, investment, hours worked, inflation, federal funds rate, consumer confidence, and the S&P stock price index. The sample period ranges from 1961:Q1 to 2010:Q4, with the starting point restricted by the data availability of the consumer confidence index. The model contains 4 lags and an intercept term with no time trend. To deal with a large number of coefficients, the VAR is estimated after imposing Minnesota priors.[8]

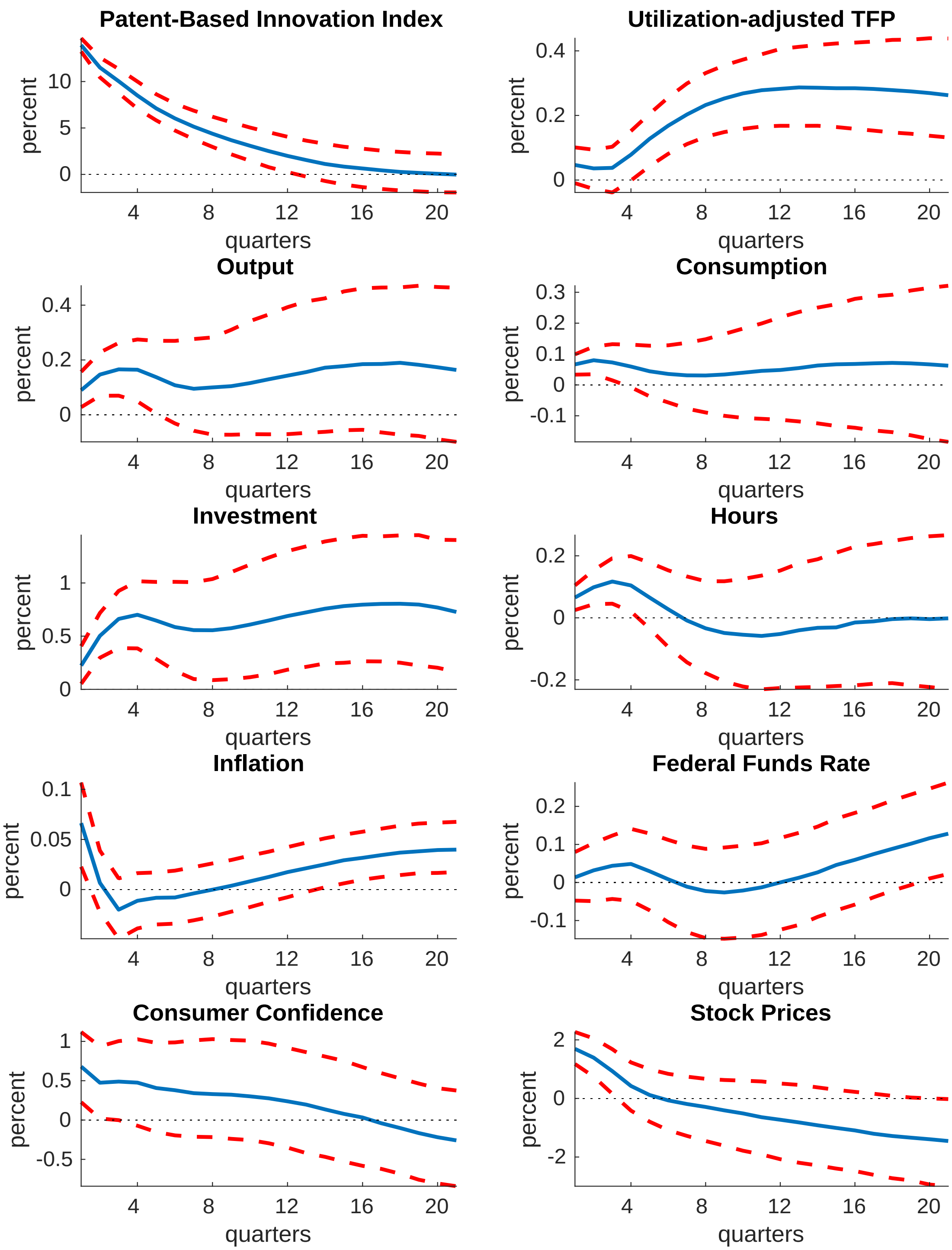
1.2.2 Impulse Responses to News Shocks

Figure 1.1 shows the impulse responses of macroeconomic variables included in the VAR system to a positive patent-based news shock. In response to a one percentage point increase in the patent-based news shock, which amounts to a 14% increase in the patent-based innovation index, the utilization-adjusted TFP rises insignificantly on impact and declines for about a year before it slowly picks up and reaches a new and higher level. The peak response occurs in the third year for an almost 0.3% increase.

The patent-based news shock induces a clear co-movement in output, consumption, investment, and hours in the short run. In particular, they all rise on impact and display hump-

[8] For more details, see Cascaldi-Garcia and Vukotić (2022) and their data appendix.

Figure 1.1. Impulse responses to patent-based news shocks



Notes: The solid lines are the estimated impulse responses to a patent-based news shock and correspond to the posterior median estimates. The dashed lines represent 1 standard deviation coverage bands computed using 1,000 draws from the posterior distribution. The sample period is from 1961:Q1 to 2010:Q4.

shaped responses. Output increases by about 0.1% on impact and stays higher over a five-year period, but the coverage bands do not rule out a zero long-run effect. The increase in consumption is limited and less prominent, with a peak increase of less than 0.1% in the second quarter. Most of the output response comes from the investment response. Investment rises slightly on impact, accelerates for about a year, and stays significantly higher for at least another four years. The hump-shaped response of investment reaches its early peak at the end of the first year before TFP reaches its new level. Such significant and persistent response in investment draws my attention to the strong anticipation effects of news shock, which is worthy of further investigation at both household and firm level.[9] The response of hours worked is positive on impact but soon becomes around zero in less than two years.

Interestingly, the response of inflation is *positive* on impact but quickly drops to around zero in the first year. Starting from the third year, inflation picks up above zero again and keeps rising at a slow pace. The federal funds rate responds similarly except that its coverage bands do not rule out a zero on-impact response. Consumer confidence rises on impact and slowly recovers over the five-year period. The response of stock price, another forward-looking variable, is positive on impact but is short-lived. The stock market boom only lasts for a year and the coverage bands seem to suggest that the stock market boom will be followed by a bust after the fifth year.

1.2.3 Advantages over TFP News Shocks

Compared to TFP-based identification, the advantage of patent-based identification is at least threefold. First, the impulse responses of TFP to the identified news shocks in Figure 1.1 closely fit into the definition of technological news shocks without imposing restrictive assumptions. Following a positive shock to the patent-based innovation index, the on-impact response of TFP is positive but not statistically significant, it decreases a little for about a year before it slowly picks up and converges to a new and higher level. Unlike Barsky and Sims (2011), the insignificant on-impact response of TFP is not due to zero-impact restriction. The short-run negative impact of news shocks is also found to be present in Kurmann

[9] The investment series in Cascaldi-Garcia and Vukotić (2022) is defined as the sum of the real personal consumption expenditure on durables and gross private domestic investment.

and Sims (2021) after removing the zero-impact restriction.[10] The patent-based identification still differs from Kurmann and Sims (2021) in that the latter imposes the assumption that the identified news shock should explain most of the FEV of TFP at a long horizon. Without imposing any of these assumptions, the patent-based news shock explains almost no short-run variations in TFP, while it accounts for a substantial part of the longer-run variations in TFP.[11]

Second, patent-based identification helps circumvent a couple of empirical puzzles in the TFP news shock literature. After Barsky and Sims (2011), several studies have pointed out the sensitivity of news shock identification to data vintages of TFP measure. Based on a different and direct measure of technological improvements, patent-based news shock is identified independent of the movements in TFP and thus should be immune to the measurement issues in TFP. More importantly, most of the empirical literature based on TFP news shock shows a persistently negative response of inflation following a favorable news shock, while we no longer find such a "disinflation puzzle" with patent-based news shock, which is consistent with a New Keynesian model without relying on additional features. Barsky and Sims (2009) show that on-impact inflation response becomes negative if real wage rigidity is high enough or the weight on output growth in the monetary policy rule is large enough. Görtz et al. (2022) provide a discussion of the inflation responses to TFP news shock in a dynamic stochastic general equilibrium (DSGE) model under financial friction. They show that inflation can increase or decrease on impact depending on whether the short-term increase of real marginal cost due to the initial outburst of activity in anticipation of the future increase in productivity outstrips or falls short of the medium- to the long-term decline in real marginal cost after productivity has improved.

Third, the patent-based news shock is identified based on the patent-based innovation index, which is calculated as the sum of the private economic value of patent grants to firms from different industries. By breaking up the aggregate innovation index into industry

[10] Cascaldi-Garcia and Vukotić (2022) attribute the initial decline to the force of creative destruction documented by Kogan et al. (2017) at the firm level. Innovation by competing firms has a negative effect – either directly through business stealing or indirectly through movements in factor prices.

[11] Cascaldi-Garcia and Vukotić (2022) show that the patent-based news shocks account for less than 2% (median) of the FEV of TFP at the 4th quarter, but more than 15% (median) at the 20th quarter.

indices and replacing it with an industry index in the SVAR analysis, we can parallelly construct sector-specific patent-based news shocks. Cascaldi-Garcia and Vukotić (2022) find that the news shock from the manufacturing sector is most relevant for explaining the observed aggregate movements, followed by business services which account for 93% of all services patents. Table 1.1 shows the summary statistics along with a correlation matrix of the patent-based news shock at both aggregate and sector levels. The mean and standard deviation of all three series are close to zero and one. The pairwise correlation coefficients show that manufacturing news is the main driver of the aggregate news, and is also highly correlated with business services news. In the firm-level analysis, I further look into the heterogeneous investment response to sector-specific news shock and see if the heterogeneity by firm characteristics depends on the source of news.

Table 1.1. Summary statistics of patent-based news shocks

(a) Marginal distribution

Variable	**Obs**	**Mean**	**Std. Dev.**	**Median**	**95th Percentile**
Aggregate News Shocks	196	0.004	0.979	0.064	1.415
– Manufacturing	196	0.004	0.979	0.025	1.497
– Business Services	196	0.004	0.979	0.007	1.408

(b) Correlation matrix

	Aggregate News	Manufacturing News	Business Services News
Aggregate News	1.000		
(p value)			
Manufacturing News	0.986	1.000	
	(0.000)		
Business Services News	0.759	0.686	1.000
	(0.000)	(0.000)	

Notes: The definition of sectors: manufacturing sector includes industries with 2-digit SIC code in the range of [20, 39], business services sector includes industries with 2-digit SIC code of 73. Sample period: 1962:Q1 - 2010:Q4. Shock series are obtained from Cascaldi-Garcia and Vukotić (2022).

These advantages of patent-based identification over TFP-based ones convince me to pick the patent-based news shock as the baseline shock measure. However, we need to understand that patent-based identification is not perfect either. First and foremost, not all innovation activity is patented, and the patent-based innovation index only considers publicly listed firms. Therefore, the baseline results should be treated as the minimum effects of news shocks. Second, the patent-based innovation index does not fully capture the knowledge spillover effects and business steeling effects of patents on other firms and industries. In Appendix 1.8.1, I also illustrate the identification strategies of TFP news shocks and repeat the firm-level analysis with TFP news shock measures.

1.3 Empirical Strategy

As evident in the previous section, the hump-shaped and persistent responses of aggregate investment draw much of my attention. From now on, I will look into the firm-level evidence and explore the potential heterogeneity in physical capital investment across firms. To do so, I first merge the firm-level financial variables with the identified patent-based news shocks. Then I employ the panel local projection technique to measure how the semi-elasticity of cumulative physical capital growth with respect to the news shock depends on various firm characteristics. Following the literature on investment dynamics, I mainly focus on the firm characteristics related to financial frictions, including leverage, liquid asset ratio, firm age, size, cash flow to asset ratio, and dividend status.

1.3.1 Firm-Level Data, Sample Selection, and Variable Description

The firm-level financial variables are drawn from the quarterly Compustat dataset. The main firm-level variables of interest include capital stock, firm characteristics, and other firm-level controls such as real sales growth and current assets as a share of total assets. The definition of each variable closely follows Data Appendix 1.8.2.

I follow Compustat's timing convention and denote the capital stock at the end of quarter t as $k_{i,t}$. As is standard in production estimation literature, I construct the measure of physical capital using the perpetual inventory method. Basically, I set the first value of k_{it} to the *deflated* level of gross plants, property, and equipment (*ppegtq*) in the first period in which it's reported in Compustat. From this period onwards, I compute the evolution of k_{it+1} using the accumulation equation:

$$k_{it+1} = k_{it} + (ppentq_{t+1} - ppentq_t) \tag{1.1}$$

which is essentially using the change of *deflated* net plants, property, and equipment (*ppentq*) to substitute for net investment $I_t - \delta k_{jt}$. I deflate *ppegtq* and *ppentq* by the implied price index for gross value added in the U.S. nonfarm business sector following Jeenas (2019).[12] If a firm has a missing observation of *ppentq*, I impute the missing value using linear interpolation. To include more observations in the sample, I do not limit the imputation of *ppentq* to non-missing adjacent values. Due to the persistent effects of news shocks, I allow for long horizons in the panel regression, so it is beneficial to have more observations in each year-quarter cell. In Section 1.5, I show that my baseline results are robust to adjacent imputation, which greatly reduces the sample size.

The firm characteristics that I focus on in this paper are liquid assets holdings, leverage, firm size, age, and cash flow. As the measure of the liquid asset holdings of a firm, following Jeenas (2019), I use the ratio of the Compustat variable *Cash and Short-term Investments (cheq)* to total assets (*atq*). As the measure of a firm's leverage, following Ottonello and Winberry (2020) and others, I use its total debts (*dlcq* + *dlttq*) divided by its total assets, both measured at book values. Firm size is the logarithm of deflated total book assets. Firm age is the number of years preceding the observation year that the firm has a non-missing share price (*prccq*) since 1961. Cash flow is defined to be operating income plus depreciation (*ibq* + *dpq*). In the following analysis, I use the cash flow to asset ratio to control for firm size in the comparison across firms.

Following the sample selection procedure described in Data Appendix 1.8.2, the firm-level dataset consists of more than 354 thousand firm-quarter observations between 1972:Q1-2007:Q3. The sample is restricted to the pre-financial crisis era to avoid potential structural changes in investment dynamics and monetary policies (e.g., zero lower bound) as well as the big swings in patent-based news shock during the Global Financial Crisis.[13] Table 1.2 panel (a) shows the summary statistics of the main firm-level variables of interest. I only include data from firms that are observed for at least 40 quarters in the regressions to improve precision and alleviate issues of endogeneity.

[12]The implied price index for gross value added is downloaded from U.S. Bureau of Economic Analysis (BEA) NIPA Table 1.3.4 Line 3.

[13]The time series of the patent-based news shock is plotted in Appendix Figure 1.16. In Section 1.5, I show that the results are robust to the inclusion of the recent financial crisis.

Table 1.2. Summary statistics of firm-level variables

(a) Marginal distribution

Variable	Obs	Mean	Std. Dev.	Median	95th Percentile
Investment ($\triangle \log k$)	354022	0.013	0.076	0.003	0.118
Liquid asset ratio (*liq*)	354022	0.130	0.170	0.058	0.513
Leverage (*lev*)	335294	0.266	0.290	0.224	0.682
Size (in log)	354022	5.240	2.236	5.210	9.001
Age	354022	13.971	9.583	12.250	33.500
Cash flow ratio (*cfr*)	300884	0.014	0.341	0.022	0.064
Pay dividends or not (*paydvd*)	354022	0.147	0.354	0	1

(b) Correlation matrix

	liq	*lev*	*size*	*age*	*cfr*	*paydvd*
liq	1.00					
(p-value)						
lev	-0.30	1.00				
	(0.00)					
size	-0.23	0.01	1.00			
	(0.00)	(0.00)				
age	-0.16	0.02	0.44	1.00		
	(0.00)	(0.00)	(0.00)			
cfr	-0.01	-0.03	0.03	0.01	1.00	
	(0.00)	(0.00)	(0.00)	(0.00)		
paydvd	-0.12	0.12	0.19	0.10	-0.01	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	

Notes: Liquid asset ratio (*liq*) is measured as cash and short-term investments to assets ratio, leverage as total debt to assets, size as deflated book assets in logarithm; age as the number of years preceding the observation year that the firm has a non-missing share price since 1961, cash flow ratio as the ratio of cash flow to total assets. In the later regression analysis, these variables will be winsorized at 0.5% and 99.5% cutoff. Pay dividend or not is a dummy variable with value 1 if the firm report a non-missing positive dividend. Sample period: 1972:Q1 - 2007:Q3.

Since the sample only contains public firms, the average size is large, about $1431 million over the sample period, while the median is only $112 million. The highly right-skewed size distribution motivates the usage of the logarithm of deflated total assets as the measure of size in regressions. The mean of the liquid asset ratio is about 13% and the mean of the leverage ratio is approximately 27%. Both exhibit considerable variations in the cross-section, with standard deviations of 17% and 29%, respectively. The mean of the firm age is almost 14 years, this is relative to the first quarter of non-missing stock price since 1961, the earliest year Compustat reports quarterly firm balance sheet data. The cash flow to asset ratio is quite small for most firms in the sample with huge variations in the cross-section.

Based on cross-sectional correlations in Table 1.2 panel (b), firms with *higher* leverage ratio also tend to hold *fewer* liquid assets as a fraction of their balance sheet assets. In contrast with Jeenas (2019), in my sample, larger firms tend to have slightly higher leverage but lower liquid asset ratios, although the correlation between size and leverage is quite weak. Also, younger firms tend to hold more liquid assets. As pointed out by Hadlock and Pierce (2010), the fact that a firm chooses to hold a high level of cash may be an indication that the firm is constrained and is holding cash for precautionary reasons. Jeenas (2019) also emphasizes that one must be careful in interpreting the liquid asset holdings as an effective measure of liquidity *per se*. Higher cash flow as a fraction of assets is positively correlated with firm size and age, and negatively correlated with leverage and liquid asset ratio, the correlations are quite weak though. The negative correlation between the liquid asset ratio and cash flow ratio suggests that firms that are not able to generate high cash flow tend to hold more liquid assets, although both are often taken as indicators of how much liquidity a firm carries to pay back debts or costs.

Firm size and age show the strongest pairwise correlation among all characteristics, i.e., young firms tend to be small and small firms are primarily in the early stage of development. Although it is still an ongoing debate in the corporate finance literature about how these financial positions are related to the financial constraints faced by the firms, it is widely accepted that small and young firms face serious financial constraints. Hadlock and Pierce (2010) find that firm size and age are particularly useful predictors of financial constraint levels, and propose a measure of financial constraints that is based solely on these two firm characteristics. They argue that their SA index would appear to be a reasonable choice for measuring financial constraints in many contexts. A firm faces more financial constraints if this index is higher. See Data Appendix 1.8.2 for its definition.

1.3.2 Panel Local Projection Specifications

To estimate how a firm's capital stocks k_{it+h} respond to the news shocks at time t conditional on the firm's financial position prior to the shock $x_{i,t-1}$, I estimate local projections in the spirit of Jordà (2005a). That is, I regress the cumulative difference $\triangle_h \log(k_{i,t+h})$, defined as $\log k_{i,t+h} - \log k_{i,t-1}$, on interaction terms of the firm's financial position at time $t-1$ and the news shock at time t alongside a set of firm-level control variables and fixed effects.

The general form of the baseline specification is as follows:

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x) x_{i,t-1} + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{i,t-1} + e_{ith}, \forall h = 0, 1, \cdots, H \tag{1.2}$$

where α_{ih} denotes firm fixed effect, α_{sth} is the sector-by-quarter fixed effect, ϵ_t^{news} is some quarterly measure of the news shock, $x \in X$ is the set of firm characteristics, and $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter.[14] I have standardized x_{it} over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firm and time to account for correlation within firms and within quarters. Our main coefficient of interest is β_h^x, which measures how the semi-elasticity of cumulative capital growth $\triangle_h \log(k_{i,t+h})$ between t and $t+h$ with respect to the news shocks depends on the firm's financial position $x_{i,t-1}$.

To visually show the differential responses across firm groups classified by different characteristics, I follow Cloyne et al. (2018) and estimate group-specific impulse response functions using the following panel regressions:

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{i,t-1} + \boldsymbol{B}_h(L) \boldsymbol{Y}_{t-1} + e_{ith} \tag{1.3}$$

$\forall h = 0, 1, \cdots, H$, where α_{ih} is firm fixed effect, α_{sqh} is sector-by-quarter seasonal fixed effect, the indicator $I[\boldsymbol{x}_{i,t-1} \in g]$ can be multi-dimensional. $\boldsymbol{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L) \boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level

[14] The definition of sectoral dummies follows Ottonello and Winberry (2020), see Data Appendix 1.8.2 for more details.

controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firm and time. In essence, this is a non-parametric way of estimating the heterogeneous effects of news shocks by multivariate firm characteristics. Without imposing linearity in the interaction, this allows us to estimate conditional impulse response functions as flexibly as possible, which is especially helpful when the firm characteristic is binary or categorical. Different from their baseline specification, I also include other firm-level controls and a few macro-level controls.

To control for outliers that might significantly affect the estimates, I drop investment rate observations below the 0.5th and above the 99.5th percentile. Besides, prior to the estimation for any given h, I winsorize $\triangle_h \log k_{i,t+h}$ and $x_{i,t}$ at 0.5% and 99.5% cutoffs. To eliminate seasonality in the key financial ratios, coming from either the numerator or denominator, I measure them as the past four quarter rolling means instead, except for age. Any reference to a firm's empirical leverage ratio, liquid asset ratio, or cash flow-asset ratio later refers to the corresponding moving average $\frac{1}{4}\sum_{j=0}^{3} x_{i,t-j}$. I conduct the estimation of firms' responses up to the horizon of $H = 20$ quarters to depict the anticipation effects before the news about future productivity actually materializes.

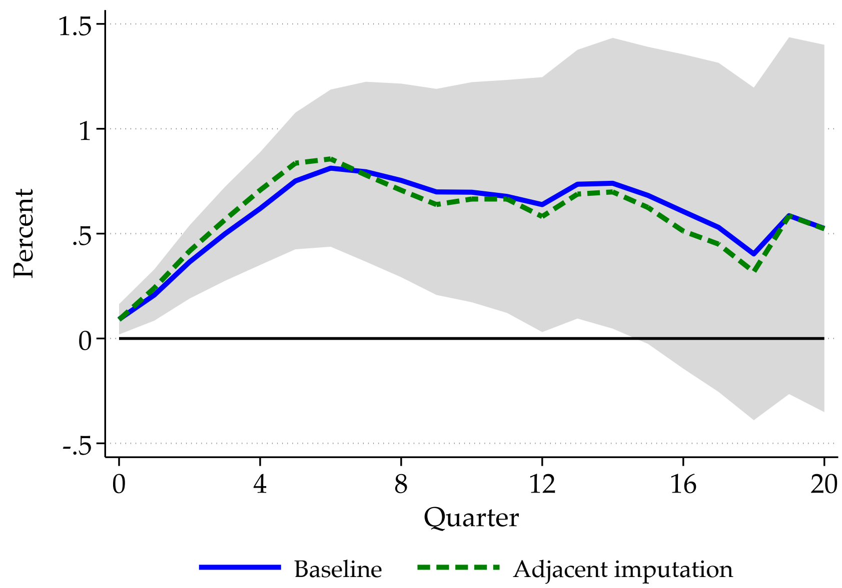
1.3.3 The Average Effect

Before discussing the heterogeneous effects of news shocks on investment across firms, it is useful to show the dynamics of the average effect in the firm-level panel data. The average effect will shed light on the contribution of the average listed firm to the dynamics of investment response based on aggregate data. Besides, the average effect will provide a benchmark against which we can evaluate the differences made by firm heterogeneity.

To estimate the average effect, following Ottonello and Winberry (2020), I replace the sector-by-quarter fixed effect α_{sth} with sector-by-quarter seasonal fixed effect α_{sqh}. The results are robust if we remove α_{sqh}. Also, I include lag polynomials of a vector of macro-level controls such as lagged GDP growth, inflation rate and unemployment rate up to 4 lags. The results are presented in Figure 1.2.

Admittedly, we should not expect the firm-level average effect to be exactly aligned with

Figure 1.2. Average effect of patent-based news shocks on investment



Notes: Reports the coefficient estimates of β_h over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \beta_h \epsilon_t^{news} + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is the news shock, $\boldsymbol{Z}_{it-1}$ is a vector of firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firm and time. The green dashed line is the average effect if we limit the imputation of *ppentq* to non-missing adjacent values, while the baseline further applies linear interpolation when constructing capital stock k.

the impulse response of investment in Figure 1.1 for at least two reasons. First, the investment series in the estimation of macro-level impulse responses measures much broader investment activities than firm-level physical capital investment, which is only part of the nonresidential fixed investment. Second, even within the nonresidential fixed investment, Compustat only includes publicly traded firms, these are larger and more established firms with potentially different financial positions when making their investment decisions. Indeed, the peak response of the average effect is slightly larger and delayed than that in Figure 1.1. Despite the difference in the timing and magnitude of the peak response, the hump-shaped dynamics of the average effect line up with the macro response of investment reasonably well. Regarding the concern about the linear interpolation when constructing firm-level capital stock, the estimated average effect of patent-based news shock if we limit the imputation to adjacent non-missing values almost overlaps the baseline result.

1.4 Heterogeneity in Investment Response across Firms

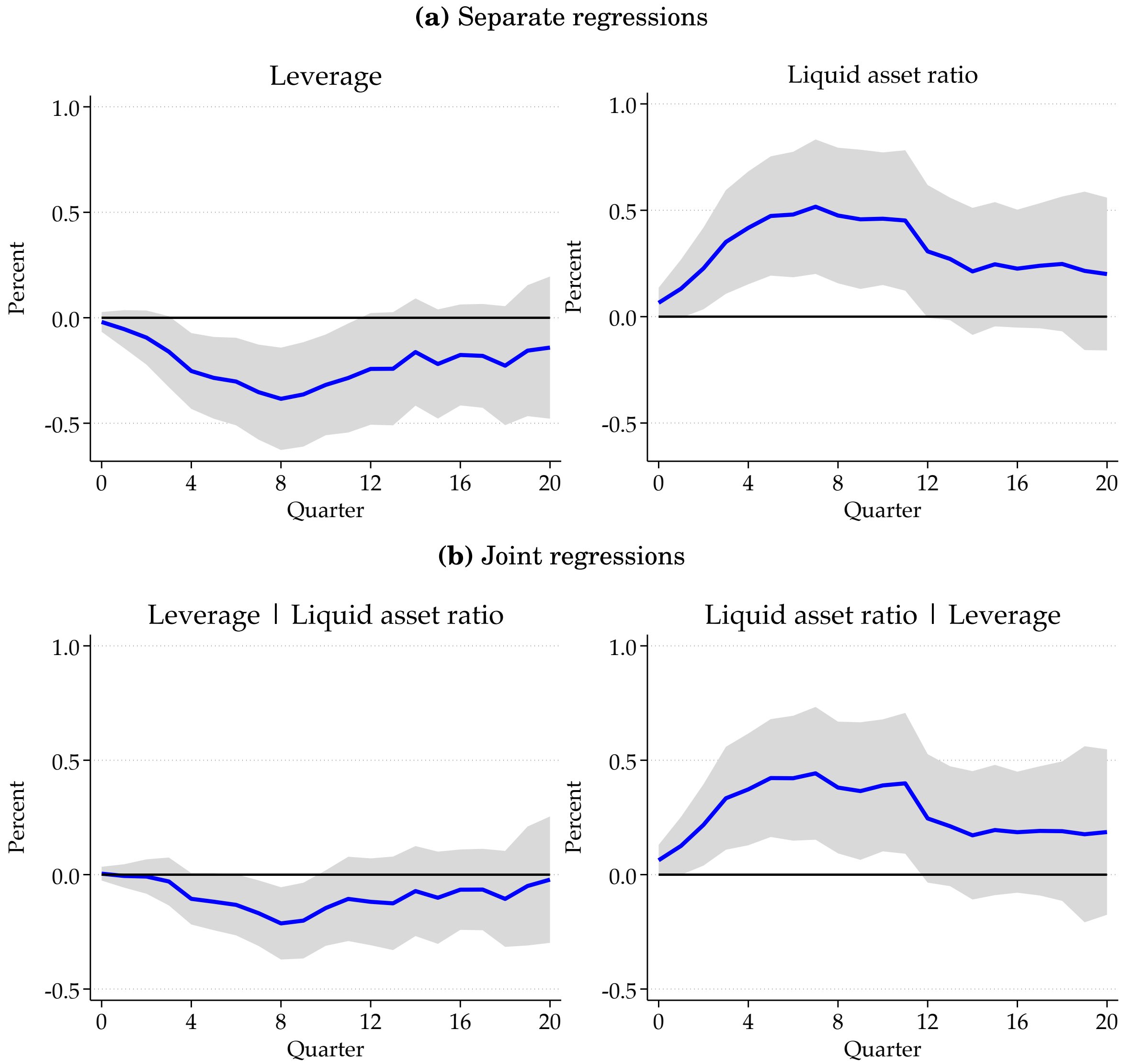
The main goal of this section is to establish which firms are more responsive to technology news shocks, and how well different firm characteristics predict heterogeneity in the response of investment. I start from two hotly debated financial positions, i.e., leverage and liquid asset ratio, followed by two commonly used indicators of financial frictions, i.e., firm age and size, to examine how a firm's capital stocks behave in response to news shock conditional on its financial positions just before the shock. Then I extend the analysis to cash flow and dividend status and test their relevance.

1.4.1 Results Based on Leverage and Liquid Asset Ratio

Figure 1.3 panel (a) shows the estimates for β_h^x from the separate estimation of regression (1.2) with x being either leverage (*lev*) or liquid asset ratio (*liq*). One can see that firms with lower leverage at the time of a favorable news shock tend to experience relatively higher capital growth in the years to follow. The differences become statistically significant starting from 1 year later, start to revert at around the 8th quarter, and stay significant for about 3 years. The differences in fixed capital accumulation are quite persistent and are still present at the end of the fifth year. Quantitatively, the estimates imply that in response to 1 pp favorable news shock, which predicts that the future aggregate TFP will rise by 0.3 pp on average in the next 5 years, one standard deviation *lower* leverage predicts 0.38 pp *higher* fixed capital growth over the two years following the shock. Given the cross-sectional standard deviation of leverage in Table 1.2 panel (a), 10 pp lower leverage (almost one-third of its standard deviation) predicts about 0.13 pp higher fixed capital growth in two years.

Analogously, the right column of the panel (a) shows that firms with higher liquid asset holdings increase their capital stock relative to others after a favorable news shock. The general dynamics are similar to those conditional on leverage. The differences become statistically significant almost immediately after the news, reach a peak 2 years later and stay significant for another year. Quantitatively, the point estimates imply that in response to a

Figure 1.3. Dynamic effects of patent-based news shocks on investment by leverage or/and liquidity



Notes: Panel (a) reports the estimates of coefficient β_h^x over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x) x_{it-1} + \mathbf{\Gamma}'_h \mathbf{Z}_{i,t-1} + e_{ith}$$

where $x \in \{lev, liq\}$, while panel (b) reports the estimates of coefficient β_h^{lev} and β_h^{liq} over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^{lev} \epsilon_t^{news} + \gamma_h^{lev}) lev_{i,t-1} + (\beta_h^{liq} \epsilon_t^{news} + \gamma_h^{liq}) liq_{i,t-1} + \mathbf{\Gamma}'_h \mathbf{Z}_{i,t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sth} is a sector-by-quarter fixed effect, ϵ_t^{news} is the news shock, and $\mathbf{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for fiscal quarter. I have standardized lev_{it} and liq_{it} over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands.

1 pp favorable news shock, a one standard deviation increase in liquid asset ratio predicts an approximately 0.52 pp stronger cumulative growth in the firm's capital stock over the two years after a 1 pp favorable news shock. Given the cross-sectional standard deviation of liquid asset ratios in Table 1.2 panel (a), a 10 pp higher liquid asset ratio (more than half of its standard deviation) predicts about 0.3 pp higher fixed capital growth over the two and half years following the shock.

These results thus show that firms with lower leverage and higher liquid asset holdings are more responsive to technology news shocks for at least 5 years. Yet as shown in panel (b) of Table 1.2 , firms with lower leverage also tend to hold more liquid assets in the cross-section. Following Jeenas (2019), to explore the possibility that the estimates from separate regressions could suffer from omitted variable bias, I include both controls for leverage and liquid assets in estimating specification (1.2). The panel (b) in Figure 1.3 presents the estimates for β_h^{lev} and β_h^{liq} from the joint regressions.

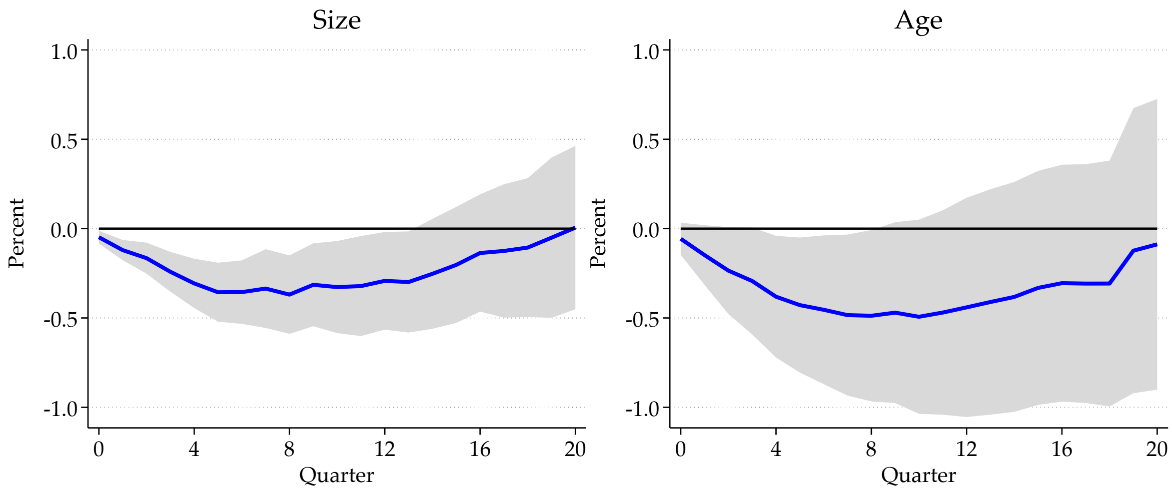
When simultaneously controlling for liquid asset ratio, the relevance of leverage becomes less prominent in explaining the differential responses of firms' capital accumulation to news shocks over the medium run. The peak difference is cut almost in half, that is, a one standard deviation decrease in leverage predicts a 0.21 pp stronger cumulative capital growth in two years. In contrast, the estimates in the right column of the panel (b) indicate that there's no significant change in the explanatory power of liquid asset holdings in characterizing heterogeneity among firms' capital stock responses. In particular, a one standard deviation increase in liquid asset ratio corresponds to a 0.44 pp faster capital accumulation in two years.

1.4.2 Results Based on Size and Age

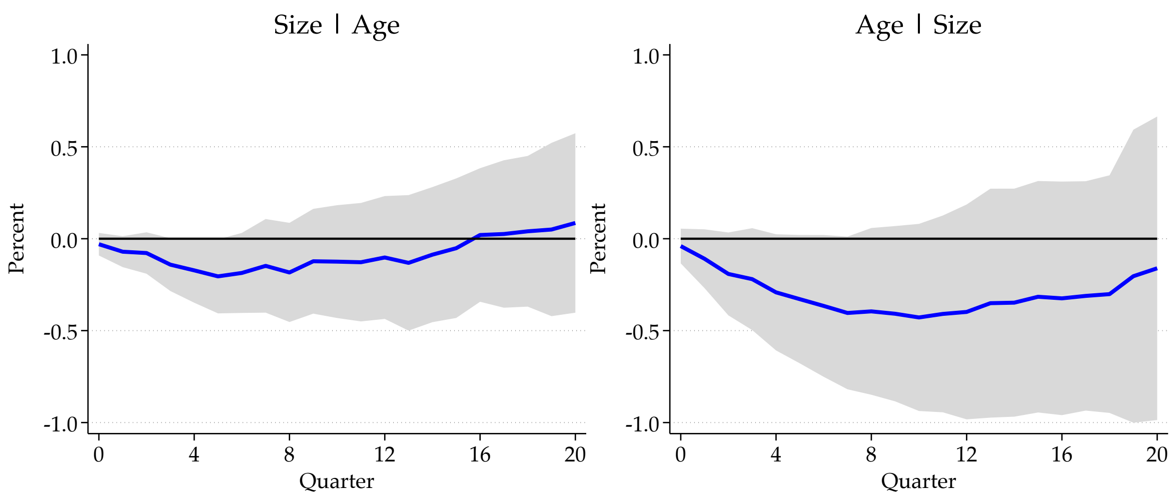
Figure 1.4 panel (a) shows the estimates of β_h^x from the separate regression with x being either firm size or age. One can see that *smaller* firms tend to increase physical capital investment by more in response to positive technology news over a five-year period. The difference is statistically significant on impact and stays significant for at least three years. Quantitatively, firms with one standard deviation smaller size will increase fixed investment by at most 0.37 pp in the 8th quarter after the shock. Similarly, *younger* firms are

Figure 1.4. Dynamic effects of patent-based news shocks on investment by size or/and age

(a) Separate regressions



(b) Joint regressions



Notes: Panel (a) reports the estimates of coefficient β_h^x over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x) x_{it-1} + \boldsymbol{\Gamma}'_h \mathbf{Z}_{i,t-1} + e_{ith}$$

where $x \in \{size, age\}$, while panel (b) reports the estimates of coefficient β_h^{size} and β_h^{age} over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^{size} \epsilon_t^{news} + \gamma_h^{size}) size_{i,t-1} + (\beta_h^{age} \epsilon_t^{news} + \gamma_h^{age}) age_{i,t-1} + \boldsymbol{\Gamma}'_h \mathbf{Z}_{i,t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sth} is a sector-by-quarter fixed effect, ϵ_t^{news} is the news shock, and $\mathbf{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for fiscal quarter. I have standardized $size_{it}$ and age_{it} over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands.

more responsive to technology news shock for the first five years. The dynamics of the semi-elasticities over time are not precisely estimated and are only marginally significant.

As evident from the correlation matrix in Table 1.2, firm size and age present the highest pairwise correlation (0.44) among all the firm characteristics. Smaller firms tend to be also younger in my sample. The joint regression result in panel (b) shows that most of the heterogeneity by size is essentially due to its correlation with firm age. When simultaneously controlling for age, the relevance of firm size almost disappears in accounting for the heterogeneity in investment response to technology news shocks. The point estimates are greatly shrunk and become only marginally significant for the first and a half years. In contrast, the estimates of the difference in semi-elasticities of cumulative capital growth by firm age stay more or less the same.

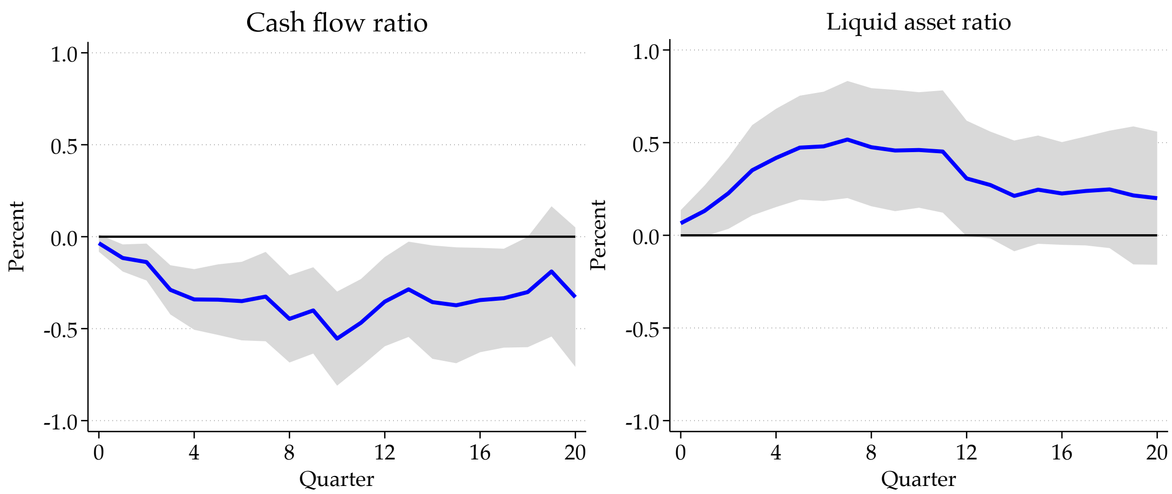
Based on firm size and age, Hadlock and Pierce (2010) propose a size-age (SA) index to measure how financially constrained a firm is. The higher this index is, the more financial constraints the firm is facing. In Appendix Figure 1.17, I present the results using the SA index as the firm characteristics. One can find a parallel pattern in the dynamics of the estimates of β_h^x as those for firm size and age. Firms with higher SA indexes invest more in physical capital in response to favorable technology news over the five-year period.

1.4.3 Results Based on Cash Flow to Asset Ratio

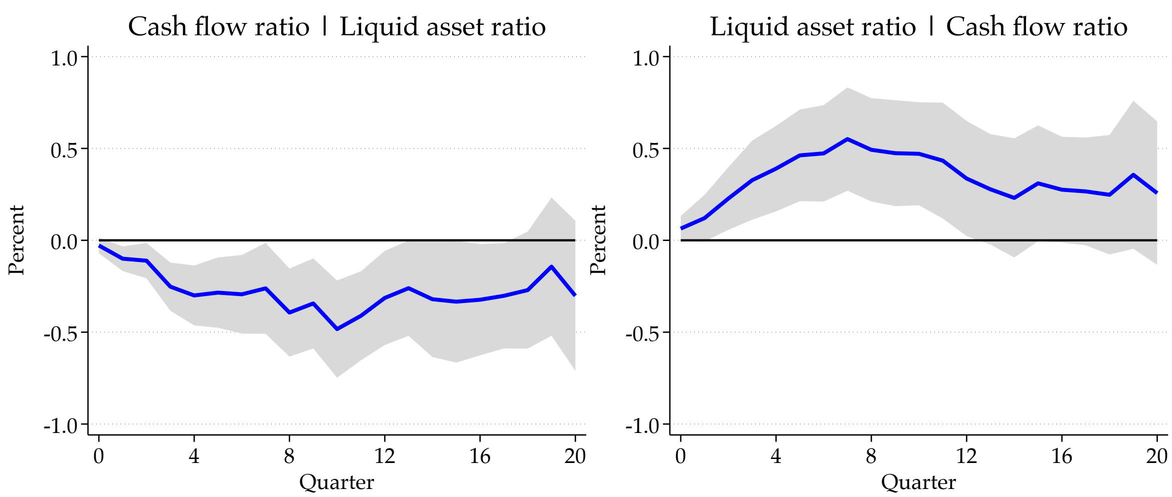
The weak cross-sectional correlations between cash flow ratio and other characteristics in Table 1.2 panel (b) suggest that cash flow to asset ratio could be less subject to omitted variable bias. Figure 1.5 shows the estimates of β_h^x from the separate regression, and β_h^x and β_h^{liq} from the joint regression along with liquid asset ratio, with x being cash flow to asset ratio. The reason why I bring cash flow and liquid asset holdings into comparison is that both of them can provide liquidity to firms to meet their financial needs such as wage and debt payments, etc. The fact that they are weakly and negatively correlated points to the possibility that firms that are unable to generate abundant cash flow regularly may hold extra liquid assets for precautionary measure. This echoes the viewpoint of Jeenas (2019) who warns that one must be careful in interpreting the liquid asset holdings as an effective measure of liquidity *per se*.

Figure 1.5. Dynamic effects of patent-based news shocks on investment by cash flow or/and liquidity

(a) Separate regressions



(b) Joint regressions



Notes: Panel (a) reports the estimates of coefficient β_h^x over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x) x_{it-1} + \mathbf{\Gamma}'_h \mathbf{Z}_{i,t-1} + e_{ith}$$

where $x \in \{cfr, liq\}$, while panel (b) reports the estimates of coefficient β_h^{cfr} and β_h^{liq} over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^{cfr} \epsilon_t^{news} + \gamma_h^{cfr}) cfr_{i,t-1} + (\beta_h^{liq} \epsilon_t^{news} + \gamma_h^{liq}) liq_{i,t-1} + \mathbf{\Gamma}'_h \mathbf{Z}_{i,t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sth} is a sector-by-quarter fixed effect, ϵ_t^{news} is the news shock, and $\mathbf{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter. I have standardized cfr_{it} and liq_{it} over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands.

The panel regression estimates of the heterogeneity by cash flow are very stable regardless of whether we simultaneously control for the liquid asset ratio or not. Firms with lower cash flow to asset ratio are more responsive in capital accumulation to technology news shocks. In particular, one standard deviation lower cash flow to asset ratio predicts approximately 0.5 pp stronger capital growth over three years, and the differential responses across firms with different cash flow to asset ratios are quite persistent.

1.4.4 Results Based on Dividend Status

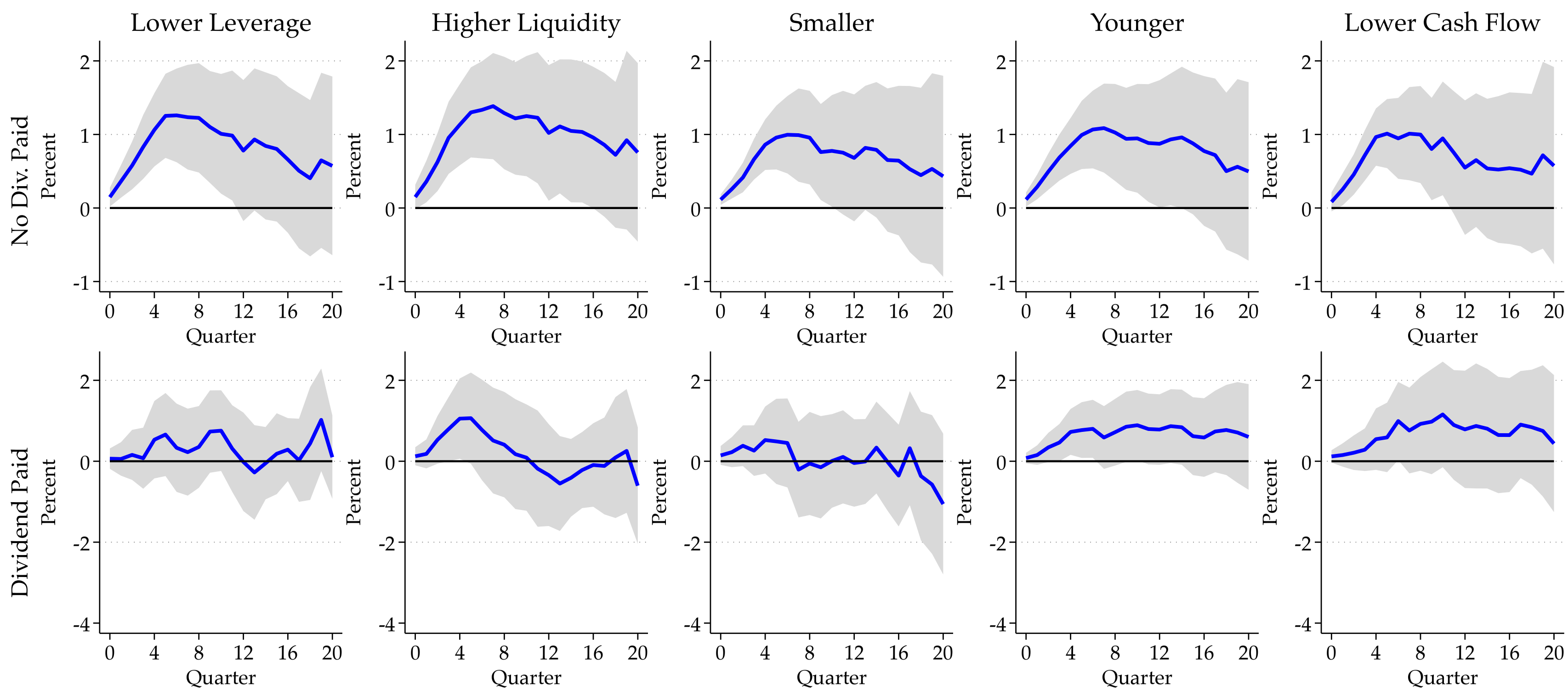
Dividends provide continuous, year-to-year indications of a company's growth and profitability. Paying dividends constantly usually sends a strong message that the firm is quite mature and financially flexible, and is less likely to have its financial stability threatened by temporary market or economic downturns. But it is also a commitment very few firms can keep. As shown in Table 1.2, less than 15% observations in the sample report positive values of *dvpq* during the sample periods. Compared with firms paying no dividends or with missing *dvpq*, firms paying dividends are older, and larger, with higher leverage but lower liquid asset ratio. Its correlation with the cash flow to asset ratio is very weak.

To take the cross-sectional correlation into account, I follow Cloyne et al. (2018) and estimate the group-specific impulse responses of cumulative capital stock by paying dividends or not interacted with other firm characteristics. Dividend status is binary, and I define the second dimension of interest using the more responsive quartile for liquidity, leverage, size, age, and cash flow to asset ratio. For example, I've shown that firms with more liquid asset holdings are more responsive to technology news shocks, so the second dimension is the higher liquidity firms above the upper quartile of the distribution of liquid asset ratio in the previous year. For leverage, size, and cash flow to asset ratio, I use their lower quartiles as cutoffs. In terms of age, "Younger" refers to age being less than 15 years, which is one year above the sample average of 14 years.[15] The estimation results are presented in Figure 1.6.

Among the firms with lower leverage, higher liquidity or smaller size, it is quite evident that firms paying no dividends are more responsive to news shocks. The hump-shaped

[15] The cutoff of 15 years is following Cloyne et al. (2018), but my definition of age is different from theirs. They define firm age as the number of years since incorporation.

Figure 1.6. Dynamic effects of patent-based news shocks on investment by dividend status and others



Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands.

responses of investment for non-dividend payers in the first row closely resemble the average effect of a technology news shock in Figure 1.2, implying that these firms drive the results. Among firms with lower leverage or smaller sizes, dividend payers' response is almost muted. Among firms with more liquidity, the investment response of dividend payers picks up in the first year, but the point estimates are marginally significant and soon drop to close to zero. Among the firms that are younger or more able to generate cash flow, dividend payers' response is persistent but less pronounced in the first two years, the point estimates are marginally significant at most.

In summary, the firm-level evidence on the heterogeneity of investment response by firm characteristics shows that firms with lower leverage or higher liquid asset ratio respond more to patent-based news shocks. However, about half of the differential response by leverage is due to the fact that lower-leverage firms also hold more liquid assets. As two commonly used proxies for financial frictions, firm age and size are also useful predictors of firms' heterogeneous investment responses. I find that younger and smaller firms are more responsive, but the relevance of firm size almost disappears after simultaneously controlling for age. Interestingly, the cash flow to asset ratio turns out to be another financial position different from liquid asset holdings that predicts clear heterogeneity in firms' capital accumulation following a news shock. In particular, firms with low cash flow to asset ratio respond by more. Last but not least, dividend status is also a strong predictor of significant heterogeneity in firms' investment dynamics. Non-dividend payers respond the most even within the firm groups that are more responsive. Given that the fraction of firms paying dividends is small (less than 15% of observations), it is not surprising that the results are mainly driven by the non-dividend payers.

1.5 Robustness and Further Results

In this section, I show that my main results – that younger and smaller non-dividend payers with lower leverage, more liquid asset holdings, or lower cash flow adjust investment the most following a technology news shock – are robust to a range of robustness checks. In particular, I consider whether the results are robust to (i) alternative news shock measures based on the movements in TFP, (ii) adjacent imputation of *ppentq* when constructing the firm-level capital stock, (iii) inclusion of the Global Financial Crisis, (iv) sector-specific news shocks. At the end of this section, I present further results discussing whether the findings come from the permanent differences across firms or within-firm variation.

1.5.1 Heterogeneity in Investment Response to TFP News Shocks

This section considers whether the results are robust to alternative measures of technology news shocks. In Appendix 1.8.1, I first review two major and related identification strategies based on exploiting the movements in the utilization-adjusted TFP. Then I use a self-collected dataset to construct the TFP news shock series in a nine-variable VAR. The corresponding impulse responses of macro variables are close to their counterparts in the literature, but the estimated average effect on firm investment convinces me to favor the earlier approach by Barsky and Sims (2011).[16]

Appendix Figure 1.21 shows the heterogeneity in investment response to a TFP news shock by the firm characteristics considered above except dividend status. Given that the correlation between the TFP news shock by Barsky and Sims (2011) and the patent-based news shock is relatively weak ($\rho = 0.13$), it is surprising to find out that the majority of the results survive. The first two columns show that firms with lower leverage or high liquid asset holdings are more responsive to a TFP news shock. When simultaneously controlling for liquidity, however, leverage loses much of its explanatory power with only half the peak difference in semi-elasticities and insignificant confidence bands. In the middle two columns, I show that younger or smaller firms are also more responsive to a TFP news shock. Between the two, firm age is a more robust predictor of the observed heterogeneity, and firm size becomes less relevant conditional on age. The difference lies in the last column about the role played by cash flow. Firms with low cash flow adjust investment slightly more in the first couple of years following a TFP news shock, but it is not as clear as what we've seen with the patent-based news shock in Figure 1.5.

In Appendix Figure 1.22, I show that dividend status remains another important dimension in which firms differ in their investment response to a TFP news shock. Among the firms

[16] I use self-collected data to compare two identification strategies, but the results are very similar if we use the Barsky and Sim's news shock provided by Cascaldi-Garcia and Vukotić (2022).

with higher liquidity, non-dividend payers adjust capital stock much more over the five-year horizon than dividend payers. Similarly, among younger or smaller firms, paying dividends predicts weaker investment responses. The difference among the firms with lower leverage or lower cash flow is small in the longer horizon, but in the first two years, non-dividend payers are still more responsive to a TFP news shock.

1.5.2 Imputation in Constructing Capital Stock

In this section, I show that my main results are not particularly sensitive to the massive linear interpolation in the construction of the firm-level capital stock. Following Ottonello and Winberry (2020), I consider limiting the imputation of *ppentq* to non-missing adjacent values and repeat the firm-level analysis. It is noteworthy that the resulting sample contains only half as many observations as the baseline, which can be treated as a refined subsample with more precisely measured capital stock.

Appendix Table 1.5 reports the summary statistics of the firm-level variables for the refined subsample. Almost half of the observations are dropped because of missing capital stock values using the perpetual inventory method, the dataset consists of nearly 190 thousand observations. Compared with my baseline sample, these firms are smaller, younger, less leveraged, and hold more liquid assets on average. Fewer of them pay dividends and the variation in cash flow to asset ratio across firms is larger. In terms of cross-sectional correlation, the liquid asset ratio is less correlated with size and age, and the cash flow to asset ratio is still not correlated with other proxies. Besides, this subsample contains observations populated in more recent periods as the information on *ppentq* is limited prior to 1982.

Appendix Figure 1.23 shows the heterogeneity in investment response to a patent-based news shock by the firm characteristics considered above except dividend status. It is still true that firms with lower leverage or higher liquidity are more responsive to news shocks, but it is no longer the case that leverage becomes less important after controlling for the liquid asset ratio in joint regression. Both play their roles in predicting firms' heterogeneous investment response. Firm size is not as important as in the baseline regardless of whether controlling for age or not. The heterogeneity by firm age stays marginally significant, and younger firms respond more to technology news. Given that these firms are significantly smaller and younger, the fact that size and age are less relevant becomes less of a concern. The heterogeneity by cash flow becomes less significant, and it's hard to tell why from the comparison of summary statistics. Despite that, the qualitative conclusion remains true, that is, firms with lower cash flow to asset ratio respond more to patent-based news shock.

Dividend status, shown in Appendix Figure 1.24, remains another dimension we should look into, especially among firms with lower leverage or smaller sizes. The difference between dividend payers and non-dividend payers is small in the first year among firms with higher liquidity, but dividend payers are less responsive thereafter. In contrast, among younger firms or firms with lower cash flow, dividend payers respond more in the longer horizons than non-dividend payers, who respond more in shorter horizons.

1.5.3 Including Recent Financial Crisis

This section considers whether the results are sensitive to including the recent financial crisis up till 2010:Q4 when the patent-based news shock series ends. The number of observations in the firm-level data increases by more than 12% from about 354 thousand to 398 thousand. In Appendix Figure 1.25, I find relatively stronger investment adjustment for firms that are younger, or smaller, with lower leverage, higher liquidity, or lower cash flow. Leverage and size lose most of their predictive power when simultaneously controlling for liquidity and age, respectively. In Appendix Figure 1.26, I show that among the more responsive firms, firms paying no dividends drive the results. All these results confirm that the financial crisis is unlikely to have overturned the heterogeneity found in Section 1.4.

1.5.4 Sector-Specific News Shocks

The last set of robustness checks looks at the heterogeneity in investment response to sector-specific news shocks. Cascaldi-Garcia and Vukotić (2022) show that patenting activity in only a few industries is predominantly responsible for explaining future movements in TFP. In particular, the electronic and electrical equipment industries within the manufacturing sector, and the business services sector drive the results. Given the strong correlation between the aggregate news shock and the manufacturing news shock in Table 1.1, it is not surprising that news from the manufacturing sector contributes the most to the observed heterogeneity by firm characteristics. The more interesting practice is to see if the results are robust to the news from important sub-industries. Here I consider the electronic and electrical equipment (EEE) industries and business services industries.

Appendix Figure 1.27 and 1.28 (1.29 and 1.30) show the heterogeneity in investment response to sector-specific news shock from EEE (business services) industries by firm characteristics. In both cases, the main takeaway still holds. That is, firms that are younger, smaller, with lower leverage, higher liquidity, lower cash flow, and paying no dividends adjust investment the most. The differences compared to the aggregate news shock are as follows: first, the average effect of a news shock is less persistent if it is originated from business services industries even among the more responsive firms. Second, the persistent heterogeneity by liquidity in Figure 1.3 is mainly driven by the EEE news rather than business services news. The heterogeneity by liquidity almost disappears at the end of the fifth year in response to business services news. Third, the role of cash flow is less pronounced in predicting the heterogeneous investment responses to the news from business services industries than that from EEE industries.

1.5.5 Within-Firm Variation

Following Ottonello and Winberry (2020), instead of the *level* of financial position $x_{i,t-1}$ in quarter $t-1$, I interact the news shock ϵ_t^{news} with the *demeaned* financial position $x_{i,t-1} - E_i[x_{it}]$. The local projection specification is as follows:

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x)(x_{i,t-1} - E_i[x_{it}]) + \mathbf{\Gamma}_h' \boldsymbol{Z}_{i,t-1} + e_{ith} \tag{1.4}$$

$\forall h = 0, 1, \cdots, H$, where α_{ih} denotes a firm fixed effect, α_{sth} is a sector-by-quarter fixed effect, ϵ_t^{news} is some quarterly measure of the news shock, $x \in X$ is the set of firm characteristics, and $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter. I have standardized $x_{it} - E_i[x_{it}]$ over the entire sample.

By demeaning financial position within firms, the estimates β_h^x are instead driven by how a given firm responds to news shocks when it has higher x than usual. If β_h^x becomes insignificant, then the heterogeneity is mainly driven by *ex-ante* or permanent difference across firms. Appendix Figure 1.31 and 1.32 present the main results. Qualitatively, it is still the firms that are younger, smaller, with lower leverage, higher liquidity, or lower cash flow that adjust investment the most. However, the bands suggest that the heterogeneous investment responses by leverage, liquidity, and size are mainly driven by permanent differences across firms, not within-firm variation. The heterogeneous investment responses by age and cash flow are still present, with the former more pronounced in the shorter horizons and the latter more pronounced in longer horizons. Among the more responsive firms, non-dividend payers drive the observed heterogeneity by leverage, size, and age.

1.6 Transmission Mechanism

In the previous section, we saw that various firm characteristics are useful predictors of firms' heterogeneous investment responses to technology news shocks. Among them, liquidity, age, and dividend status are quite robust. In particular, younger non-dividend payers with higher liquidity adjust their investment in physical capital the most. To the best of my knowledge, this is the first paper focusing on the firm heterogeneity in investment response to technology news shocks, and little is known about the transmission channels through which news about future technology affects firms' investment decisions in a business cycle frequency. The main goal of this section is, therefore, to investigate how technology news affects firm finance that leads to the observed heterogeneity in investment.

1.6.1 Is This a Financial Accelerator Story?

The firm characteristics studied above are frequently used as proxies for financial frictions, but it is still an ongoing debate in corporate finance literature about how they are related to the financial constraints faced by the firms. Despite this "proxy chaos", it is widely accepted that smaller and younger firms face more severe financial constraints. In this sense, the previous finding that younger and smaller firms are more responsive to news shocks suggests a financial accelerator story where more financially constrained firms are more responsive. This sounds quite familiar, especially in the study of monetary policies. The firm balance sheet channel of monetary policy in the tradition of Kiyotaki and Moore (1997) and Bernanke et al. (1999) emphasizes that financial frictions could amplify the transmission of monetary policy. Along the same vein, earlier theoretical studies in news shock literature have also emphasized the importance of financial frictions in driving the aggregate responses, see e.g., Kobayashi et al. (2012); Chen and Song (2013); Walentin (2014), etc.

In terms of leverage and liquidity, the findings that firms with lower leverage or higher liquidity respond more seem to be at odds with the financial accelerator story as these firms are usually considered not financially constrained. However, recent studies point out that one has to be careful in interpreting these two proxies, especially liquidity. Firms may hold more liquid assets because it has no access to stable external finance (e.g., debt instruments), and in case of sudden financial need, they have the option to use their precautionary saving. For leverage, a firm has low leverage because it has limited access to external finance thus financially constrained, or it relies on other sources of finance. It's hard to tell without knowing other dimensions of a firm's financial condition.

According to Hadlock and Pierce (2010), firms with high debt-to-asset ratios are more likely to be constrained, while firms with more cash are more likely to be constrained. If this is the case on average, then we would observe a positive correlation between leverage and liquidity. In my sample, the more constrained smaller and younger firms tend to hold more liquid assets, but liquidity is *negatively* correlated with leverage. Therefore, if we accept the view that firms with high liquidity and low leverage are more constrained, then a financial accelerator story seems not bad. Even if the firms with low leverage are not financially constrained, liquidity should still be more important in explaining the heterogeneity in capital expenditure as leverage becomes less relevant when we simultaneously control for liquidity in Figure 1.3.

Regarding the cash flow, it is usually the case that firms with more cash flow are more financially stable and typically preferred by financial intermediaries when issuing debts, which is also confirmed by Hadlock and Pierce (2010). The fact that firms with lower cash flow to asset ratio respond more to a technology news shock is also consistent with the financial accelerator mechanism.

Lastly, paying dividends constantly usually sends a strong signal about a firm's financial stability. However, the financial well-being of a non-dividend payer is less clear. Hadlock and Pierce (2010) find mixed evidence about the relationship between dividends and whether a firm is financially constrained or not. Nevertheless, the finding that (non-)dividend payers adjust their investment (more) less is not against the financial accelerator mechanism.

When firms are aware of favorable news about future technology, how they finance their investment projects greatly depends on their debt availability and net worth. Previous literature has mostly focused on the asset pricing channel. From the macro-level impulse responses to TFP news shocks in Appendix Figure 1.18 and 1.19, we saw a long-lasting increase in stock prices, raising firms' market values. This alleviates the asymmetric information problems in financial markets, resulting in more investment. In the rest of this section, I look at how firms' borrowing and net worth respond to understand how they finance their investment projects. In particular, I estimate the heterogeneity in the response of short-term and long-term debts, and book and market values of equity following a patent-based news shock. The sales and earnings are also inspected.

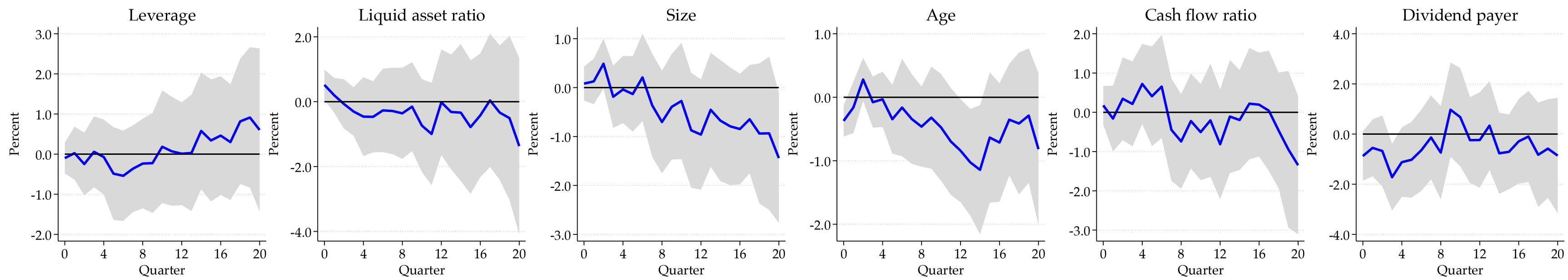
1.6.2 Borrowing: Short-Term vs. Long-Term Debts

Figure 1.7 shows the dynamics of the heterogeneity in borrowing following a patent-based news shock by firm characteristics. The results for short-term debts are collected in panel (a) and the results for long-term debts are listed in panel (b). Since investment projects usually last for years if not for decades, firms definitely rely on long-term debts more than short-term debts. Indeed, most of the heterogeneity in borrowing across firms lies in long-term debts, we find no significant heterogeneity in the response of short-term debts. However, this does not imply firms' short-term debts do not respond to news shocks. In fact, Figure 1.8 shows that firms demand more short-term debts after a positive news shock, most of the response occurs in the first two years when the real economic activities respond the most. Later, we will see that firms' sales pump up in the first year due to the high consumption demand following a favorable news shock shown in Figure 1.1. To meet the higher demand, firms may have to borrow more short-term debts to pay for wages and other input costs.

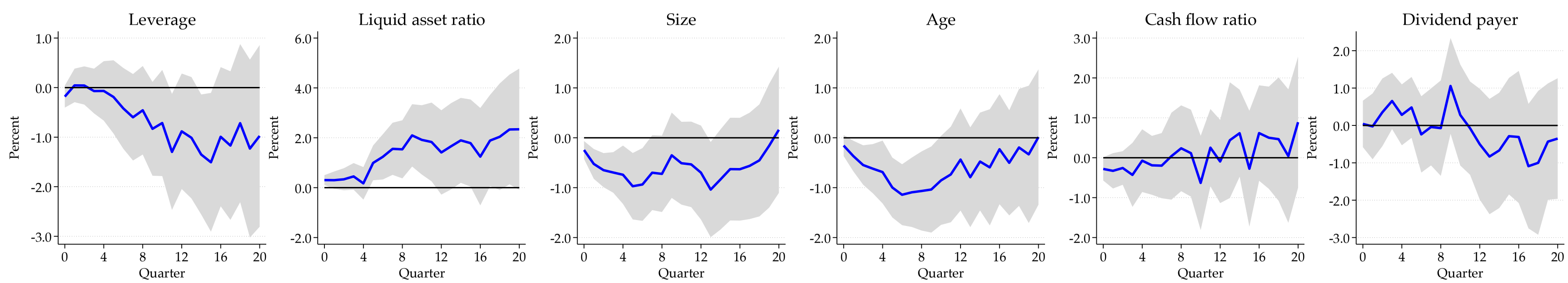
In terms of the observed heterogeneity in long-term debts, we find similar patterns as in investment. Younger and smaller firms with lower leverage or higher liquidity raise significantly more long-term debts in response to a positive news shock. The results by cash flow and dividend status are more homogeneous across firms. In Figure 1.9, we see a more

Figure 1.7. Dynamic effects of patent-based news shocks on borrowing by firm characteristics

(a) Short-term debts



(b) Long-term debts

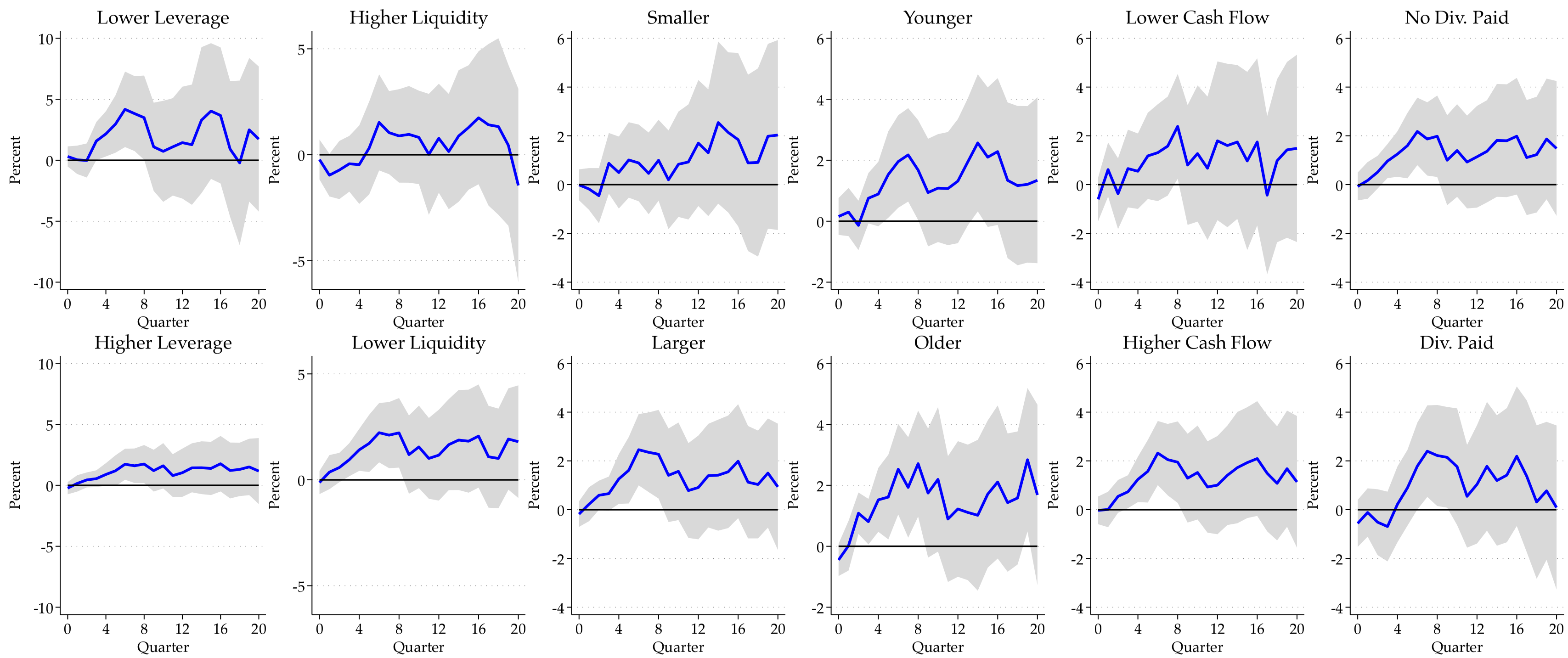


Notes: Reports the estimates of coefficient β_h^x over quarters h from

$$\triangle_h \log(y_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x) x_{it-1} + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

where y is short-term debts in panel (a) and long-term debts in panel (b), both are deflated by CPI. α_{ih} is a firm fixed effect, α_{sth} is a sector-by-quarter fixed effect, ϵ_t^{news} is the news shock, and $\mathbf{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter. I have standardized lev_{it} and liq_{it} over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firms and time. The grey areas indicate 90% confidence bands.

Figure 1.8. Dynamic effects of patent-based news shocks on short-term debts by firm characteristics

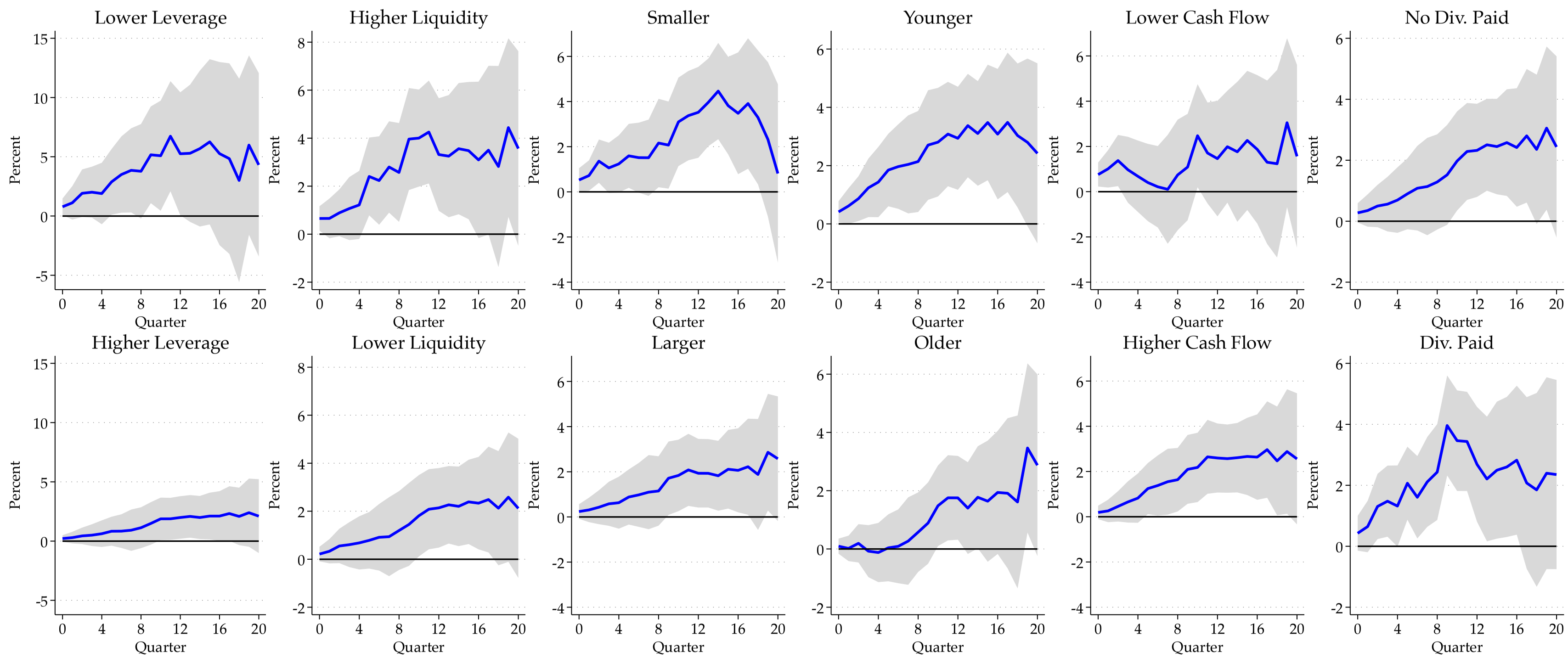


Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(y_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}'_h \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where y is short-term debt deflated by CPI, α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firms and time. The grey areas indicate 90% confidence bands.

Figure 1.9. Dynamic effects of patent-based news shocks on long-term debts by firm characteristics



Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(y_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where y is long-term debt deflated by CPI, α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firms and time. The grey areas indicate 90% confidence bands.

stable increase in long-term debts for firms with more cash flow. Both dividend-payers and non-payers increase long-term borrowing by a similar amount.

1.6.3 Net Worth: Market Value of Equity vs. Book Value of Equity

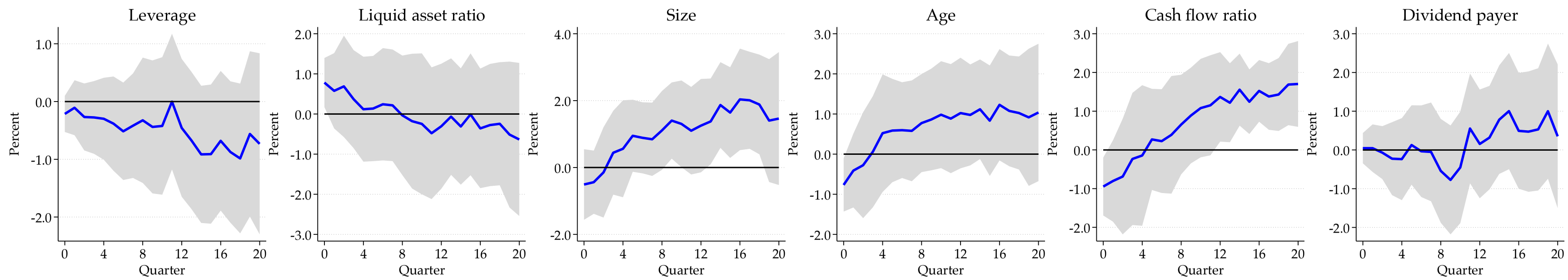
Before looking at the firm-level evidence, it is noteworthy that in Figure 1.1 the aggregate stock prices (deflated by CPI) only rise in the first year following a positive patent-based news shock. Besides the "disinflation puzzle", this is another big difference between patent-based news shock and TFP-based news shock. As shown in Appendix Figure 1.18 and 1.19, the stock market boom is very persistent after a positive TFP news shock, somewhat too persistent to be true. Appendix Figure 1.33 shows the average effect on firm-level share prices of patent-baed news shocks in panel (a) and TFP news shocks in panel (b). Both panels deliver a similar message, that is, the rise in stock prices is immediate but short-lived. If anything, the effect is negative at the end of the fifth year. It is thus the patent-based news shock that delivers consistent evidence at both macro and micro levels, and firms' market values of equity only rise temporarily followed by a quick reversal.

Figure 1.10 panel (a) shows the heterogeneity in the response of market equity by firm characteristics. One has to be careful in interpreting the signs. Firms with different positions in leverage, liquidity, or dividend status do not seem to differ much in their market equity response. The results by size, age, and cash flow are easier to read if we look at the group-specific impulse responses in Figure 1.11. The market equity of a smaller, younger firm with a lower cash flow to asset ratio jumps up by slightly more on impact, but the following drop is also more severe than the rest of the market.

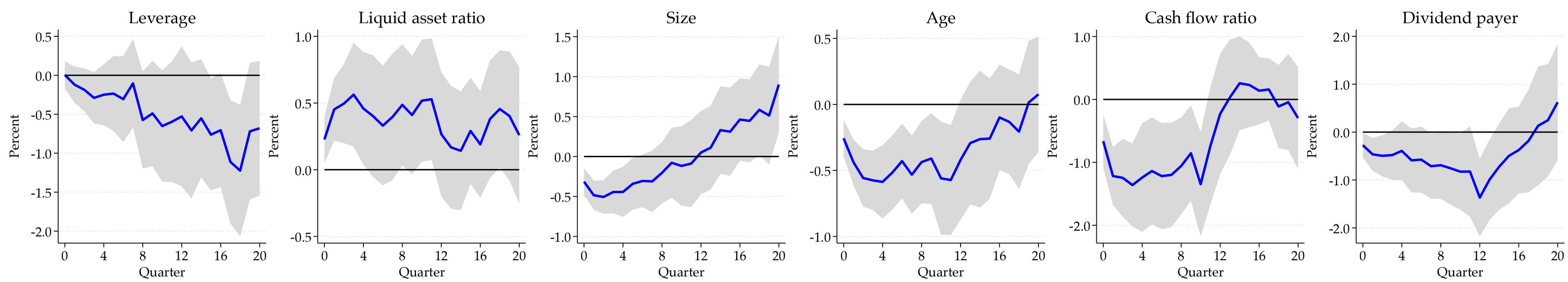
In contrast, the heterogeneity in the response of book equity by firm characteristics, shown in Figure 1.10 panel (b), is very similar to the heterogeneity in investment response. Younger, smaller non-dividend payers with lower leverage, higher liquidity, or lower cash flow display a much faster growth in book equity following positive news about future technology. In Figure 1.12, all groups witness large and significant increases in book equity at least in the first couple of years. The effects are more pronounced and persistent for younger firms with lower leverage or higher liquidity. The increase in firms' asset values or net worth helps relax their borrowing constraints and enables them to finance externally from financial intermediaries.

Figure 1.10. Dynamic effects of patent-based news shocks on net worth by firm characteristics

(a) Market equity



(b) Book equity

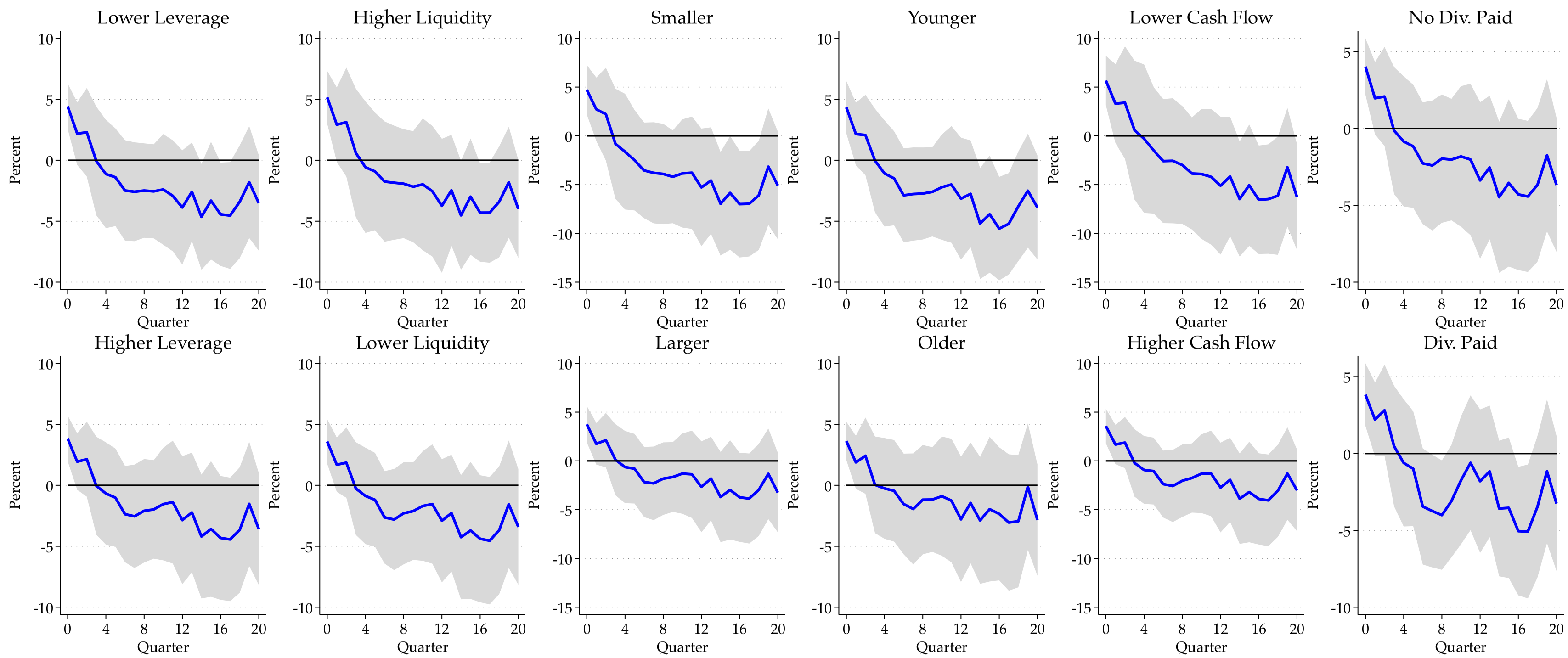


Notes: Reports the estimates of coefficient β_h^x over quarters h from

$$\triangle_h \log(y_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x) x_{it-1} + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

where y is the market equity in panel (a) and book equity in panel (b), both are deflated by CPI. α_{ih} is a firm fixed effect, α_{sth} is a sector-by-quarter fixed effect, ϵ_t^{news} is the news shock, and $\mathbf{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter. I have standardized lev_{it} and liq_{it} over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firms and time. The grey areas indicate 90% confidence bands.

Figure 1.11. Dynamic effects of patent-based news shocks on market equity by firm characteristics

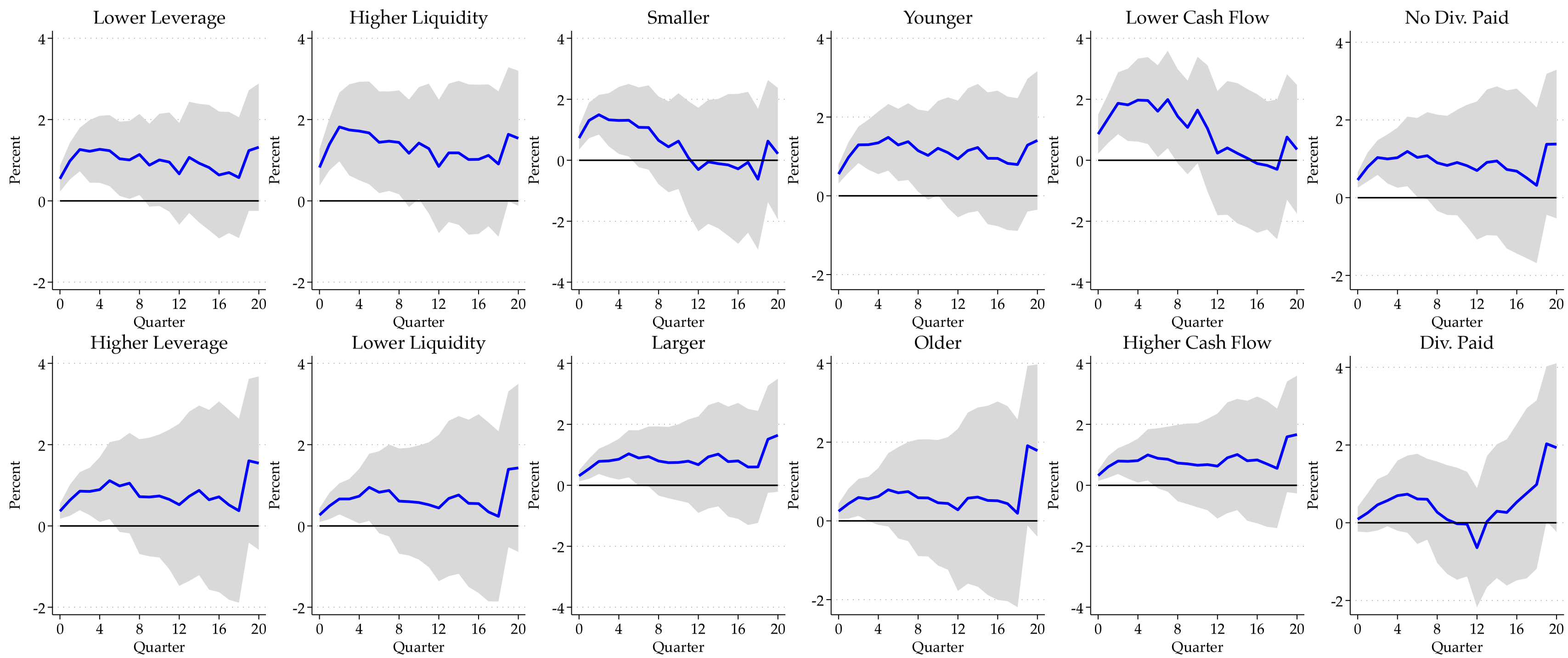


Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(y_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where y is the market value of equity deflated by CPI, α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firms and time. The grey areas indicate 90% confidence bands.

Figure 1.12. Dynamic effects of patent-based news shocks on book equity by firm characteristics



Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(y_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}'_h \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where y is the book value of equity deflated by CPI, α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firms and time. The grey areas indicate 90% confidence bands.

1.6.4 Cash Flow: Sales and Earnings

The last set of results looks at the cash flow channel. Unlike monetary policy shock, news shock does not immediately affect firms' borrowing costs, but consumption demand rises significantly on impact due to the anticipation effect, which will improve firms' sales and earnings. If a significant fraction of firm debt was secured against earnings prior to the shock, then firms will become more capable of borrowing to invest after good news. The observed increase in long-term debts may be a result of the improvement in both net worth and cash flow.

Figure 1.14 and 1.15 confirm the previous conjecture that following good news about future technology, all firms enjoy higher sales and earnings due to the consumption boom in the first two years. However, the heterogeneity by firm characteristics shown in Figure 1.13 is not so clear, and neither is stable over time. I interpret the relatively uniform responses of sales and earnings across firm groups as evidence that the cash flow channel is limited in accounting for the observed heterogeneity in investment response.

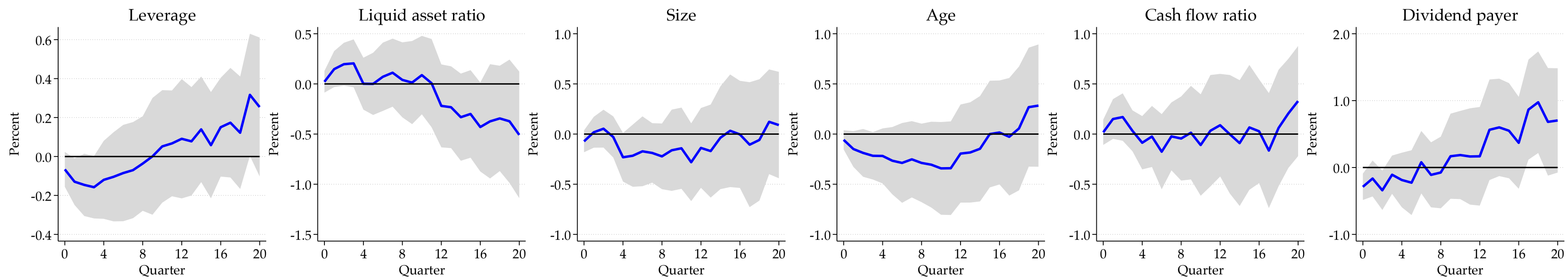
1.6.5 Summary

So far this section has provided a detailed and rich analysis of how technology news shocks affect different sources of firm finance to help understand the heterogeneity in the investment response. The main findings are: (i) the borrowing of long-term debts increases significantly for all firms, especially for younger, smaller, less indebted, and more liquid firms, while the increase in short-term debts is more uniform and temporary; (ii) the firms that adjust investment the most also display a significantly larger increase in book equity, while the boom in market equity is limited and short-lived; (iii) the rise in sales or earnings is short-lived and more homogeneous across firms.

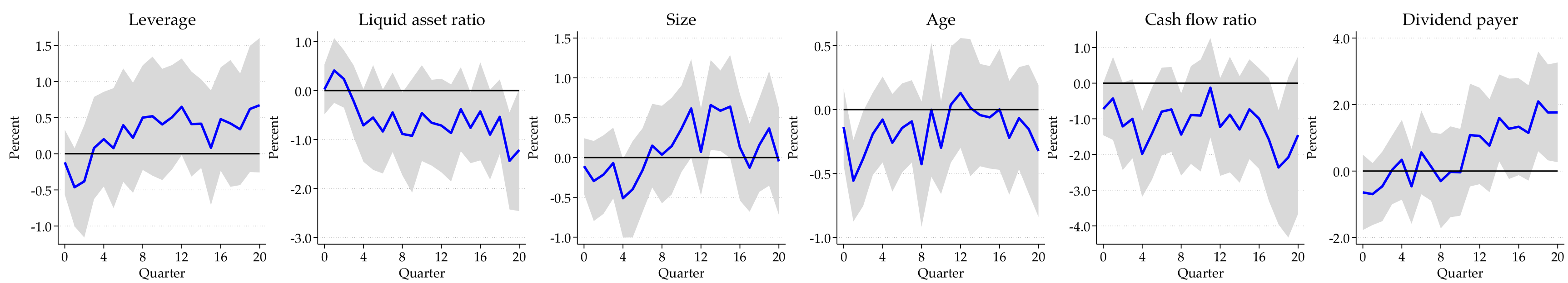
All these findings are consistent with the theoretical framework in which a firm's ability to borrow is related to its net worth. The good news about future technology fosters a consumption boom, sending a signal to firms to expand production. To meet the higher demand, firms borrow more short-term debts to pay for wages and input costs. All of these occur for most firms, the difference lies in how firms' net worth responds relative to before

Figure 1.13. Dynamic effects of patent-based news shocks on cash flow by firm characteristics

(a) Sales



(b) Earnings

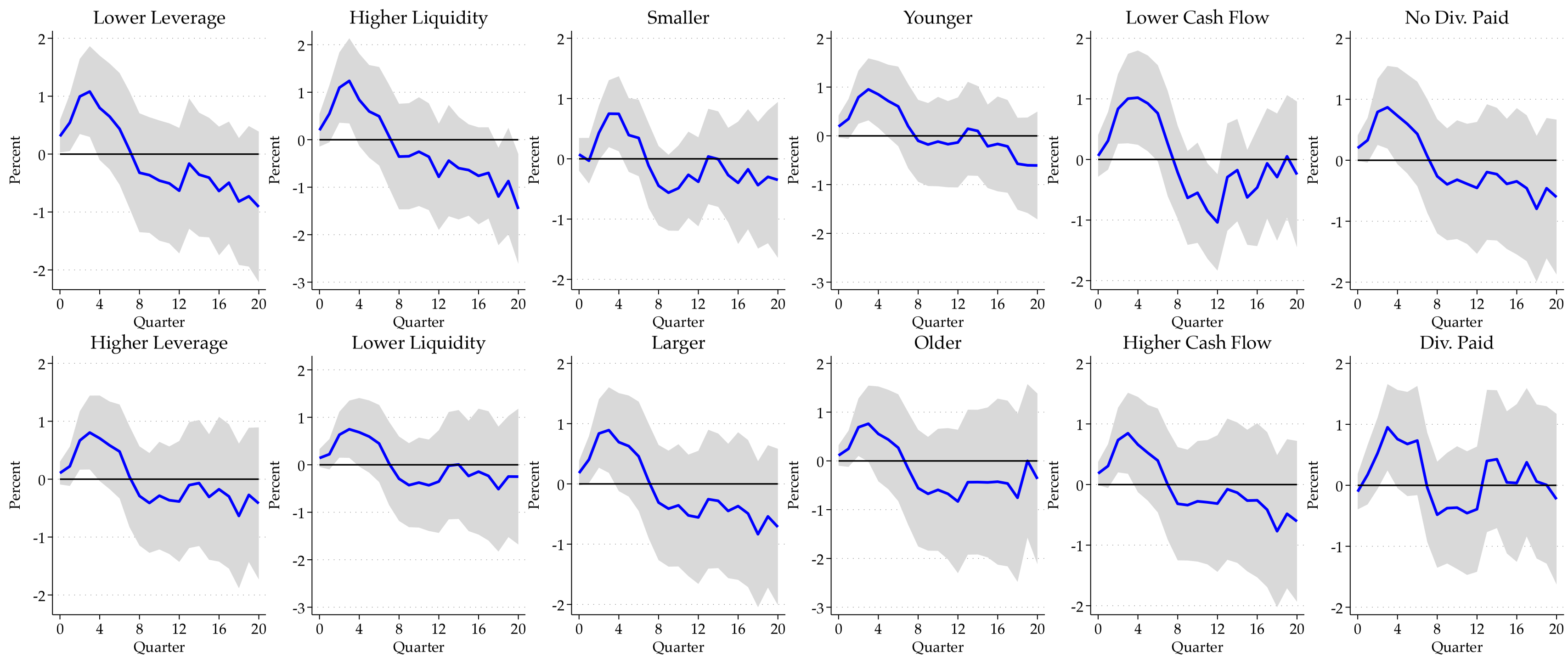


Notes: Reports the estimates of coefficient β_h^x over quarters h from

$$\triangle_h \log(y_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x) x_{it-1} + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{i,t-1} + e_{ith}$$

where y is the sales in panel (a) and earnings in panel (b), both are deflated by CPI. α_{ih} is a firm fixed effect, α_{sth} is a sector-by-quarter fixed effect, ϵ_t^{news} is the news shock, and $\boldsymbol{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter. I have standardized lev_{it} and liq_{it} over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firms and time. The grey areas indicate 90% confidence bands.

Figure 1.14. Dynamic effects of patent-based news shocks on sales by firm characteristics

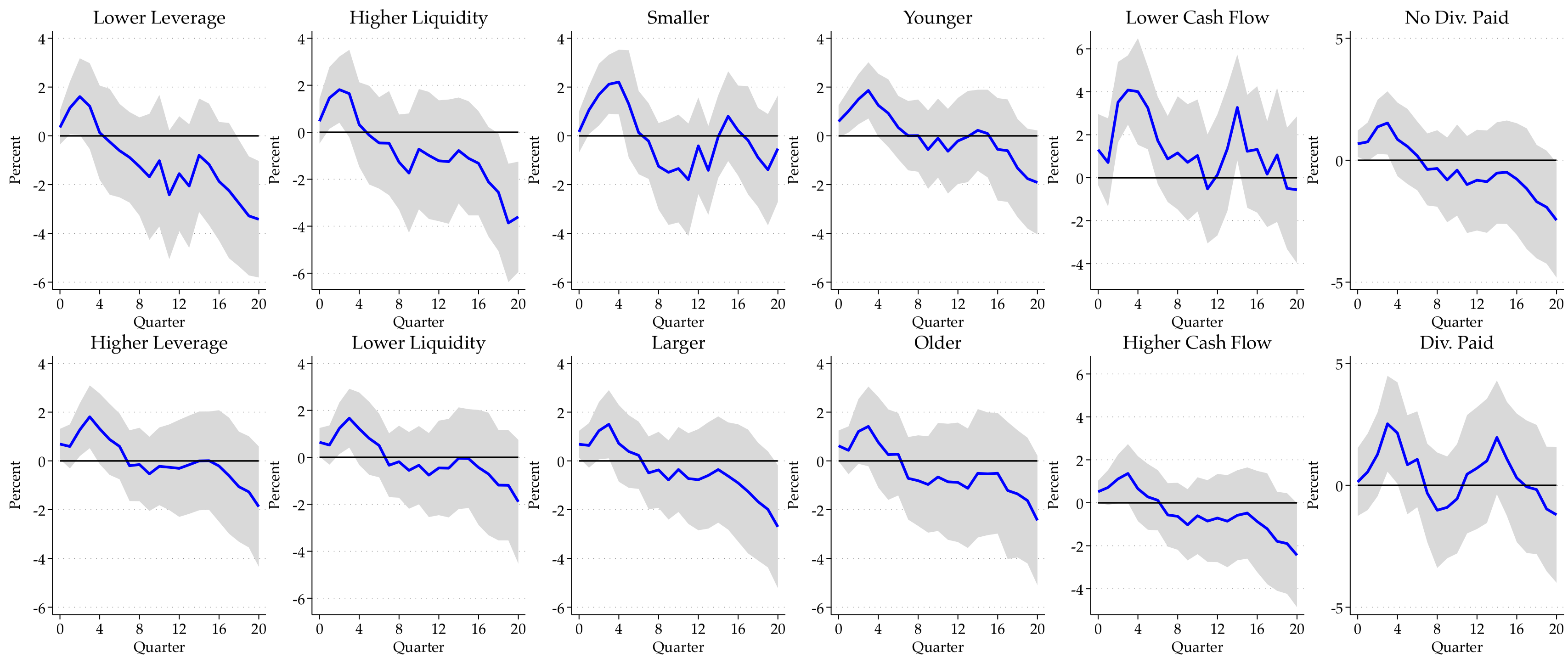


Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(y_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where y is Sales (Net) deflated by CPI, α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firms and time. The grey areas indicate 90% confidence bands.

Figure 1.15. Dynamic effects of patent-based news shocks on earnings by firm characteristics



Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(y_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where y is earnings (EBITDA) deflated by CPI, α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firms and time. The grey areas indicate 90% confidence bands.

the shock. It turns out that firms that are more responsive in investment adjustment also show a more pronounced increase in net worth. This enables them to borrow more long-term debts from external sources as they face less severe asymmetric information problems, thus less severe financial frictions. Contrary to the previous literature, I find very limited and short-lived increases in firms' market values of equity and share prices, which are commonly used to justify the asset pricing channel in the news shock literature.

1.7 Conclusion

In this paper, I have combined the identified technology news shocks with panel local projection techniques for micro-level data to study the relevance of several proxies for financial frictions in explaining firms' heterogeneous capital accumulation responses to news about future technology. I find that firms that are younger, smaller, with lower leverage, higher liquidity, or lower cash flow respond the most to news shocks over a five-year horizon. After taking into account the correlation among different proxies, liquidity, age, and cash flow remain powerful in predicting the heterogeneous investment adjustment. Dividend status also turns out to be useful. Among the more responsive firm groups, non-dividend payers drive the results.

The main results are shown to be largely robust to (i) alternative news shock measures based on the movements in TFP, (ii) limited imputation when constructing firm-level capital stock, (iii) inclusion of the GFC, (iv) sector-specific news shocks. By focusing on the within-firm variation, I show that the heterogeneous investment responses by leverage, liquidity, and size are mainly due to permanent differences across firms, while the heterogeneity by age and cash flow is still present, with the former more pronounced in the shorter horizons and the latter more pronounced in longer horizons.

This paper also provides suggestive evidence on the transmission mechanism of technology news shocks to investment dynamics through the impact on firm finance. Contrary to the previous literature, I find very limited and short-lived increases in firms' market values of equity and share prices, which are commonly used to justify the asset pricing channel upon the arrival of good news about future productivity. In contrast, the book values of firms' equity respond more persistently and prominently over the decade, which enables firms to have more pledgeability over their investment projects to finance with external funding. Indeed, the borrowing of long-term debts increases significantly for all firms, especially for younger, smaller, less indebted, and more liquid firms. The cash flow channel is less important in explaining the heterogeneity in investment response as the rise in sales or earnings is short-lived and more homogeneous across firms.

The results of this paper are particularly important for at least two reasons. First, in the face of widespread pessimism about future productivity and growth during the pandemic, which group of firms should the government target when providing urgent financial support to restore business confidence greatly depends on how firm heterogeneity matters in investment responses. To the best of my knowledge, this is the first paper to study the firm heterogeneity in the transmission of technology news shocks to firms' investment dynamics. Second, micro-level evidence provides further tests on the stance of the news shock identified based on strict short-run and/or long-run assumptions and the mechanism through which news about future technology affects the real economy.

1.8 Appendix

1.8.1 Identification of TFP News Shocks

In the news shock literature, most studies have been focused on exploiting the movements in the utilization-adjusted TFP by Fernald (2014*a*) to identify the news about future productivity, i.e., the so-called TFP news shocks. In this appendix, I first review two related identification strategies proposed by Barsky and Sims (2011) and Kurmann and Sims (2021). Then I use a self-collected dataset to construct the TFP news shock series in a nine-variable VAR. The corresponding impulse responses of macro variables are close to their counterparts in the literature, but the estimated firm-level average effect on investment convinces me to favor the earlier approach by Barsky and Sims (2011). The resulting TFP news shock series is used in Section 1.5 for robustness check.

Identification Strategies Barsky and Sims (2011) identify the TFP news shock as the innovation that accounts for most of the forecast error variance (FEV) of utilization-adjusted TFP *over* a ten-year horizon but has no contemporaneous effect on TFP.

Let A denote the level of total factor productivity (TFP) and the stochastic structure can be expressed using the following MA representation:

$$\ln A_t = v(L)\epsilon_{1,t} + d(L)\epsilon_{2,t}$$

where $\epsilon_{1,t}$ is the conventional TFP surprise shock while $\epsilon_{2,t}$ is the news shock. The only restriction is that $d(0) = 0$, i.e. news shocks have no contemporaneous effect on TFP.

In a VAR featuring a utilization-adjusted measure of aggregate TFP and several forward-looking variables, a conventional surprise technology shock is identified as the reduced form innovation in TFP, while news shock is then identified as the shock orthogonal to TFP innovation that best explains future variation in measured TFP. In particular, it maximizes the sum of the contributions to the FEV of TFP over a finite horizon. This identification strategy is closely related to the maximum FEV approach by Francis et al. (2007), which builds on Faust (1998) and Uhlig (2003, 2004).

Let $\boldsymbol{y}_t$ be a $k \times 1$ vector of observables of length T. One can form the reduced form MA representation in levels of the observables by estimating either a stationary VECM or an unrestricted VAR in levels:

$$\boldsymbol{y}_t = \boldsymbol{B}(L)\boldsymbol{u}_t$$

where $\boldsymbol{B}(L) = \boldsymbol{I} - \boldsymbol{B}_1 L - \boldsymbol{B}_2 L^2 - \cdots$. Assuming there exists a linear mapping between innovations $\boldsymbol{u}_t$ and structural shocks $\boldsymbol{\epsilon}_t$, $\boldsymbol{u}_t = \boldsymbol{A}_0\boldsymbol{\epsilon}_t$, this implies that

$$\boldsymbol{y}_t = \boldsymbol{C}(L)\boldsymbol{\epsilon}_t$$

where $\boldsymbol{C}(L) = \boldsymbol{B}(L)\boldsymbol{A}_0$ and $\boldsymbol{\epsilon}_t = \boldsymbol{A}_0^{-1}\boldsymbol{u}_t$. The impact matrix $\boldsymbol{A}_0$ must satisfy $\boldsymbol{A_0}\boldsymbol{A_0'} = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is the variance-covariance matrix of innovations, i.e., $\boldsymbol{\Sigma} = E[\boldsymbol{u}_t\boldsymbol{u}_t']$. $\boldsymbol{A}_0$ is not unique. There exists some alternative matrix $\tilde{\boldsymbol{A}}_0$ (e.g., a Choleski decomposition), s.t. $\tilde{\boldsymbol{A}}_0\boldsymbol{Q} = \boldsymbol{A}_0$, where $\boldsymbol{Q}$ is a $k \times k$ orthonormal matrix ($\boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{I}$).

Therefore,

$$\boldsymbol{y}_t = \boldsymbol{C}(L)\boldsymbol{\epsilon}_t = \boldsymbol{B}(L)\boldsymbol{A}_0\boldsymbol{\epsilon}_t = \sum_{\ell=0}^{\infty}\boldsymbol{B}_\ell\boldsymbol{A}_0\boldsymbol{\epsilon}_{t-\ell} = \sum_{\ell=0}^{\infty}\boldsymbol{B}_\ell\tilde{\boldsymbol{A}}_0\boldsymbol{Q}\boldsymbol{\epsilon}_{t-\ell}$$

The h step ahead forecast error is

$$\boldsymbol{y}_{t+h} - E_{t-1}\boldsymbol{y}_{t+h} = \sum_{\ell=0}^{h}\boldsymbol{B}_\ell\tilde{\boldsymbol{A}}_0\boldsymbol{Q}\boldsymbol{\epsilon}_{t+h-\ell}$$

so that the share of the forecast error variance of variable i attributable to structural shock j at horizon h is given as

$$\Omega_{i,j}(h) = \frac{\boldsymbol{e}_i'\left(\sum_{\ell=0}^{h}\boldsymbol{B}_\ell\tilde{\boldsymbol{A}}_0\boldsymbol{Q}\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{Q}'\tilde{\boldsymbol{A}}_0'\boldsymbol{B}_\ell'\right)\boldsymbol{e}_i}{\boldsymbol{e}_i'\left(\sum_{\ell=0}^{h}\boldsymbol{B}_\ell\boldsymbol{\Sigma}\boldsymbol{B}_\ell'\right)\boldsymbol{e}_i}$$

where $\boldsymbol{e}_i$'s are selection vectors with one in the ith entry and zeros elsewhere, the $\boldsymbol{e}_i$'s outside the parentheses pick out the ith row, while the $\boldsymbol{e}_j$'s inside the parentheses pick out the jth column. Let $\boldsymbol{\gamma} = \boldsymbol{Q}\boldsymbol{e}_j$ and $\boldsymbol{B}_{i,\ell} = \boldsymbol{e}_i'\boldsymbol{B}_\ell$, we have

$$\Omega_{i,j}(h) = \frac{\sum_{\ell=0}^{h}\boldsymbol{B}_{i,\ell}\tilde{\boldsymbol{A}}_0\boldsymbol{\gamma}\boldsymbol{\gamma}'\tilde{\boldsymbol{A}}_0'\boldsymbol{B}_{i,\ell}'}{\sum_{\ell=0}^{h}\boldsymbol{B}_{i,\ell}\boldsymbol{\Sigma}\boldsymbol{B}_{i,\ell}'}$$

Let TFP be the first in the VAR system, and denote the unanticipated shock by index 1 and the news shock by index 2. According to the stochastic structure of TFP, we know the surprise shock and news shock should account for all the variations in TFP, i.e.

$$\Omega_{1,1}(h) + \Omega_{1,2}(h) = 1, \forall h$$

However, in a multivariate VAR setting, this equality may not hold for all h, so instead, they propose to pick the impact matrix $\boldsymbol{A}_0$ to make this equality as close to hold as possible. It is equivalent to maximizing the contributions to $\Omega_{1,2}(h)$ over h since the surprise shock is identified in a way that $\Omega_{1,1}(h)$ is invariant at all h. Hence, we only need to solve the following optmization problem:

$$\max_{\boldsymbol{\gamma}} \quad \sum_{h=0}^{H} \Omega_{1,2}(h) = \sum_{h=0}^{H} \frac{\sum_{\ell=0}^{h} \boldsymbol{B}_{1,\ell} \tilde{\boldsymbol{A}}_0 \boldsymbol{\gamma}\boldsymbol{\gamma}' \tilde{\boldsymbol{A}}_0{}' \boldsymbol{B}_{1,\ell}'}{\sum_{\ell=0}^{h} \boldsymbol{B}_{1,\ell} \boldsymbol{\Sigma} \boldsymbol{B}_{1,\ell}'}$$

$$s.t. \quad \boldsymbol{\gamma}(1,1) = 0, \quad \boldsymbol{\gamma}'\boldsymbol{\gamma} = 1$$

where $\boldsymbol{\gamma}$ is the second column of $\boldsymbol{Q}$ and $\tilde{\boldsymbol{A}}_0$ is the Choleski decomposition of $\boldsymbol{\Sigma}$. With the optimal $\boldsymbol{\gamma}^*$, we can back out the news shock by taking the second structural shock, $\epsilon_{2t} = \boldsymbol{e}_2'\boldsymbol{\epsilon}_t = \boldsymbol{e}_2'\boldsymbol{A}_0^{-1}\boldsymbol{u}_t$, since $\boldsymbol{A}_0 = \tilde{\boldsymbol{A}}_0\boldsymbol{Q}, \boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{I}$, and $\boldsymbol{Q}\boldsymbol{e}_2 = \boldsymbol{\gamma}$, we can express the news shock in terms of reduce-form innovations as

$$\epsilon_{2t} = \boldsymbol{e}_2'\boldsymbol{Q}^{-1}\tilde{\boldsymbol{A}}_0^{-1}\boldsymbol{u}_t = \boldsymbol{e}_2'(\boldsymbol{Q}')\tilde{\boldsymbol{A}}_0^{-1}\boldsymbol{u}_t = \boldsymbol{\gamma}^{*'}\tilde{\boldsymbol{A}}_0^{-1}\boldsymbol{u}_t$$

We can get estimates of $\boldsymbol{B}_\ell, \boldsymbol{u}_t, \boldsymbol{\Sigma}$, and $\tilde{\boldsymbol{A}}_0$ from VAR(p) and solve the optimization problem for $\boldsymbol{\gamma}^*$ to get an estimate of $\boldsymbol{\epsilon}_{2t}, \forall t = p+1, \cdots, T$.

This identification strategy is a partial identification strategy based on the common assumption in the news shock literature that a limited number of shocks lead to movements in aggregate technology. The zero impact restriction is originated from Beaudry and Portier (2006), who take two structural VAR specifications, one is orthogonal to the current TFP shock, the other has long-run effects on TFP, and find that they are highly correlated. Long run identification is problematic especially in finite samples, see Barsky and Sims (2011) and the reference therein for more discussion about the merits of their approach compared to Beaudry and Portier (2006) and Beaudry and Lucke (2010).

In their original paper, Barsky and Sims (2011) find that a favorable news shock leads to an increase in consumption but declines in hours, investment, and output on impact, consistent with the prediction of their standard real business cycle model augmented with both news and surprise technology shocks. However, several recent papers apply their identification scheme but yield different results using different vintages of utilization-adjusted TFP series, see e.g., Chen et al. (2018).

Kurmann and Sims (2021) give a profound discussion about the measurement issues with TFP that will cause trouble for Barsky and Sims (2011) approach and propose an alternative and arguably more robust identification method based on Francis et al. (2007), i.e. Max-Share approach without the zero impact restriction. They look for the shock that accounts for the maximum FEV share of TFP at one long horizon H and drop the zero impact restriction.

Formally, their identification is given by solving the following optimization problem:

$$\max_{\boldsymbol{\gamma}} \quad \Omega_{1,2}(H) = \frac{\sum_{\ell=0}^{H} \boldsymbol{B}_{1,\ell} \tilde{\boldsymbol{A}}_0 \boldsymbol{\gamma} \boldsymbol{\gamma}' \tilde{\boldsymbol{A}}_0{}' \boldsymbol{B}_{1,\ell}'}{\sum_{\ell=0}^{H} \boldsymbol{B}_{1,\ell} \boldsymbol{\Sigma} \boldsymbol{B}_{1,\ell}'} \quad s.t. \quad \boldsymbol{\gamma}' \boldsymbol{\gamma} = 1$$

where they set H to be 80 quarters. Using the US data, they show that their identification method is robust to using different vintages of utilization-adjusted TFP series.

VAR specification The information set in the VAR system consists of 9 variables, including the utilization-adjusted TFP, real output, real consumption, real investment, real S&P composition index, federal funds rate, inflation rate, interest rate spread, and business confidence index. The first four series are representative aggregate macro variables and the rest are forward-looking variables that are believed to contain information about future TFP variations. The sample period in my baseline analysis is 1960:Q1-2007:Q3, where the starting point is picked because of the data availability of confidence index, and the end point is chosen so as to avoid possible systematic structural changes in the behavior of micro agents in the firm-level analysis after the Global Financial Crisis. It is not a problem with the identification of TFP news shock in the literature *per se*, and the identified TFP news shocks are quite stable if we extend the sample to more recent episodes.

The VAR is estimated in levels with 4 lags of each variable, an intercept term, but no time trend. To improve precision, I impose a Minnesota prior on the estimation and compute error bands by drawing from the posterior. In the structural VAR system, I order utilization-adjusted TFP the first, followed by real consumption, real GDP, real investment, real S&P

composition index, inflation, federal funds rate, interest rate spread, and confidence. The macro aggregates are all logged and where applicable in real chain-weighted terms and population adjusted. For inflation, I use the growth rate of the GDP deflator. For interest rate spread, I use the difference between the yield of the 5-year treasury bond and the federal funds rate. The results are robust if we use the yield of the 3-month treasury bill for the short rate. The data sources are listed in Appendix Table 1.4.

A description of identified news shocks Appendix Table 1.3 shows the summary statistics of the identified news shocks and Appendix Figure 1.16 plot the time series against the NBER recession quarters. The “good" and “bad" news shocks seem to balance out in the long run and are quite volatile in certain periods, eg. the 1970s and 1980s. There are periods when TFP news shocks experience big swings between favorable news and bad news about future productivity. In the post-1990 sample, the TFP-based news shocks show limited variations, while the patent-based news shock experiences big swings before and after the dot-com bubble and the Global Financial Crisis. On average, the patent-based news shock is the most volatile among the three.

Based on the movements in utilization-adjusted TFP, two TFP news shock series are highly correlated with a correlation coefficient of 0.74. In contrast, the patent-based news shock is only weakly correlated with both TFP news shocks. In particular, the correlation coefficient between patent-based and “TFP-based (KS)” is only 0.05, and is not statistically significant.

Impulse responses to TFP news shocks To show that the TFP news shock identified following Barsky and Sims (2011) is plausible and comparable to the literature, I plot the impulse response functions (IRFs) of macro variables in response to a TFP news shock in Appendix Figure 1.18 over a ten-year horizon. In response to a 1 pp positive TFP news shock, TFP itself does not change on impact by assumption, then it slowly increases by 0.4 percent over a five-year horizon and stays high thereafter. Consumption jumps up by 0.2 percent on impact and increases further to a permanently higher level. Output increases insignificantly on impact but then starts to grow quickly. By the end of the third year, real output has increased by 0.7 percent. The gross private investment drops slightly on impact but starts to grow quickly. After reaching its peak at the end of the second year, investment adjusts back a little bit before reaching its new level. The hump shape and overshooting in investment are quite common in the news shock literature. As evident from the impulse responses of real macroeconomic variables, most actions have taken place in the first 5 years following a TFP news shock.

The deflated stock prices jump up high on impact, and then slowly recovers. The effect is still present at the end of the decade. In terms of the responses of inflation and federal funds rate, we find the so-called "disinflation" puzzle where inflation drops in response to a technology news shock, and the federal funds rate (FFR) closely traces out the inflation response. The slope of term structure increases on impact due to the sharp decline in short rate (FFR), in line with Kurmann and Otrok (2013). Business confidence rises on impact and gradually recovers. All of these results are qualitatively and quantitatively close to Barsky and Sims (2011), Kurmann and Otrok (2013) and Kurmann and Sims (2021) except the IRFs of inflation using the growth rate of GDP deflator. Instead of the hump-shaped response shown here, they find a dramatic decline on impact. If I use the growth rate of CPI instead, the inflation response will be very close to their results.

The impulse responses to TFP news shock identified by Kurmann and Sims (2021) are shown in Figure 1.19. They are roughly in line with their eight-variable VAR impulse responses. Without the zero impact restrictions, using the latest vintage of utilization-adjusted TFP series, the TFP immediately jumps up on impact and responds in an "S" shape as horizons extend. Investment drops more significantly on impact. The deflated stock prices jump up on impact and evolve in a hump shape fashion. The inflation rate declines much more dramatically on impact, and the federal funds rate stays lower for longer.

The average effect of TFP news shocks Appendix Figure 1.20 shows the dynamics of the firm-level average effect of TFP news shocks on cumulative capital growth. In panel (a), the average effect of TFP news shocks identified by Barsky and Sims (2011) looks very similar to the result for the patent-based news shocks, especially if we construct capital stock using linear interpolation. Note that the magnitude of the average effect across different news shock measures is not comparable, as evident from their macro impact on real aggregates. The adjacent imputation greatly reduces the sample size, the resulting sample can be treated as a refined subsample with more precisely measured capital stock. The average effect in this subsample is almost twice as large as the baseline result with similar hump-shaped dynamics over the five-year horizon. In contrast, the average effect of TFP news shocks identified by Kurmann and Sims (2021) is mostly muted under linear interpolation in panel (b). In the refined subsample, the hump-shaped dynamics show up but are still short-lived and less pronounced than in panel (a). Given that the two TFP news shock series are highly correlated, the reason behind the distinct firm-level investment evidence requires further investigation in future research. In the robustness check, I mainly focus on the identification method by Barsky and Sims (2011).

1.8.2 Data Appendix

This appendix contains additional information related to the empirical analysis of this paper, including the data sources of macro-level time series used to identify TFP news shocks in Appendix 1.8.1, the definitions of firm financial variables, and a detailed sample selection process of my firm-level dataset used in the panel regression analysis.

Macro-level data sources In the construction of TFP news shocks, I collect macro-level data from various public sources listed in Appendix Table 1.4.

Construction of firm-level variables

1. Lquid asset ratio: defined as the ratio of cash and short-term investments (`cheq`) to total assets (`atq`).

2. Leverage: defined as the ratio of total debt (`dlcq + dlttq`) to total assets (`atq`).

3. Size: measured as the log of total assets (`atq`) deflated by GDP deflator.

4. Age: defined as the number of years preceding the observation year that the firm has a non-missing share price (`prccq`) since 1961.

5. SA index: calculated following Hadlock and Pierce (2010). Specifically,

$$SA = 0.043 \cdot Size^2 - 0.737 \cdot Size - 0.040 \cdot Age$$

   where size is the log of GDP deflator-adjusted book assets, and age is the number of years the firm has been on Compustat with a non-missing stock price. In calculating this index, size is replaced with log($4.5 billion) and age with thirty-seven years if the actual values exceed these thresholds.

6. Cash flow: defined as income before extraordinary items (`ibq`) plus depreciation and amortization (`dpq`).

7. Cash flow to asset ratio: defined as the ratio of cash flow to total assets.

8. Real sales growth: measured as log-difference in sales (`saleq`) deflated by CPI.

9. Dividend payer: defined as a dummy variable taking a value of one in firm-quarter observations in which the firm paid dividends to preferred stock (`dvpq` is positive).

10. Market equity: defined as closing price (`prccq`) times common shares outstanding (`cshoq`) deflated by CPI.

11. Book equity: measured as shareholders' equity (`seqq`), plus balance sheet deferred taxes and investment tax credit (`txditcq`) if available, minus the book value of the preferred stock (`pstkq`) deflated by CPI.

12. Earnings or EBITDA (earnings before interest, taxes, depreciation, and amortization): the sum of Sales - Net (`saleq`) minus Cost of Goods Sold (`cogsq`) minus Selling, General & Administrative Expense (`xsgaq`)

13. Sectoral dummies: I consider the following sectors: (1) agriculture, forestry, and fishing: 2-digit $sic < 10$; (2) mining: 2-digit $sic \in [10,14]$; (3) construction: 2-digit $sic \in [15,17]$; (4) manufacturing: 2-digit $sic \in [20,39]$; (5) transportation, communications, electric, gas, and sanitary services: 2-digit $sic \in [40,49]$; (6) wholesale trade: 2-digit $sic \in [50,51]$; (7) retail trade: 2-digit $sic \in [52,59]$; (8) service: 2-digit $sic \in [70,89]$.

Sample selection The sample selection process is similar to Jeenas (2019), I exclude all firm-quarters for which:

1. The firm is not incorporated in the U.S., i.e. FIC needs to be "USA".
2. The firm is in the utilities industries (2-digit SIC code is 49), finance, insurance, and real estate industries (2-digit SIC code is between 60 to 69). For firms in the utilities industries, they are heavily regulated and therefore assumptions about cost minimization or profit maximization are unlikely to hold, while for the latter two industries, their balance sheets are quite different from other industries. I also exclude public administration industries (2-digit SIC code is between 90 to 99) since they're not particularly interesting in this context.
3. The measurements of total assets `atq`, sales `saleq`, constructed capital stock `K` are missing or not positive; the measurements of cash and short-term investments `cheq` are missing or negative.

Following Ottonello and Winberry (2020), I also exclude firm-quarter observations that satisfy one of the following conditions in order to exclude extreme observations:

1. Aquisitions `aqcy` is larger than 5% of total assets `atq`.
2. Investment rate is in the top and bottom 0.5% of the distribution.
3. Leverage is higher than 10.
4. Net current assets as a share of total assets is higher than 10 or below -10.
5. Quarterly real sales growth exceeds 1 or below -1.

After applying these sample selection operations, I also drop all firms with cash flow to asset ratio below -1. To improve precision and avoid endogeneity issues in the regression, I drop all firms which are observed before 2007:Q3 for less than 40 quarters and drop the years 1970 and 1971 with too few (< 10) observations left.

1.8.3 Appendix Tables and Figures

Table 1.3. Summary statistics of news shocks by identification methods

(a) Marginal distribution

Variable	Obs	Mean	Std. Dev.	Median	95th Percentile
TFP-based (BS)	183	−0.002	0.736	−0.015	1.076
TFP-based (KS)	183	0.002	0.805	0.032	1.154
Patent-based	183	−0.001	0.86	0.039	1.333

(b) Correlation matrix

	TFP-based (BS)	TFP-based (KS)	Patent-based
TFP-based (BS)	1.00		
(p-value)			
TFP-based (KS)	0.74	1.00	
	(0.00)		
Patent-based	0.13	0.05	1.00
	(0.07)	(0.51)	

Notes: Sample period: 1962:Q1 - 2007:Q3. "TFP-based (BS)" is the TFP news shock series identified by Barsky and Sims (2011), "TFP-based (KS)" is the TFP news shock series identified by Kurmann and Sims (2021), Patent-based shock series is obtained from Cascaldi-Garcia and Vukotić (2022).

Table 1.4. Macro-level data sources

Time Series Description	Code	Source
Gross domestic product	GDPC1	BEA
Real personal consumption expenditure	PCECC96	BEA
Real gross private domestic investment	GPDIC1	BEA
Civilian noninstitutional population	CNP16OV	BLS
Non-farm business sector: hours		BLS
Non-farm business sector: employment		BLS
GDP deflator	GDPDEF	BEA
CPI	CPIAUCSL	BEA
Utilization-adjusted TFP	ltfp_util	BFK-Fernald (2014)
Effective federal funds rate	FEDFUNDS	BG FED
5Y treasury constant maturity rate	GS5	FRED
3M treasury bill: secondary market rate	TB3MS	FRED
S&P 500 index		Robert Shiller's data
E5Y confidence index	Table 29	Michigan Survey

Notes: BEA = U.S. Department of Commerce: Bureau of Economic Analysis, BLS = U.S. Department of Labor: Bureau of Labor Statistics, BG FED= Board of Governors of the Federal Reserve System. FRED: Federal Reserve Economic Data, created and maintained by the Research Department at the Federal Reserve Bank of St. Louis.

Table 1.5. Summary statistics of firm-level variables: refined subsample

(a) Marginal distribution

Variable	Obs	Mean	Std. Dev.	Median	95th Percentile
Investment ($\triangle \log k$)	189356	0.016	0.092	0.002	0.143
Liquid asset ratio (*liq*)	189356	0.155	0.191	0.072	0.581
Leverage (*lev*)	180983	0.258	0.315	0.201	0.705
Size (in log)	189356	4.532	2.057	4.464	7.997
Age	189356	9.927	6.59	9	22.25
Cash flow ratio (*cfr*)	166979	0.011	0.446	0.021	0.068
Pay dividends or not (*paydvd*)	189356	0.1	0.3	0	1

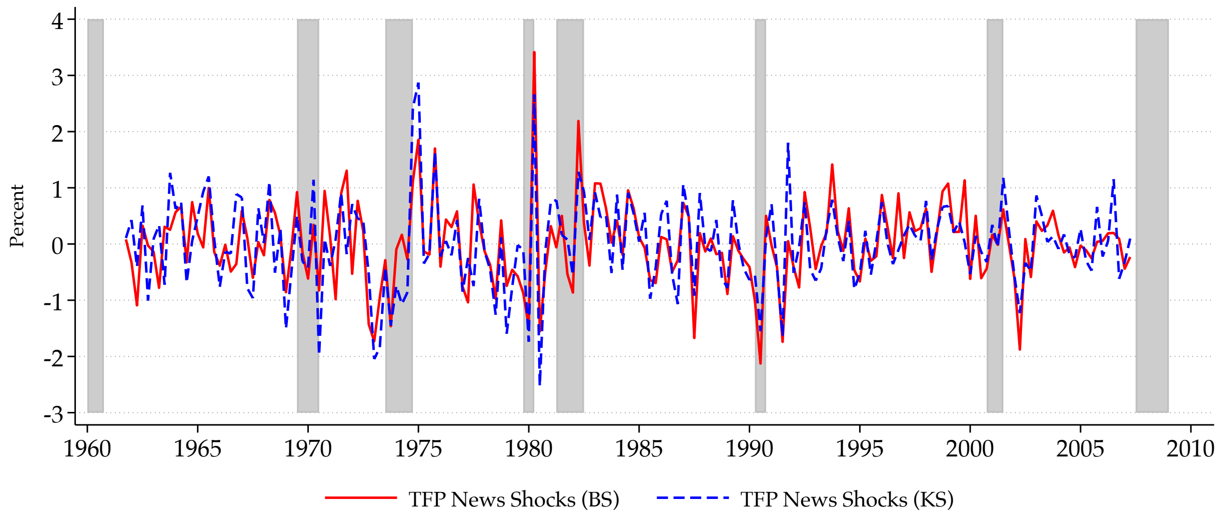
(b) Correlation matrix

	liq	*lev*	*size*	*age*	*cfr*	*paydvd*
liq	1.00					
(p-value)						
lev	-0.31	1.00				
	(0.00)					
size	-0.16	-0.00	1.00			
	(0.00)	(0.91)				
age	-0.06	0.01	0.19	1.00		
	(0.00)	(0.00)	(0.00)			
cfr	-0.00	-0.03	0.03	0.01	1.00	
	(0.08)	(0.00)	(0.00)	(0.02)		
paydvd	-0.08	0.12	0.06	0.04	-0.01	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	

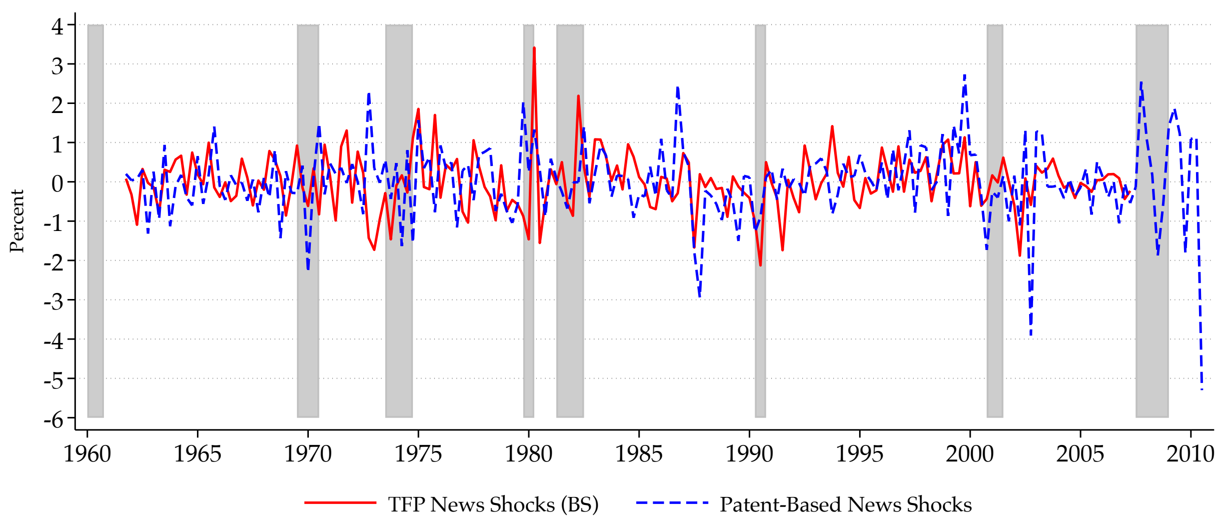
Notes: Liquid asset ratio (*liq*) is measured as cash and short-term investments to assets ratio, leverage as total debt to assets, size as deflated book assets in logarithm; age as the number of years preceding the observation year that the firm has a non-missing share price since 1961, cash flow ratio as the ratio of cash flow to total assets. In the later regression analysis, these variables will be winsorized at 0.5% and 99.5% cutoff. Pay dividend or not is a dummy variable with value 1 if the firm report a non-missing positive dividend. Sample period: 1972:Q1 - 2007:Q3. In the construction of capital stock, I limit the imputation of *ppentq* to non-missing adjacent values.

Figure 1.16. Evolution of news shocks by identification methods

(a) TFP news shocks

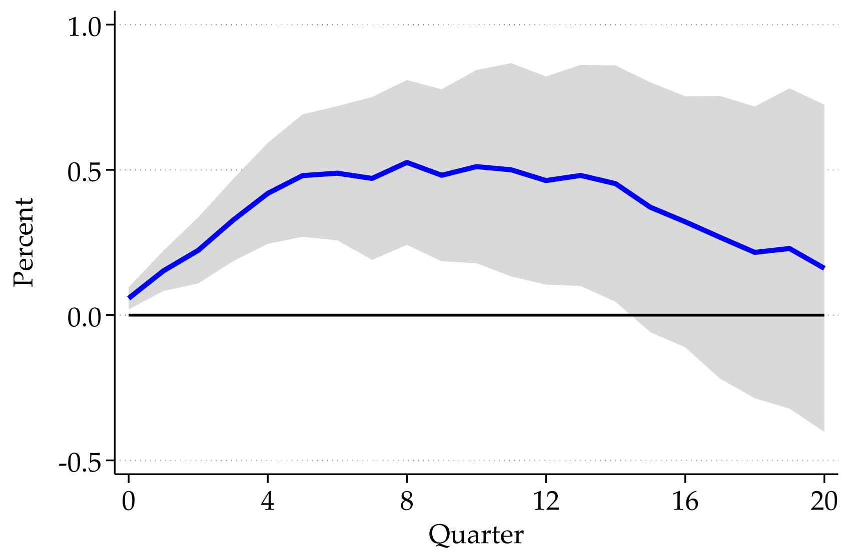


(b) TFP news shocks vs. patent-based news shocks



Notes: Panel (a) compares two TFP news shock series following Barsky and Sims (2011) (BS) and Kurmann and Sims (2021) (KS); panel (b) compares the identified TFP news shock with patent-based news shock. Sample period: 1962:Q1 - 2007:Q3. Grey areas denote the NBER recession quarters.

Figure 1.17. Dynamic effects of patent-based news shocks on investment by SA index

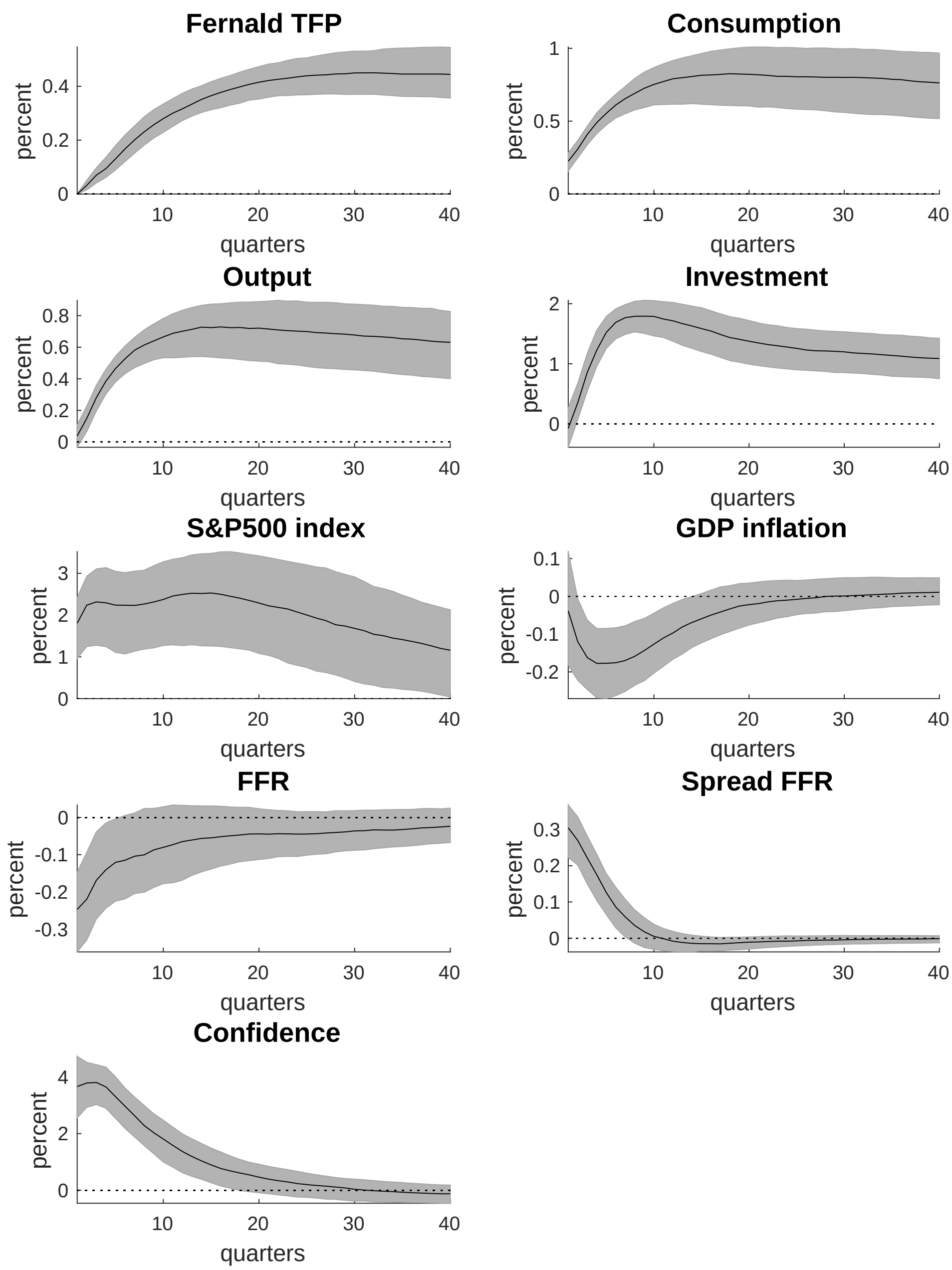


Notes: Reports the estimates of coefficient β_h^x over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x) x_{it-1} + \boldsymbol{\Gamma}'_h \boldsymbol{Z}_{i,t-1} + e_{ith}$$

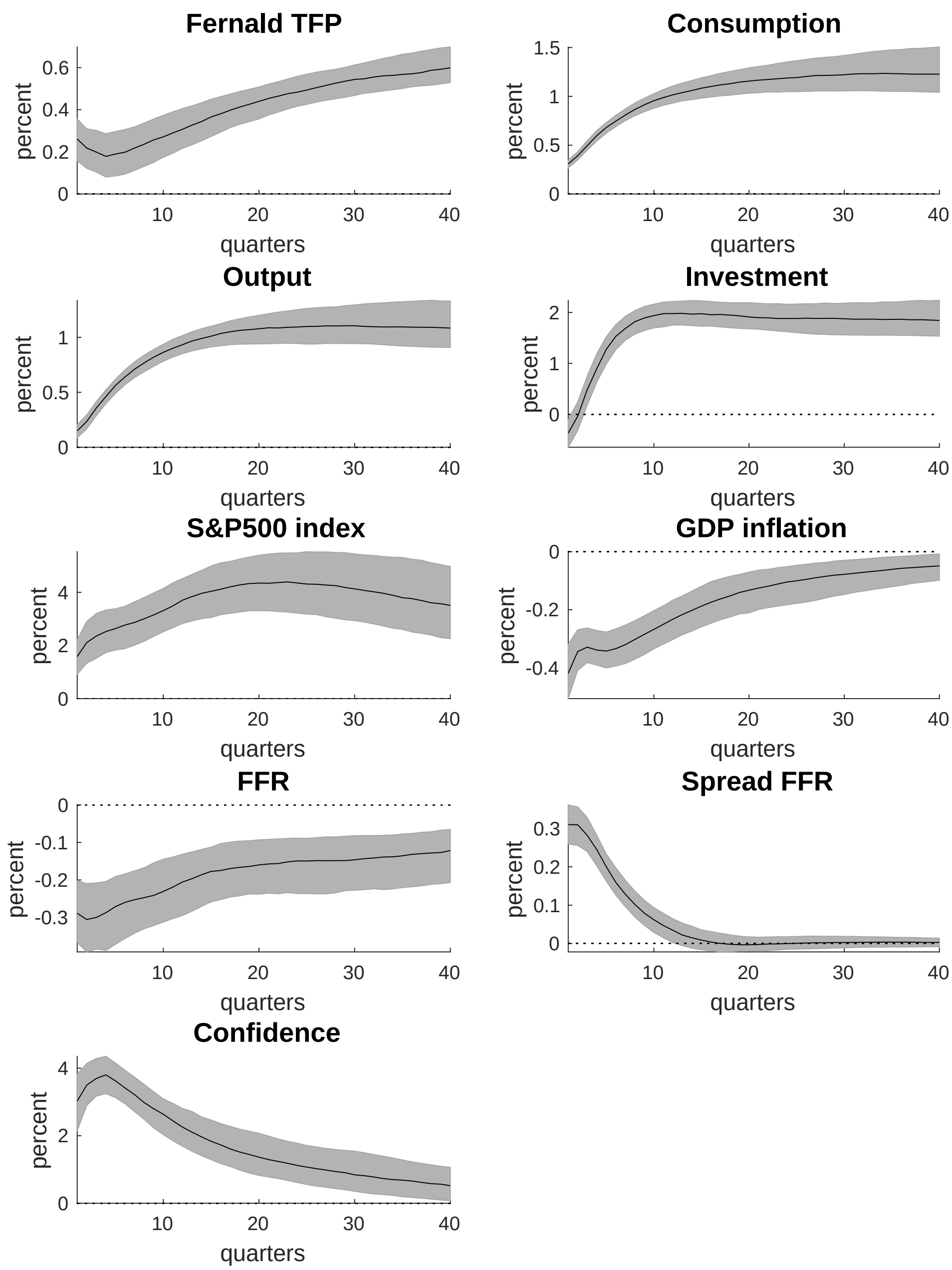
where x is the SA index, α_{ih} is a firm fixed effect, α_{sth} is a sector-by-quarter fixed effect, ϵ_t^{news} is the TFP news shock, and $\boldsymbol{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter. I have standardized $size_{it}$ and age_{it} over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands.

Figure 1.18. Impulse responses to TFP news shocks: BS method



Notes: Identification follows Barsky and Sims (2011). H is set to 40 quarters. Solid black lines are the impulse responses to a 1 pp positive TFP news shocks and grey areas are 68% coverage bands. The units are in percentages. The variable description is in Appendix 1.4.

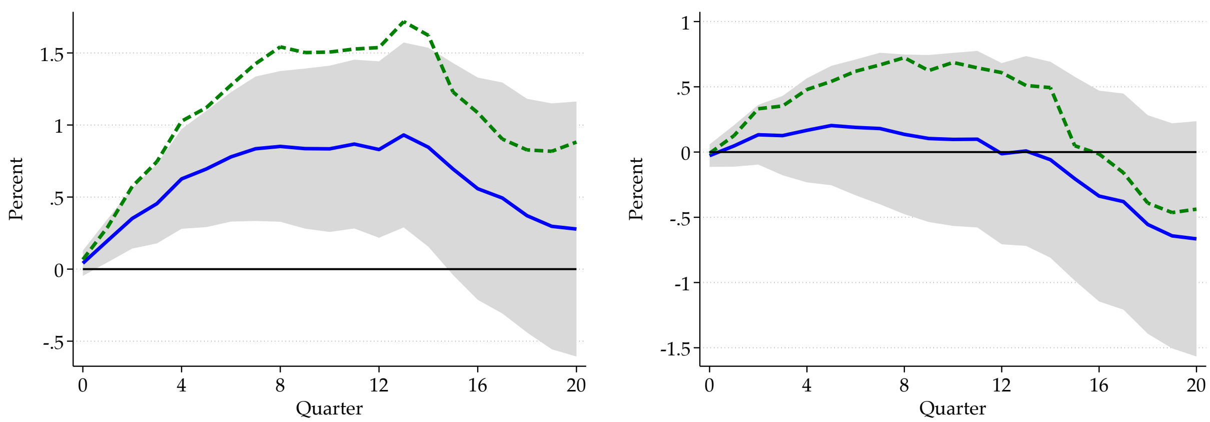
Figure 1.19. Impulse responses to TFP news shocks: KS method



Notes: Identification follows Kurmann and Sims (2021). H is set to 80 quarters. Solid black lines are the impulse responses to a 1 percent positive TFP news shocks and grey areas are 68% confidence intervals. The units are in percentages.

Figure 1.20. Average effect of TFP news shocks on investment by identification methods

(a) Barsky and Sims (2011) **(b)** Kurmann and Sims (2021)

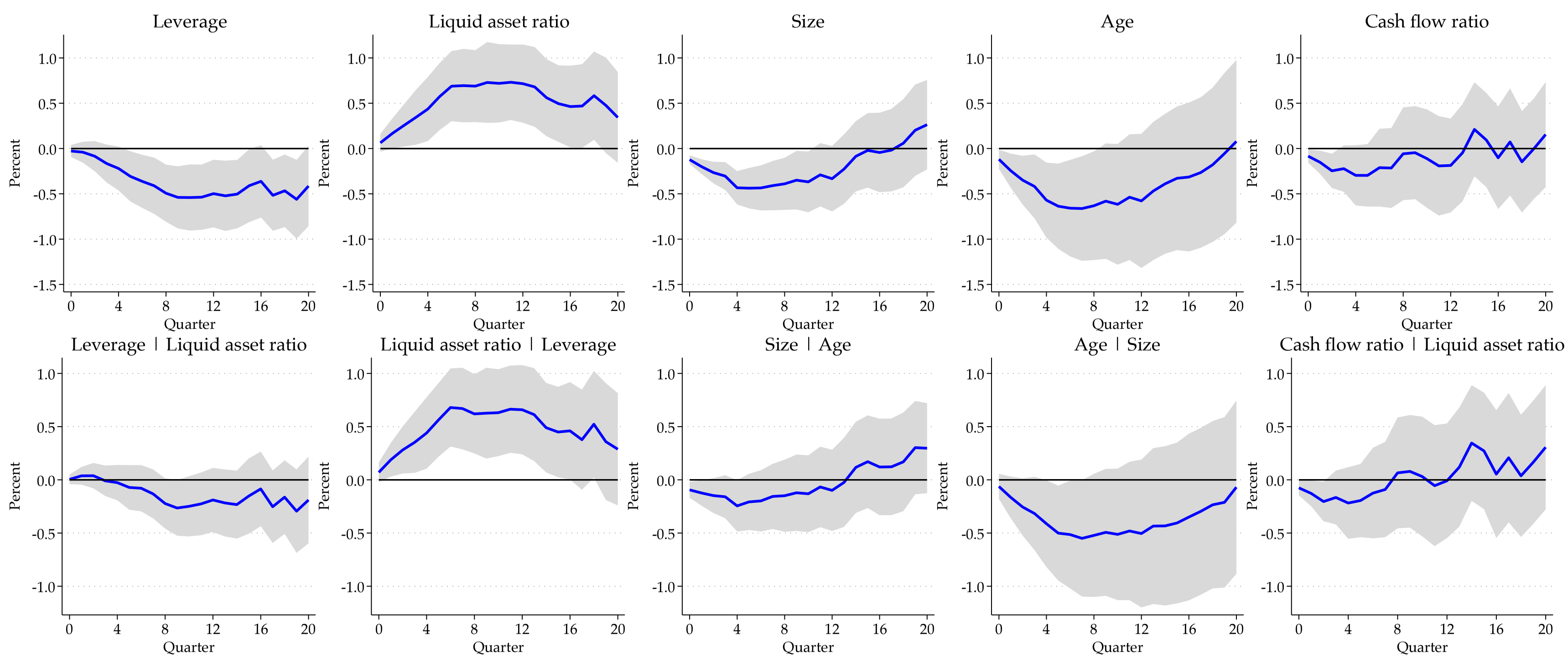


Notes: Reports the coefficient estimates of β_h over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \beta_h \epsilon_t^{news} + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is the news shock, $\boldsymbol{Z}_{it-1}$ is a vector of firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. The blue solid line and the grey areas show the point estimates and 90% confidence bands in the baseline with linear interpolation, while the green dashed line is the average effect if we limit the imputation to non-missing adjacent values when constructing capital stock k.

Figure 1.21. Dynamic effects of TFP news shocks on investment by firm characteristics



Notes: Panel (a) reports the estimates of coefficient β_h^x over quarters h from the separate regression

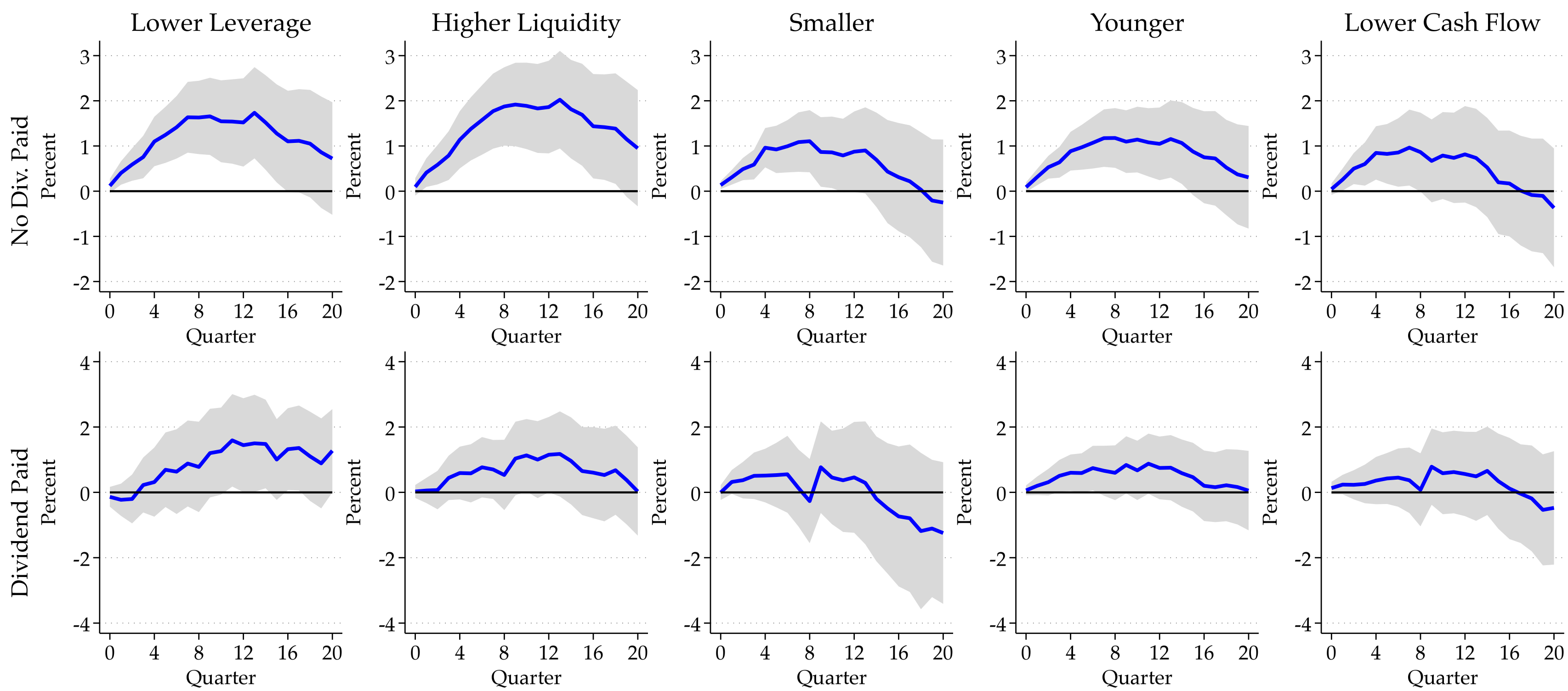
$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x) x_{it-1} + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

while panel (b) reports the estimates of coefficients $\beta_h^{x_1}$ and $\beta_h^{x_2}$ over quarters h from the joint regression

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^{x_1} \epsilon_t^{news} + \gamma_h^{x_1}) x_{1,i,t-1} + (\beta_h^{x_2} \epsilon_t^{news} + \gamma_h^{x_2}) x_{2,i,t-1} + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

where α_{ih} is firm fixed effect, α_{sth} is sector-by-quarter fixed effect, x, x_1, x_2 denote various firm characteristics. ϵ_t^{news} is the news shock, and $\mathbf{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets, and an indicator for the fiscal quarter. I have standardized firm characteristics over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands.

Figure 1.22. Dynamic effects of TFP news shocks on investment by dividend status and others

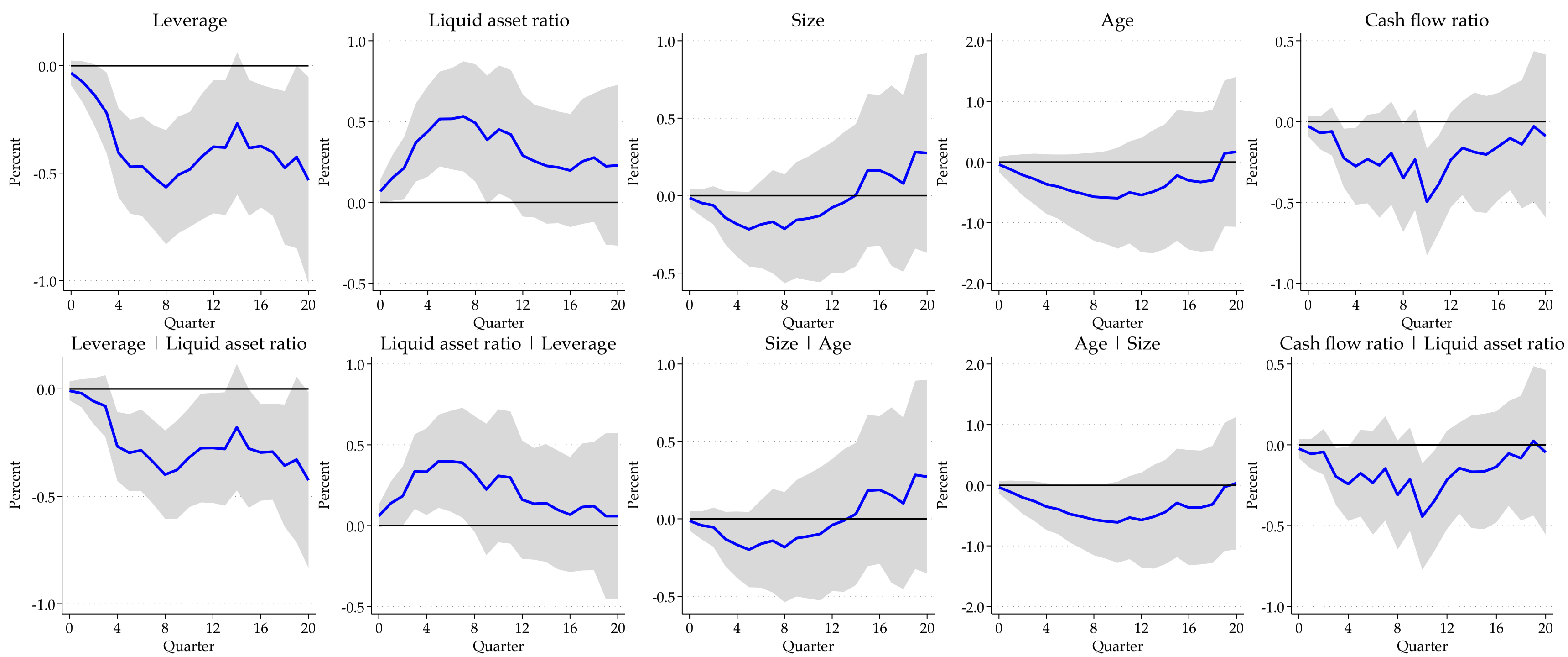


Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands.

Figure 1.23. Dynamic effects of patent-based news shocks on investment by firm characteristics: refined subsample



Notes: Panel (a) reports the estimates of coefficient β_h^x over quarters h from the separate regression

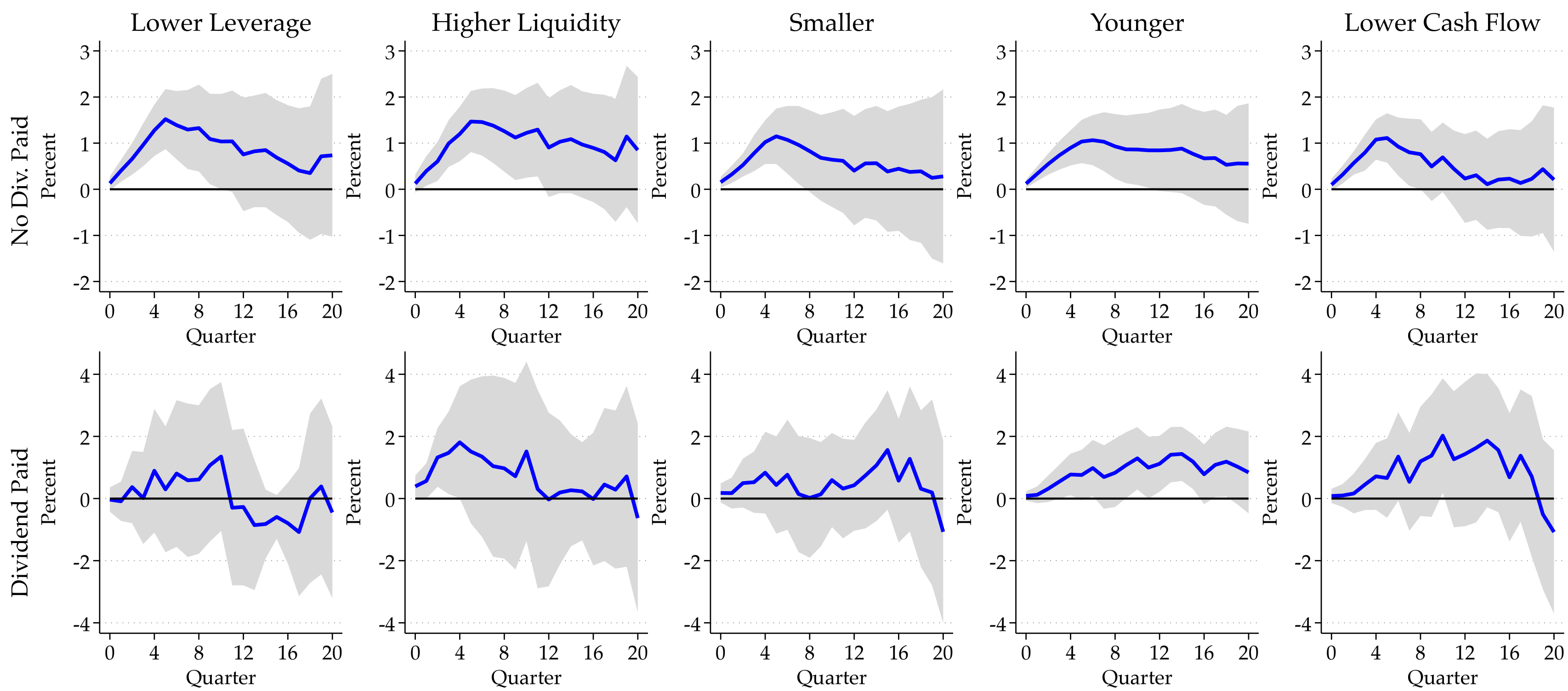
$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x) x_{it-1} + \boldsymbol{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

while panel (b) reports the estimates of coefficients $\beta_h^{x_1}$ and $\beta_h^{x_2}$ over quarters h from the joint regression

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^{x_1} \epsilon_t^{news} + \gamma_h^{x_1}) x_{1,i,t-1} + (\beta_h^{x_2} \epsilon_t^{news} + \gamma_h^{x_2}) x_{2,i,t-1} + \boldsymbol{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

where α_{ih} is firm fixed effect, α_{sth} is sector-by-quarter fixed effect, x, x_1, x_2 denote various firm characteristics. ϵ_t^{news} is the news shock, and $\mathbf{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets, and an indicator for the fiscal quarter. I have standardized firm characteristics over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands. In the construction of capital stock, I limit the imputation of *ppentq* to non-missing adjacent values.

Figure 1.24. Dynamic effects of patent-based news shocks on investment by dividend status and others: refined subsample

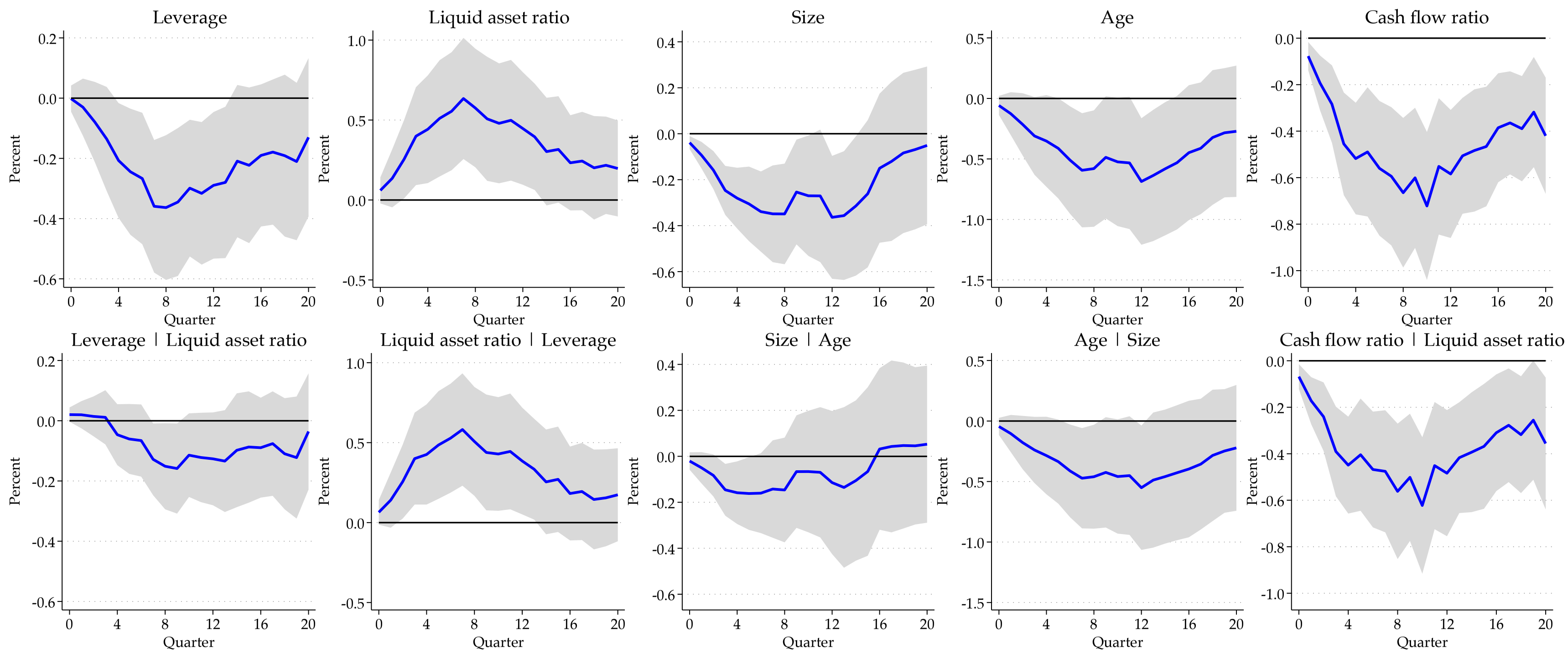


Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands. In the construction of capital stock, I limit the imputation of *ppentq* to non-missing adjacent values.

Figure 1.25. Dynamic effects of patent-based news shocks on investment by firm characteristics: including GFC



Notes: Panel (a) reports the estimates of coefficient β_h^x over quarters h from the separate regression

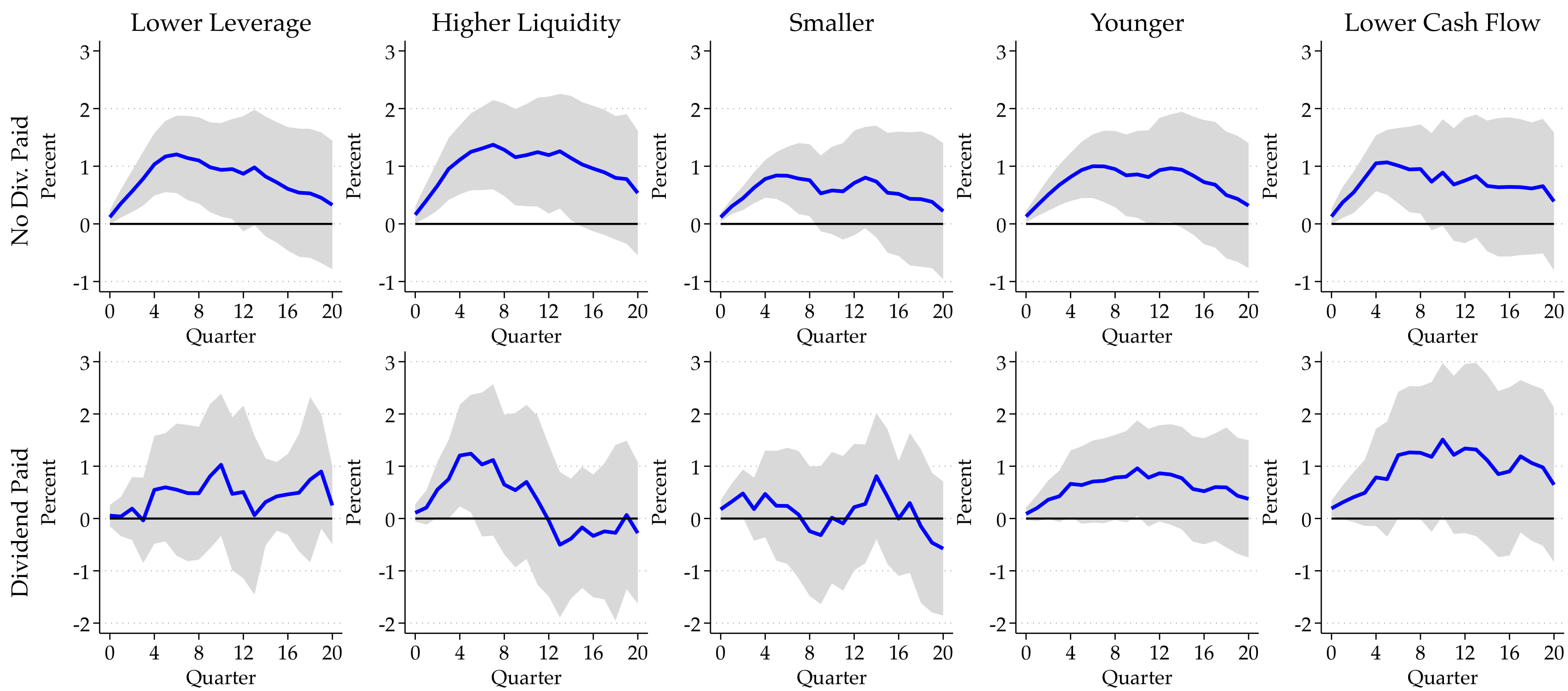
$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x)x_{it-1} + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

while panel (b) reports the estimates of coefficients $\beta_h^{x_1}$ and $\beta_h^{x_2}$ over quarters h from the joint regression

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^{x_1} \epsilon_t^{news} + \gamma_h^{x_1})x_{1,i,t-1} + (\beta_h^{x_2} \epsilon_t^{news} + \gamma_h^{x_2})x_{2,i,t-1} + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

where α_{ih} is firm fixed effect, α_{sth} is sector-by-quarter fixed effect, x, x_1, x_2 denote various firm characteristics. ϵ_t^{news} is the news shock, and $\mathbf{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets, and an indicator for the fiscal quarter. I have standardized firm characteristics over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands. Sample period: 1961:Q1 - 2010:Q4.

Figure 1.26. Dynamic effects of patent-based news shocks on investment by dividend status and others: including GFC

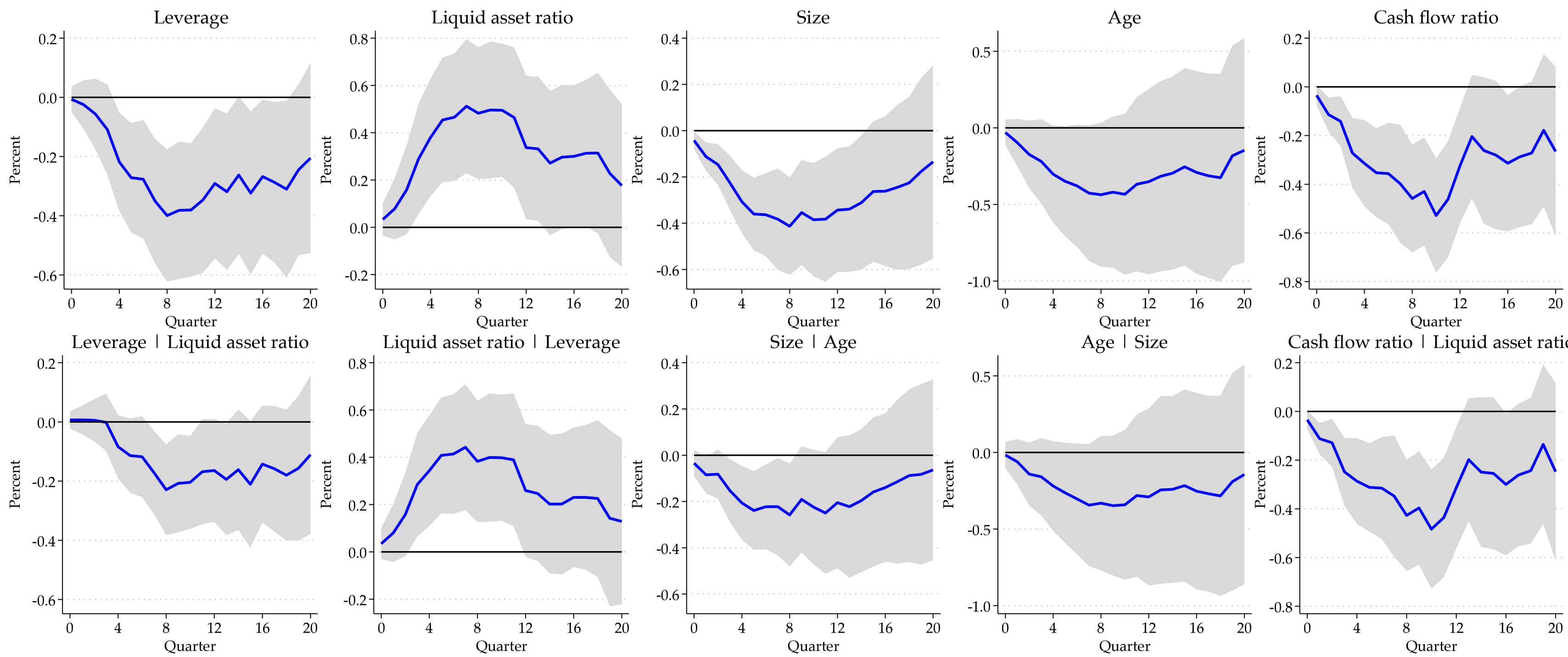


Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands. Sample period: 1961:Q1 - 2010:Q4.

Figure 1.27. Dynamic effects of patent-based news shocks from EEE industries on investment by firm characteristics



Notes: Panel (a) reports the estimates of coefficient β_h^x over quarters h from the separate regression

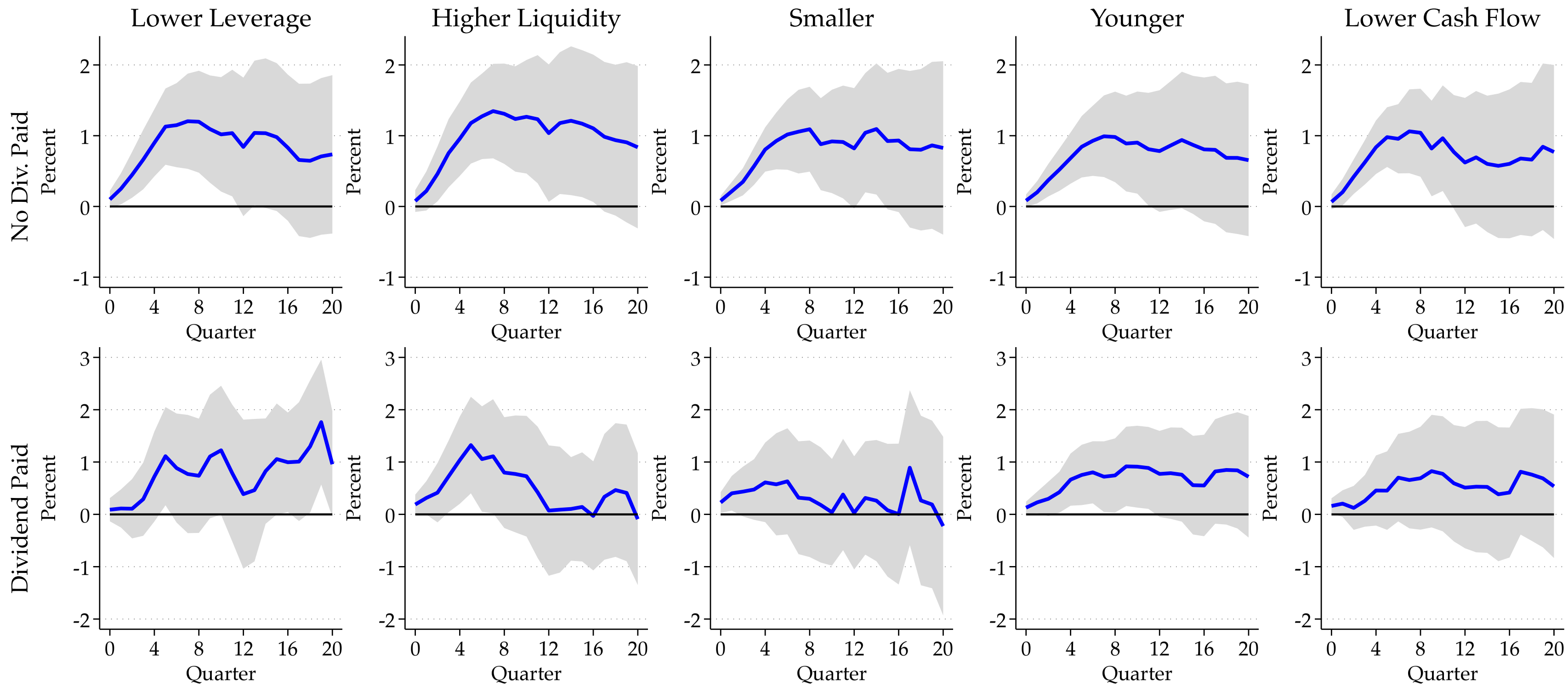
$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x)x_{it-1} + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

while panel (b) reports the estimates of coefficients $\beta_h^{x_1}$ and $\beta_h^{x_2}$ over quarters h from the joint regression

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^{x_1} \epsilon_t^{news} + \gamma_h^{x_1})x_{1,i,t-1} + (\beta_h^{x_2} \epsilon_t^{news} + \gamma_h^{x_2})x_{2,i,t-1} + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

where α_{ih} is firm fixed effect, α_{sth} is sector-by-quarter fixed effect, x, x_1, x_2 denote various firm characteristics. ϵ_t^{news} is the news shock, and $\mathbf{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets, and an indicator for the fiscal quarter. I have standardized firm characteristics over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands. EEE industries refer to the "electronic and electrical equipment" industries.

Figure 1.28. Dynamic effects of patent-based news shocks from EEE industries on investment by dividend status and others

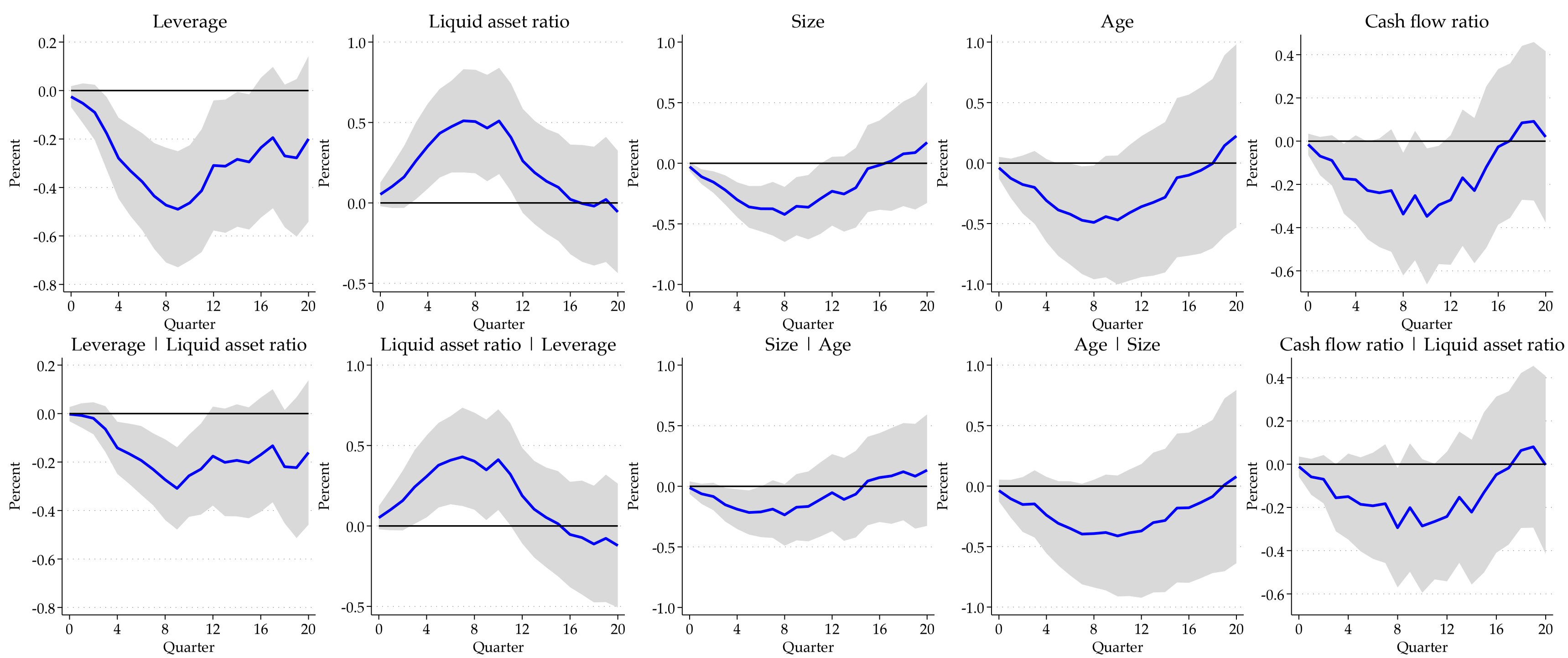


Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands. EEE industries refer to the "electronic and electrical equipment" industries.

Figure 1.29. Dynamic effects of patent-based news shocks from business services on investment by firm characteristics



Notes: Panel (a) reports the estimates of coefficient β_h^x over quarters h from the separate regression

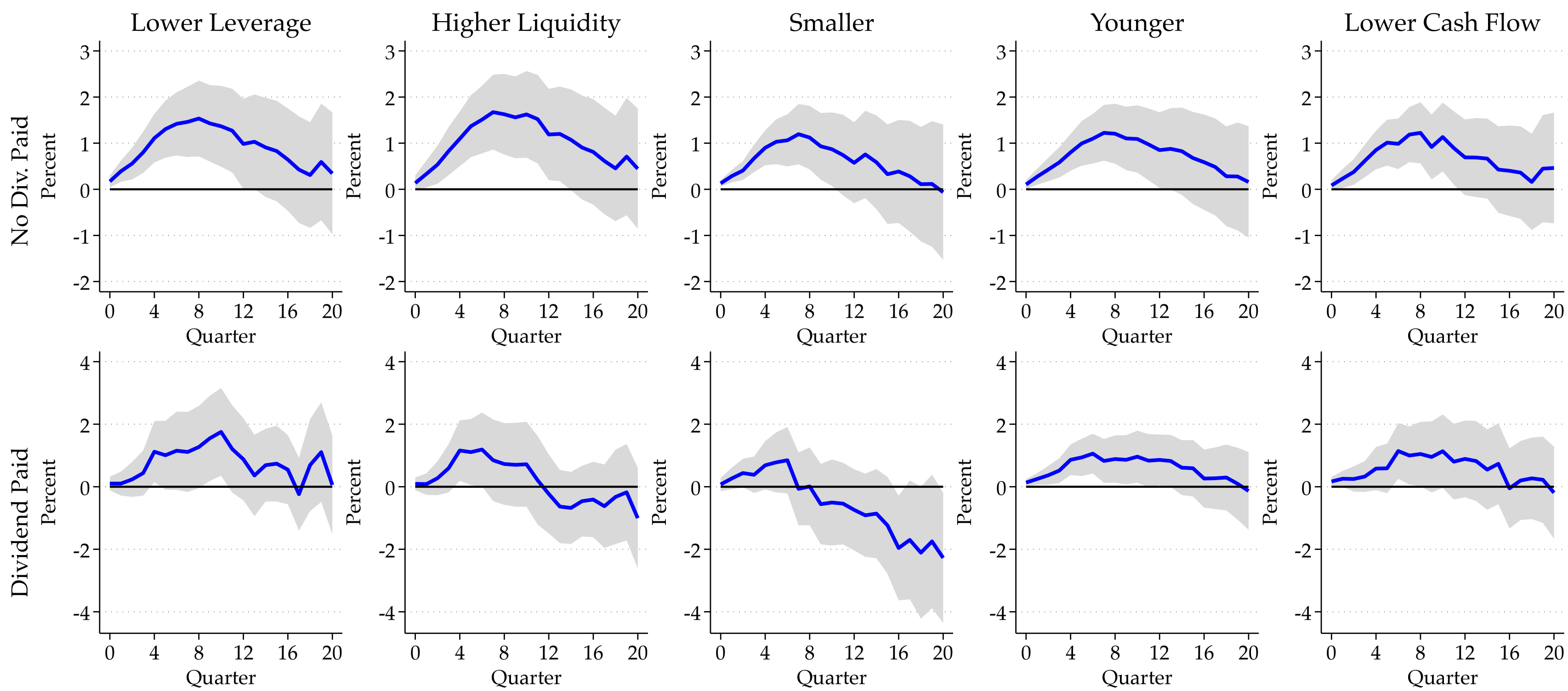
$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x) x_{it-1} + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

while panel (b) reports the estimates of coefficients $\beta_h^{x_1}$ and $\beta_h^{x_2}$ over quarters h from joint regression

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^{x_1} \epsilon_t^{news} + \gamma_h^{x_1}) x_{1,i,t-1} + (\beta_h^{x_2} \epsilon_t^{news} + \gamma_h^{x_2}) x_{2,i,t-1} + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

where α_{ih} is firm fixed effect, α_{sth} is sector-by-quarter fixed effect, x, x_1, x_2 denote various firm characteristics. ϵ_t^{news} is the news shock, and $\mathbf{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets, and an indicator for the fiscal quarter. I have standardized firm characteristics over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands.

Figure 1.30. Dynamic effects of patent-based news shocks from business services on investment by dividend status and others

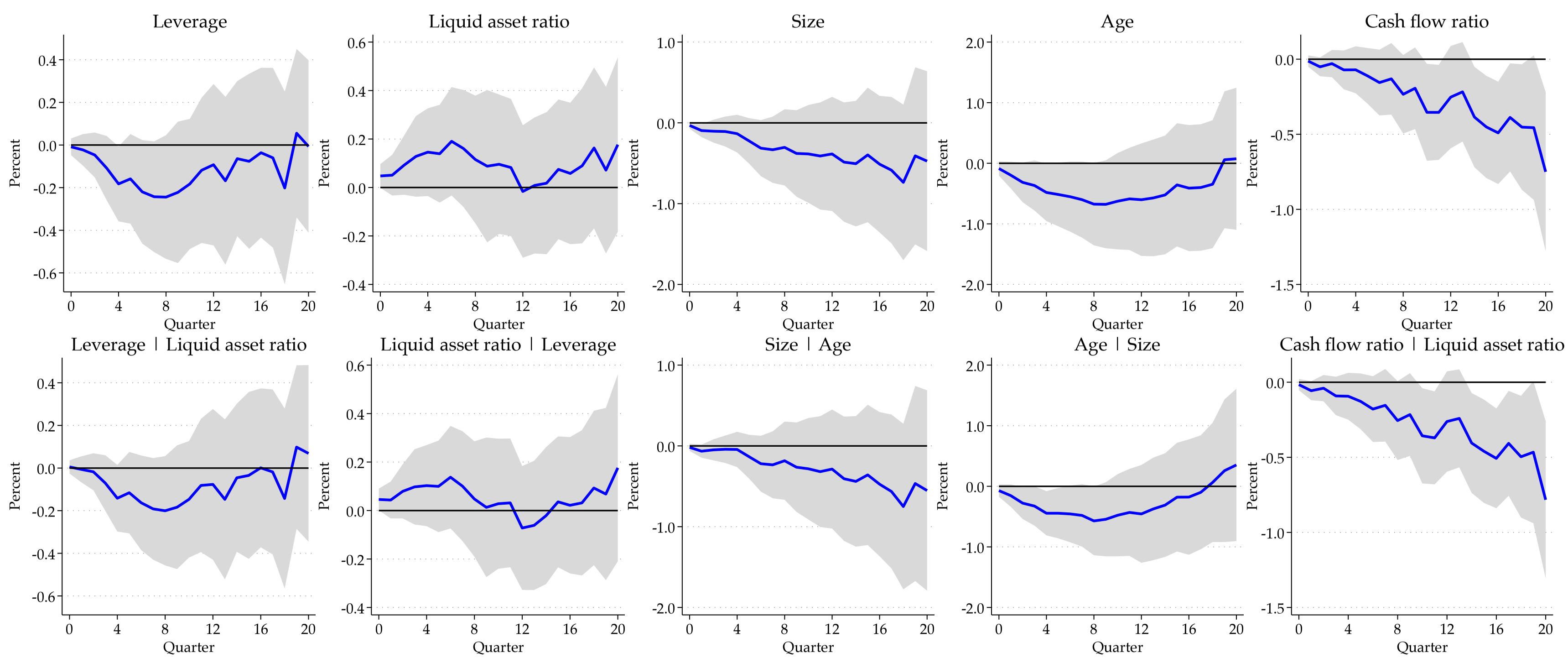


Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of leverage, size, and cash flow to asset ratio in the previous year. Younger firms refer to those with ages less than 15 years. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands.

Figure 1.31. Dynamic effects of patent-based news shocks on investment by firm characteristics: within-firm variation



Notes: Panel (a) reports the estimates of coefficient β_h^x over quarters h from the separate regression

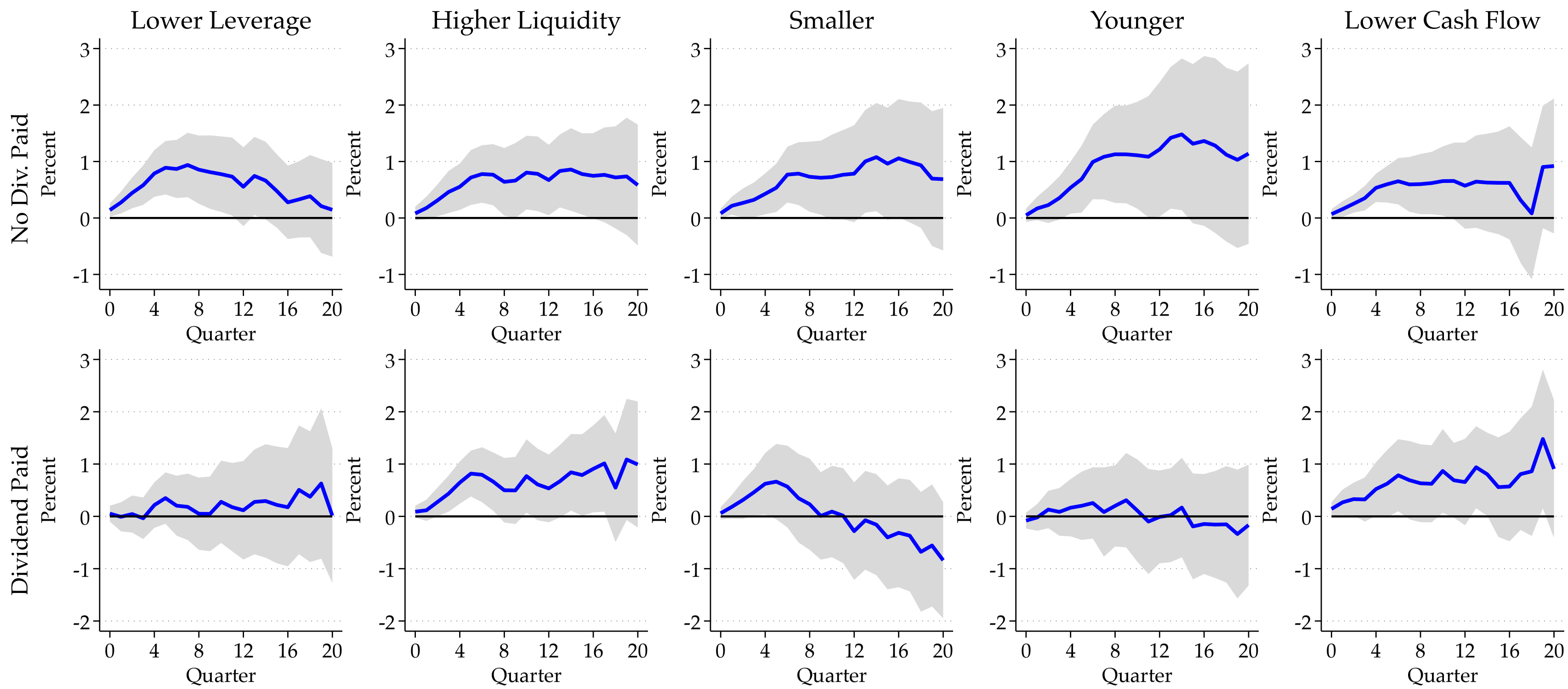
$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^x \epsilon_t^{news} + \gamma_h^x)(x_{i,t-1} - E_i[x_{it}]) + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

while panel (b) reports the estimates of coefficients $\beta_h^{x_1}$ and $\beta_h^{x_2}$ over quarters h from the joint regression

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sth} + (\beta_h^{x_1} \epsilon_t^{news} + \gamma_h^{x_1})(x_{1,i,t-1} - E_i[x_{1,it}]) + (\beta_h^{x_2} \epsilon_t^{news} + \gamma_h^{x_2})(x_{2,i,t-1} - E_i[x_{2,it}]) + \mathbf{\Gamma}_h' \mathbf{Z}_{i,t-1} + e_{ith}$$

where α_{ih} is firm fixed effect, α_{sth} is sector-by-quarter fixed effect, x, x_1, x_2 denote various firm characteristics. ϵ_t^{news} is the news shock, and $\mathbf{Z}_{i,t-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets, and an indicator for the fiscal quarter. I have standardized firm characteristics over the entire sample, so their units are in standard deviations relative to the mean. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands.

Figure 1.32. Dynamic effects of patent-based news shocks on investment by dividend status and others: within-firm variation



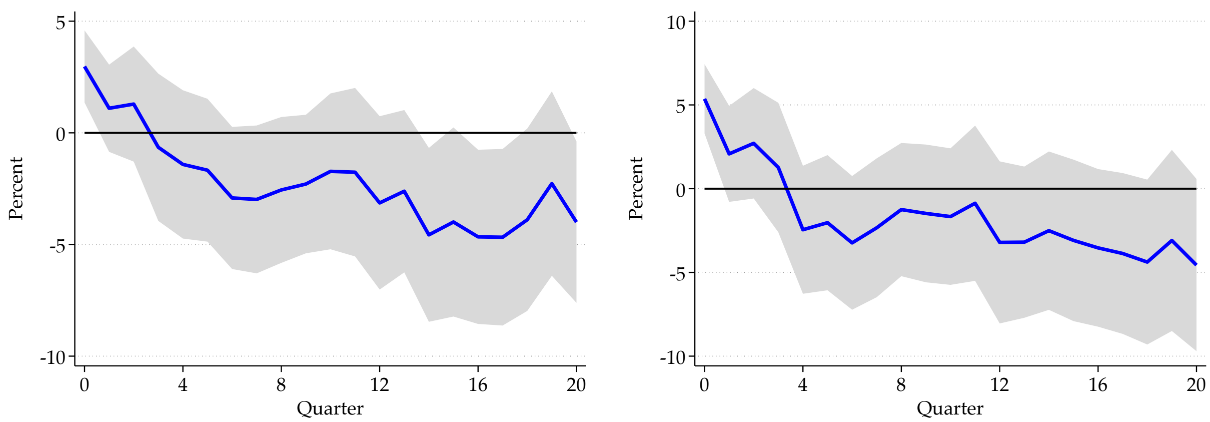
Notes: Reports the coefficient β_h^g over quarters h from

$$\triangle_h \log(k_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \sum_{g=1}^{G} (\beta_h^g \epsilon_t^{news} + \gamma_h^g) \cdot I[\boldsymbol{x}_{i,t-1} \in g] + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is interacted with binary indicators for dividend payer, leverage, liquidity, size, age and cash flow to asset ratio. Higher liquidity firms refer to those in the upper quartile of the distribution of a *given firm*'s liquid asset ratio in the previous year. Lower leverage, smaller, and lower cash flow firms refer to those in the lower quartile of the distribution of a *given firm*'s leverage, size, age, and cash flow to asset ratio in the previous year. $\boldsymbol{Z}_{it-1}$ is a vector of lagged firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. Standard errors are two-way clustered by firm and time. The grey areas indicate 90% confidence bands.

Figure 1.33. Average effect of news shocks on share prices by identification methods

(a) Patent-based news shocks **(b)** TFP news shocks



Notes: Reports the coefficient estimates of β_h over quarters h from

$$\triangle_h \log(y_{i,t+h}) = \alpha_{ih} + \alpha_{sqh} + \beta_h \epsilon_t^{news} + \boldsymbol{\Gamma}_h' \boldsymbol{Z}_{it-1} + \boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1} + e_{ith}$$

where y is the share price deflated by CPI. α_{ih} is a firm fixed effect, α_{sqh} is a sector-by-quarter seasonal fixed effect, ϵ_t^{news} is the news shock, $\boldsymbol{Z}_{it-1}$ is a vector of firm-level controls containing sales growth, size, current assets as a share of total assets and an indicator for the fiscal quarter, and $\boldsymbol{B}_h(L)\boldsymbol{Y}_{t-1}$ is lag polynomials of a vector of macro-level controls including lagged GDP growth, inflation rate and unemployment rate up to 4 lags. The blue solid line and the grey areas show the point estimates and 90% confidence bands. The TFP news shock is identified following Barsky and Sims (2011).

Chapter 2

TFP Growth Regimes and the State Dependence of the Slope of the Phillips Curve

The connection between the slack in the economy or unemployment and inflation was very strong if you go back 50 years, and it's got weaker and weaker and weaker to the point where it's a faint heartbeat that you can hear now. ...... At the end of the day, there has to be a connection because low (un)employment will drive wages up, and ultimately higher wages will drive inflation (up).

Jerome Powell (July 2019)

2.1 Introduction

The trade-off between inflation and economic slack is traditionally inferred from the Phillips curve and is at the core of monetary policymaking. Recent episodes of high and fast-growing inflation rates bring the Phillips curve back to public attention and policy debate. Thanks to the massive policy measures put forward by both fiscal and monetary authorities in the last couple of years, the pandemic recession turned out to be short, but the economy is now faced with a dilemma: stagflation. Although the Federal Reserve had taken increasingly bigger steps to raise interest rates, yet the inflation rate reached a four-decade high level of 9.1% in June 2022 and stays high till today. If the trade-off between inflation and economic slack is really weak, or in other words, the Phillips curve is almost flat, the contractionary monetary policy that aims to cool down the aggregate demand is more likely to push the economy towards recession without effectively bringing down the inflation rate.

Many scholars have documented that the slope of the Phillips curve has kept flattening over time, in terms of both reduced-form "Phillips correlation" and structural estimation.[1] The literature on the "missing disinflation" puzzle during the Great Recession and "missing inflation" puzzle in the slow recovery thereafter have led some economists to go so far as to declare that the Phillips curve is dead or at least "hibernating".[2] In the policy circle, policymakers also started to worry less about the inflation pressure from low unemployment due to the weak trade-off prior to the pandemic. The quote from Federal Reserve Chairman Jerome Powell in his testimony before the House Financial Services Committee in July 2019 is one of the examples.

While the stylized fact of a flatter Phillips curve has been reasonably well established, the precise reasons for this change are not well understood. Anchoring of inflation expectations due to the successful conduct of monetary policy and globalization are two possibilities in early discussions. This paper attempts to provide an alternative explanation and considers the slope change jointly with the secular trends of productivity slowdown amid declining business dynamism in major advanced economies. Recent literature on the macroeconomic implications of the rising market power and declining business dynamism points to a new possibility. It has been shown that the impact of market concentration on the flattening of the Phillips curve can be quite sizable.[3] The analyses along this vein either take market structure as given or abstract away the determination of market structure by long-run factors. However, we do observe that the declining business dynamism since the 1980s, and more strikingly, since the 2000s is accompanied by productivity slowdown among major advanced economies.

[1] Among others, see Kuttner and Robinson (2010), Stock and Watson (2019).

[2] For brief discussions about the possible explanations, see Coibion and Gorodnichenko (2015) on the "missing disinflation" puzzle, see Ratner and Sim (2022) and Heise et al. (2022) on the "missing inflation" puzzle. See e.g., Hooper et al. (2020) for a discussion on the "hibernating" Phillips curve.

[3] In a calibrated model, Wang and Werning (2022) show that going from monopolistic competition to an oligopoly with $n = 3$ firms doubles the half-life of the price level following monetary shocks, or equivalently divides the slope of the Phillips curve by four. See also Baqaee et al. (2021); Fujiwara and Matsuyama (2022).

Is the flattening Phillips curve a by-product of the same structural changes that have led to the recent secular trends in both market structure and productivity growth? Are these events merely coincidentally observed in recent years? Or, have they been stably linked for a long time? If the latter is true, we should expect the slope of the Phillips curve to be state-dependent on the trend productivity growth.

To identify the slope difference of the Phillips curve in periods of high versus low productivity growth, I rely on the long-run cross-country variations. The reason for the coverage of a long history is twofold. For one thing, it usually takes decades for an economy to experience a full round of both high and low productivity growth. For another, we need multiple switches between two regimes or states to identify a robust state-dependent relationship. Productivity growth in the modern economy experiences rises and falls constantly due to changes in both long-run and short-run factors. My focus is on the low-frequency component that describes the technological waves. Following the standard practice in the growth accounting literature, I assume that an economy is alternating between two "growth regimes": periods of high and low growth in trend productivity. Take the post-WW2 U.S. economy for example, the regime classification result in Kahn and Rich (2007) or Fernald (2014*b*) features a handful of regime switches over more than fifty years.

Furthermore, time series variations alone might not be rich enough for our purpose. In the single-country case, if two regime variables say the level of inflation and the trend productivity growth, comove in the time horizon, then the time series variations will not allow us to separate the effects of different "state" variables. This is indeed the case for the post-WW2 U.S. economy prior to the Global Financial Crisis. The low-growth regime in the 1970s and early 1980s coincides with the "Great Inflation" era, and the high-growth regime from the mid-1990s to mid-2000s coincides with the "Great Moderation" era. Existing literature has argued that the level of inflation may affect firms' pricing behavior, hence the slope of the Phillips curve may differ across growth regimes because of the difference in inflation levels.[4] Therefore, it is quite data-demanding to study the growth regime dependence using

[4] Gabriel (2021) uses newly assembled data for 18 advanced economies between 1870 and 2020 and shows that the slope of the wage Phillips curve is flatter in a low inflation environment. Ramos-Francia and Torres (2008) estimate the hybrid New Keynesian Phillips curve for the low inflation sub-sample 1997-2006 in Mexico

the time series data of a single country. In contrast, cross-country variations in the time path of growth regimes and other macro variables could provide important heterogeneity for us to identify the different roles played by different regime variables.

Fortunately, two high-quality longitudinal databases meet the purpose of this study and are ready for public access. The first database is the *Long-Term Productivity Database* from Bergeaud, Cette and Lecat (2016). They offer data on Total Factor Productivity (TFP), labor productivity, capital intensity, and GDP per capita for 23 economies since 1890.[5] The second database is the *Macrohistory Database* from Jordà, Schularick and Taylor (2017). It captures rich macroeconomic and asset price dynamics for the near-universe of advanced economies since 1870.[6] I use the *Long-Term Productivity Database* to classify growth regimes and merge the classification results with the *Macrohistory Database* to estimate the Phillips curve across regimes. The merged dataset consists of 17 advanced economies that are arguably close in technological advancement and economic institutions for the sample period between 1890 and 2012.

In terms of the empirical strategy to estimate the New Keynesian Phillips curve (NKPC), I rely on the state-of-the-art method of estimating structural forward-looking macroeconomic equations proposed by Barnichon and Mesters (2020). To deal with the endogeneity issues pervasive in the estimation of forward-looking macro equations, Barnichon and Mesters' approach consists of projecting the structural equation of interest on the space spanned by the present and past values of some well-chosen structural shocks. They argue that identified monetary policy shocks are valid instruments to estimate the NKPC.

Instead of focusing on robust inference, this paper adapts the two-step approach implied by their theoretical justification and pays special attention to the exogeneity condition of valid instruments in a panel setting. In particular, the identified monetary policy shocks may still contain the information that the policymakers and the public know but the econometricians don't. Also, monetary policy shocks may influence inflation through channels other

and find that prices on average remain fixed for a longer horizon.

[5]The data are available at http://www.longtermproductivity.com.

[6]The data are available at https://www.macrohistory.net/database.

than domestic demand.[7] It is therefore recommended that we control for as much information available to the agents as possible in the projection system. The identified monetary policy shocks used in this paper are based on the trilemma mechanism proposed by Jordà, Schularick and Taylor (2020*a*). For peg countries that allow free capital mobility, the base country's interest rate policy surprises can function as a source of natural experiments in domestic monetary policy. For the base countries, I use the narrative measures from Romer and Romer (2004) and Cloyne and Hürtgen (2016) for the U.S. and U.K., respectively.

The main empirical findings of this paper are as follows: first and foremost, the estimation of the NKPC suggests that the Phillips curve is *steeper* in *high* productivity growth regime. The difference in slope estimates across regimes is quantitatively large, especially when we measure the output gap using a low-pass HP filter. Second, growth regime classified using the cross-country variations over a long history is only weakly correlated with other regime indicators for whether the economy is in a boom or slump, in low or high inflation, and in a credit boom or slowdown, indicating that stratifying the data by productivity growth regime provides new and interesting perspectives. Third, monetary policy is subject to clear and different trade-offs between its nominal and real effects across productivity growth regimes. In response to contractionary monetary policy shocks, prices fall dramatically in the high-growth regime while they are almost muted in the low-growth regime in the first four years. In contrast, output responses are relatively stronger in the low-growth regime. We don't find such clear trade-offs in the state-dependent analyses of other regimes.

To the best of my knowledge, this paper is the first to establish the empirical connection between the slope of the Phillips curve and long-run productivity growth. The conventional wisdom is that short-run fluctuations and long-run growth are determined separately by assuming classical dichotomy. In a standard New Keynesian model with constant elasticity of substitution (CES) preferences and monopolistic competition, market structure and growth are irrelevant to firms' pricing behavior, thus the slope of the Phillips curve. Recent literature starts to relax the standard assumptions to study the impact of the market

[7] see e.g., Razin and Binyamini (2007); Borio and Filardo (2007). The idea is that as a larger proportion of the rise in domestic demand is satisfied through imports, rather than domestic production, increases in the domestic output gap will have a smaller impact on domestic marginal costs, and hence on inflation.

concentration on the slope of the Phillips curve. Wang and Werning (2022) generalize the New Keynesian model by allowing for dynamic oligopolistic competition between any finite number of firms in each sector. Fujiwara and Matsuyama (2022) extend the canonical New Keynesian model by introducing endogenous entry and Homothetic Single Aggregator (HSA) demand systems. While both successfully demonstrate that market concentration flattens the Phillips curve, it remains unclear to us how these two events are related to the secular trend of productivity slowdown. Therefore, the empirical findings of this paper also have strong theoretical implications.

To explain the growth regime dependence of the slope of the Phillips curve, I propose a mechanism that links productivity growth, market structure, and the slope of the Phillips curve. Using a calibrated endogenous growth model with CES preferences and oligopolistic competition, I show that the structural changes that are attributed to higher productivity growth could also lead to a more competitive market. If the price elasticity of demand rises with market competition, then the pass-through of marginal costs due to fluctuations in short-run demand will be larger, indicating a steeper Phillips curve, and vice versa. This mechanism is qualitatively in tune with the recent secular trends of a flattening Phillips curve and productivity slowdown amid rising market concentration and markups in many advanced economies.

Related Literature. This paper relates to at least three broad strands of literature. First, this paper joins the discussion of state dependence of the efficacy of monetary policy. At the aggregate level, previous literature has well-recognized that monetary policy has asymmetric effects in the boom versus the slump. Tenreyro and Thwaites (2016) find that the effects of monetary policy are less powerful in recessions. And contractionary policy shocks are more powerful than expansionary shocks. A similar message is delivered by Barnichon and Matthes (2018). At the disaggregated level, leverage is another factor that affects the efficacy of monetary policies. Cloyne et al. (2020) work on the household mortgage indebtedness, and Ottonello and Winberry (2020) focus on the firm leverage. Jordà, Schularick and Taylor (2020*a*) conduct a comprehensive state-dependent analysis using the *Macrohistory Database* and trilemma monetary shocks. They confirm that output response appears to be quite strong in booms, when inflation is above 2% and during the mortgate credit boom, but considerably weaker in slumps, low inflation episodes, and low growth in mortgage credit. This paper contributes to this literature by introducing a new source of state dependence, i.e., productivity growth regime, and shows that monetary policy faces different trade-offs between its nominal and real effects across growth regimes. The asymmetry of monetary policy effects constitutes the direct reason for the slope difference in the estimated NKPC across growth regimes.

Second, this paper links the recent studies on the declining business dynamism and productivity slowdown to the flattening Phillips curve. Akcigit and Ates (2021) summarize the ten trends related to the declining business dynamism in the U.S. since the 1980s. The entry rate of new businesses, the job reallocation rate, and the labor share have all been decreasing, yet the profit share, market concentration, and markups have all been rising, see the reference therein for an extensive discussion. Some of these trends, if not all, have been examined to be also present in some other OECD countries. Calligaris et al. (2018) find strong evidence that markups are increasing over the period 2001-2014 using Orbis data for 26 high-income economies. Bajgar et al. (2019) document a clear increase in industry concentration in Europe as well as in North America between 2000 and 2014. Calvino et al. (2020) highlight that declines in business dynamism have been pervasive in many OECD countries and are driven by dynamics occurring at a disaggregated sectoral level, rather than reallocation across sectors. Koltay et al. (2022) find an overall tendency towards oligopolistic structure and a sustained increase in aggregate profitability over the recent decades for European economies.

Efforts have been made to link these secular trends and study their causes and consequences. Aghion et al. (2019) propose a theory in which the driving force is falling overhead costs of spanning multiple products or a rising efficiency advantage of large firms. Akcigit and Ates (2019) use a general equilibrium model of endogenous firm dynamics to assess the relative importance of multiple potential mechanisms that can drive the observed trends in business dynamism, and their results highlight the dominant role of a decline in the intensity of knowledge diffusion from the frontier firms to the laggard ones. Bearing these trends in mind, I establish new stylized facts linking the flattening Phillips curve with productivity slowdown and present a conceptual framework to rationalize the growth regime dependence of the slope of the Phillips curve through the changes in market structure due to structural changes. The proposed channel is also inspired by the extensive literature on innovation, growth, and market competition, see Aghion et al. (2014) and references therein.

In addition to the literature discussed above, my work builds on the vast literature on the Phillips curve. This strand of literature starts from Phillips (1958) and Samuelson and Solow (1960), followed by massive efforts to build theoretical models to rationalize the empirical relationship, e.g., Friedman (1968), Phelps (1967). After Lucas (1972), New Keynesian economists in particular started to incorporate nominal rigidity into rational expectation models. Since then, the NKPC has gained popularity from its appealing theoretical micro-foundations and what appeared to be early empirical success (Mavroeidis et al., 2014). More recent papers working on more flexible and robust estimation of the NKPC using time series data include Ball and Mazumder (2019), Stock and Watson (2019), Barnichon and Mesters (2020), and Del Negro et al. (2020). For the Phillips curve estimation using regional data, see Hazell et al. (2022) and the references therein. My paper applies Barnichon and Mesters (2020) method in a panel setting and estimates the slope of the "average" NKPC for the group of advanced economies using trilemma monetary shocks as instruments.

The remainder of the paper proceeds as follows. In Section 2.2, I present the baseline growth regime classification results using the regime-switching regression model. The section ends with a comparison between the growth regime and other regime variables considered in the literature. In Section 2.3, I provide more details on the estimation procedure of the NKPC and present the estimation results across growth regimes. In Section 2.4, I show that the baseline results are robust to alternative regime classification methods, output gap measures, projection horizons, and sample selections. At the end of the section, I compare the asymmetric monetary policy effects across different regime variables and provide further support for the unique growth regime perspectives. Section 2.5 discusses the potential mechanism that could square with the empirical results. Section 2.6 concludes.

2.2 Growth Regime Classification

The first part of my empirical analysis is to classify the growth regime for a wide group of advanced economies over a long history. In the growth accounting literature, scholars usually focus on labor productivity growth, which can be decomposed into capital deepening, improvement in labor quality, and growth in total factor productivity (TFP), see Fernald (2014*a*). This paper classifies growth regimes based on the low-frequency component of TFP growth for the following two reasons: first, TFP growth is mainly driven by the medium and long-run factors such as innovation and technological progress, which is the main focus of my study. Improvement in labor quality is quite stable over time and is mainly driven by secular trends in labor supply. Capital deepening is mainly driven by capital accumulation, which varies mostly at business cycle frequency. Second, the growth regime classification by Fernald (2014*b*) for the postwar U.S. economy shows that the main difference in labor productivity growth across growth regimes comes from the variations in TFP growth, while the contributions of capital deepening and labor quality are quite stable in both regimes over time.

To classify the growth regime, I incorporate the data from Bergeaud, Cette and Lecat (2016). Their *Long Term Productivity Database* measures aggregate productivity for a rich panel of countries over a long period of time. In version 2.4 published in August 2020, the database provides consistent measures of labor productivity, TFP, GDP per capita, capital intensity, and the average age of equipment capital stock for 23 countries dating back to 1890. Given the wide coverage of countries and sample periods, I am able to classify growth regimes with different methods using their TFP measure, and then merge the classification results with Jordà-Schularick-Taylor *Macrohistory Database* to estimate the slope of the Phillips curve for 17 advanced economies over the sample period 1890-2012.

2.2.1 Baseline Classification: Regime-Switching Model

Existing literature has proposed a couple of methods for detecting changes in trend productivity growth, both of which were first applied to study the U.S. experience. One is the regime-switching dynamic factor model proposed by Kahn and Rich (2007), and the other is based on Bai and Perron (1998, 2003) test for multiple structural changes applied in Fernald (2014*b*). My baseline classification method is closely related to Kahn and Rich (2007) in that I also assume that the trend component of TFP switches between high-growth and low-growth regimes with some probability at any point in time. However, instead of building more complicated models allowing for regime changes in both permanent and transitory components, I first filter out the cyclical component and then apply the regime-switching regression model on the trend component for each country.

There are 17 countries in the merged dataset, the 3-digit ISO country codes are given as follows: { AUS, BEL, CAN, CHE, DEU, DNK, ESP, FIN, FRA, GBR, ITA, JPN, NLD, NOR, PRT, SWE, USA}. Let's denote $A_{c,t}$ the TFP of country c in year t. It can be decomposed into a trend component and a cyclical component as follows:

$$a_{c,t} = \tau_{c,t} + cyc_{c,t} + \epsilon_{c,t} \tag{2.1}$$

where $a_{c,t} = \ln A_{c,t}$, $\tau_{c,t}$ is the trend component, and $cyc_{c,t}$ is the cyclical component. Following the standard practice in the literature, I estimate the trend component using a low-pass HP filter with a smoothing parameter of 100.

Let $S_{c,t}$ denote the underlying regime index for country c in year t. The trend TFP growth is assumed to be alternating between two average levels:

$$g^{\tau}_{c,t} = \mu(S_{c,t}) + v_{c,t}, \quad v_{c,t} \sim N(0,1) \tag{2.2}$$

where

$$\mu(S_{c,t}) = \begin{cases} \mu_{L,c} & \text{if } S_{c,t} = L \\ \mu_{H,c} & \text{if } S_{c,t} = H \end{cases}$$

with $\mu_{L,c} < \mu_{H,c}$ and a transition probability matrix of

$$Pr[S_{c,t} = L | S_{c,t-1} = L] = q_{LL,c}; \qquad Pr[S_{c,t} = H | S_{c,t-1} = L] = 1 - q_{LL,c}.$$
$$Pr[S_{c,t} = H | S_{c,t-1} = H] = q_{HH,c}; \qquad Pr[S_{c,t} = L | S_{c,t-1} = H] = 1 - q_{HH,c}.$$

After estimating the model for each country, I calculate its probability of being in a high-growth regime $\Pr\{S_{c,t} = H\}$ over time, and then define the high-growth regime samples as $\{t_c : \Pr(S_{c,t} = H) \geq 0.5\}$, and the low-growth regime samples if otherwise.

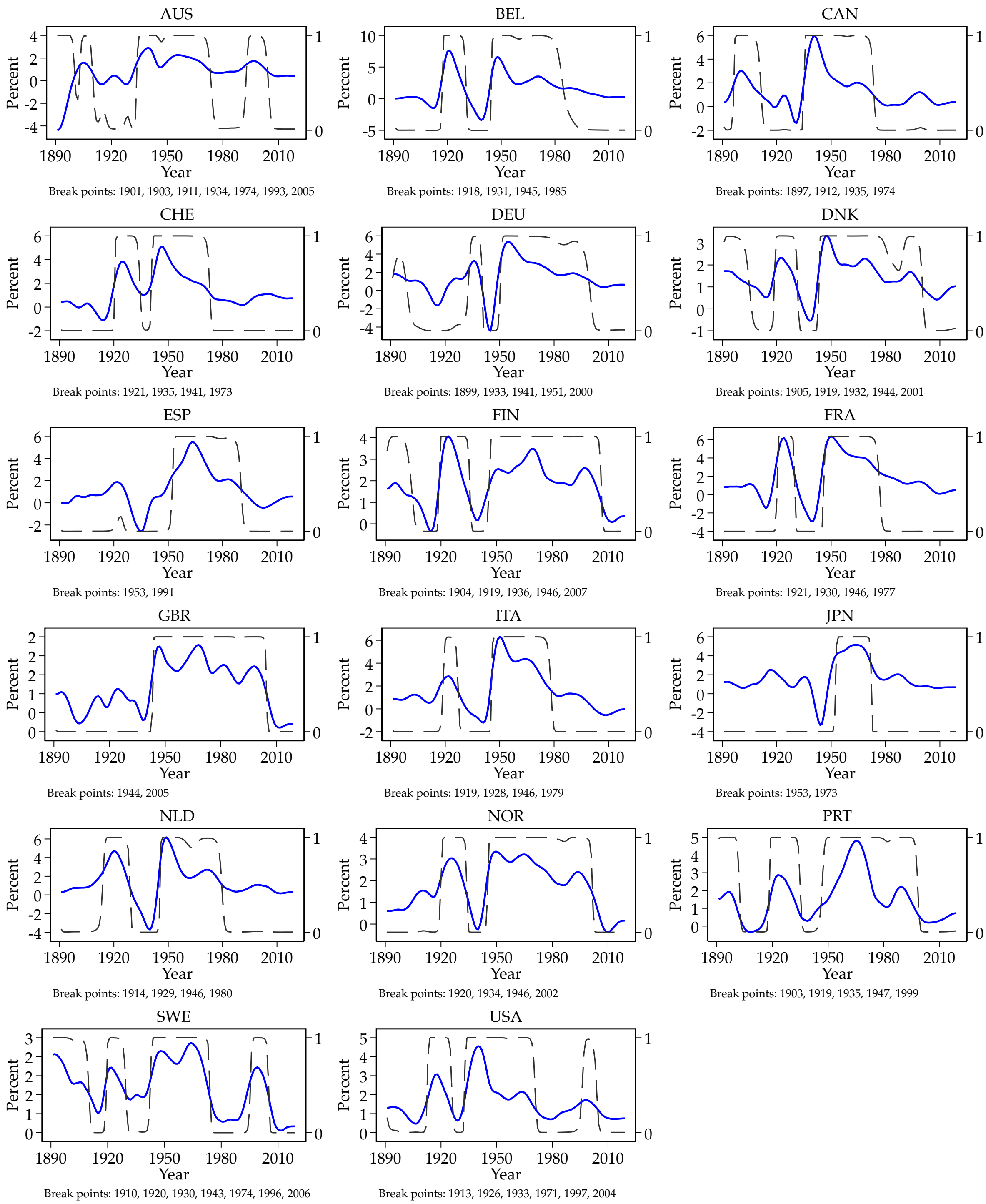
Since the estimates of average growth rates μ_L and μ_H may vary with the choice of the sample period, sample selection matters for the detection of turning points. As we shall see in the data, European countries experienced low productivity growth (some even experienced productivity loss) during two world wars, and productivity booms after the wars. If we exclude war periods, then the average growth rate in low-growth regimes will be estimated slightly higher, and more samples from normal years will be classified as in the low-growth regime. In the baseline results, I use the entire sample period to estimate the model and classify growth regimes.

Figure 2.1 collects the results of the regime classification for all 17 countries using the regime-switching approach. The solid blue line shows each country's trend TFP growth over time (left axis), while the dashed grey line indicates its probability of being in the high-growth regime (right axis) over the entire sample period of 1890-2019. The breakpoints at the bottom of each chart list the tuning points of regime switches.

Let's take the U.S. economy as an example. Unlike most of the European countries, the U.S. economy underwent productivity booms during the two world wars. Between 1913 and 1926, TFP growth in the U.S. is relatively high. Thanks to a series of economic measures to pull the economy out of the quagmire of the Great Depression, the U.S. economy started another round of rapid technological progress in 1933 and sustained at about 2% annual growth for the decade after the Second World War. Similar to the conclusions of Kahn and Rich (2007) and Fernald (2014*b*), the post-WW2 U.S. economy features "high-low-high-low" growth regimes alternating, with the recent high-growth regime from the mid-1990s to mid-2000s, and the recent low-growth regime started before the Global Financial Crisis.

For European countries, despite the geographical proximity and similar technological advancement, we can still observe considerable heterogeneity in their productivity growth profiles, especially in the post-WW2 period. Almost all of these countries suffered a lot from two world wars (1914-1919, 1939-1947) and the Great Depression in between (1929-1933) with low productivity growth or even productivity loss. Many rebounded quickly between wars and after the Second World War, but the specific timing and phases of recovery differ

Figure 2.1. Regime classification results: regime-switching model



Notes: Solid blue lines: trend TFP growth over time (left axis); dashed grey line: the probability of being in high-growth regimes (right axis) over the entire sample period 1890-2019 including wartime.

across countries. In the post-WW2 period, some European countries remained in the high-growth regime for longer (eg. NOR, PRT, DEU, DNK, FIN, GBR, NLD), some lost growth momentum before the 1990s and grew slowly thereafter (eg. BEL, CHE, ESP, FRA, ITA), while the others went through another high-growth regime from mid-1990s to mid-2000s, similar to the U.S. (eg. AUS, SWE).

2.2.2 Correlation: Growth Regime vs. Other Regimes

How do other regime variables correlate with the TFP growth regime? If the regime classified by another factor aligns closely with the growth regime, then it's hard to tell which regime variable is the relevant true state that matters for the relationship of interest. The other regime variables considered in this paper include output gap, level of inflation, and credit growth. I follow Jordà, Schularick and Taylor (2020*a*) and define these regime dummies as follows:

- Output boom: if a country's output gap is non-negative, where the output gap is measured using the HP filter with smoothing parameter equals 100.
- High inflation: if a country's inflation rate is greater or equal to 2% per year, excluding the hyperinflation periods when the annual inflation rate exceeds 45%.
- Credit boom (total): if a country's 3-year mean changes in *total* credit over GDP is above its historic mean changes.
- Credit boom (mortgage): if a country's 3-year mean changes in *mortgage* credit over GDP is above its historic mean changes.
- Credit boom (non-mortgage): if a country's 3-year mean changes in *non-mortgage* credit over GDP is above its historic mean changes.

Table 2.1 presents the conditional and unconditional means along with their correlation coefficients of regime dummies defined above for the full sample (Panel A) and post-WW2 sample (Panel B). The first two columns of each panel compare the means conditional on TFP growth regimes. If the means of another regime dummy across two growth regimes turn out to be very close, then we are assured that the classification of these two regimes does not correlate systematically with one another. The correlation coefficient (ρ) between regime dummies delivers similar information.

Table 2.1. Descriptive statistics by regimes

Regime dummies	**Panel A**: *Full sample*				**Panel B**: *Post-WW2 sample*			
	LG	HG	Total	ρ	LG	HG	Total	ρ
High TFP growth	0.00	1.00	0.58	1.00	0.00	1.00	0.66	1.00
Output boom	0.49	0.53	0.51	0.04	0.49	0.53	0.52	0.04
High inflation	0.48	0.65	0.58	0.17	0.70	0.75	0.73	0.06
Credit boom (total)	0.56	0.50	0.53	-0.06	0.51	0.49	0.49	-0.02
Credit boom (mortgage)	0.55	0.39	0.45	-0.16	0.55	0.35	0.42	-0.19
Credit boom (non-mortgage)	0.52	0.55	0.54	0.02	0.42	0.54	0.50	0.11

Notes: Full sample: 1890-2006 excluding world wars (1914-1919 and 1939-1947). Post-WW2 sample: 1948-2006. For each sample, the first two columns (LG/HG) list the means of regime dummies conditional on being in the low/high TFP growth regime, the third column (Total) lists the unconditional means of regime dummies, and the last column (ρ) lists the correlation coefficients between each regime dummy and the TFP growth regime dummy.

It is evident that none of the other regimes aligns very well with the TFP growth regime, the pairwise correlation coefficients in absolute values are all below 0.2. Among the other five regime dummies, "Output boom" and "Credit boom (total)" correlate with the TFP growth regime dummy the least. Take the output boom dummy for example. About an equal fraction of the sample are in output boom conditional on the growth regime, although it is slightly more likely to observe an output boom in the high-growth regime. The correlation coefficient between the growth regime dummy and output boom dummy is very small, about 0.04 in both sample results.

On the other hand, the inflation regime correlates with the growth regime relatively strongly in the full sample results. We tend to observe an annual inflation rate above 2% more often in the high-growth regime (65%) than in the low-growth regime (48%). However, in the post-WW2 period, both growth regimes saw high inflation quite often, and the difference in means is small. The non-mortgage credit boom dummy correlates with the growth regime dummy relatively strongly in the post-WW2 sample results. It is more likely to observe a non-mortgage credit boom in the high-growth regime (54%) than in the low-growth regime

(42%), while the difference is small in the full sample results. The mortgage credit growth regime is a little bit worrisome. In both panels, the probability of observing high mortgage credit growth is less than 40% in the high-growth regime. In Section 2.4 and Appendix 2.7.2, I argue that monetary policy has different state-dependent effects on prices/inflation and output/output gap across different regimes.

2.3 Phillips Curve Estimation and its Growth Regime Dependence

The second part of my empirical analysis is to estimate the slope of the Phillips curve and study its state dependence on the classified TFP growth regime. Although the Phillips curve was first introduced by Phillips (1958) to describe the empirical relationship between wage inflation and unemployment rate, macroeconomic models, especially the New-Keynesian models, have been using it to describe the relationship between the price inflation and output gap. It summarizes a firm's optimal price-setting condition facing nominal rigidity in the short run. Despite its theoretical foundation, the empirical estimation of the New Keynesian Phillips curve is notoriously challenging.

In this section, I first review the identification issues embedded in the estimation of the New-Keynesian Phillips curve, followed by a discussion of the validity of using monetary policy shocks as instrumental variables. Then I adapt the two-step approach proposed by Barnichon and Mesters (2020) in a panel setting and estimate the average Phillips curve for the group of advanced economies. At last, I stratify the sample by TFP growth regime and study the growth regime dependence of the slope of the Phillips curve.

2.3.1 Identification Using Monetary Policy Shocks as Instruments

Consider the hybrid New Keynesian Phillips curve (e.g., Galí and Gertler, 1999) given by

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t[\pi_{t+1}] + \lambda x_t + \epsilon_t^s \tag{2.3}$$

where π_t is inflation, $x_t \equiv y_t - y_t^n$ is the output gap which depends on the natural/potential level of output y_t^n and ϵ_t^s denotes the cost-push shock. The parameters of interest γ_b, γ_f, and λ are functions of deep structural parameters of an underlying model.

Despite decades of research, the estimation of the Phillips curve is still notoriously difficult due to pervasive endogeneity issues (Barnichon and Mesters, 2020). To see this, let's re-write Equation (2.3) as follows

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f \pi_{t+1} + \lambda \hat{x}_t + \underbrace{\epsilon_t^s - \gamma_f(\pi_{t+1} - E_t[\pi_{t+1}]) - \lambda(\hat{x}_t - x_t)}_{u_t} \tag{2.4}$$

where $\hat{x}_t$ is some empirical measure of output gap. We can see that there are three sources of endogeneity problems: (i) cost-push factors (ϵ_t^s) that can simultaneously affect inflation and output gap through the systematic response of monetary policy to inflation. In response to a negative cost-push shock that drives inflation up high (e.g., the oil price shock in the 1970s), monetary authorities may raise interest rates which will reduce the short-run output. (ii) measurement error in the output gap ($\hat{x}_t - x_t$) since the potential output is unobservable, and (iii) forecast error in inflation ($\pi_{t+1} - E_t[\pi_{t+1}]$) as inflation expectation is not observed.

Barnichon and Mesters (2020) argue that monetary policy shocks $\boldsymbol{\xi}^m_{t:t-H} \equiv (\xi^m_t, \xi^m_{t-1}, \cdots, \xi^m_{t-H})'$ are valid instruments to identify the Phillips curve in that they satisfy the following two conditions:

(i) Exogeneity condition: $E[\boldsymbol{\xi}^m_{t:t-H} u_t] = \mathbf{0}$ holds if monetary policy shocks are orthogonal to the three sources of endogeneity problems.

(ii) Relevance or rank condition: $E[\boldsymbol{\xi}^m_{t:t-H}(\pi_{t-1}, \pi_{t+1}, \hat{x}_t)]$ is of full column rank. That is, monetary policy shocks should be able to affect the output gap and inflation.

One concern for the exogeneity of monetary policy shocks is that policymakers respond to information beyond what has been considered when constructing the monetary policy shocks, eg. Greenbook forecasts for Romer and Romer (2004) shocks. It is often suggested that we add more macroeconomic controls as extra information set to clean the shocks.

Another concern is that monetary policy shocks may affect inflation through channels other than the domestic output gap. This is particularly worrisome in our context in that the monetary policy of the base country could also affect the global aggregation demand, e.g., the U.S. in the post-WW2 era. The globalization-related view of the flattening Phillips curve suggests that as part of the domestic demand is satisfied through imports, if the rise in domestic demand due to expansionary monetary policy in the base country coincides with the rise in global demand, then the inflation pressure "leakage" will be small, inflation will be higher than if global demand is not sensitive to the monetary policy of the base country. So we should also control for the global business cycle effects.

Let's re-write Equation (2.4) h periods ahead, multiply both sides by the monetary policy shock ξ_t^m and take expectations conditional on extra information set $\boldsymbol{z}_t$,

$$E[\xi_t^m \pi_{t+h}|\boldsymbol{z}_t] = \gamma_b E[\xi_t^m \pi_{t+h-1}|\boldsymbol{z}_t] + \gamma_f E[\xi_t^m \pi_{t+h+1}|\boldsymbol{z}_t] + \lambda E[\xi_t^m \hat{x}_{t+h}|\boldsymbol{z}_t] + E[\xi_t^m u_{t+h}|\boldsymbol{z}_t]$$

Dividing both sides by $E[(\xi_t^m)^2|\boldsymbol{z}_t]$, and denoting the projection of variable w on monetary policy shock as $\mathscr{R}_h^w$, $w \in \{\pi, \hat{x}, u\}$, we get

$$\mathscr{R}_h^\pi = \gamma_b \mathscr{R}_{h-1}^\pi + \gamma_f \mathscr{R}_{h+1}^\pi + \lambda \mathscr{R}_h^{\hat{x}} + \mathscr{R}_h^u, \forall h = 0, 1, \cdots, H \tag{2.5}$$

where we assume that all variables are stationary and that the endogenous variables and residuals can be written as linear functions of monetary policy shocks. If monetary policy shocks are autocorrelated, then we also include their lags in $\boldsymbol{z}_t$. For the last term to be zero, we only need monetary policy shocks to be orthogonal to the three sources of endogeneity issues conditional on $\boldsymbol{z}_t$. Under these assumptions, the exogeneity and relevance conditions can be restated in impulse response space:

(i) Exogeneity condition: $\mathscr{R}_h^u = 0, \forall h = 0, 1, \cdots, H$.

(ii) Relevance condition: $[\mathscr{R}_{h-1}^\pi, \mathscr{R}_{h+1}^\pi, \mathscr{R}_h^{\hat{x}}]_{h=0}^H$ is linearly independent.

Equation (2.5) therefore implies that to estimate the Phillips curve, we can take a **two-step approach**: (1) estimate the impulse responses of inflation and output gap to the monetary policy shocks and obtain $\hat{\mathscr{R}}_h^\pi$ and $\hat{\mathscr{R}}_h^{\hat{x}}$ for $h = 0, \cdots, H$; (2) run the following linear regression using the estimated impulse responses: $\hat{\mathscr{R}}_h^\pi = \gamma_b \hat{\mathscr{R}}_{h-1}^\pi + \gamma_f \hat{\mathscr{R}}_{h+1}^\pi + \lambda \mathscr{R}_h^{\hat{x}} + e_h$, where e_h is a linear combination of estimation errors. To identify all three coefficients, rank condition (ii) requires that the dynamics of the impulse response functions (IRFs) have to be rich enough.

In empirical studies, $\gamma_b + \gamma_f = 1$ is often imposed and it is consistent with a vertical long-run Phillips curve and money neutrality in the long run (Barnichon and Mesters, 2020). A simple case study can be shown to demonstrate the improvement in the slope estimates with versus without imposing this restriction, especially when the rank condition fails. Take the simple three-equation New-Keynesian model with pure forward-looking NKPC for example where $\gamma_b = 0, \gamma_f = \beta$. The theoretical impulse responses of inflation expectation and output are perfectly colinear. Therefore, we cannot identify β and λ separately, and will over-estimate λ if monetary shock is autocorrelated, i.e., $\rho_\xi \neq 0$. Following the standard practice in the literature, if we calibrate $\beta = 0.99$ and set $\rho_\xi = 0.2$, we can calculate that the slope coefficient will be inflated by about 25%, see Appendix 2.7.1 for more details.

In this simple case, since $\gamma_b = 0, \gamma_f = \beta = 0.99$, then $\gamma_b + \gamma_f = 0.99 \approx 1$, the long-run restriction is very close to the truth. Imposing this restriction actually provides useful information to help identify the underlying parameters. Using the theoretical IRFs, I show in Appendix 2.7.1 that imposing the long-run restriction ($\gamma_b + \gamma_f = 1$), along with the range inequality constraints ($0 < \gamma_b, \gamma_f, \lambda < 1$), is able to get the slope parameter close to its true value.[8] Therefore, in the empirical practice that follows, I draw my main conclusions based on the results with the restrictions imposed.

2.3.2 Estimation of Impulse Responses of Inflation and Output Gap

In the first step, I estimate the impulse responses of inflation and output gap to monetary policy shocks using the merged dataset since *1890* (the start of the productivity data) in the spirit of Jordà (2005*b*) and using the external instrument for exogenous monetary policy fluctuations based on trilemma mechanism (Jordà, Schularick and Taylor, 2020*a*). Compared with the local projection on the identified shocks directly, LP-IV is more robust to instrument problems and measurement errors, which is helpful in inference[9]. For the specification, I stay close to Jordà, Schularick and Taylor (2020*a*), but adjust the controls that are related to the response variables since they are no longer price and real output levels, but inflation and output gap.

[8] $\lambda < 1$ is imposed in that most of the empirical estimates of the slope of the Phillips curve are less than 1.

[9] I also check the local projection results with monetary policy shocks as regressors, they differ mainly in magnitude as expected.

The baseline local projection specification is as follows:

$$y_{c,t+h} - y_{c,t-1} = \alpha_{c,h} + \triangle i_{c,t}\beta_h + \boldsymbol{z}'_{c,t}\boldsymbol{\gamma}_h + u_{c,t+h}, \quad h = 0,1,\cdots,H \tag{2.6}$$

using a longitudinal sample where $c = 1,\cdots,N$ and $t = 1,\cdots,T$.

The response variable $y_{c,t+h} - y_{c,t-1}$ is the cumulative change in inflation or output gap measured in percentage deviations relative to its initial value in year $t-1$ computed as log change times 100. Inflation is measured by the first difference of log CPI, and the output gap is measured using Hodrick-Prescott (HP) filter with a smoothing parameter of 100. In Section 2.4, I show that the main conclusion holds qualitatively if we choose other measures of output gap.[10] The policy interest rate change $\triangle i_{c,t}$ will be defined as the one-year change in the short-term interest rate in year t, and normalized to a 1 percentage point increase. The instrumental variable is the base country's policy surprises conditional on a rich set of macroeconomic controls. For most countries in the sample, U.S. and U.K. serve as their base country, their policy surprises come from two seminal papers: Romer and Romer (2004) and Cloyne and Hürtgen (2016) for the post-WW2 period.

The macroeconomic variables $\boldsymbol{z}_{c,t}$ include the first difference of the contemporaneous values, and up to 2 lags of the first difference of the variables from the following list: log real consumption per capita; log real investment per capita; short-term interest rate (usually a 3-month government bond); long-term interest rate (usually a 5-year government bond); log real house prices; log real stock prices; and the credit to GDP ratio.[11] In addition, I control for the first difference of the contemporaneous values of the log real GDP per capita (log CPI) and output gap (inflation) in the regression for inflation (output gap) response, respectively.[12] For both regressions, I also add up to 2 lags of the first difference of the log real GDP per capita, output gap, log CPI, and inflation. What's more, to partially address the "price puzzle" issue, I interact all these controls with oil crisis dummy to allow the controls to take on a potentially different coefficient for the subsample period of the oil crisis (1973-1980). To control for the global business cycle effects, it is crucial to add global GDP as one

[10]Existing literature has proposed different measures of the output gap, but this paper will not take a stand.
[11]The credit to GDP ratio is defined as the ratio of total loans to the nonfinancial private sector to GDP.
[12]This is equivalent to controlling for the potential output in the output gap regression.

of the explanatory variables.[13]

Figure 2.2 presents the LP-IV estimates of the impulse responses of the output gap and inflation to a 1 percentage point increase in short-term interest rate. In Panel (a), the sample spans 1890-2006 but excludes two world wars (WW1: 1914-1919, WW2: 1939-1947) and short windows around the wars. We end up with the full sample covering 1890-1908, 1921-1933, and 1948-2006. Panel (b) reports the results for the post-WW2 sample prior to the Global Financial Crisis: 1948-2006. In Section 2.4, I extend the analysis to include more recent sample till 2012, the latest year for which the trilemma monetary policy shocks are provided at the time of writing.

The impulse responses estimated using the full sample look quite similar to those using the post-WW2 sample. In response to a 1 percentage point increase in the policy rate, the output gap decreases moderately for a peak response of 0.9% in the fourth year in the full sample results, and 0.8% in the second year in the post-WW2 sample results, respectively. In comparison, the post-WW2 output gap responses have a relatively shorter half-life. The inflation rate declines for a peak response of 1% in both cases and the peak response occurs in the third year. After reaching the peak response, impulse responses decay to zero slowly. Consistent with Jordà, Singh and Taylor (2020*b*), monetary policy effects last for a prolonged period of time.

2.3.3 Estimation of the Phillips Curve by Constrained NLS

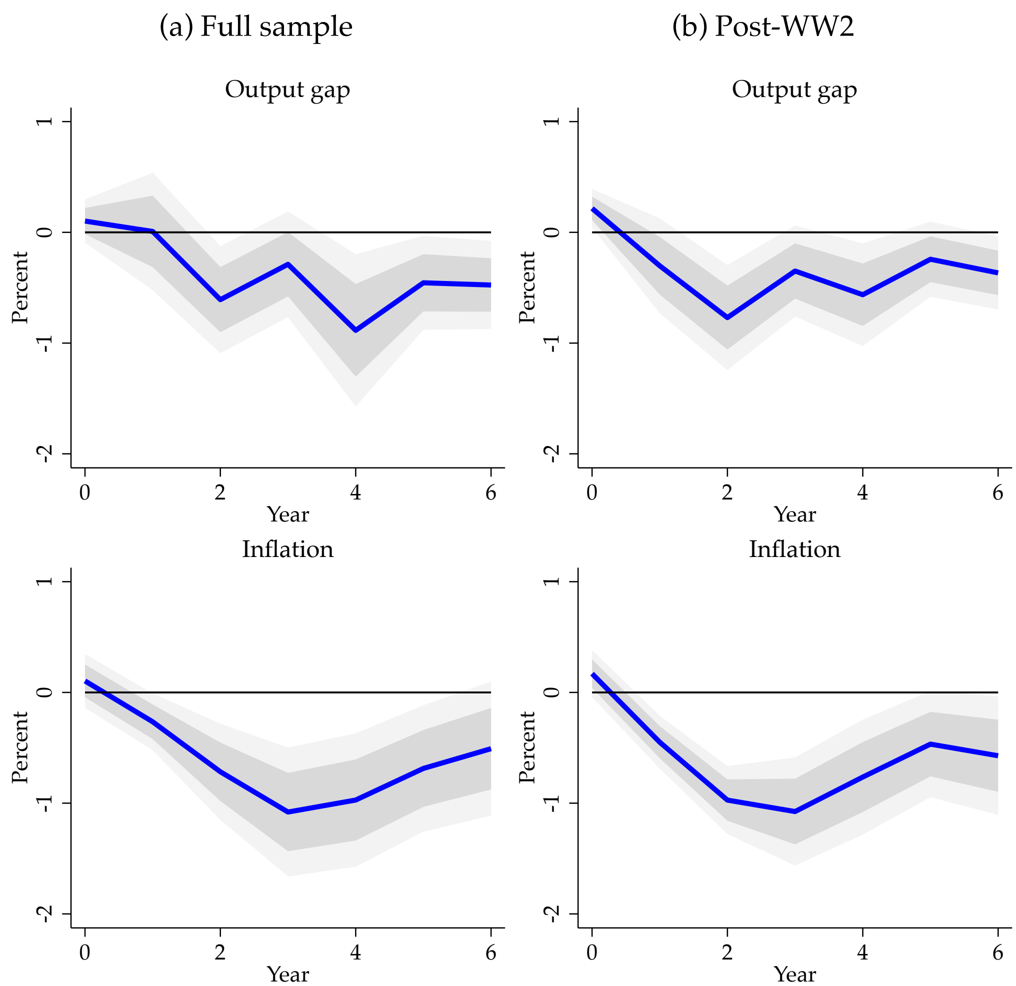
In the second step, I run the following constrained Nonlinear Least Squares (NLS) regression model using the estimated impulse responses from the first step:

$$\hat{\mathscr{R}}_h^{\pi} = \gamma_b \hat{\mathscr{R}}_{h-1}^{\pi} + \gamma_f \hat{\mathscr{R}}_{h+1}^{\pi} + \lambda \hat{\mathscr{R}}_h^{\hat{x}} + e_h, \forall h = 0, \cdots, H \tag{2.7}$$

where I restricted all three coefficients to be between 0 and 1, i.e., $0 < \gamma_b, \gamma_f, \lambda < 1$. Because the range of the inverse logit function is the interval $(0, 1)$, we can use the inverse logit function to set this restriction and estimate the coefficients using the NLS method. The initial values of the coefficients are set to match the empirical estimates reported in Barnichon and Mesters (2020) on the U.S. Phillips curve, that is, $\gamma_b \approx 0.6, \gamma_f \approx 0.4$, and $\lambda \approx 0.3$.

[13]Global GDP is defined as the sum of all countries' real GDP measures adjusted for their PPP values.

Figure 2.2. Inflation and output gap responses to a 1 percent increase in interest rates



Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. Baseline non-state dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Sample sizes are not harmonized in estimation. The output gap is measured using the HP filter with a smooth parameter of 100.

Since there are three parameters to estimate, the effective sample size cannot go below 3. Given that we need to take the lead of the inflation responses on the right-hand side, we need H to be at least 4. As is pointed out in the identification, the rank condition requires the dynamics of the impulse responses to be rich enough, it is more likely to fail when the sample size is small. However, H should not be too large because as h increases, the effect of monetary policy shocks tends to die out quickly, especially on the output gap. Given all these concerns, I set H to be 6 in the baseline results. Section 2.4 shows the results for longer projection horizons. Because the lagged inflation is predetermined, it should be orthogonal to the shock this period so that $\mathscr{R}^{\pi}_{-1} = 0$.

Table 2.2 presents the constrained NLS estimates of the hybrid NKPC. The restricted estimates of λ range from 0.12 in the full sample results to 0.23 in the post-WW2 results. The latter is roughly in the same range as in Barnichon and Mesters (2020), and Stock and Watson (2019). To map the range of macro-level price stickiness to the micro-level frequency of price adjustment, I refer to the slope of the standard NKPC under Calvo (1983) pricing formulated as:

$$\lambda = \frac{(1-\theta)(1-\beta\theta)}{\theta}(\chi+\sigma) \tag{2.8}$$

where θ is the probability for the firm to be unable to reset the price within the period, and $1-\theta$ is the probability of price adjustment; β is the discount factor; χ is the inverse Frisch elasticity of labor supply; σ is the elasticity of substitution. Under the standard calibration of these parameters where $\beta = 0.99$[14], $\chi = \sigma = 1$, the implied degree of price rigidity θ ranges from 0.72 to 0.79. That is, only a fraction of 21% to 28% of prices will adjust each quarter, prices are quite sticky.

Table 2.2. Estimates of the Phillips curve without considering state dependence

SAMPLE	UNRESTRICTED			RESTRICTED	
	γ_f	γ_b	λ	γ_f	λ
Full	0.51 (0.16)	0.53 (0.25)	0.07 (0.39)	0.50 (0.12)	0.12 (0.13)
Post-WW2	0.28 (0.20)	0.35 (0.19)	0.71 (0.40)	0.48 (0.15)	0.23 (0.21)

Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. All coefficients are constrained to be between 0 and 1 and estimated by NLS with $H = 6$. Coloumn "RESTRICTED" imposes an additional restriction: $\gamma_b + \gamma_f = 1$. The output gap is measured using the HP filter with a smooth parameter of 100.

For the unrestricted estimates, the full sample estimates are reasonably close to their unrestricted counterparts for both γ's and λ. However, the post-WW2 estimates for γ_b and γ_f are relatively small. The lack of dynamics in the impulse responses of inflation potentially fails the rank condition. Imposing the long-run restriction deflates the slope estimate by a factor

[14]This corresponds to a 2% annual real interest rate and is adjusted for a steady state growth rate of 2% per year.

of 3. Lastly, the backward-looking (γ_b) and forward-looking inflation expectations (γ_f) are of similar importance in determining inflation dynamics, indicating that the hybrid NKPC is preferred to the purely forward-looking NKPC.

2.3.4 Growth Regime Dependence of the Slope of Phillips Curve

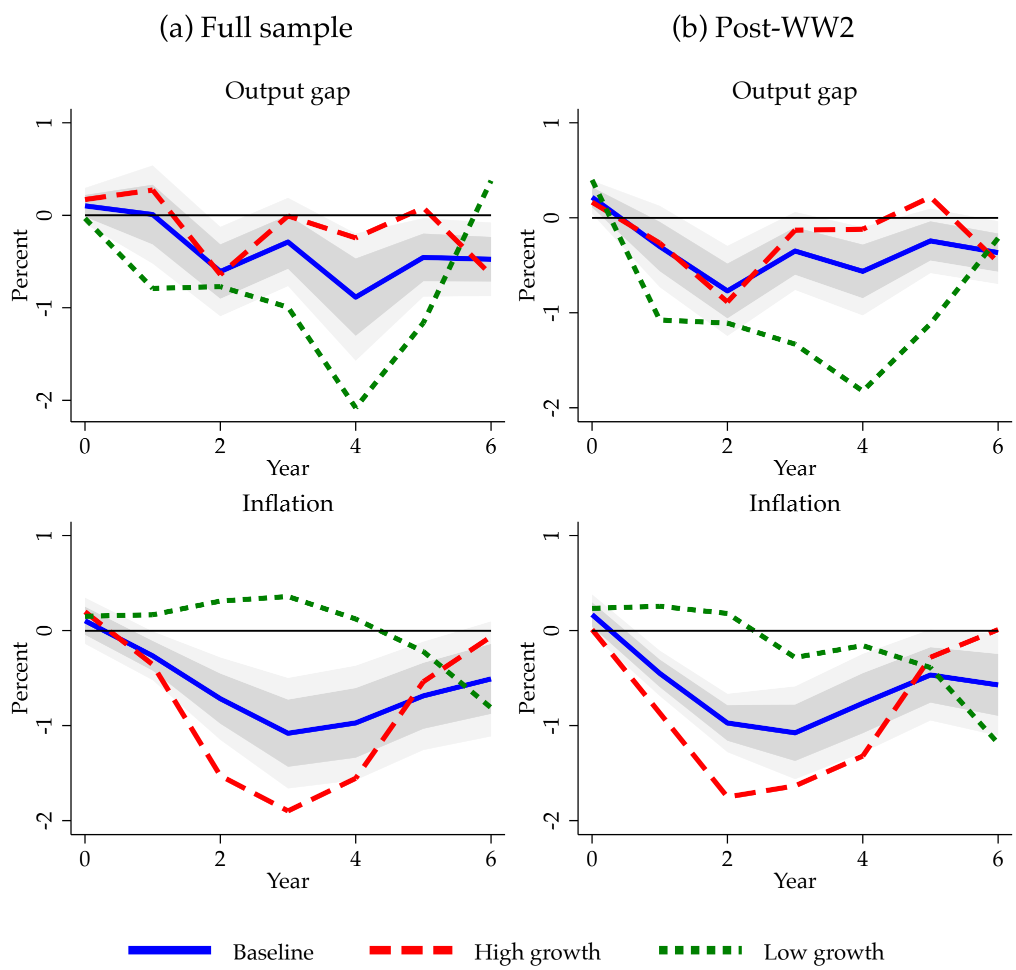
In this section, I stratify the sample by productivity growth regimes classified in Section 2.2.1 and apply the two-step approach to estimate the hybrid NKPC for each regime. To enlarge the sample size of each regime and avoid further sample loss from taking first difference, lags, and leads, I do not harmonize the sample sizes across horizons and keep the observations even if the value of the trilemma instrument equals zero.

Figure 2.3 shows the impulse response of inflation and output gap estimated using the LP-IV approach. The baseline non-state dependent estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates in the high-growth regime are displayed with a red long-dashed line whereas estimates in the low-growth regime are displayed with a green dashed line.

It is evident that monetary policy faces different trade-offs between its ability to stabilize inflation versus output across growth regimes. Inflation responses are almost muted in the low-growth regime for a few years, especially in the full sample results, but the output gap is much more responsive in the low-growth regime. The opposite is true in the high-growth regime: contractionary monetary policy is very effective in controlling inflation with a low cost of reduction in the output gap. It's noteworthy that the inflation responses in the low-growth regime are slightly positive for a few years, not statistically significant though. Given the significant negative responses of the output gap, the correlation between the impulse responses of inflation and the output gap is slightly negative, especially in the full sample results. If we impose the non-negativity constraints, the slope estimate for the low-growth regime will be bounded by zero.

The constrained NLS estimates of the Phillips curve are presented in Table 2.3. Indeed, in the full sample results, the restricted estimate of λ for the low-growth regime sample is bounded by zero, and the estimate of γ_b without imposing the long-run constraint is bounded

Figure 2.3. Asymmetric responses of inflation and output gap: regime-switching approach



Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. Baseline non-state dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high-growth regime are displayed with a red long-dashed line whereas estimates in the low-growth regime are displayed with a green dashed line. Sample sizes are not harmonized in estimation. The output gap is measured using the HP filter with a smooth parameter of 100.

by one. In contrast, the coefficient estimates are well-behaved in the high-growth regime. The slope estimate is much larger in magnitude than their counterparts in Table 2.2 for both regimes and is statistically significant. The unrestricted estimates are quite close to the restricted ones, with the slope estimate in the high-growth regime (0.83) much larger than that in the low-growth regime (0.05).

As we've seen in Table 2.2, the lack of dynamics in inflation after WW2 renders a serious challenge to the data to distinguish the effects of lagged and leading inflation terms, especially in the low-growth regime. Despite this, both unrestricted and restricted estimates suggest that the Phillips curve is steeper in the high-growth regime. The restricted estimates suggest that λ ranges from 0.8 to 0.9 in the high-growth regime, which implies that the degree of price stickiness is approximately 0.54. That is, about half of the prices adjust each quarter, the price adjustment is more than twice as frequent as the average. Comparing the importance of two types of inflation expectation terms, I find that the backward-looking term plays a slightly more important role in inflation dynamics than the forward-looking term in the low-growth regime.

Table 2.3. Estimates of the Phillips curve by TFP growth regime

SAMPLE	BIN	UNRESTRICTED			RESTRICTED	
		γ_f	γ_b	λ	γ_f	λ
Full	HG	0.49 (0.11)	0.56 (0.11)	0.83 (0.41)	0.47 (0.08)	0.89 (0.33)
Full	LG	0.39 (0.07)	1.00 (.)	0.05 (0.03)	0.38 (0.07)	0.00 (.)
Post-WW2	HG	0.30 (0.08)	0.58 (0.08)	1.00 (.)	0.39 (0.08)	0.80 (0.25)
Post-WW2	LG	0.31 (0.16)	0.34 (0.43)	0.00 (.)	0.31 (0.25)	0.00 (0.10)

Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. The growth regime is classified based on the regime-switching model. In Column "BIN", "HG" denotes the high-growth regime, "LG" denotes the low-growth regime. All coefficients are constrained to be between 0 and 1 and estimated by NLS with $H = 6$. Coloumn "RESTRICTED" imposes an additional restriction: $\gamma_b + \gamma_f = 1$. The output gap is measured using the HP filter with a smooth parameter of 100.

2.4 Robustness Checks and Further Results

In this section, I consider a number of robustness checks to see if the baseline results are robust to a variety of changes in the state-dependent analysis of the slope of the Phillips curve. The first set of robustness checks uses alternative growth regime classifications that are more intuitive without imposing any assumptions on the underlying process of trend TFP. Next, I show that my baseline conclusions survive at least qualitatively even if we choose other measures of the output gap and longer projection horizons. Third, I consider alternative sample selection criteria: include more recent samples till 2012 and the subsample of European countries. Lastly, I compare the state-dependent effects of monetary policy across different regimes to provide further evidence of the relevance and uniqueness of the TFP growth regime. Below I will focus on the discussion of conclusions, with further detail and results shown in the Appendix.

2.4.1 Alternative Regime Classification Methods

Without taking a stand on which classification result makes more sense, it's worth checking if the baseline conclusion still holds when we classify growth regimes differently. In this part, a brief illustration of three alternative classification methods is followed by a discussion of regime classification results. Compared with the regime-switching approach, these simple methods share some *pros and cons*. Then I repeat the state-dependent analysis as in the previous section. Here are three alternative classification methods sorted by complexity.

1. ***Above universal cutoff.*** We can simply define the high-growth regime as the years when a country's trend TFP growth rate is above some universal cutoff level, i.e., $S_{c,t} = H$ if $g^{\tau}_{c,t} \geq \bar{g}$, and $S_{c,t} = L$ if otherwise, where $\bar{g}$ is fixed $\forall c, t$.

What level of $\bar{g}$ should we choose? One potential candidate is the sample average of trend TFP growth, which obviously depends on the sample period that we look at. Similar to the regime-switching approach, if we include more low-growth periods in the sample (e.g., two world wars or more recent samples), the sample average TFP growth will be lower, and fewer observations will be classified to be in the low-growth regime. Besides the common problem of the sensitivity to the choice of the sample period, it is particularly a problem for this method that it tends to generate too many tuning points. This often occurs when a country's trend TFP growth rises above or falls below the cutoff level for a short period of time.

In the data, the average trend TFP growth over the period 1890-2018 for all countries is 1.47%. That is, if $g^{\tau}_{c,t} \geq 1.47\%$, then $S_{c,t} = H$ and the regime dummy equals one. Appendix Figure 2.4 collects the classification results for all 17 countries in the merged dataset. The solid blue line shows each country's trend TFP growth over time (left axis), while the dashed grey line indicates the regime dummy (right axis). The universal cutoff level is displayed with a horizontal dashed red line.

Take the U.S. economy as an example. It is quite surprising to find that this simple method is able to pick up almost all the tuning points suggested by the regime-switching approach. The breakpoints of the regime changes are quite close to my baseline results. The main difference lies in the starting point of the recent period of productivity boom. This simple method suggests that the U.S. economy entered the high-growth regime in 1994 and the regime lasts for a decade, while the regime-switching approach suggests that the recent high-growth regime in the U.S. hasn't started until 1997, and it only lasts for 8 years.

On the other hand, we do find quite different classification results for a bunch of other countries. This simple approach tends to generate too many regime changes, making each regime shorter and more segmented. When the trend TFP growth fluctuates around the cutoff level, the method will treat the ups and downs as regime switches, e.g., Sweden (SWE) and Norway (NOR).

Since this method doesn't impose the assumption that the trend productivity growth varies around two average levels over time, it may provide some new insights as well. In the case of Japan, we get two distinct high-growth regimes prior to WW2, one from 1912 to 1924, the other from 1933 to 1938; and the post-WW2 high-growth regime also lasts for almost two decades longer, from 1950 to 1992, compared to what the regime-switching approach suggests, i.e., from 1953 to 1973.

Does the main conclusion still hold under this regime classification? The LP-IV estimates of the impulse responses are displayed in Appendix Figure 2.5. The output gap responses are still stronger in low-growth regimes, especially at longer horizons. In the full sample results, the inflation responses are almost muted in the low-growth regime but are quite significant in the high-growth regime. In the post-WW2 sample results, however, it seems that the inflation responses are stronger in the low-growth regime at longer horizons.

The estimates of the Phillips curve are shown in Appendix Table 2.6 Panel A. The full sample results are quite similar in both restricted and unrestricted cases, and the slope coefficient in the low-growth regime is bounded by zero. The Phillips curve is steeper in the high-growth regime, the slope coefficient is not statistically significant though. In the post-WW2 sample results, the inflated unrestricted slope estimate for the low-growth regime gets quite close to the estimate for the high-growth regime, but the restricted estimates restore the stunning difference. Again, the sum of inflation expectation terms is far below 1 without imposing the long-run restriction.

2. ***Above country-specific average growth rate.*** Despite the closeness in the geographical, cultural, and technological environment, countries still differ in their relative distances to the productivity frontier and how fast they adopt new technologies, especially in the aftermath of two world wars. Some European countries recovered really quickly, e.g. Netherlands (NLD) achieved 4.6% and 4.9% growth rates in trend TFP after the two wars, while others grew relatively moderately.[15] The timing of regime changes might be similar across countries, but the level of trend TFP growth is less comparable. So it is reasonable to assume a country-specific cutoff level when defining the growth regime for each country, i.e., $S_{c,t} = H$ if $g^{\tau}_{c,t} \geq \bar{g}_c$, and $S_{c,t} = L$ if otherwise, where $\bar{g}_c$ is fixed for country c $\forall t$.

Using each country's sample average of trend TFP growth over the sample period of 1890-2019 as its cutoff level, the regime classification results are collected in Appendix Figure 2.6. Compared with the universal cutoff approach, more sample periods will be classified as in the low-growth regime for the countries that achieve higher average trend TFP growth than 1.47%. That is, if $\bar{g} \leq g^{\tau}_{c,t} < \bar{g}_c$, $S_{c,t}$ will adjust from H to L, e.g., FIN, JPN, NOR, SWE, U.S., etc. The opposite is true if $\bar{g}_c \leq g^{\tau}_{c,t} < \bar{g}$, e.g. AUS, GBR, etc. It is evident that the problem of too frequent regime changes is still prevalent for some countries, e.g. JPN from 1934 to 1937, from 1977 to 1982, etc. The recent high-growth regime in the U.S. is no longer

[15] The average trend TFP growth rate was 1.54% during the world war I (1917–1919), 2.51% from 1920 to 1922, 1.2% during the world war II (1939–1947), and 3.2% from 1948 to 1956.

self-evident anymore.

Does the main conclusion still hold under this regime classification? The LP-IV estimates of the impulse responses are displayed in Appendix Figure 2.7. The main conclusions remain the same, that is, in the low-growth regime, output gap responses are relatively stronger while the inflation responses are relatively weaker. The coefficient estimates of the Phillips curve are shown in Appendix Table 2.6 Panel B. All specifications deliver qualitatively similar conclusions as my baseline results.

3. ***Two percentiles with a medium high-growth regime.*** The third classification method tries to mitigate the problem of frequent regime changes shared in the previous two methods. Let's denote the two percentile levels as p_1 and p_2, $p_1 < 50 < p_2$ (e.g., 45 and 55 percentiles). In the first step, I calculate the two percentiles of each country's trend TFP growth, denoted by $\bar{g}_{p_1,c}$ and $\bar{g}_{p_2,c}$. In the second step, I classify the *raw* growth regime as follows:

$$S_{c,t} = \begin{cases} L & \text{if } g^{\tau}_{c,t} < \bar{g}_{p_1,c} \\ M & \text{if } \bar{g}_{p_1,c} \leq g^{\tau}_{c,t} < \bar{g}_{p_2,c} \\ H & \text{if } g^{\tau}_{c,t} \geq \bar{g}_{p_2,c} \end{cases}$$

It is worth pointing out that a lot of the medium-high-growth regimes are short and temporary, and mainly occur when trend TFP growth transitions from one growth regime to the other. So I take a recursive process in step 3 to reassign these transitional periods to their previous regime. For example, if the trend TFP growth switches from the low-growth regime to the high-growth regime, with less than 5 years in the medium-high-growth regime, then these transitional periods will be classified as low-growth regime. That is, if $S_{c,t} = M, S_{c,t-1} = L$, and $S_{c,t+x} = H$ where $x = 1, \cdots 5$, then $S_{c,t} = L$ after the adjustment. In this way, I essentially impose the rule that when the regime switches from L to H, the trend TFP growth should exceed the higher percentile, and vice versa.

In step 4, to cope with the temporary rises and falls in trend TFP growth, I reassign these periods to the nearby regime. For example, in the middle of a long growth regime appears a few years (less than 5) of another growth regime, then the short regime will be re-classified to be part of the more populated regime. The last step deals with the high-growth regime in which the trend TFP growth only exceeds the higher percentile slightly, i.e., $g_{p_2,c} \leq g^{\tau}_{c,t} < g_{p_2,c} + \triangle)$, where $\triangle$ is some tolerance level. I adjust these samples to be in the medium high-growth regime.

Which two percentiles should we choose? Considering the fact that medium high-growth regimes are segmented and transitional, I prefer to choose two percentiles close to the median not only to maximize the sample sizes of low and high-growth regimes but also to allow for a medium regime to absorb some less representative middle cases. Here I use each country's 45 and 55 percentiles of trend TFP growth over the sample period of 1890-2018 as the raw cutoff levels, and $\triangle = 0.1\%$. The classification results are shown in Appendix Figure 2.8. Compared with the previous two methods, this method further refines the two representative regimes, the high-growth regime subsample features a slightly higher average trend TFP growth and is less segmented. In the state-dependent analysis, I group the low-growth regime and medium high-growth regime as one group, so the comparison is essentially the high-growth regime versus the rest non-high-growth regimes.

Does the main conclusion still hold under this regime classification? The LP-IV estimates of the impulse responses are displayed in Appendix Figure 2.9 and the corresponding estimates of the hybrid NKPC are summarized in Appendix Table 2.6 Panel C. We can see that the main results still hold in this case. Inflation responses are almost muted in the non-high-growth regimes and are stronger than the baseline in the high-growth regime. The output gap responses are slightly stronger in the non-high-growth regime at longer horizons. The differences in the slope coefficients across regimes are quite evident in both sample results, and the Phillips curve is much steeper in the high-growth regime, and almost flat in the non-high-growth regime.

2.4.2 Alternative Measures of Output Gap

By definition, $y_t \approx y^n_t + x_t$, that is, the (log) real output y_t can be decomposed into a trend component y^n_t, and a cyclical component x_t, also known as the short-run output or output gap. Since y^n_t is unobservable, different ways of estimating the trend component will yield different measures of the output gap. If one attributes more variations in real output to long-run forces that change potential output, then fewer variations are left to the business cycle factors that affect the output gap. In other words, if potential output is estimated to be more volatile or less smoothed, then the output gap will be less volatile.

Does the monetary policy have long-run effects on output? The conventional wisdom is that monetary policy only affects output in the short run due to the existence of nominal or real rigidities, but money is neutral in the long run. In this case, monetary policy shocks should not affect potential output, but only the output gap. The IRFs of the output gap should be exactly the same as the IRFs of the output. However, Jordà, Singh and Taylor (2020*b*) challenge this widely accepted benchmark and find empirically that monetary policy shocks have effects on output, capital, and TFP over more than a decade!

Therefore, the extent to which potential output responds to the monetary policy shocks distinguishes some measures of output gap from others. The simplest way of estimating the potential output is to fit a linear trend over the entire sample period. The resulting residual gives us the linearly detrended measure of the output gap, denoted as x_{linear}. In the context of TFP growth regimes, it is natural to incorporate the non-linearity by allowing the growth rate of the trend component to be regime-dependent, the resulting residual $x_{linear,sd}$ should be less volatile than x_{linear}.

On the other extreme is the output gap measure using the HP filter. Depending on the degree of smoothness, the trend component can be estimated to be more or less smooth. In the baseline results, I follow Jordà, Schularick and Taylor (2020*a*) and set the smoothing parameter to 100. Compared to the previous two output gap measures, the output gap using the HP filter, denoted as x_{hp}, is the least volatile measure because the trend component is time-varying and comoves with output more closely. So we would expect that the IRFs of x_{hp} would look relatively dwarfed in magnitude.

Appendix Figure 2.10 plots the time series of different measures of the output gap for the U.S. from 1890 to 2013. Apart from x_{hp},x_{linear} and $x_{linear,sd}$, I also consider the quadratically detrended output gap x_{qt} and two linearly detrended output gap measures that allow the coefficients to take on different values before and after the WW2 when estimating the linear trend(s). The main takeaway is threefold: first, all these measures of output gap comove together most of the time; second, x_{hp} is relatively less volatile than most of the other measures most of the time, whereas x_{linear} and x_{qt} are the most volatile measures; third, huge structural changes occur during the Great Depression and WW2. Allowing for coefficients to take on different values before and after WW2 essentially adjusts the linearly detrended measures towards x_{hp}.

Appendix Figure 2.11 shows the impulse responses of output and different measures of output gap to monetary policy shocks using the LP-IV approach. In both panels, the LP-IV estimates of the output responses are displayed with a solid blue line and 68% and 90% confidence bands in grey. Various types of dashed colored lines show the estimated IRFs of different measures of the output gap. Among these output gap measures, it is expected that the IRFs of x_{linear} will overlap with the IRFs of output conditional on the same set of macroeconomic controls $\boldsymbol{z}_{c,t}$. The minor difference in the graph is due to the controls related to the output gap. In contrast, the IRFs of x_{hp} get dampened the most. The IRFs of x_{qt} are very close to the IRFs of the linearly detrended output gap x_{linear} in the full sample results, while the IRFs of the output gap assuming regime-dependent average growth rates in trend ($x_{linear,sd}$) lie close to the IRFs of x_{linear}. Panel (b) delivers quite similar results, except that the IRFs of x_{qt} and $x_{linear,sd}$ lie between the IRFs of x_{linear} and x_{hp}. In summary, x_{linear} and x_{hp} represent two extreme cases.

In Appendix Table 2.7, I collect the estimates of the hybrid NKPC using three alternative measures of the output gap. The differences in the slope estimates are quite consistent with what we've noticed from Appendix Figure 2.11. In the full sample results, the output gap measures assuming constant linear trend and quadratic trend yield quite small slope estimates, while the measures assuming growth regime-dependent linear trends fall in between. Across all measures of the output gap, both restricted and unrestricted estimates confirm that the Phillips curve is steeper in the high-growth regime.

2.4.3 Alternative Choices of Projection Horizon

In the baseline results, I chose 6 as the projection horizon for the IRFs estimated in the first step, which is also the number of observations for the regression in the second step. In this part, I will extend the projection horizon (denoted as H) and utilize more periods of IRFs in the second-step estimation. Appendix Table 2.8 displays the estimates of the Phillips curve using the impulse responses of inflation and output gap estimated for $H = 9$ in Panel A and $H = 12$ in Panel B. The IRFs are plotted in Appendix Figure 2.12 and 2.13, respectively.

The main takeaway is twofold: first, the main conclusions associated with the state dependence of the slope of the Phillips curve still hold. That is, the Phillips curve is steeper in the high-growth regime. Second, the more periods of impulse responses we include in the second step regression, the flatter the Phillips curve is estimated to be. The output gap responses decay quickly after the sixth year, especially in the low-growth regime. Inflation responses differ quite substantially across the two regimes: in the high-growth regime, inflation declines quickly in the first three years and almost recovers in the sixth year, then falls again with a smaller peak response thereafter. In the low-growth regime, however, it takes a couple of years for the inflation to fall and the peak response occurs in the sixth or seventh year. The inertia in the inflation response is also coupled with a smaller peak response.

2.4.4 Alternative Sample Selections

1. ***Including more recent samples.*** The baseline results include the sample periods prior to the Great Financial Crisis to avoid the zero lower bound by which the conduct of monetary policy is constrained. In this part, I extend the sample period to include more recent samples till 2012, the latest year for which the trilemma monetary policy shocks are available in the *Macrohistory Database*.

Appendix Figure 2.14 plots the impulse responses of inflation and output gap estimated using the LP-IV approach. With 6 more years of data, the estimated impulse responses look quite close to the baseline results. The IRFs in the high-growth regime are almost the same since almost all countries are in the low-growth regime after 2006, except for Finland (FIN), which enters the low-growth regime in 2007. The IRFs of the output gap in the low-growth regime change slightly with a smaller peak response. In Appendix Table 2.9 Panel A, we find very similar point estimates as those in Table 2.3, although the slope coefficients in the low-growth regime are bounded by zero in the full sample results due to the “price puzzle”.

2. ***The subsample of European countries.*** Considering the similarities within the group of European countries in various other dimensions beyond economic development, I conduct the same analysis using the subsample of European countries: BEL, DEU, DNK, EPS, FIN, FRA, ITA, NLD, PRT, SWE, NOR, CHE. The results are displayed in Appendix Figure 2.15 and Appendix Table 2.9 Panel B.

In the full sample results, things are qualitatively similar to their counterparts in the baseline. In the post-WW2 sample results, the inflation responses show little difference across growth regimes, but the output gap responses are much stronger in the low-growth regime. The slope estimates deliver qualitatively the same conclusion.

2.4.5 State Dependence: Growth Regime vs. Other Regimes

Although the descriptive analysis in Section 2.2.2 shows only weak correlations between the TFP growth regime and other regime dummies, there remain concerns about how and to what extent the TFP growth regime differs from other regimes in terms of the monetary policy effects. In this section, I conduct similar state-dependent analyses as in Jordà, Schularick and Taylor (2020*a*) and compare the asymmetric responses of output and price levels to monetary policy shocks under the stratifications based on inflation level, credit growth (both mortgage and non-mortgage), and trend TFP growth. This helps us visualize how differently trend TFP growth matters for the transmission of monetary policy. The LP-IV specification and IRFs are displayed and discussed in Appendix 2.7.2 with more details.

Compared with the inflation regime and credit growth regimes, the rich panel data at hand suggest that the stratification based on the trend TFP growth shows different state-dependent effects of monetary policy. The main takeaway is as follows: monetary policy is subject to clear and distinct trade-offs between its nominal effect on prices and real effect on output across TFP growth regimes. We don't find such trade-offs in the state-dependent analyses based on inflation regime in Appendix Figure 2.16 or credit growth regimes in Appendix Figure 2.17 and 2.18. In particular, Appendix Figure 2.19 shows that output response is relatively stronger in the low-growth regime than in the high-growth regime. In contrast, the price response is almost muted in the low-growth regime for the first 4 years after the policy change, but the price is very responsive in the high-growth regime. Such clear and distinct trade-offs between the nominal and real effects of monetary policy suggest that agents might face different degrees of nominal rigidity across TFP growth regimes.

2.5 Theoretical Explanations

So far the empirical estimation of the slope of the NKPC using the merged panel data on 17 advanced economies over a century suggests that the short-run trade-off between inflation and output gap depends on the productivity growth regimes. In particular, the Phillips curve is steeper when the trend productivity growth is faster. Since the growth regimes are classified based on trend growth in TFP, and NKPC summarizes a firm's optimal pricing behavior when faced with short-run nominal rigidities, my empirical finding implies that the long-run technological development matters for the firm's pricing behavior in the frictional short run. This, to the best of my knowledge, remains a gap in the macroeconomic theories.

In this section, I propose one potential channel through which the slope coefficient will exhibit the observed growth regime dependence. The proposed mechanism bridges the recent literature on endogenous growth, market structure, and the slope of the Phillips curve. On the one hand, it has been successfully established by recent studies that rising market concentration flattens the slope of the Phillips curve, see e.g., Wang and Werning (2022), and Fujiwara and Matsuyama (2022). On the other hand, market structure and productivity growth are interdependent in the endogenous growth framework. To match my empirical findings, I argue that structural changes that contribute to higher productivity growth historically also lead to a more competitive market, and vice versa. That is, a productivity boom (slowdown) is often accompanied by lower (higher) market concentration and markups.

To illustrate the mechanism, let me start with the standard NKPC. Under the assumptions of CES preference and monopolistic competition as in Dixit and Stiglitz (1977), and price adjustment costs à la Rotemberg (1982), the slope of the NKPC is formulated as

$$\lambda = \frac{e-1}{\phi_R} \times \frac{\partial mc}{\partial x} \tag{2.9}$$

where e is the price elasticity of demand, ϕ_R denotes the degree of price adjustment costs and $\frac{\partial mc}{\partial x}$ captures the sensitivity of marginal cost with respect to output gap.[16]

[16]The Calvo (1983) assumption of staggered price-setting leads to $\lambda = \kappa(\theta) \times \frac{\partial mc}{\partial x}$ as in equation (2.8). If

Under the assumptions of CES preference and monopolistic competition, each firm is atomic and too small to affect the price or output of the entire market, and the price elasticity of demand is determined solely by the elasticity of substitution between products/varieties, which is assumed to be constant. Therefore, market structure is irrelevant to the slope of the Phillips curve.

However, in a general setting, the sensitivity of λ with respect to the trend growth of productivity g is given by

$$\frac{\partial \ln \lambda}{\partial \ln g} = \underbrace{\frac{e}{e-1}\frac{\partial \ln e}{\partial \ln g}}_{\text{real rigidity}} - \underbrace{\frac{\partial \ln \phi_R}{\partial \ln g}}_{\text{nominal rigidity}} + \frac{\partial}{\partial \ln g}\ln\left(\frac{\partial mc}{\partial x}\right) \tag{2.10}$$

The three terms on the right-hand side represent three potential channels through which the slope of the Phillips curve might exhibit growth regime dependence: the first term captures the "real rigidity" channel through the price elasticity of demand given the form of market competition and preference system; the second term captures the "nominal rigidity" channel through the price adjustment costs, and the last term captures the channel through the marginal cost variations with respect to short-run demand conditions.

My proposed mechanism breaks the irrelevance through the "real rigidity" channel by linking the price elasticity of demand to the market structure and endogenizing the market structure by firm entry and exit in the long run. In particular, the long-run forces that determine firms' innovation incentive and profitability will also affect the number of firms operating in the economy, thus the slope of the Phillips curve through the change in the price elasticity of demand.

There are at least two approaches in the literature to establish the linkage between the price elasticity of demand (e) and the market structure, or the number of firms (N) in the symmetric case: one relies on oligopolistic competition, and the other relies on non-CES preferences. For the first approach, Smulders and Van de Klundert (1995, 1997) build an endogenous growth model featuring CES preference and oligopolistic competition. Different varieties are imperfect substitutes with a constant elasticity of substitution $\epsilon > 1$, and the

$\kappa(\theta) = \frac{e-1}{\phi_R}$, it will be equivalent to the NKPC under Rotemberg (1982) assumption.

number of firms/varieties is perceived as fixed in the short run but is endogenously determined by the zero-profit condition in the long run. The perceived price elasticity of demand is increasing in the number of firms, i.e., $e'(N) > 0$, so if N increases – more competition – when the economy transitions to the high growth regime, then this mechanism can qualitatively explain the observed fact.

In Appendix 2.7.3, I calibrate their model along the balanced growth path and conduct a simple comparative statics analysis based on two structural changes proposed in the literature to explain the productivity boom in the mid-1990s and the productivity slowdown in the mid-2000s. One is the decline in the overhead costs due to the arrival of new Generous Purpose Technology (GPT), e.g., Information and Communication technology (ICT), see Aghion et al. (2019). The other is the rise in research productivity as the new GPT frees out the research resources in the overly-crowded traditional sector and reallocates them into booming sectors. Anzoategui et al. (2019) and Bloom et al. (2020) point out that the decline in R&D efficiency had led to the slowdown in productivity prior to the Great Recession.

In Appendix Figure 2.20 Panel (c), the comparative statics analysis shows that a fall in the overhead costs and a simultaneous rise in the research productivity result in a new long-run equilibrium with higher productivity growth and a more competitive market. In contrast to Aghion et al. (2019), the decline in the overhead costs or rise in the process efficiency alone in Panel (a) attracts more firms to enter but discourages innovation as each firm's market share is lower. Meanwhile, the rise in research productivity in Panel (b) encourages firms to invest more in firm-specific knowledge, and the increase in the fixed R&D costs is roughly offset by the increase in profits, leaving the number of firms almost unchanged. With both structural changes in action, the model is able to link productivity boom to more competition.

However, the model is unable to generate quantitatively large enough change in the slope. To see this, let's take Bertrand competition for example. The perceived price elasticity of demand is formulated as $e = \epsilon - (\epsilon - 1)\frac{1}{N}$, where ϵ denotes the elasticity of substitution between goods, and the "real rigidity" channel can be simplified as $\frac{e}{e-1}\frac{\partial \ln e}{\partial \ln g} = \frac{1}{N-1}\frac{\partial \ln N}{\partial \ln g}$. If N is not so small, then this mechanism only plays a secondary role.

My empirical estimate of the slope coefficient in the baseline for the post-WW2 sample increases by a factor of 3.5 from 0.23 on average to 0.8 in the high-growth regime. Given the parameter calibration $\epsilon = 6$ and the initial equilibrium number of firms $N \approx 5$, a back-of-the-envelope calculation implies that the price elasticity of demand has to increase by a factor of 3, from 5 on average to 15 in high growth regime. However, in the calibrated model, e is at most $\epsilon = 6$, which corresponds to a 25% increase in the slope even if the number of firms goes from 5 to infinity.[17]

Alternatively, the non-CES homothetic preference system is also able to introduce state dependence into the slope of the Phillips curve through the "real rigidity" channel. One such preference uses the translog expenditure function proposed by Feenstra (2003), and another features exponential love-of-variety. For both preference specifications, the symmetric price elasticity of demand is formulated as $e = 1 + \gamma N$, where $\gamma > 0$ is a free parameter, and N is the number of available varieties (see Bilbiie, Ghironi and Melitz, 2019).

Unlike the previous approach, the non-CES preference system is able to generate much larger variations in the slope of the Phillips curve. We can show that $\frac{e}{e-1}\frac{\partial \ln e}{\partial \ln g} = \frac{\partial \ln N}{\partial \ln g}$, which is $N-1$ times the magnitude in the previous case. Given the same initial values for e and N, to match my empirical results, we need the number of firms to increase by a factor of 3.5, from 5 on average to 17.5 in the high growth regime. From here we can infer that models with more flexible non-CES preference under monopolistic competition or oligopolistic competition are potentially able to generate an even larger effect on the slope coefficient through this channel.[18]

Besides the plausible support in theory, the sample correlation between trend productivity growth and markups also confirms my conjunction that *higher* trend growth is associated with *lower* markups at least for the post-1980 sample. By incorporating the markup measures from De Loecker and Eeckhout (2018) since 1980, I find that the average markup ratio for the advanced economies in my sample is 1.208 in the high-growth regime and 1.265 in

[17] e goes from 5 to 6 as N increases from 5 to infinity, the percentage change in the slope is $\frac{\epsilon - e}{e-1} = 25\%$.

[18] Wang and Werning (2022) build a model that features non-CES preference (e.g. Kimball) and oligopolistic competition with Calvo pricing. They show that going from monopolistic competition to oligopolistic competition with $N = 3$ divides the slope of the Phillips curve by four.

the low-growth regime.

Although I mainly rely on the "real rigidity" channel to explain the growth regime dependence of the slope of the Phillips curve, other channels are by no means less important nor less promising. In terms of the "nominal rigidity" channel, empirical evidence on the price adjustment costs alone is quite rare though. Survey studies that look at the frequency and duration of price changes reflect both nominal and real rigidities.[19] It would be helpful to empirically distinguish one from the other, and study how technological changes affect the price adjustment costs. Regarding the sensitivity of marginal costs with respect to demand changes, globalization-related views suggest that this term has been decreasing due to the increasing fraction of imports as shares of GDP in advanced economies since the 1970s. But its state dependence on the growth regime is not clear to us.[20] There still remain a number of gaps in this topic that would benefit from future research.

2.6 Conclusion

What is the relationship between long-run productivity growth and the short-run trade-off between inflation and economic slack? This paper investigates the state dependence of the slope of the Phillips curve on the trend productivity growth. By merging the Bergeaud-Cette-Lecat *Long-Term Productivity Database* with Jordà-Schularick-Taylor *Macrohistory Database*, this paper stratifies the sample based on the level of the trend TFP growth for 17 advanced economies over more than a century. The rich cross-country variations, together with the long sample period, enable us to distinguish the different roles played by different regime variables and the productivity growth regime is shown to provide new perspectives, especially in terms of the asymmetric monetary policy effects. Monetary policy faces clear and different trade-offs between its nominal and real effects across growth regimes. Following the state-of-the-art method of estimating structural forward-looking macroeconomic equations proposed by Barnichon and Mesters (2020), empirical estimation of the slope of the NKPC suggests that the Phillips curve is *steeper* in the *high* productivity growth regime.

[19]see Nakamura and Steinsson (2013) and references therein for an extensive discussion.

[20]The relationship between trade openness and economic growth is ambiguous from both theoretical and empirical points of view. See Silajdzic and Mehic (2018) for an extensive discussion.

To explain these findings, this paper proposes a mechanism that bridges the literature on endogenous growth, market structure, and the slope of the Phillips curve. In a calibrated endogenous growth model with CES preference and oligopolistic competition, an exogenous fall in overhead costs and a simultaneous rise in research productivity could lead to higher trend growth and a more competitive market populated with more varieties/firms. If the price elasticity of demand rises with market competition, then the pass-through of marginal costs due to short-run demand changes will be larger, indicating a steeper Phillips curve, and vice versa. This mechanism is qualitatively in tune with the recent secular trends of a flattening Phillips curve and productivity slowdown amid rising market concentration and markups in many advanced economies. Quantitatively, however, an endogenous growth model with non-CES preference and oligopolistic competition should be a good path to pursue for future research. Admittedly, the model in this paper does not have nominal rigidity, so my future work will seek to link the Phillips curve and this mechanism in a unified micro-founded manner.

The policy implications of my paper are at least twofold. First, when the potency of monetary policy to steer inflation is limited in the low productivity growth regime, central banks should be more alert to inflation shocks and take decisive actions to rein in inflation before expectation de-anchors. Second, and more importantly, good growth perspectives and active business dynamism enhance the potency of monetary policy to stabilize inflation, therefore, structural reforms that can improve productivity and restore business dynamism not only help alleviate supply constraints but also improve the potency of monetary policy to stabilize inflation in the long run. Such structural reforms include more strict enforcement of antitrust laws to reduce the market power of big firms, competition-enhancing policies to reduce entry barriers, and growth-enhancing policies to foster more efficient R&D.

2.7 Appendix

2.7.1 Phillips Curve Estimation Using Theoretical IRFs

The simple three-equation New-Keynesian model (see e.g., Galí, 2015, Chapter 3) yields the following theoretical IRFs to monetary policy shocks ξ_t:

$$\hat{\pi}_t = -\lambda\Lambda_\xi\xi_t$$

$$\hat{x}_t = -(1-\beta\rho_\xi)\Lambda_\xi\xi_t$$

where $\hat{y}_t = y_t - y_{ss}, y \in \{\pi, x\}$, is measured by the percentage deviations from the steady state, and

$$\Lambda_\xi = \frac{1}{(1-\beta\rho_\xi)[\sigma(1-\rho_\xi)+\phi_y]+\lambda(\phi_\pi-\rho_\xi)}$$

β is the discount factor, σ is the elasticity of substitution, ρ_ξ is the AR(1) parameter of the monetary policy shock, ϕ_π and ϕ_y are the weight parameters in the monetary policy rule for inflation gap and the output gap. λ is the slope of the Phillips curve, and it is a function of deep model parameters as well,

$$\lambda = \frac{(1-\theta)(1-\beta\theta)}{\theta}(\chi+\sigma)$$

where θ is the probability of price adjustment each period, χ is the Frisch elasticity of labor supply. Given the following calibrations: $\beta = 0.99, \sigma = 1, \chi = 1, \theta = 0.8, \phi_\pi = 1.5, \phi_y = 0$, and $\rho_\xi = 0.2$, we have $\Lambda_\xi = 1.2873$ and the coefficients of the "hybrid" NKPC are $\gamma_b = 0, \gamma_f = \beta = 0.99$, and $\lambda = 0.104$.

Consider a one-period innovation in monetary policy shock,

$$\xi_t = \rho_\xi\xi_{t-1} + \epsilon_t^\xi$$

and $\epsilon_0^\xi = 1, \epsilon_t^\xi = 0, \forall t \neq 0$ or $\xi_t = \rho_\xi^t, \forall t \geq 0$. The theoretical IRFs can be derived as follows:

$$\mathscr{R}_h^\pi = \frac{E[(\pi_{t+h}-\pi_{t-1})\xi_t]}{E[\xi_t^2]} = \frac{E[(\pi_{t+h}-\pi_{ss}-(\pi_{t-1}-\pi_{ss}))\xi_t]}{E[\xi_t^2]} = \frac{E[(\hat{\pi}_{t+h}-\hat{\pi}_{t-1})\xi_t]}{E[\xi_t^2]} = -\lambda\Lambda_\xi\rho_\xi^h$$

where $E[\hat{\pi}_{t-1}\xi_t] = 0$ as $\hat{\pi}_{t-1}$ is predetermined when the shock occurs at time t.

Similarly, we can derive the IRFs of the output gap and get

$$\mathscr{R}_h^x = -(1-\beta\rho_\xi)\Lambda_\xi\rho_\xi^h$$

Notice that

$$\mathscr{R}_h^{\pi} = \frac{\lambda}{1-\beta\rho_{\xi}}\mathscr{R}_h^{x}$$

and

$$\mathscr{R}_{h+1}^{\pi} = \frac{\lambda\rho_{\xi}}{1-\beta\rho_{\xi}}\mathscr{R}_h^{x} \propto \mathscr{R}_h^{x}$$

If we simply regress $\mathscr{R}_h^{\pi}$ on $\mathscr{R}_{h-1}^{\pi}$, $\mathscr{R}_{h+1}^{\pi}$ and $\mathscr{R}_h^{x}$, we will have perfect collinearity problem.

The series of monetary policy shock ξ_t and the theoretical IRFs of inflation $\mathscr{R}_t^{\pi}$ and output gap $\mathscr{R}_t^{x}$ are presented in Table 2.4. We can see that as h increases, the IRFs revert back to zero exponentially. So the horizon over which we run regression (with no constant term) in the second step should not be too large, here I choose $H = 5$.

Table 2.4. Theoretical IRFs to monetary policy shocks

h	ξ_h	$\mathscr{R}_h^{\pi}$	$\mathscr{R}_h^{x}$	$\mathscr{R}_{h-1}^{\pi}$	$\mathscr{R}_{h+1}^{\pi}$
0	1	-0.13388	-1.03244	0	-0.02678
1	0.2	-0.02678	-0.20649	-0.13388	-0.00536
2	0.04	-0.00536	-0.04130	-0.02678	-0.00107
3	0.008	-0.00107	-0.00826	-0.00536	-0.00021
4	0.0016	-0.00021	-0.00165	-0.00107	-0.00004
5	0.00032	-0.00004	-0.00033	-0.00021	-0.00001

The unrestricted OLS estimates of the Phillips curve using the theoretical IRFs shows that γ_b is essentially zero, the expected inflation term is omitted due to perfect collinearity, and the slope coefficient is estimated to be 0.1297, which is approximately 25% larger than the true value. If we impose the long-run restriction that $\gamma_b + \gamma_f = 1$, the point estimate is 1 for γ_f, approximately zero for γ_b, but the slope coefficient is omitted due to perfect collinearity.

To estimate the coefficients that match their ranges in theoretical models, I further constrain three coefficients to be between 0 and 1, i.e., $0 < \gamma_b, \gamma_f, \lambda < 1$. Because the range of the inverse logit function is the interval $(0, 1)$, we can use the inverse logit function to set this restriction and estimate the coefficients using Nonlinear Least Squares (NLS) method.

The coefficient estimates without imposing the long-run restriction depend on the initial values we choose for the coefficients when running the NLS method. The point estimate of γ_b is always approximately zero, whereas the point estimate of γ_f will rest at its initial value, suggesting that there're insufficient variations in the "data" – theoretical IRFs – to identify all three coefficients. For example, if we choose the initial value of γ_f to be approximately 0, or the logit function to be some large negative number (say -20), then the point estimate of λ remains at 0.1297. But if the initial value of γ_f is set to be 0.5, or the logit function to be at 0, then the point estimate of λ is 0.1167, closer to the true value. And if we instead choose the initial value of γ_f to be approximately 1, or the logit function to be some large positive number (say 20), then λ is estimated to be 0.1037, which is almost at the true value.

Alternatively, we can impose the long-run restriction $\gamma_b + \gamma_f = 1$ to improve the point estimates. After imposing the long-run restriction, the slope estimate becomes 0.1037, and γ_f is bounded by 1. What's more, if we knew and imposed the true long-run restriction $\gamma_b + \gamma_f = 0.99$ under this calibration, we could actually recover the true coefficients.

2.7.2 State-Dependent Analysis of Monetary Policy Effects

Following Jordà, Schularick and Taylor (2020*a*), in this section I conduct state-dependent analyses for the monetary policy effects across different regimes by estimating the impulse responses of price and real output using local projection with the trilemma monetary shocks as external instruments. The baseline LP-IV specification follows equation (2.6) except that the response variables are real GDP per capita and log CPI, and the macroeconomic control variables on the right-hand side do not include variables related to output gap nor inflation.

Inflation Regimes

First, let me present the state dependence of the impulse response functions by stratifying the data based on the level of inflation for the sample period starting from 1890. I use an annual 2% CPI inflation rate cutoff to define the high/low inflation regime after excluding a few hyperinflation periods (greater than 45% annual rate, occurred e.g., in Germany after the first world war).

Figure 2.16 shows the asymmetric impulse responses of real GDP per capita and price level to a 1 percentage point increase in short-term interest rate for the baseline (solid blue line), the low inflation regime (dashed green line) and the high inflation regime (long dashed red line). The analysis is conducted using both the full sample (1890-2006 excluding world wars) and the post-WW2 sample (1948-2006). The response of output to monetary policy is relatively stronger when the inflation is above 2%. The full sample results show a more significant difference between the two regimes, and it takes longer time for the monetary policy to have real effects when the inflation is below 2%.

The response of price level to monetary policy changed somehow after WW2, and the difference between the two regimes became less observable. In the full core sample results, we find almost no effect on the price level for the first three years when the inflation is below 2%, while in the post-WW2 sample, the effect in the low inflation regime is even stronger than that in the high inflation regime for the first two years. Put two pairs of impulse responses together, we don't find the typical trade-off between nominal and real effects of monetary policy implied by a nominal rigidity story.

Credit Growth Regimes

Next, let's turn to the state dependence of the impulse responses by stratifying the data based on the credit growth for the sample period starting from 1890. The high mortgage / non-mortgage credit growth regime is when a country's 3-year mean changes in mortgage / non-mortgage credit over GDP (y_{it}) is above its historic mean changes, i.e., $y_{it} > \bar{y}_i$ or $y_{it} - \bar{y}_i > 0$.

Figure 2.17 shows the asymmetric impulse responses of real GDP per capita and price level to a 1 percentage point increase in short-term interest rate for the baseline (solid blue line), the low mortgage credit growth regime (dashed green line) and the high mortgage credit growth regime (long dashed red line). The left panel shows the results for the full sample (1890-2006 excluding world wars), whereas the right panel shows the results for the post-WW2 sample (1948-2006).

The main takeaway is twofold: first, the price responses show little asymmetry across regimes relative to the baseline, especially after the second world war; second, the effect of interest rates on output differs a lot depending on whether mortgage credit growth is fast or slow, monetary policy has much stronger real effects when the mortgage credit grows faster. In response to the same 1% interest rate increase, the decline in output is about 3 percentage points higher 4 years after the intervention.

Things are quite the opposite with the non-mortgage credit stratification. Figure 2.18 shows the asymmetric impulse responses of output and price level to a 1 percentage point increase in short-term interest rate for the baseline (solid blue line), the low non-mortgage credit growth regime (dashed green line) and the high non-mortgage credit growth regime (long dashed red line). On the one hand, the output responses show little asymmetry relative to the baseline in both regimes at least for the first three years after the policy change, the regime-specific impulse responses lie within the 90% confidence bands of the baseline. On the other hand, the effect of interest rates on the price level differs a bit depending on whether non-mortgage credit growth is fast or slow, especially in the full sample results. In response to the same 1% interest rate increase, the decline in the price level is about 2.5 percentage points higher 4 years after the intervention.

TFP Growth Regime

Now let's switch to the main focus of this paper concerning how the monetary policy may have different effects in episodes of high and low productivity growth. Figure 2.19 shows the asymmetric impulse responses of real GDP per capita and price level to a percentage point increase in short-term interest rate for the baseline (solid blue line), the low growth regimes (dashed green line) and the high growth regimes (long dashed red line). The TFP growth regimes are classified by the baseline regime-switching approach.

The output response to monetary policy is relatively stronger in the low-growth regime than in the high-growth regime. In terms of magnitude, in the full sample results, at the end of year 4, in response to a 100 bps increase in short-term interest rate, output declines by approximately 2.3% in the baseline, 2.5% in the low-growth regime and 1.0% in the high-growth regime. The post-WW2 sample delivers similar results with slightly smaller magnitudes, output decreases by 1.9% in the baseline and the low-growth regime, and by 0.8% in the high-growth regime.

In contrast, the price response is almost muted in the low growth regime for the first 4 years after the change in interest rates, while price responds quite strongly in a period of high productivity growth. In the full sample results, the price level drops by more than 5% at the end of year 4, and in post-WW2 sample results, the price level drops by more than 4%. Unlike the previous discussion about inflation regimes or credit growth regimes, there appears to be a clear trade-off between output and price level responses when we stratify the data by TFP growth regimes.

In relation to the correlation among different regime dummies, we do observe that the asymmetric responses of output look qualitatively similar to those when we stratify the data based on mortgage credit growth, whereas the asymmetric responses of price look qualitatively similar to those when we stratify the data based on inflation. In particular, the TFP growth regime dummy correlates positively with the inflation regime dummy (esp. for the full sample) but negatively with the mortgage credit growth regime dummy, and we observe that the effect of interest rates on output is stronger in *low* TFP growth regime and *high* mortgage credit growth regime, while the effect on price is weaker/muted in *low* TFP growth regime and *low* inflation regime. But all of these regimes differ in the combination of nominal and real effects of monetary policy.

2.7.3 Model Appendix

The model in Section 2.5 follows Smulders and Van de Klundert (1995, 1997) in discrete time. It combines elements from endogenous growth theory and industrial organization literature on innovation. The exercise here is to calibrate their model and show that the long-run forces that change the trend growth along the balanced growth path may alter the number of firms in the market, which will affect firms' perceived price elasticity of demand. The model has no nominal rigidity and treats entry and exit in an old-fashioned way. That is, the number of firms is fixed in the short run, but in the long run, there is free entry and exit. I only focus on comparative statics analysis along the balanced growth path.

There are two sectors: a high-tech sector with differentiated products and a traditional sector with a homogenous good. The traditional sector is competitive while the high-tech sector is oligopolistic, with a finite number of imperfectly substitutable product types. The number of firms/brands N is determined in the long run by free entry condition. Agents have rational expectations and perfect foresight. Time is discrete and infinite.

Households

The representative household's preferences are given by

$$U = E_t \sum_{s=0}^{\infty} \beta^s \frac{C_{t+s}^{1-\sigma}}{1-\sigma} \tag{2.11}$$

Consumers trade off future consumption for present consumption according to a CRRA utility function with an elasticity of intertemporal substitution $1/\sigma$ and discount factor β. The final consumption good is a composite good, according to equation (2.12), of the homogeneous good Z and the differentiated good X, with η representing the expenditure share of X-goods:

$$C_t = X_t^{\eta} Z_t^{1-\eta}, \quad 0 < \eta < 1 \tag{2.12}$$

The X-goods is a bundle of N varieties which are imperfect substitutes with constant elasticity of substitution ϵ:

$$X_t = \left(\sum_{j=1}^{N} x_{jt}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}}, \quad \epsilon > 1 \tag{2.13}$$

Besides consumption, households supply labor inelastically for a nominal wage rate of W_t, and the total labor supply is denoted as L. The period budget constraint is

$$P_t^C C_t + Q_{t+1} B_{t+1} \leq B_t + W_t L + T_t \tag{2.14}$$

where P_t^C is the ideal price index for the final consumption good, B_{t+1} represents purchases of one-period bonds that will mature in period $t+1$ at a price of $Q_{t+1} \equiv \frac{1}{1+i_{t+1}}$, W_t is the nominal wage rate, and T_t is a lump-sum component of income (which may include, among other things, the dividends from ownership of firms).

Proposition 1. *Household behavior can be summarized by the following optimality conditions:*

i. The optimal consumption/savings decision is described by

$$Q_{t+1} = \beta E_t \left[\left(\frac{C_{t+1}}{C_t} \right)^{-\sigma} \frac{P_t^C}{P_{t+1}^C} \right] \tag{2.15}$$

ii. The optimal composition of sectoral goods satisfies:

$$X_t = \eta \frac{P_t^C}{P_t^X} C_t \tag{2.16}$$

$$Z_t = (1-\eta) \frac{P_t^C}{P_t^Z} C_t \tag{2.17}$$

iii. The optimal composition of differentiated goods satisfies:

$$x_{jt} = X_t \left(\frac{p_{jt}^x}{P_t^X} \right)^{-\epsilon} \tag{2.18}$$

or

$$x_{jt} = \eta C_t \left(\frac{P_t^X}{P_t^C} \right)^{-1} \left(\frac{p_{jt}^x}{P_t^X} \right)^{-\epsilon} \tag{2.19}$$

iv. The price indices

$$P_t^C = \left(\frac{P_t^X}{\eta} \right)^{\eta} \left(\frac{P_t^Z}{1-\eta} \right)^{1-\eta} \tag{2.20}$$

$$P_t^X = \left(\sum_{j=1}^{N} \left(p_{jt}^x \right)^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}} \tag{2.21}$$

$$P_t^X X_t = \sum_{j=1}^{N} p_{jt}^x x_{jt} \tag{2.22}$$

Firms

The homogeneous good Z is produced subject to a linear technology using labor only. By choice of units, one unit of labor is required to produce one unit of Z:

$$Z = L_{Zt} \tag{2.23}$$

Perfect competition implies that the price of the homogeneous good equals the wage rate, i.e.,

$$P_t^Z = W_t \tag{2.24}$$

Each of the differentiated products $j \in \{1, \cdots, N\}$ is produced by one manufacturer, using labor and knowledge as inputs. The production and management of each variety/brand require its own product/firm-specific knowledge h_{jt}. The technology to produce differentiated products is given by

$$x_{jt} = h_{jt} L_{xjt} \tag{2.25}$$

where L_{xjt} is production labor time and h_{jt} is labor productivity.

The law of motion of firm-specific knowledge is formulated as

$$h_{j,t+1} - h_{jt} = \xi \left(h_{jt}^{1-\alpha} H_t^{\alpha} \right) L_{rjt} \tag{2.26}$$

Firms can employ labor L_{rjt} to conduct in-house R&D activities to accumulate new firm-specific knowledge. The knowledge base in R&D activities is represented by the term $h_{jt}^{1-\alpha} H_t^{\alpha}$ in equation (2.26), where $H_t \equiv \frac{1}{N} \sum_{j=1}^{N} h_{jt}$ is the average knowledge level in the economy and $0 < \alpha < 1$ represents the degree of knowledge spillover. In particular, innovation is mainly based on its own firm-specific knowledge but may be subject to diminishing returns as $1 - \alpha < 1$. Furthermore, research may benefit from spillover from the knowledge developed for other products and from generally applicable public knowledge as indicated by H_t.

Besides labor costs for production L_{xjt} and R&D activities L_{rjt}, firms also have to incur a fixed overhead cost L_f for any level of production every period, which captures the costs of management, marketing, and coordination. Her period profit flow is given by

$$D_{jt} = p_{jt}^{x} x_{jt} - (L_{xjt} + L_{rjt} + L_f) W_t \tag{2.27}$$

Each producer in the high-tech sector maximizes her firm value by choosing L_{xjt}, L_{rjt} and p_{jt}^{x}, taking as given the demand from the final sector for consumption, the aggregates $\Upsilon_t \equiv \{P_t^C, N, H_t, C_t\}$, the production function (2.25), and the law of motion of firm-specific knowledge (2.26).

Let's denote the value of the firm j as $V_t^j(\cdot) \equiv V^j(h_{jt}; \Upsilon_t)$ such that

$$V^j(h_{jt}; \Upsilon_t) = \max_{L_{xjt}, L_{rjt}, h_{j,t+1}, p_{jt}^{x}} \quad D_{jt} + E_t \left[Q_{t,t+1} V^j(h_{j,t+1}; \Upsilon_{t+1}) \right] \tag{2.28}$$

subject to equation (2.19), (2.25), and (2.26). Note that with oligopolistic competition at the sector level, P_t^X is a function of own price p_{jt}, price of other varieties $p_{-j,t}$, and the number of varieties N.

Proposition 2. *Differentiated goods producer's optimization problem can be summarized by the following optimality conditions:*

i. Optimal pricing rule

$$p_{jt}^x = \frac{e_{jt}}{e_{jt}-1}\frac{W_t}{h_{jt}}$$

ii. Optimal investment in R&D activities:

$$W_t = \Psi_{jt}\xi h_{jt}^{1-\alpha}H_t^\alpha \tag{2.29}$$

where Ψ_{jt} denotes the marginal cost/shadow price of h_{jt}.

iii. The no-arbitrage condition between investing in the capital market versus creating new knowledge:

$$\Psi_{jt} = E_t\left\{Q_{t,t+1}\left[p_{j,t+1}^x\left(1-\frac{1}{e_{j,t+1}}\right)L_{x,j,t+1} + \Psi_{j,t+1}(1+\xi(1-\alpha)h_{j,t+1}^{-\alpha}H_{t+1}^\alpha L_{r,j,t+1})\right]\right\} \tag{2.30}$$

Symmetric Balanced Growth Equilibrium

Assuming perfect foresight of agents and symmetry across firms, we may drop the subscript j. Each firm has the same level of productivity and firm-specific knowledge which consequently equals the average knowledge level: $h_{jt} = H_t$. The economic growth comes from the accumulation of firm-specific knowledge, let $g_t = H_t/H_{t-1} - 1$ be the growth rate of knowledge. Assume a constant number of varieties N along the balanced growth path.

Proposition 3. *The general equilibrium balanced growth path for a given number of varieties N where $g_{t+1} = g_t = g$ and $\pi_t^x \equiv \frac{p_t^x}{p_{t-1}^x} - 1 = \pi^x, \pi_t^C \equiv \frac{P_t^C}{P_{t-1}^C} - 1 = \pi^C, \forall t$, can be characterized as follows:*

i. Growth rates are related:

$$g \equiv g_H = g_X = \frac{1}{\eta}g_C = \frac{1}{\eta}g_Y = \frac{1}{\eta}g_w = \frac{1}{1-\eta}(\pi^C - \pi^x) = g_W - \pi^x \tag{2.31}$$

ii. Constant markup pricing:

$$p_t^x = \frac{e}{e-1}\frac{W_t}{H_t} \tag{2.32}$$

where $e \equiv e(\epsilon, N)$ defined as

$$e^B \equiv \epsilon - (\epsilon - 1)\frac{1}{N}$$

for Bertrand (price) competition, and

$$e^C \equiv \frac{\epsilon}{1 + (\epsilon - 1)\frac{1}{N}}$$

for Cournot (quantity) competition.

iii. Labor allocations L_Z, L_x, L_r are constant over time.

$$L_r = \frac{g}{\xi} \tag{2.33}$$

$$L_x = \frac{\eta(e-1)}{e-\eta}\left(\frac{L}{N} - L_f - L_r\right) \tag{2.34}$$

$$L_Z = \frac{1-\eta}{\eta}\frac{e}{e-1}NL_x \tag{2.35}$$

iv. "Preference" line (supply of savings) summarizes the household's intertemporal choice:

$$r_x = \rho + (\sigma\eta + 1 - \eta)g \tag{2.36}$$

where $r_x \equiv E_t[i_{t+1}] - \pi_{ss}^x$, and $\rho \equiv \frac{1}{\beta} - 1$.

v. "Technology" line (demand for savings) summarizes the firm's investment decision:

$$r_x = \left(\frac{e(1-\eta)}{e-\eta} - \alpha\right)g + \xi\frac{\eta(e-1)}{e-\eta}\left(\frac{L}{N} - L_f\right) \tag{2.37}$$

vi. The interaction between "Preference" line and "Technology" line yields the short-run equilibrium (SRE) line along which the demand and supply of savings are equalized:

$$g = \frac{\xi b\left(\frac{L}{N} - L_f\right) - \rho}{(\sigma-1)\eta + \alpha + b} \tag{2.38}$$

where $b = 1 - \frac{e(1-\eta)}{e-\eta} = 1 - \frac{1-\eta}{1-\eta/e}$.

vii. Increasing concentration (smaller N) is conducive to growth.

With the entry and exit of firms in the differentiated goods sector, (2.38) is no longer a complete characterization of the long-run equilibrium. A positive net present value of profits induces the entry of new firms until no firm can earn a profit by entering the market. This endogenizes the number of firms. To make a solution tractable, it is assumed that new firms have the same structure and productivity as incumbent firms and that there is no sunk cost of entry.

Profits in the equilibrium along the balanced growth path are defined as

$$D_{jt} = p_t^x x_t - W_t(L_x + L_f + L_r) = W_t\left(\frac{1}{e-1}L_x - L_f - L_r\right)$$

Substituting in the expressions for L_x and L_r, we find the zero-profit condition implies

$$D_{jt} = W_t\left[\frac{\eta}{e-\eta}\frac{L}{N} - \frac{e}{e-\eta}(L_f + L_r)\right] = \frac{w_t}{e-\eta}\left[\eta\frac{L}{N} - e(L_f + \frac{g}{\xi})\right] = 0 \iff g = \xi\left(\frac{\eta}{e}\frac{L}{N} - L_f\right) \tag{2.39}$$

Note that e is increasing in N, so g **is decreasing in** N. Intuitively, a higher rate of innovation (g) implies higher fixed R&D costs ($w_t L_{rt}$ is independent of x_t), so there is less room for firms. Under free entry, growth is higher and the number of firms in the market is smaller relative to the short-run equilibrium with fixed N.

Proposition 4. *The general equilibrium balanced growth path with entry and exit, in the long run, is characterized by the interaction between the short-run equilibrium (SRE) line (2.38) and the zero-profit (ZP) line (2.39) in the $g-N$ space. This equilibrium is stable when the SRE line is flatter than the ZP line or the ZP line cuts the SRE line from above. See Smulders and Van de Klundert (1995) and the references therein.*

Calibration

The model is calibrated in such a way that the main statistics are kept close to their post-WW2 sample averages if applicable. σ in the utility function is inversely related to the elasticity of intertemporal substitution (EIS), and the range for EIS is still quite debatable in the literature. I follow the results from Attanasio and Weber (1995) that EIS is between 0.1 and 0.5, which implies $\sigma \in [2, 10]$, and I choose $\sigma = 8$. σ only affects the SRE line, higher σ flattens out the SRE line quite significantly, and increases the sensitivity of the number of firms with respect to shifts in the ZP line. The value of ϵ governs the elasticity of substitution between differentiated products and is also related to the price elasticity of demand and markup ratio. I choose $\epsilon = 6$ so that the gross markup is 1.25 on average. $1-\eta$ is the expenditure share of homogeneous products that don't experience much technological progress, and I refer to the share of the primary sector in the economy, around 5%. α measures the knowledge spillover, and I set it to 0.9 following Kung (2015). The discount factor β is set to be 0.99, and the labor supply L is normalized to 100.[21] L_f and ξ are the two treatment variables I will consider later. When choosing f and ξ, I target the quarterly growth rate of knowledge g to be 0.54%, which corresponds to a 2.04% average annual growth rate in the post-WW2 sample.[22] One such pair of (f, ξ) is $(2.05, 0.0035)$.

Table 2.5. Parameter calibration

Parameter	Description	Value
β	discount factor	0.99
σ	inverse of EIS	8
ϵ	elasticity of substitution between varieties	6
η	share of differentiated goods sector	0.95
α	knowledge spillover	0.9
L	labor supply	100
L_f	overhead labor	2.05
ξ	research productivity	0.0035

Comparative Statics Analysis

This paper focuses on the two structural changes that recent literature has pointed out to explain the reasons behind the productivity boom in the mid-1990s: a fall in the overhead costs (L_f) and a rise in the efficiency of R&D (ξ). Figure 2.20 presents three cases of comparative statics: a one-time fall in overhead costs L_f in Panel (a), a one-time increase

[21]For post-WW2 sample, the average growth rate is set to be the same as TFP growth rate, which is about 2.04% on average. The real short-term interest rate is about 2.2%, so quarterly calibration of $\beta = 1/(1+r^*)(1+g) \approx 0.99$.

[22]TFP growth in this model is η fraction of the knowledge growth rate g.

in research productivity ξ in Panel (b), and both changes in Panel (c). Under the current calibration, the SRE line (in blue) is almost invariant to the change in L_f, and the new long-run equilibrium in Panel (a) features lower growth and more firms. The intuition is simple: lower overhead costs L_f encourage more firms to enter the market. However, this lowers the market share of existing firms, discouraging them from innovation, therefore, growth declines. Unlike Aghion et al. (2019), this model is unable to generate a productivity boom in transition but a productivity slowdown in the new long-run equilibrium.

On the contrary, research productivity mainly affects the growth rate, while the number of products/firms stays more or less the same. The new equilibrium in Panel (b) exhibits much higher growth with almost the same number of firms. The intuition is as follows: more efficient research encourages existing firms to invest more in firm-specific knowledge, improving their profitability. However, a high rate of innovation also implies higher fixed R&D costs, leaving less room for firms. Under current calibration, these two effects on entry and exit nearly cancel, leaving the number of firms almost the same.

In Panel (c), both changes occur simultaneously and I specify the new parameter values such that the growth rate along the new balanced growth path is targeted at the average trend TFP growth rate in the post-WW2 sample, which is 2.61%. The number of firms increases from 5 to almost 9, and the implied slope of the Phillips curve should be steeper.

Qualitatively, this model is able to generate correct predictions on the slope of the Phillips curve: a fall in overhead costs and a simultaneous rise in research productivity lead to higher growth and a steeper slope, the impact on the slope of the Phillips curve is quantitatively small though. The limit of this mechanism is restricted by the range of e. Even if $N \to \infty$, $e \to \epsilon$, the percentage change in $\frac{e-1}{\phi_R}$ term amounts to 25%[23], whereas the empirical estimate for the high growth regime can be more than twice the average slope.

[23] When $\epsilon = 6$ and $N = 5$, $e = 5$. Percentage change in $\frac{e-1}{\phi_R}$ in the extrem case is equal to $\frac{\epsilon-e}{e-1} = 0.25$.

2.7.4 Appendix Tables and Figures

Table 2.6. Estimates of the Phillips curve: alternative growth regime classifications

SAMPLE	BIN	UNRESTRICTED			RESTRICTED	
		γ_f	γ_b	λ	γ_f	λ
Panel A: *Above universal cutoff*						
Full	HG	0.51 (0.17)	0.55 (0.16)	0.28 (0.41)	0.48 (0.12)	0.35 (0.30)
Full	LG	0.50 (0.29)	0.66 (0.48)	0.00 (.)	0.39 (0.47)	0.07 (0.19)
Post-WW2	HG	0.47 (0.16)	0.58 (0.15)	0.65 (0.37)	0.44 (0.12)	0.67 (0.32)
Post-WW2	LG	0.20 (0.31)	0.22 (0.42)	0.50 (0.46)	0.40 (0.27)	0.01 (0.21)
Panel B: *Above country-specific mean growth*						
Full	HG	0.54 (0.20)	0.57 (0.17)	0.04 (0.58)	0.46 (0.15)	0.22 (0.49)
Full	LG	0.60 (0.21)	0.64 (0.40)	0.00 (.)	0.56 (0.19)	0.00 (.)
Post-WW2	HG	0.42 (0.18)	0.66 (0.17)	0.59 (0.47)	0.38 (0.14)	0.62 (0.43)
Post-WW2	LG	0.19 (0.24)	0.40 (0.41)	0.13 (0.15)	0.23 (0.24)	0.04 (0.12)
Panel C: *45 and 55 percentiles*						
Full	HG	0.56 (0.17)	0.56 (0.18)	0.12 (0.79)	0.50 (0.14)	0.30 (0.67)
Full	LG	0.40 (0.38)	0.00 (.)	0.00 (.)	0.50 (0.30)	0.00 (.)
Post-WW2	HG	0.49 (0.16)	0.54 (0.16)	0.61 (0.50)	0.47 (0.12)	0.64 (0.41)
Post-WW2	LG	0.00 (0.26)	0.00 (.)	0.14 (0.08)	0.28 (0.53)	0.03 (0.14)

Notes: Different panels apply alternative growth regime classification methods, see text. Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. All coefficients are constrained to be between 0 and 1 and estimated using NLS with $H = 6$. Coloumn "RESTRICTED" imposes an additional restriction: $\gamma_b + \gamma_f = 1$. In Column "BIN", "HG" denotes the high-growth regime, "LG" denotes the low-growth regime. The output gap is measured using the HP filter with a smoothing parameter of 100.

Table 2.7. Estimates of the Phillips curve: alternative measures of output gap

SPEC	BIN	UNRESTRICTED			RESTRICTED	
		γ_f	γ_b	λ	γ_f	λ
Panel B: *Full sample*						
linear	HG	0.45 (0.15)	0.30 (0.23)	0.65 (0.43)	0.53 (0.11)	0.31 (0.19)
linear	LG	0.47 (0.15)	0.48 (0.42)	0.00 (.)	0.47 (0.13)	0.00 (.)
linear, sd	HG	0.41 (0.11)	0.39 (0.12)	0.93 (0.31)	0.50 (0.10)	0.60 (0.23)
linear, sd	LG	0.19 (0.41)	0.94 (0.70)	0.05 (0.07)	0.28 (0.28)	0.03 (0.06)
quadratic	HG	0.45 (0.15)	0.29 (0.24)	0.67 (0.46)	0.53 (0.11)	0.33 (0.20)
quadratic	LG	0.39 (0.91)	0.45 (1.34)	0.02 (0.11)	0.16 (0.49)	0.04 (0.07)
Panel B: *Post-WW2 sample*						
linear	HG	0.37 (0.28)	0.30 (0.35)	0.43 (0.52)	0.51 (0.14)	0.13 (0.15)
linear	LG	0.39 (0.14)	0.48 (0.32)	0.00 (.)	0.38 (0.13)	0.00 (.)
linear, sd	HG	0.15 (0.07)	0.76 (0.05)	0.87 (0.12)	0.23 (0.06)	0.78 (0.12)
linear, sd	LG	0.26 (0.30)	0.52 (0.42)	0.08 (0.16)	0.28 (0.27)	0.04 (0.14)
quadratic	HG	0.31 (0.24)	0.32 (0.23)	0.66 (0.49)	0.49 (0.13)	0.22 (0.19)
quadratic	LG	0.37 (0.30)	0.49 (0.41)	0.00 (0.12)	0.36 (0.13)	0.00 (.)

Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. All coefficients are constrained to be between 0 and 1 and estimated by NLS with $H = 6$. Coloumn "RESTRICTED" imposes an additional restriction: $\gamma_b + \gamma_f = 1$. Column "SPEC" specifies the measure of the output gap. In Column "BIN", "HG" denotes the high-growth regime, "LG" denotes the low-growth regime. Regimes are classified by the regime-switching approach. See the text for the definitions of different measures of the output gap.

Table 2.8. Estimates of the Phillips curve: alternative projection horizons

SAMPLE	BIN	UNRESTRICTED			RESTRICTED	
		γ_f	γ_b	λ	γ_f	λ
Panel A: $H = 9$						
Full	HG	0.53 (0.18)	0.55 (0.17)	0.12 (0.51)	0.49 (0.14)	0.22 (0.42)
Full	LG	0.50 (0.06)	0.76 (0.07)	0.00 (.)	0.42 (0.09)	0.00 (.)
Post-WW2	HG	0.35 (0.18)	0.56 (0.15)	0.67 (0.45)	0.41 (0.13)	0.54 (0.35)
Post-WW2	LG	0.51 (0.22)	0.50 (0.22)	0.00 (.)	0.51 (0.18)	0.00 (.)
Panel B: $H = 12$						
Full	HG	0.50 (0.15)	0.57 (0.15)	0.00 (.)	0.47 (0.12)	0.00 (.)
Full	LG	0.49 (0.14)	0.59 (0.14)	0.00 (.)	0.45 (0.12)	0.00 (.)
Post-WW2	HG	0.39 (0.18)	0.63 (0.14)	0.35 (0.35)	0.38 (0.13)	0.37 (0.30)
Post-WW2	LG	0.47 (0.19)	0.56 (0.19)	0.00 (.)	0.46 (0.16)	0.00 (.)

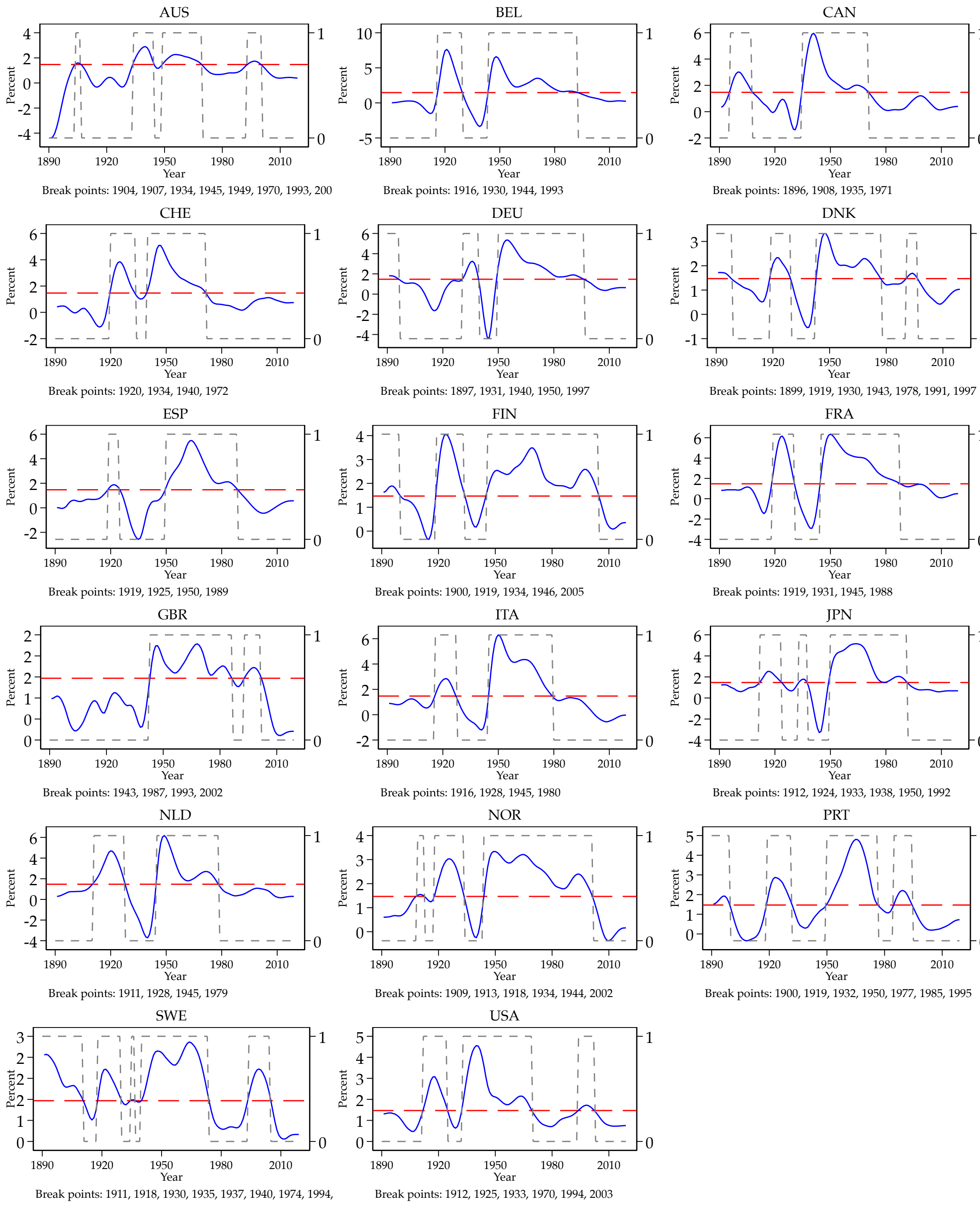
Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. All coefficients are constrained to be between 0 and 1 and estimated using NLS with $H = 9$ in Panel A, $H = 12$ in Panel B. Column “RESTRICTED” imposes an additional restriction: $\gamma_b + \gamma_f = 1$. In Column “BIN”, “HG” denotes the high-growth regime, “LG” denotes the low-growth regime. The output gap is measured using the HP filter with a smoothing parameter of 100. See text.

Table 2.9. Estimates of the Phillips curve: alternative sample selections

SAMPLE	BIN	UNRESTRICTED			RESTRICTED	
		γ_f	γ_b	λ	γ_f	λ
Panel A: *Include more recent samples*						
Full	HG	0.52 (0.10)	0.55 (0.10)	0.91 (0.41)	0.49 (0.08)	0.98 (0.36)
Full	LG	0.29 (0.37)	0.00 (.)	0.00 (.)	0.38 (0.32)	0.00 (.)
Post-WW2	HG	0.33 (0.08)	0.58 (0.08)	1.00 (.)	0.38 (0.08)	0.96 (0.27)
Post-WW2	LG	0.25 (0.44)	0.27 (0.70)	0.04 (0.24)	0.28 (0.42)	0.09 (0.22)
Panel B: *Subsample of European countries*						
Full	HG	0.49 (0.17)	0.22 (0.29)	0.65 (0.46)	0.57 (0.13)	0.33 (0.21)
Full	LG	0.17 (0.12)	0.31 (0.21)	0.20 (0.06)	0.32 (0.12)	0.11 (0.03)
Post-WW2	HG	0.47 (0.09)	0.14 (0.09)	1.00 (.)	0.59 (0.13)	0.56 (0.26)
Post-WW2	LG	0.45 (0.33)	0.49 (1.64)	0.01 (0.48)	0.46 (0.22)	0.00 (.)

Notes: In Panel A, 17 advanced economies are included. Full sample: 1890-1908, 1921-1933, and 1948-2012. Post-WW2 sample: 1948-2012. In Panel B, only European countries are included. Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. All coefficients are constrained to be between 0 and 1 and estimated by NLS with $H = 6$. Coloumn "RESTRICTED" imposes an additional restriction: $\gamma_b + \gamma_f = 1$. In Column "BIN", "HG" denotes the high-growth regime, "LG" denotes the low-growth regime. The output gap is measured using the HP filter with a smoothing parameter of 100. Regimes are classified by the "regime-switching" approach. See text.

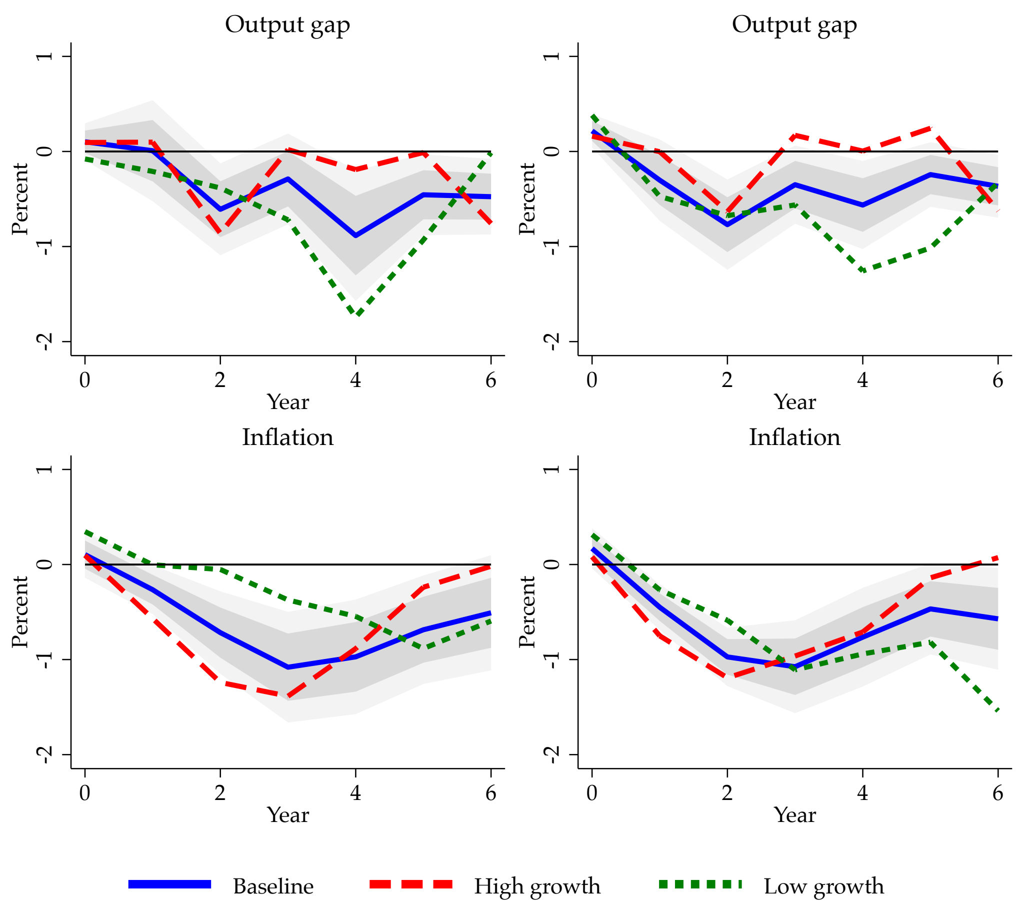
Figure 2.4. Regime classification results: above universal cutoff



Notes: Solid blue lines: trend TFP growth over time (left axis); dashed grey lines: regime dummy (right axis), equals 1 in the high-growth regime, 0 in the low-growth regime; red dashed lines: sample average. Sample period: 1890-2019.

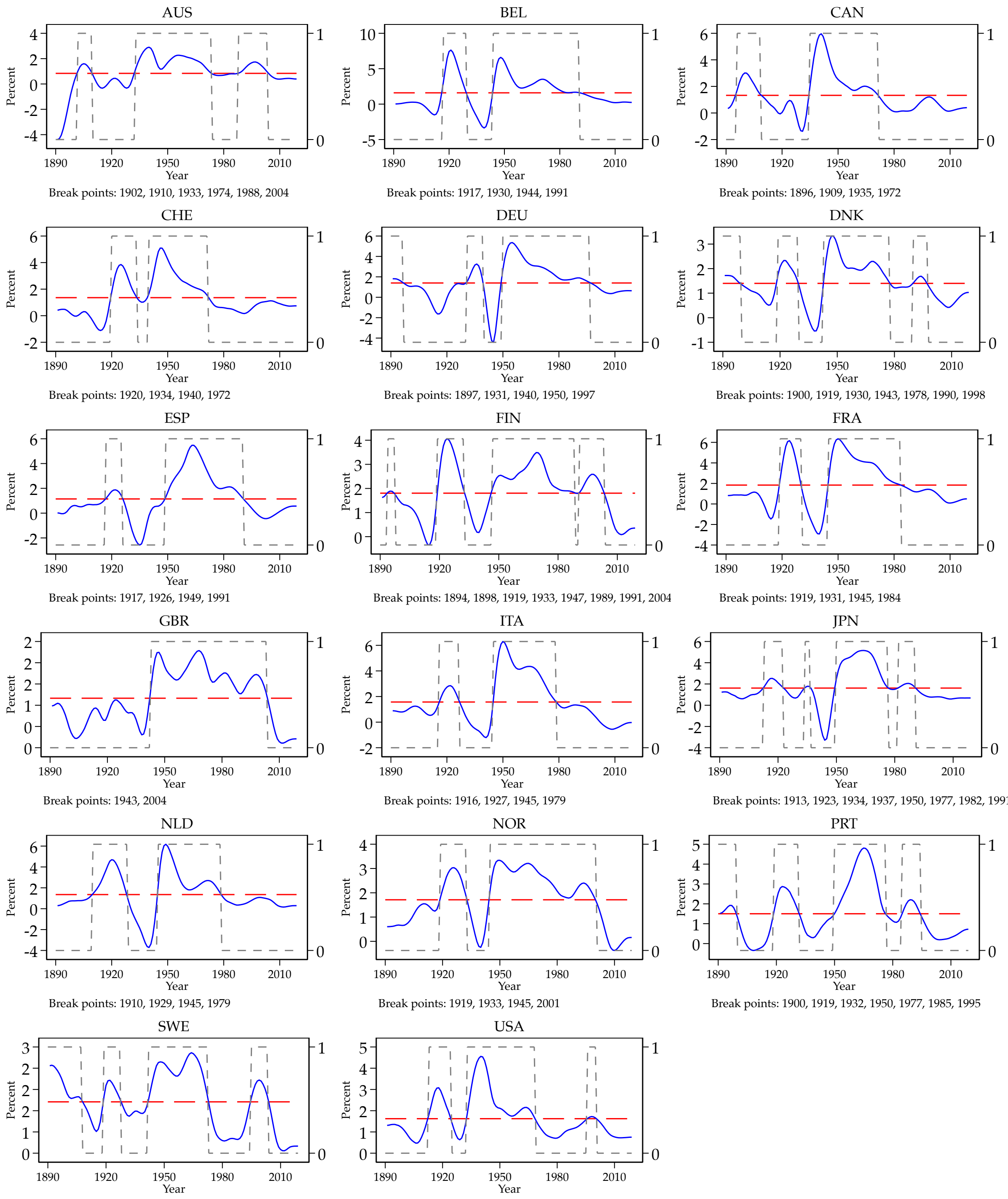
Figure 2.5. Asymmetric responses of inflation and output gap: above universal cutoff

(a) Full sample (b) Post-WW2



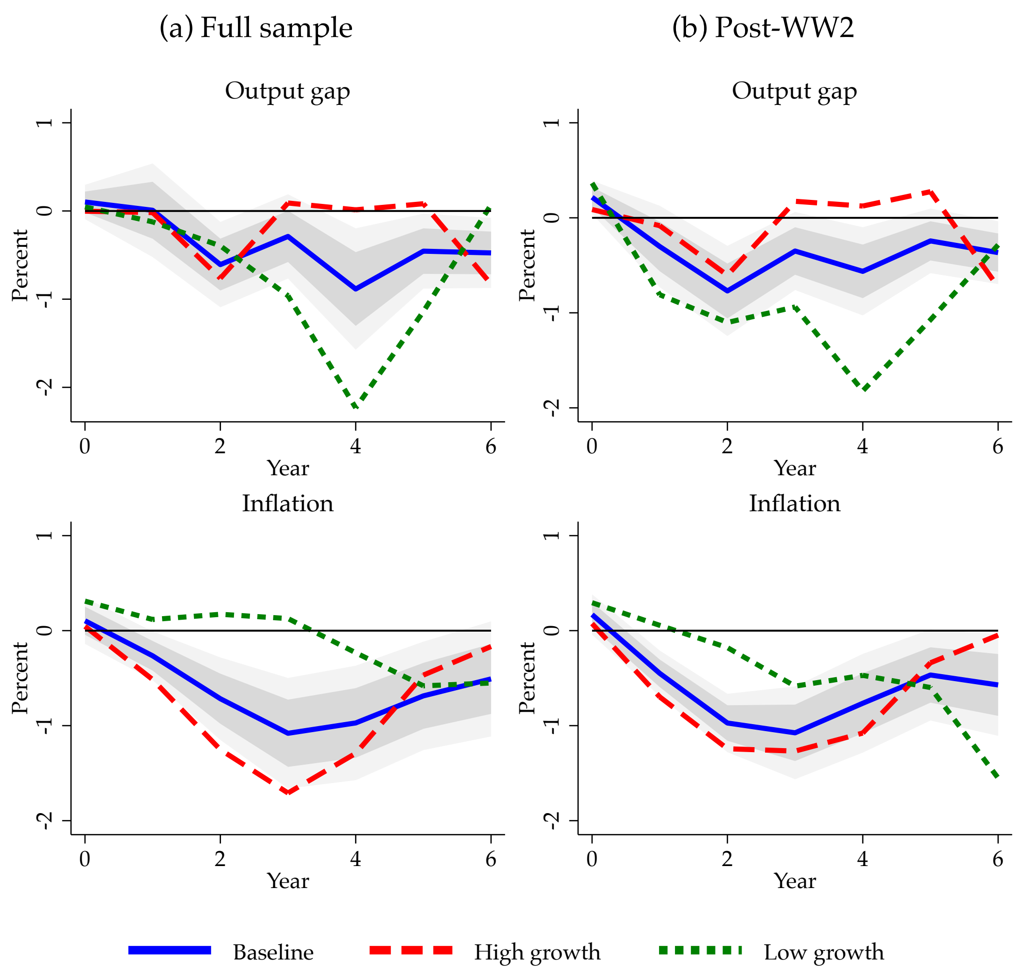
Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. Baseline non-state dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high (low) growth regime are displayed with a red long-dashed (green dashed) line. The output gap is measured using the HP filter with a smoothing parameter of 100. Sample sizes are not harmonized across horizons.

Figure 2.6. Regime classification results: above country-specific mean



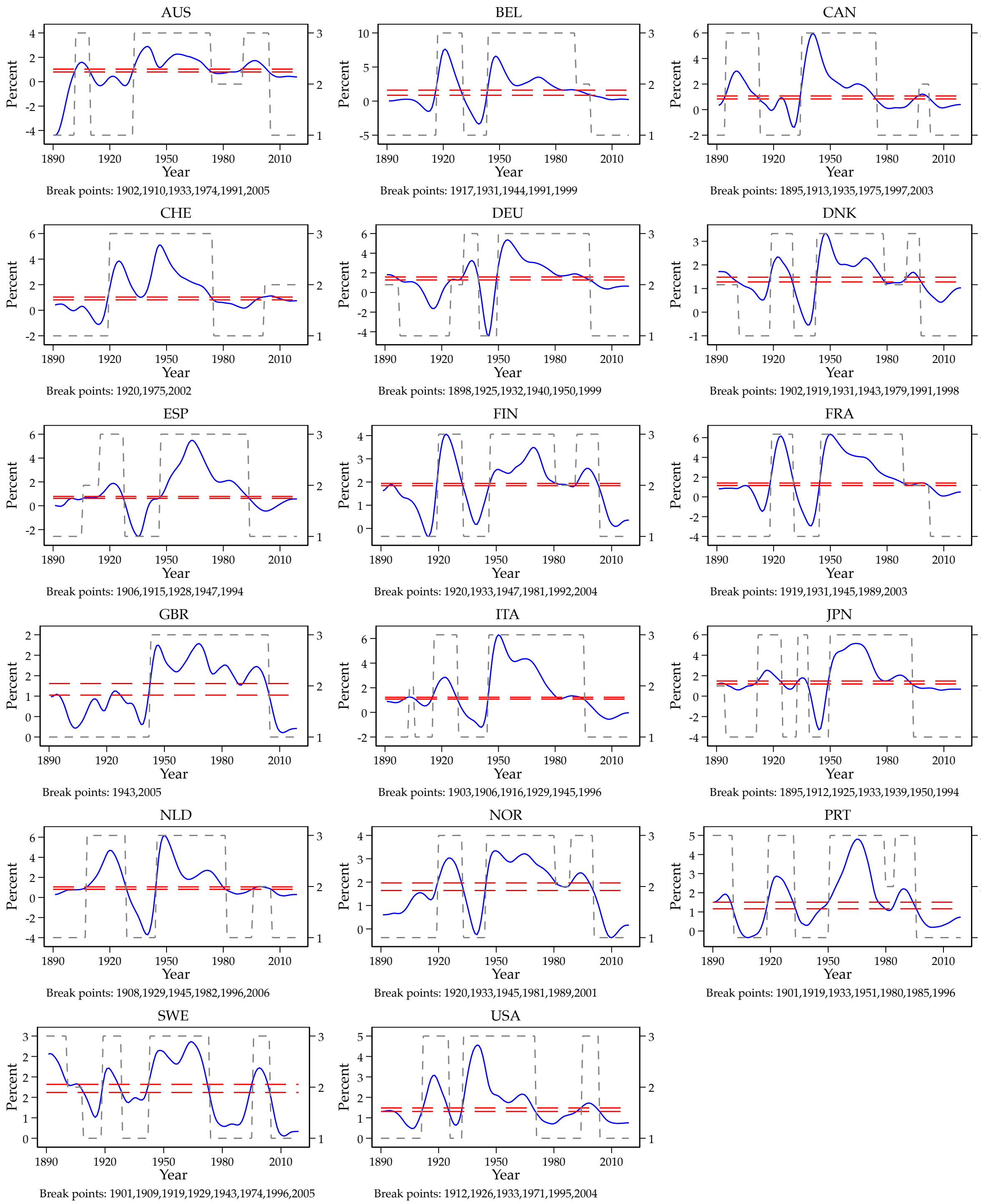
Notes: Solid blue lines: trend TFP growth over time (left axis); dashed grey lines: regime dummy (right axis), equals 1 in the high-growth regime, 0 in the low-growth regime; red dashed lines: country-specific averages. Sample period: 1890-2019.

Figure 2.7. Asymmetric responses of inflation and output gap: above country-specific mean



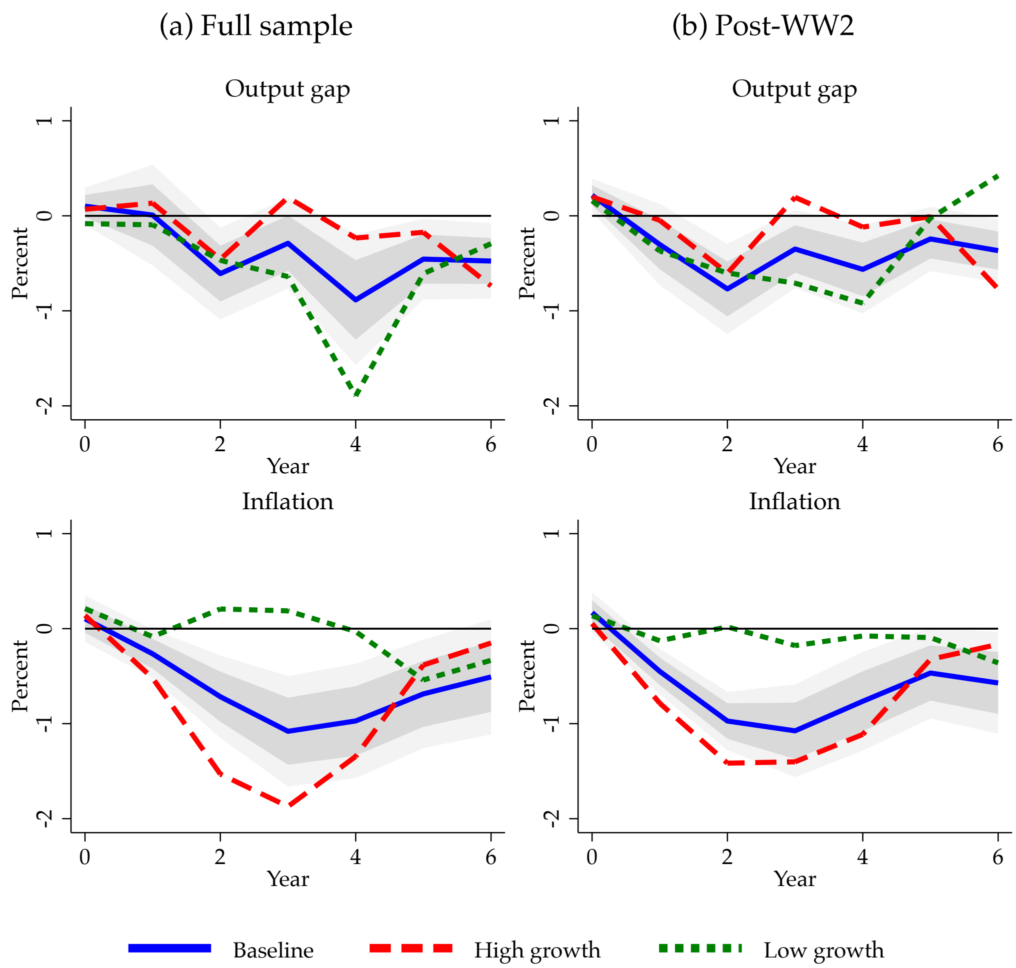
Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. Baseline non-state dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high (low) growth regime are displayed with a red long-dashed (green dashed) line. Sample sizes are not harmonized across horizons. The output gap is measured using the HP filter with a smoothing parameter of 100.

Figure 2.8. Regime classification results: 45 and 55 percentiles



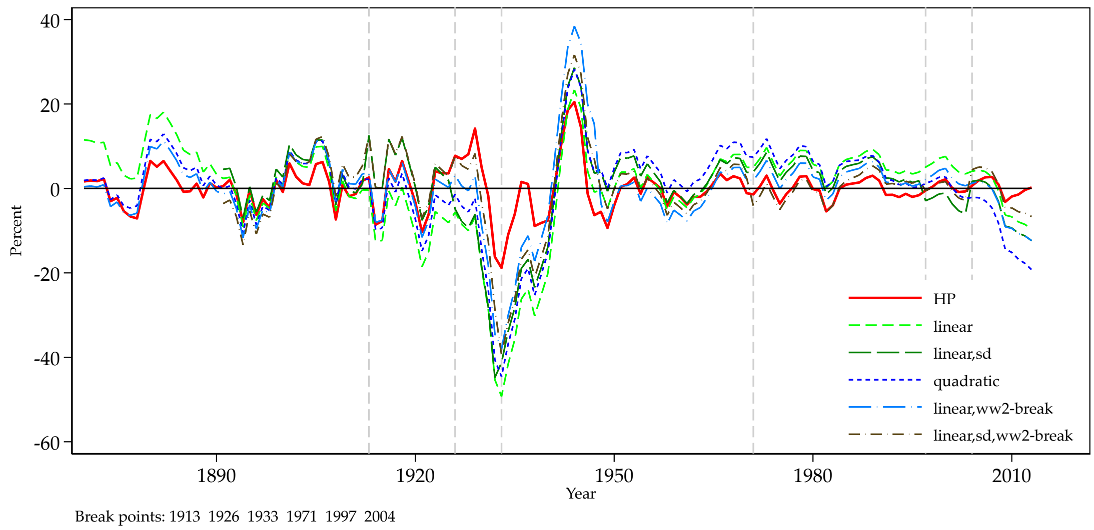
Notes: Solid blue lines: trend TFP growth over time (left axis); dashed grey lines: regime dummy (right axis), 3 for the high-growth regime, 2 for the medium high-growth regime, 1 for the low-growth regime; red dashed lines: country-specific 45 and 55 percentiles. Sample period: 1890-2019.

Figure 2.9. Asymmetric responses of inflation and output gap: 45 and 55 percentiles



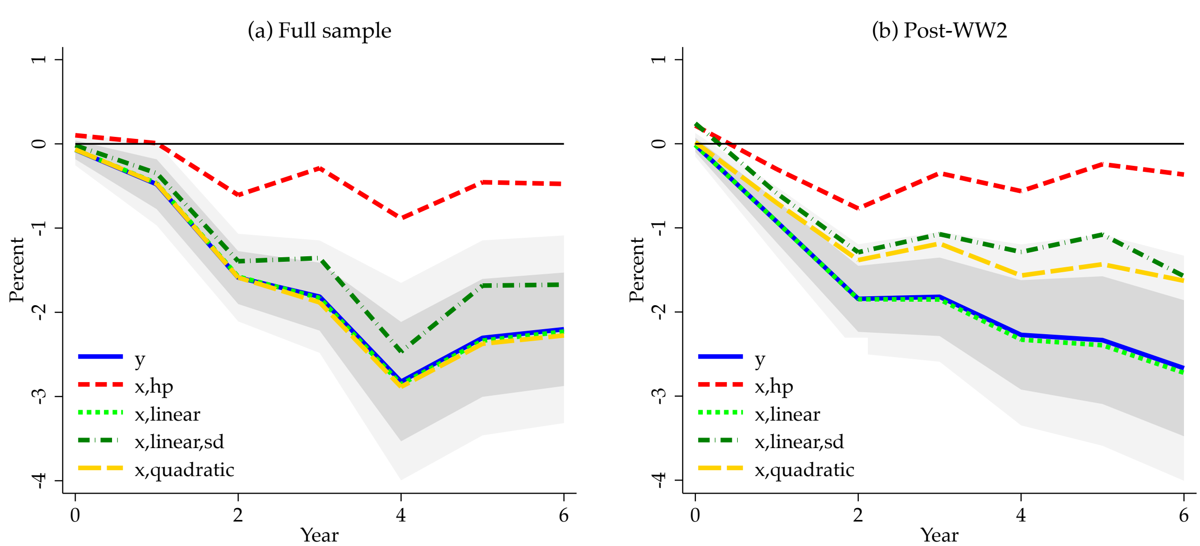
Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. Baseline non-state dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high (low) growth regime are displayed with a red long-dashed (green dashed) line. Sample sizes are not harmonized across horizons. The output gap is measured using the HP filter with a smoothing parameter of 100. The low regime here also includes the medium-high growth regime. See text.

Figure 2.10. Different measures of the output gap for the U.S.



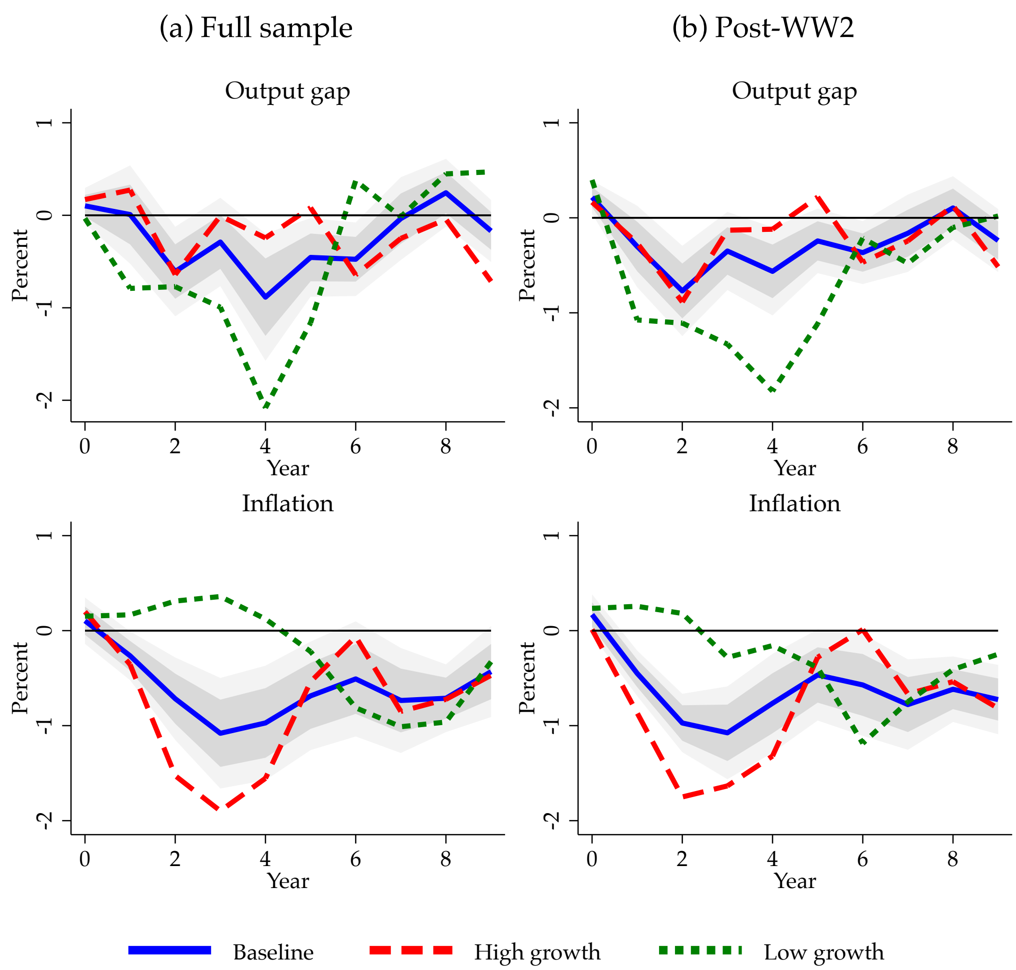
Notes: "HP": output gap using HP filter with a smoothing parameter of 100; "linear": linearly detrended output gap; "linear, sd": growth regime-dependent linearly detrended output gap; "ww2-break": the coefficients in the linear trend regression can take on different values before and after the WW2; "quadratic": quadratically detrended output gap. The breakpoints are the tuning points of TFP growth estimated using the regime-switching approach. See text.

Figure 2.11. Impulse responses of output and different measures of output gap



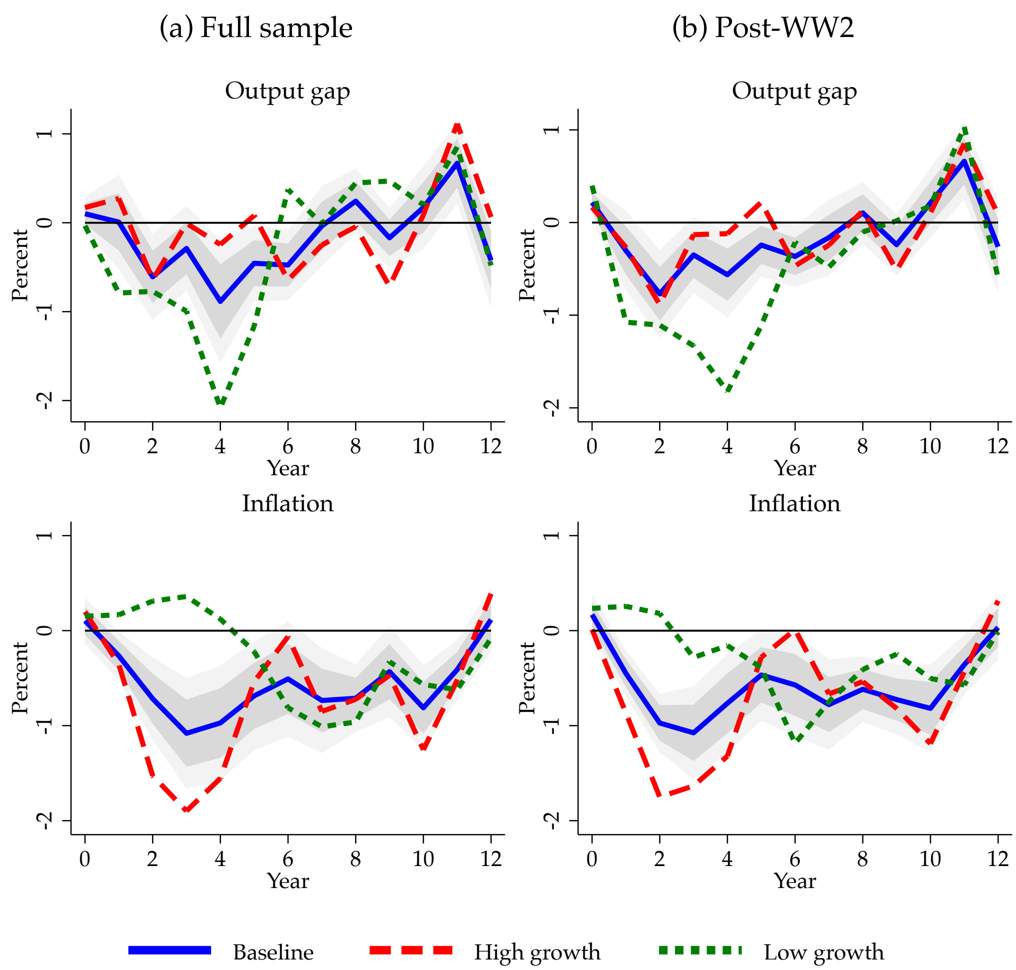
Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. LP-IV estimates of the IRFs of output are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Dashed lines are IRFs of different measures of the output gap, relative to the IRFs of output, the controls also include up to 2 lags of the first difference of the output gap. Sample sizes are not harmonized across horizons.

Figure 2.12. Asymmetric responses of inflation and output gap: $H = 9$



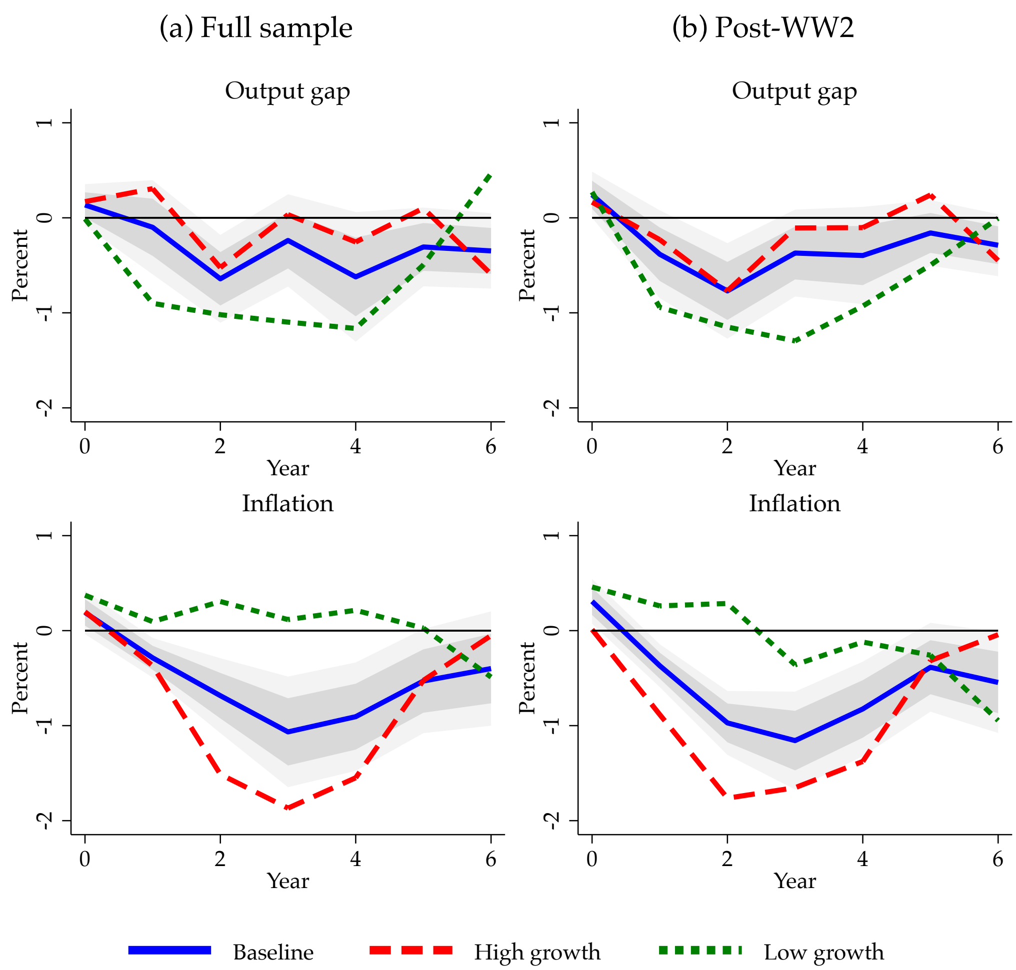
Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. Baseline non-state dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high (low) growth regime are displayed with a red long-dashed (green dashed) line. Sample sizes are not harmonized across horizons. The output gap is measured using the HP filter with a smoothing parameter of 100.

Figure 2.13. Asymmetric responses of inflation and output gap: $H = 12$



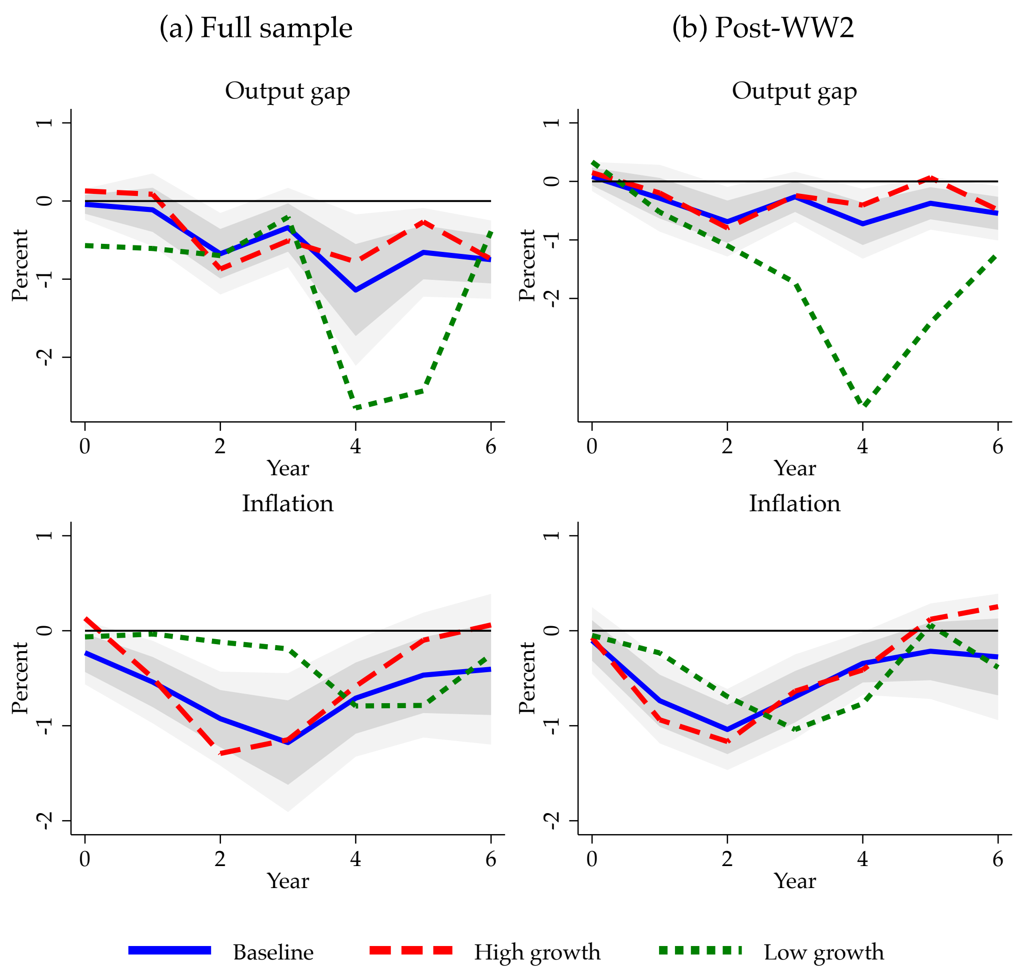
Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. Baseline non-state dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high (low) growth regime are displayed with a red long-dashed (green dashed) line. Sample sizes are not harmonized across horizons. The output gap is measured using the HP filter with a smoothing parameter of 100.

Figure 2.14. Asymmetric responses of inflation and output gap: include GFC



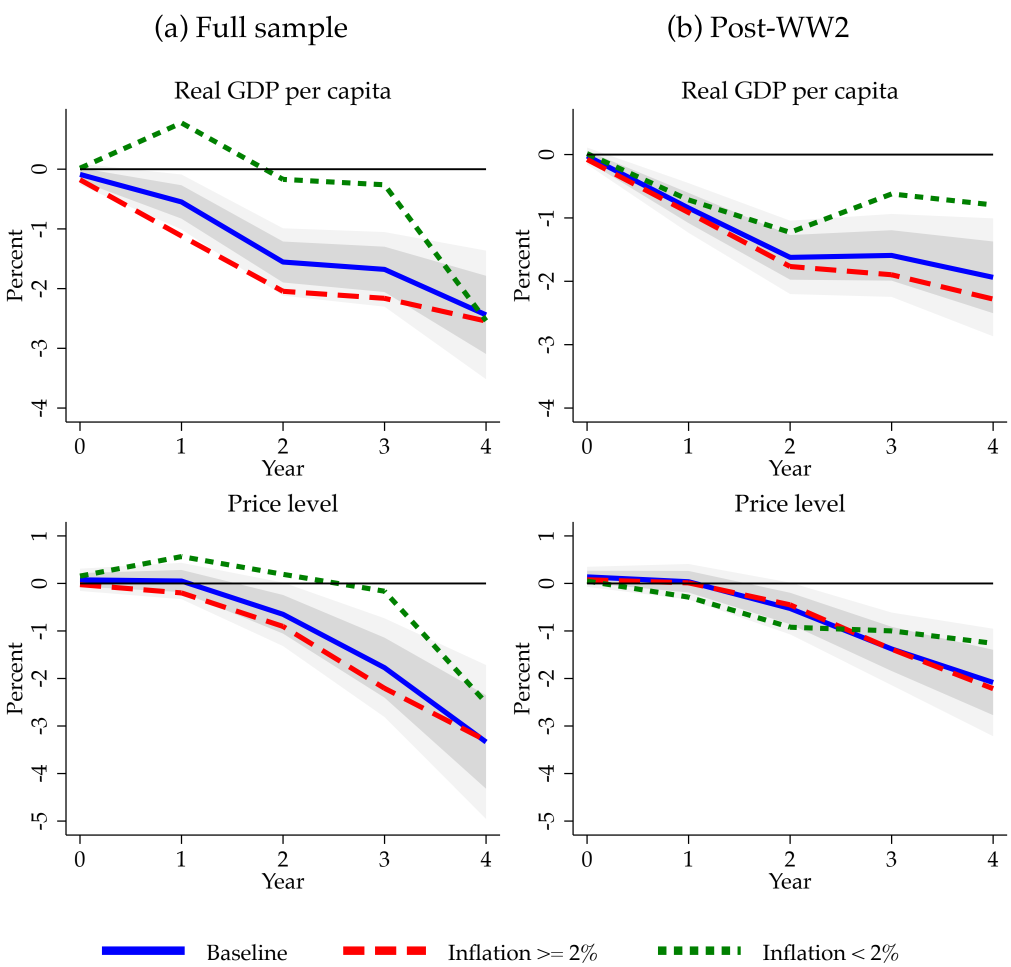
Notes: Full sample: 1890-1908, 1921-1933, and 1948-2012. Post-WW2 sample: 1948-2012. Baseline non-state dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high (low) growth regime are displayed with a red long-dashed (green dashed) line. Regimes are classified by the "regime-switching" approach. Sample sizes are not harmonized across horizons. The output gap is measured using the HP filter with a smoothing parameter of 100.

Figure 2.15. Asymmetric responses of inflation and output gap: European countries



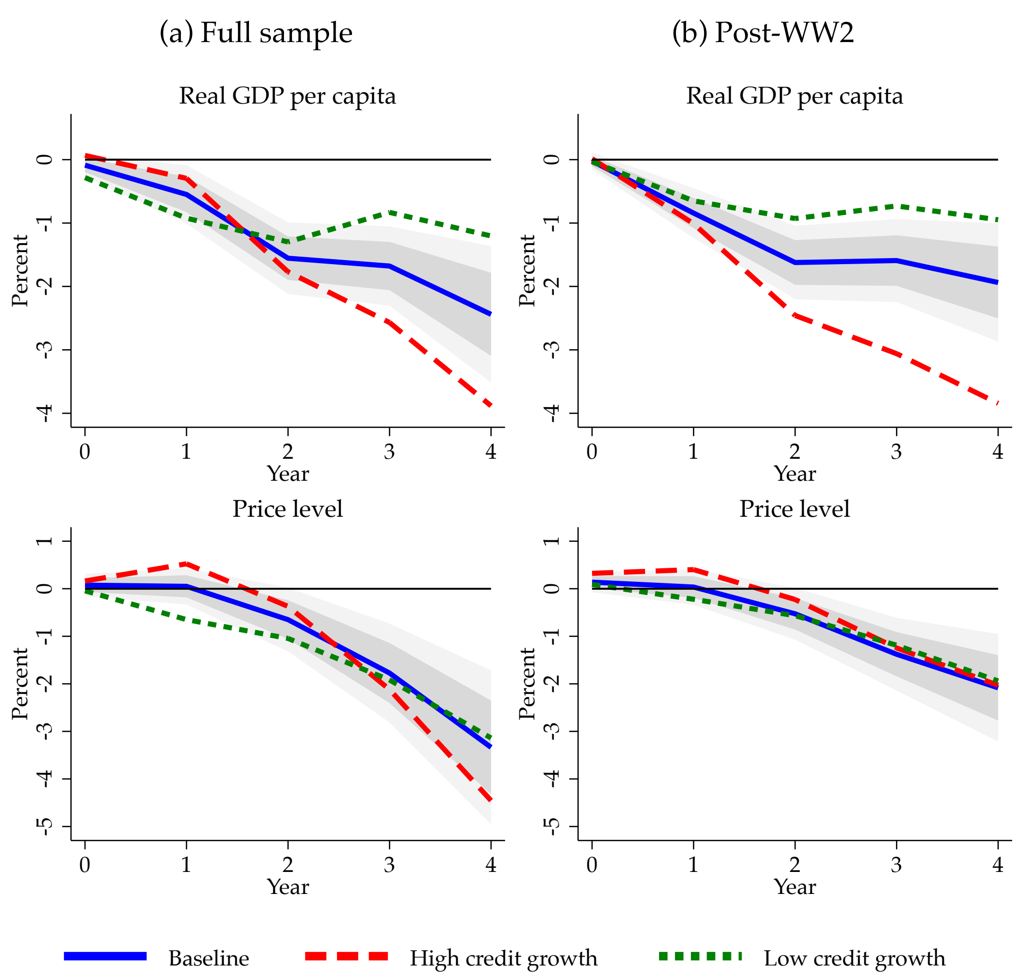
Notes: Full sample: 1890-1908, 1921-1933, and 1948-2006. Post-WW2 sample: 1948-2006. Only European countries are included. Baseline non-state-dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high (low) growth regime are displayed with a red long-dashed (green dashed) line. Sample sizes are not harmonized across horizons. The output gap is measured using the HP filter with a smoothing parameter of 100. Regimes are classified by the "regime-switching" approach.

Figure 2.16. Impulse responses of output and price level by inflation regimes



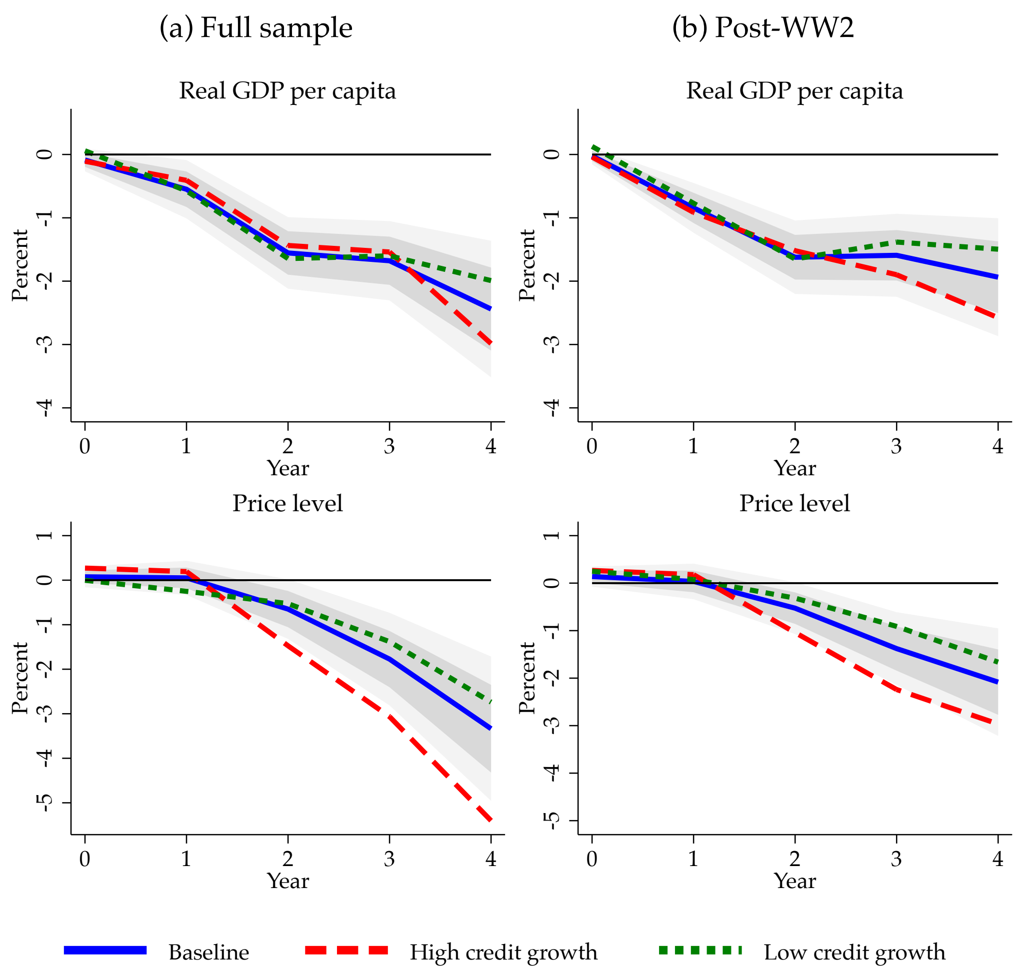
Notes: Full sample: 1890-2006 excluding world wars (1914-1919 and 1939-1947). Post-WW2 sample: 1948-2006. Baseline non-state-dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high inflation regime are displayed with a red long-dashed line whereas estimates in the low inflation regime are displayed with a green dashed line. Sample sizes are not harmonized when estimating IRFs.

Figure 2.17. Impulse responses of output and price level by mortgage credit regimes



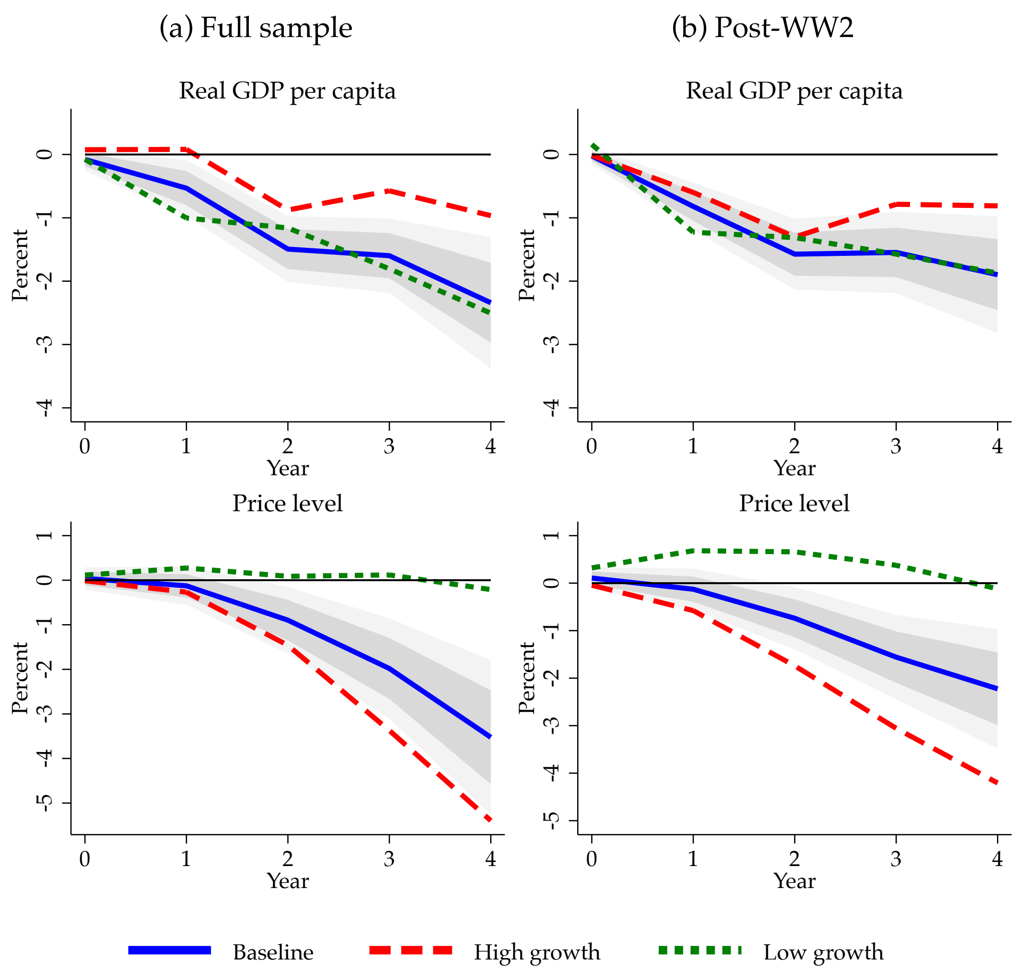
Notes: Full sample: 1890-2006 excluding world wars (1914-1919 and 1939-1947). Post-WW2 sample: 1948-2006. Baseline non-state-dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high credit growth regime are displayed with a red long-dashed line whereas estimates in the low credit growth regime are displayed with a green dashed line. Sample sizes are not harmonized across horizons.

Figure 2.18. Impulse responses of output and price level by non-mortgage credit regimes



Notes: Full sample: 1890-2006 excluding world wars (1914-1919 and 1939-1947). Post-WW2 sample: 1948-2006. Baseline non-state-dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high credit growth regime are displayed with a red long-dashed line whereas estimates in the low credit growth regime are displayed with a green dashed line. Sample sizes are not harmonized across horizons.

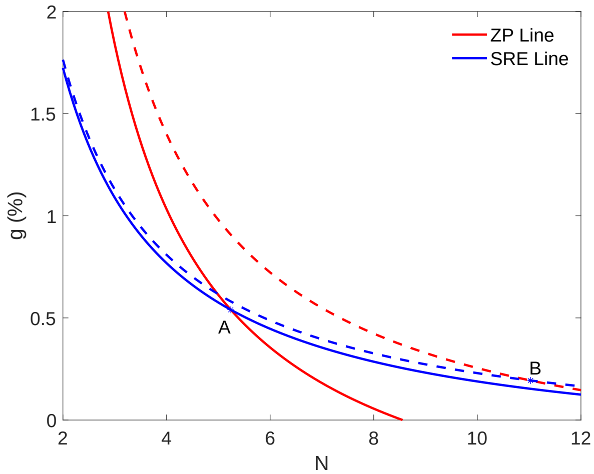
Figure 2.19. Impulse responses of output and price level by TFP growth regimes



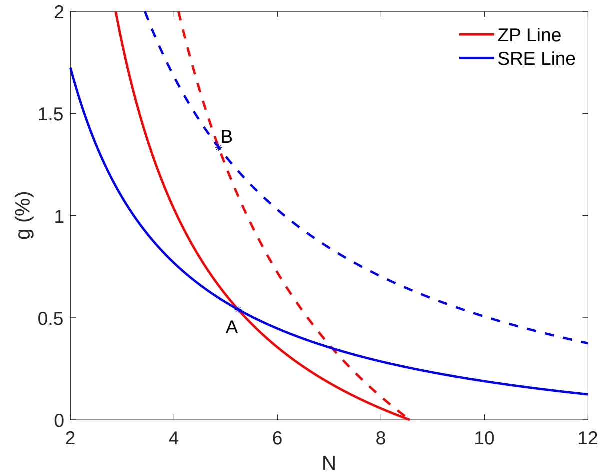
Notes: Full sample: 1890-2006 excluding world wars (1914-1919 and 1939-1947). Post-WW2 sample: 1948-2006. Baseline non-state-dependent LP-IV estimates are displayed with a solid blue line and 68% and 90% confidence bands in grey areas. Estimates stratified by the high TFP growth regime are displayed with a red long-dashed line whereas estimates in the low TFP growth regime are displayed with a green dashed line. Sample sizes are not harmonized across horizons.

Figure 2.20. Comparative statics analysis

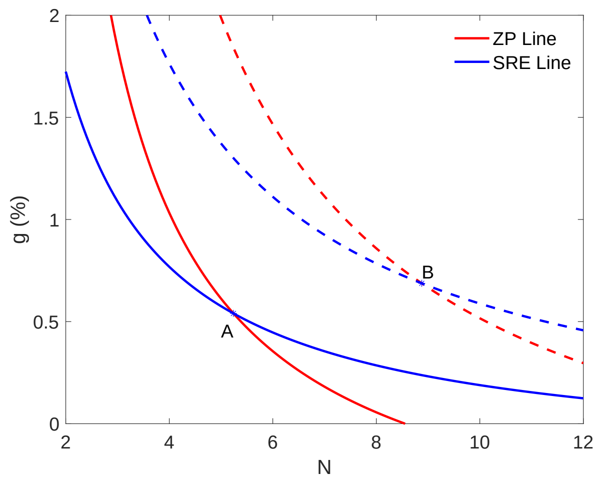
(a) Overhead costs $L_f \downarrow$



(b) Research productivity $\xi \uparrow$



(c) Both $\xi \uparrow$ and $L_f \downarrow$



Notes: Number of firms N on the x-axis, growth rate of knowledge g on the y-axis. In Panel (a), overhead costs L_f falls from 2.05 to 1 unit. In Panel (b), research productivity ξ rises from 0.0035 to 0.0071. The ZP line is in red and the SRE line is in blue, and solid lines describe the initial equilibrium and the dashed lines describe the new equilibrium after the parameter changes.

Chapter 3

Pitfalls and Caveats in the Estimation of the New Keynesian Phillips Curve

3.1 Introduction

The Phillips curve is a formal description of the trade-off relation between inflation and economic slack: when the demand is higher than normal, workers will be provoked to ask for higher wages and firms to raise their prices. It has been widely accepted as the guidance of monetary policy making by central banks since its introduction by Phillips (1958). However, the debate surrounding the stability and validity of the Phillips curve to guide inflation forecast remains one of the most active topics in macroeconomics over the last several decades. Meanwhile, the formulation of the Phillips curve has also undergone three major stages of development.[1] This paper focuses on the most widely used formulation, the New Keynesian Phillips curve (NKPC), as it has gained popularity from its appealing theoretical micro-foundations and what appeared to be early empirical success (Mavroeidis et al., 2014). NKPC postulates that inflation is determined by three factors: inflation expectation, the output or unemployment gap – the difference between the actual level of real economic activities and its natural level under flexible price –, and the supply shocks. Distinct from other formulations, NKPC explicitly emphasizes the central role played by *forward-looking* inflation expectation, which is certainly endogenous and unobservable to econometricians.

[1] See "Related Literature" for a brief review.

With little understanding of how agents form their inflation expectations, early empirical research on the NKPC has identified a battery of shortcomings using aggregate macroeconomic time series. Mavroeidis et al. (2014) give a comprehensive survey of the main identification strategies and empirical evidence on the estimation of the NKPC using limited-information econometric methods.[2] They point out that there are substantial specification uncertainty and sampling uncertainty due to weak identification, that is, there isn't enough variation in the aggregate data to separately identify the coefficients on the expected inflation and unemployment or output gap.

Continuous efforts have been made in primarily three aspects to resolve the identification problems since then. The first two approaches aim to distinguish the contributions of expected inflation and real economic activities in determining inflation and try to control for the endogenous inflation expectation. In contrast, the third approach directly speaks to the endogeneity issues pervasive in the estimation of structural forward-looking macroeconomic equations including the Phillips curve, and introduces a new source of aggregate variations to address the identification problems.

Specifically, the first approach focuses on adopting more flexible specifications or survey measures of inflation expectation to allow for potential regime changes in the formation of inflation expectation over time, especially since the Volcker disinflation era. Ball and Mazumder (2019), and Jorgensen and Lansing (2022) are two representative examples. The second approach seeks help from cross-sectional data to difference out the long-run inflation expectation and aggregate supply factors, see Hazell et al. (2022) and the references therein. The third approach, pioneered by Barnichon and Mesters (2020), attempts to bring in a new identification scheme based on aggregate structural shocks. In particular, independently identified aggregate demand shocks such as monetary shocks are valid instruments to estimate the NKPC under relatively mild assumptions. In relation to the traditional generalized instrumental variable (GIV) estimators, the third approach also replaces expected future inflation with its realization but uses sequences of monetary shocks as instruments

[2] Including generalized instrumental variables (GIV) estimator, and VAR estimator implemented by maximum likelihood (ML) or minimum distance (MD), all of which can be implemented using generalized method of moments (GMM).

instead of predetermined macroeconomic variables as instruments.[3]

This paper directly speaks to the third approach without taking a stand on how agents form their inflation expectations and how inflation expectations evolve over time. In spite of the rising popularity, state-of-the-art estimators based on monetary shocks are not necessarily able to alleviate all the problems pervasive in the estimation of the Phillips curve. In fact, they might be confronted with similar pitfalls in the empirical estimation due to instrument and data issues. For one thing, the exogeneity condition may fail due to imperfect measures of monetary shocks or the existence of other confounding channels through which monetary shocks may affect inflation. For another, the lack of variations in the data may lead to the failure of rank conditions, and it could be very hard to obtain estimates with the right sign and magnitude.

Furthermore, weak and many instruments problems are still notoriously difficult to deal with. Using different methodologies, recent findings from Gorodnichenko and Lee (2020) and Plagborg-Møller and Wolf (2022) show that monetary shocks can only explain a small fraction of the variations in inflation and the output gap, especially at short horizons.[4] How these two problems distort the point estimates remains an interesting research question in this literature. What's worse, macroeconomic datasets are typically short due to limited data availability. In particular, structural monetary shocks are not readily observable, and their proxy measures are only available for short samples. The narrative monetary shocks are only available at a quarterly frequency from 1969 for the U.S. (Romer and Romer, 2004), and 1975 for the U.K. (Cloyne and Hürtgen, 2016). The high frequency identified monetary shocks (Kuttner, 2001) are only available after the 1990s. Despite the long coverage of over a century, the trilemma monetary shocks by Jordà, Schularick and Taylor (2020*a*) are available at an annual frequency. It thus becomes even more challenging to identify the Phillips curve using monetary shocks as instruments in small samples.

[3]The GIV estimator, first introduced by Hansen and Singleton (1982) in the estimation of Euler equation models, has been popularized in the estimation of NKPC by the seminal contributions of Roberts (1995) and Galí and Gertler (1999).

[4]For instance, Gorodnichenko and Lee (2020) find that the monetary shocks can explain at least between 10% and 20% of the variation in output, and between 10% and 40% of the variation in inflation at the long horizon, but little at the shorter horizon. Plagborg-Møller and Wolf (2022) find that monetary shocks can explain at most 31% of the variation in output and 8% of the variation in inflation.

Using carefully designed simulation studies, I evaluate the strengths and weaknesses of different IV estimators in the face of the following challenges: violation of the exogeneity condition, failure of the rank condition, weak instruments problem, many instruments problem, and small sample bias. Besides the traditional GIV estimator that uses predetermined macroeconomic variables as instruments, this paper pays special attention to the state-of-the-art estimators using proxies for monetary shocks as instruments. The latter can be further divided into three categories: the 2SLS estimator that uses lag sequences of monetary shocks as instruments, the Almon-restricted instrumental variable (ARIV) estimator that reduces the number of instruments to three by applying Almon (1965) parameterization, and the two-step estimators in the spirit of "regression in impulse response space". In theory, the two-step approach is most general in that the 2SLS estimator is one special case where econometricians estimate the impulse responses using the distributed lag model in the first step. Based upon the 2SLS estimator, the ARIV estimator – proposed by Barnichon and Mesters (2020) – was meant to alleviate the many instruments problem.

The main takeaway of this paper is as follows: first and foremost, the flexibility of the two-step approach to allow for extra controls in estimating the impulse responses gains its additional advantage over the other IV estimators under at least two scenarios: (i) proxies for monetary shocks are correlated with lags of inflation and/or output gap; (ii) monetary shocks affect domestic inflation through channels other than domestic output gap. Under both scenarios, the measured monetary shocks are valid instruments only after conditioning on extra information. Second, imposing range inequality constraints along with the long-run constraint brings the point estimates close to their true values when the rank condition fails. Third, the many instruments and weak instruments problems are now intertwined together in that including more lags of monetary shocks or longer projection horizon provides more variations for identification, but endogenous variables become harder to predict at longer horizons. The natural tension between relevance and many weak instruments remains one of the main concerns. Lastly, among all IV estimators based on monetary shocks, the ARIV estimator exhibits smaller biases when the instruments are constructed using fewer lags of monetary shocks. However, it becomes unstable and skewed when the instruments are constructed using a long sequence of monetary shocks.

Related Literature This paper closely relates to the vast literature on the empirics of the Phillips curve. The empirical specification of the Phillips curve greatly depends on its formulation backed by the underlying theory, both of which have been through a tremendous development over time. The original Phillips curve, named after New Zealand economist A. William Phillips was introduced in 1958 to describe the empirical inverse relationship between wage inflation and the unemployment rate in the U.K. from 1861 to 1913. It became well-known after Samuelson and Solow (1960) provided further support using the U.S. data from 1934 to 1958. This seemly simple and stable relationship led many policymakers in the 1960s to believe that they could fine-tune the economy by choosing a pair of desirable levels of unemployment and inflation. As long as the unemployment costs of lowering inflation, later on, are bearable, it is politically attractive to sustain mild inflation to reduce unemployment in the short run.

However, this view was not uniformly accepted even at that time. Friedman (1969) and Phelps (1967) both argued for the "natural rate hypothesis" suggesting a vertical long-run Phillips curve. When the monetary authority tries to exploit the trade-off relation by creating higher inflation, agents will adjust their expectations slowly and the "expectation-augmented" Phillips curve will shift until the expectations are fully fulfilled. It was not until the stagflation period in the 1970s did mainstream macroeconomists start to rethink the macroeconomic models behind the Phillips curve. After Lucas (1972), New Keynesian economists in particular started to incorporate sticky prices and wages into rational expectation models. The linearization of the optimality condition for firms' profit maximization problem delivers the NKPC, which contains three drivers of inflation: forward-looking inflation expectation, demand-pull factors, and cost-push factors. The forward-looking feature of inflation greatly limits the scope for policy exploitation.

Despite decades of research, the consistent estimation of the Phillips curve remains one of the most challenging empirical tasks in macroeconomic studies due to pervasive endogeneity issues. The early seminal papers such as Roberts (1995) and Galí and Gertler (1999) addressed the endogeneity concerns by using lagged macroeconomic observables as instruments. The validity of this traditional approach requires that the cost-push factors are not autocorrelated, which is not guaranteed in practice. Besides, lagged macro instruments are found to be weak instruments, which can lead to substantial specification uncertainty and sampling uncertainty, see e.g., Mavroeidis et al. (2014) for a comprehensive survey study.

Recent papers in this literature attempt to resolve the identification problems through at least three approaches. The first approach focuses on adopting more flexible specifications or survey measures of expected inflation to control for the endogenous expected inflation using aggregate data. Informed by the evolution of long-term SPF inflation expectations, Ball and Mazumder (2019) explicitly model the expected inflation as the weighted average of a backward-looking component and an anchored component and allow for a break in the weighting parameter. They find that expected inflation was backward-looking until the late 1990s, but then became strongly anchored at the Federal Reserve's target. The anchored inflation expectation can partially explain the "missing deflation" puzzle during the Great Recession. Jorgensen and Lansing (2022) further develop a theoretical model of expectation anchoring to allow for endogenous evolution in the degree of anchoring of agents' inflation forecast over time. Using GMM with lagged variables as instruments, their empirical estimation recovers a stable structural slope coefficient for the U.S. over the period from 1960 to 2019, arguing against the flattening of the Phillips curve. Coibion and Gorodnichenko (2015) find no "missing disinflation" after taking into account the rise in household and presumably firm's expected inflation from 2009 to 2011 due to the surge in oil prices.

The second approach seeks help from cross-sectional data to control inflation expectations, see Hazell et al. (2022) and the references therein. The idea is that the micro-level panel data on price dynamics can difference out the common variations in long-run inflation expectation and aggregate supply shocks, and shift-share instruments can be used to deal with regional supply shocks. Besides the limited availability of regional data on price dynamics, this approach also relies on theoretical assumptions to simplify the inference of the aggregate Phillips curve from the regional Phillips curve. Nevertheless, it sheds light on how to circumvent the endogeneity issues in the estimation of the NKPC using micro-level data and techniques.

The last approach, pioneered by Barnichon and Mesters (2020), attempts to bring in a new identification scheme based on aggregate structural shocks. Specifically, independently identified aggregate demand shocks, e.g., monetary shocks, are valid instruments to estimate the NKPC under relatively mild assumptions. By projecting inflation and the forcing variable (unemployment or output gap) on the lag sequence of monetary shocks, we can project out three sources of endogeneity issues: (i) cost-push shocks, (ii) forecast error in inflation, (iii) measurement error in forcing variable. This paper directly speaks to the last approach and tries to evaluate the performance of different IV estimators based on monetary shocks in the face of various challenges.

One recent paper closely related to this one is by Lewis and Mertens (2022). Their System Projections on Instrumental Variable (SP-IV) is equivalent to estimating dynamic structural macroeconomic equations using impulse responses obtained from local projections (LPs) or vector autoregressions (VARs), which is in the same spirit as the two-step approach in this paper. By transforming the problem into a Generalized Method of Moments (GMM) problem, they provide inference procedures under strong and weak identification for the SP-IV estimator. Furthermore, they present both theoretical and simulation evidence that by allowing the inclusion of lagged variables as controls, SP-IV outperforms conventional IV estimators such as the 2SLS estimator.

This paper differs from theirs in at least two aspects. First and foremost, I study the performance of a different set of estimators in the face of a broader range of empirical challenges. Besides the conventional GIV estimator using predetermined variables as instruments, we both look at the 2SLS without controls and the two-step approach with controls implemented with LPs. However, I add the ARIV estimator into the discussion but exclude the 2SLS with controls and the two-step approach implemented with VARs. On top of the exogeneity condition, weak instruments problem, and small sample bias, I also point out the threat of the failure of rank condition and how the many and weak instruments problems are intertwined when using monetary shocks as instruments.

Second, there are various reasons for including extra controls in estimating the impulse response functions with LPs. Besides their working example in which monetary shocks are often weakly correlated with lagged endogenous variables, my justification for the superiority of the two-step approach with controls originates more generally from the identification concern that monetary shocks may affect inflation through channels other than the domestic output gap. One such example is related to the international spillover effects of monetary policy. As shown by Razin and Binyamini (2007), in an open economy, domestic inflation is affected by not only the domestic output gap but also the foreign output gap. If domestic monetary policy also affects foreign demand, then using domestic monetary shocks as instruments will lead to bias in the slope estimate of the closed-economy NKPC. This is the case for large open economies like the U.S. and there is a big strand of literature studying the spillover effects of the U.S. monetary policy on other economies (e.g., Dedola et al., 2017; Gai and Tong, 2022; Georgiadis, 2016, etc). Another well-known case is related to the trilemma mechanism: base country's exogenous interest rate movements will spill into local interest rates for open pegs, see Jordà, Schularick and Taylor (2020*a*) and the references therein. I show in a simulation study that controlling for the *contemporaneous as well as lagged values* of the foreign output gap in the estimation of impulse responses will eliminate the omitted variable bias, while controlling only lagged values of the foreign output gap is not enough. This finding is quite unique in that empirical impulse response estimation by LPs usually does not include contemporaneous variables in the control set.

The remainder of the paper is organized as follows. In Section 3.2, I discuss the endogeneity issues pervasive in the estimation of the NKPC, the identification problem with the traditional solution based on predetermined instruments, and the identification assumptions necessary for the monetary shocks to be valid instruments. Such an identification scheme yields a list of IV estimators, namely, 2SLS, ARIV, and two-step estimators. In Section 3.3, I evaluate the performance of different IV estimators in five different aspects using carefully designed simulation studies. Section 3.4 concludes.

3.2 Phillips Curve Estimation Using Monetary Shocks as Instruments

In this section, I first review the endogeneity issues pervasive in the estimation of the NKPC and the reason why the traditional approach using lagged macroeconomic observables as instruments may fail. Then I discuss the identification assumptions necessary for the monetary shocks to be valid instruments to overcome the endogeneity issues. Such an identification scheme yields a list of IV estimators, namely, the 2SLS estimator, the Almon-restricted IV estimator, and the two-step estimators. The section ends with a brief comparison between the 2SLS estimator and the two-step estimators.

3.2.1 Endogeneity Issues in Estimating the Phillips Curve

Consider the following general forward-looking New Keynesian Phillips curve,

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \lambda x_t + \epsilon_t^s \tag{3.1}$$

where π_t is inflation, $x_t \equiv y_t - y_t^n$ is the output gap which is the difference between actual output y_t and the natural level of output or potential output y_t^n, and ϵ_t^s denotes the cost-push factors. The parameters of interest γ_b, γ_f, and λ are functions of deep structural parameters of an underlying model as in Galí and Gertler (1999).

Despite its sound theoretical foundation, the empirical estimation of the Phillips curve is notoriously challenging due to pervasive endogeneity issues (Barnichon and Mesters, 2020). To see this, we can rewrite (3.1) as follows

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f \pi_{t+1} + \lambda \hat{x}_t + u_t \tag{3.2}$$

where

$$u_t = \epsilon_t^s - \gamma_f(\pi_{t+1} - E_t \pi_{t+1}) - \lambda(\hat{x}_t - x_t) \tag{3.3}$$

and $\hat{x}_t$ is a proxy measure of x_t in the data. In practice, the true output gap x_t is unobserved as potential output y_t^n is unobserved, so is the expected future inflation $E_t \pi_{t+1}$. The standard practice is to replace it with the realized inflation next period, and the forecast error naturally enters the error term.

We can see that there are three sources of endogeneity problems:

(i) *simultaneous equation bias* due to the presence of ϵ_t^s. In particular, cost-push shocks ϵ_t^s can simultaneously affect inflation and output gap through the systematic response of monetary policy to inflation so that $E(\hat{x}_t u_t) \neq 0$.[5] The existence of an underlying monetary policy rule that aims at price stability necessarily leads to such a connection between $\hat{x}_t$ and ϵ_t^s.

(ii) *measurement error in the output gap* $\hat{x}_t - x_t$ due to the unobserved potential output, we have $E[\hat{x}_t u_t] \neq 0$ if $\lambda \neq 0$.

(iii) *forecast error in future inflation* $\pi_{t+1} - E_t \pi_{t+1}$ due to the unobserved expected future inflation $E_t \pi_{t+1}$, we have $E[\pi_{t+1} u_t] \neq 0$ if $\gamma_f \neq 0$.

To handle these endogeneity issues, the traditional approach in the literature uses predetermined macroeconomic variables as instruments, such as $\hat{x}_{t-1}, \hat{x}_{t-2} \cdots$ and $\pi_{t-2}, \pi_{t-3}, \cdots$, or lags of other variables with readily available data. For these variables to be valid instruments, the exogeneity condition requires that none of the components in the error term u_t are autocorrelated. Increasing the lag length helps mitigate this concern, but at the same time worsens the weak instruments problem. As demonstrated by Mavroeidis et al. (2014), weak identification could result in considerable sampling uncertainty and specification uncertainty.

3.2.2 Identification Using Monetary Shocks as Instruments

The state-of-the-art method, proposed by Barnichon and Mesters (2020), suggests that the sequences of monetary policy shocks are valid instruments to identify the coefficients in the Phillips curve. To illustrate, let ϵ_t denote the mean zero monetary shock for time period t. The sequence of monetary shocks $\boldsymbol{\epsilon}_{t:t-H} \equiv (\epsilon_t, \cdots, \epsilon_{t-H})'$ satisfies both the exogeneity and relevance conditions under mild assumptions.

The exogeneity condition $E[\boldsymbol{\epsilon}_{t:t-H} u_t] = \mathbf{0}$ is satisfied if monetary shocks are orthogonal to (i) cost-push shocks, (ii) measurement error in the output gap, and (iii) forecast error in future inflation. Since monetary shocks are innovations to the systematic conduct of monetary

[5]For example, in response to oil price shock in the 1970s, U.S. monetary policy responded by aggressively raising interest rates to stabilize inflation, which led to the big economic downturn in early 1980s.

policy, they should be orthogonal to cost-push shocks. Condition (ii) holds if monetary shocks do not correlate with potential output, or if money is neutral under flexible prices. Condition (iii) holds under rational expectation or if we use survey measures of inflation expectation, then the survey measurement error should be orthogonal to monetary shocks.[6]

The relevance condition that $E[\boldsymbol{\epsilon}_{t:t-H}(\pi_{t-1}, \pi_{t+1}, \hat{x}_t)]$ has full column rank first requires that monetary shocks affect inflation and the output gap to some extent. That is, there must exist an underlying IS curve that links the output gap to the nominal interest rate and thus to monetary shocks. Second, it also requires that the impulse responses of lagged inflation, future inflation, and the output gap are not linearly dependent. Barnichon and Mesters (2020) also provide some intuition by showing that this identification strategy amounts to a regression in "impulse response space".

By post-multiplying equation (3.2) by ϵ_{t-h}, take the expectation, and apply the exogeneity condition, we get

$$\mathscr{R}_h^{\pi} = \gamma_b \mathscr{R}_{h-1}^{\pi} + \gamma_f \mathscr{R}_{h+1}^{\pi} + \lambda \mathscr{R}_h^{\hat{x}} \quad \forall h = 0, 1, \cdots, H \tag{3.4}$$

where $\mathscr{R}_h^{y}$ is the impulse response function (IRF) of outcome variable y to monetary shocks at horizon h. The relevance condition requires that $[\mathscr{R}_{h-1}^{\pi}, \mathscr{R}_{h+1}^{\pi}, \mathscr{R}_h^{\hat{x}}]_{h=0}^{H}$ to be linearly independent. That is, the dynamics of the IRFs have to be rich enough. This turns out to be not as trivial as it appears! In Section 3.3.4, I analytically show that the rank condition always fails when estimating the hybrid Phillips curve ($\gamma_b > 0$) if the output gap follows an AR(1) process.[7] It remains a challenge in the empirical estimation of the Phillips curve when there are not sufficient variations in the IRFs of inflation and the output gap.

3.2.3 Estimators Using Monetary Shocks as Instruments

To simplify notation, we can write equation (3.2) as

$$\pi_t = \boldsymbol{w}_t' \boldsymbol{\beta} + u_t$$

[6] This is the case if the inflation expectation is formed based on the same information set that policymakers know when conducting monetary policy. In the spirit of Romer and Romer (2004), the constructed monetary shocks will be orthogonal to this information set. Recent literature on the information channel of the monetary policy starts to challenge this view, see e.g., Nakamura and Steinsson (2018).

[7] I disprove Proposition 1 in the appendix of Barnichon and Mesters (2020).

where

$$\boldsymbol{w}_t = \begin{pmatrix}\pi_{t-1} & \pi_{t+1} & \hat{x}_t\end{pmatrix}', \quad \boldsymbol{\beta} = \begin{pmatrix}\gamma_b & \gamma_f & \lambda\end{pmatrix}'$$

Although structural monetary shocks $\boldsymbol{\epsilon}_{t:t-H}$ can serve as valid instruments for estimating the Phillips curve, they are not directly observed. In practice, macroeconomists have come up with various different proxies ξ_t for exogenous monetary policy changes. The well-known identification methods include narrative approach (e.g., Cloyne and Hürtgen, 2016; Romer and Romer, 2004) and high-frequency identification (HFI) approach (e.g., Kuttner, 2001; Nakamura and Steinsson, 2018, etc). Moreover, Jordà, Schularick and Taylor (2020*a*) construct monetary shocks based on the trilemma mechanism for the near universe of advanced economies since 1870.

2SLS Estimator. The first approach for estimating the Phillips curve is to take the lag sequence of proxies for monetary shocks $\boldsymbol{\xi}_{t:t-H} = (\xi_t, \cdots, \xi_{t-H})'$ as instruments and obtain the 2SLS estimator:

$$\hat{\boldsymbol{\beta}}_{2SLS} = \left(S'_{\xi w}\Omega_\xi S_{\xi w}\right)^{-1} S'_{\xi w}\Omega_\xi s_{\xi y} \tag{3.5}$$

where $S_{\xi w} = \frac{1}{T}\sum_{t=H+1}^{T}\boldsymbol{\xi}_{t:t-H}\boldsymbol{w}'_t, \Omega_\xi = \left(\frac{1}{T}\sum_{t=H+1}^{T}\boldsymbol{\xi}_{t:t-H}\boldsymbol{\xi}'_{t:t-H}\right)^{-1}, s_{\xi\pi} = \frac{1}{T}\sum_{t=H+1}^{T}\boldsymbol{\xi}_{t:t-H}\pi_t$. This is essentially the special case of the "naive" moment estimator in Barnichon and Mesters (2020) where Ω_ξ can be some positive definite weight matrix.

Almon-Restricted IV Estimator. Given that this "naive" approach suffers from many instruments problem if H is relatively large, e.g., $H = 20$ for quarterly data, Barnichon and Mesters (2020) suggest reducing the number of instruments to three using Almon (1965) parameterization.

The new set of instruments with Almon restriction is

$$\boldsymbol{z}_t = \left(\sum_{h=0}^{H}\xi_{t-h}, \sum_{h=0}^{H}h\xi_{t-h}, \sum_{h=0}^{H}h^2\xi_{t-h}\right) \tag{3.6}$$

The model becomes just-identified and the resulting Almon-restricted IV estimator is

$$\hat{\boldsymbol{\beta}}_{ARIV} = S_{zw}^{-1}s_{z\pi} \tag{3.7}$$

where $S_{zw} = \frac{1}{T}\sum_{t=H+1}^{T} \boldsymbol{z}_t \boldsymbol{w}_t', s_{z\pi} = \frac{1}{T}\sum_{t=H+1}^{T} \boldsymbol{z}_t \pi_t$. The ARIV estimator mitigates the many instruments problem but is not robust to weak instruments problem.

Two-Step Estimator. The intuition behind the identification strategy implies a two-step approach to identify the coefficients in the Phillips curve:

(i) Estimate the impulse response functions (IRFs) of inflation and output gap to the monetary shocks and obtain $\hat{\mathscr{R}}_h^{\pi}$ and $\hat{\mathscr{R}}_h^{\hat{x}}$ for $h = 0, \cdots, H$;

(ii) Run linear regression with the estimated impulse response functions to estimate $(\gamma_b, \gamma_f, \lambda)$:

$$\hat{\mathscr{R}}_h^{\pi} = \gamma_b \hat{\mathscr{R}}_{h-1}^{\pi} + \gamma_f \hat{\mathscr{R}}_{h+1}^{\pi} + \lambda \mathscr{R}_h^{\hat{x}} + e_h$$

where e_h is a linear combination of estimation errors.

Two-Step Estimator Versus 2SLS Estimator. The 2SLS estimator with the lag sequence of monetary shocks as instruments is equivalent to the two-step approach when we estimate the IRFs using the distributed lag (DL) model in the first step. However, in practice, impulse responses are rarely estimated with DL specification, but instead with VARs or LPs. More importantly, it is often the case that the instruments are contaminated in a way such that they are only valid after controlling for extra information. As a result, the ability of the two-step approach to allow for extra controls in the first step will be superior to the simple 2SLS estimator. This is the case at least in the following two scenarios:

The first scenario relates to the construction of monetary policy shocks. Despite careful construction, monetary shock series may still contain some components predictable based on predetermined information set. As a result, researchers typically include lagged macroeconomic variables as controls when estimating IRFs to monetary shocks. Interestingly, this viewpoint is also emphasized in recent independent work by Lewis and Mertens (2022).

Second, and more importantly, monetary shocks may affect inflation through channels other than the domestic output gap. One such example is closely related to the globalization view of the flattening Phillips curve, see e.g., Razin and Binyamini (2007); Borio and Filardo (2007). For an open economy, domestic inflation is affected by both domestic and foreign demand conditions. Intuitively, the increase in the foreign country's demand will raise the price of goods produced in the foreign country. When these goods get imported into the domestic economy, they will push up domestic inflation as a result. As a larger proportion of the rise in domestic demand is satisfied through imports, rather than domestic production, increases in the domestic output gap will have a smaller impact on domestic marginal costs, and hence on inflation.

To illustrate, take the pure forward-looking open economy NKPC as in Razin and Binyamini (2007) for example:

$$\pi_t = \beta E_t \pi_{t+1} + \lambda' [n x_t + (1-n) x_t^*] + e_t$$

where x_t (x_t^*) denotes the domestic (foreign country's) output gap. $1-n$ is the trade openness parameter and e_t contains terms related to the difference between domestic and foreign potential output $y_t^n - y_t^{n^*}$, exchange rate changes, and other supply-side shocks. As $1-n$ increases, or n decreases, the sensitivity of domestic inflation with respect to domestic output gap $\lambda' n$ decreases. In other words, the slope of the Phillips curve in front of the domestic output gap becomes flatter.

Putting aside the comparative statics analysis of the change in n, the connection between x_t and x_t^* may lead to severe omitted variable bias in the slope estimates if we only include the domestic output gap in the estimation of the Phillips curve for an open economy. Specifically, if the domestic monetary policy also affects the foreign country's aggregate demand, then domestic monetary shocks are no longer valid instruments to estimate the closed economy Phillips curve as they are correlated with the omitted variable x_t^*. This is the case for some large open economies like the U.S., there is a big strand of literature studying the spillover effects of the U.S. monetary policy on other economies (e.g., Dedola et al., 2017; Gai and Tong, 2022; Georgiadis, 2016). Another well-known case is related to the trilemma mechanism: base country's exogenous interest rate movements spill into local interest rates for open pegs, see Jordà, Schularick and Taylor (2020*a*) and the references therein. Controlling for x_t^* or global output gap in the first stage will clean out the part that's only correlated with x_t. In this way, the *residualized* monetary shocks should satisfy the exogeneity condition to estimate the sole effect of the domestic output gap.

3.3 Performance of Different Estimators in Simulations

How well do different estimators perform in the face of different empirical challenges? Using carefully designed simulation studies, I evaluate the performance of different estimators in the following aspects: (1) violation of exogeneity condition for predetermined variables and for raw sequence of monetary shocks; (2) failure of rank condition; (3) weak instruments problem; (4) choice of projection horizon or number of lags of monetary shocks as IVs; (5) small sample bias. In particular, part (4) involves both many and weak instruments problems pervasive in the empirical estimation of the Phillips curve.

3.3.1 Simulation Design

To preserve simplicity and tractability, I adapt the simulation design in Barnichon and Mesters (2020) and consider the following data-generating process for the system of equations for inflation π_t and output gap x_t:

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \lambda x_t + e_t \tag{3.8}$$

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \epsilon_t - e_t \tag{3.9}$$

where x_t follows an AR(2) process motivated by the hump-shaped impulse responses of inflation and the output gap. There are two shocks in the model: cost-push shock e_t and monetary policy shock ϵ_t. In equation (3.9), the sign of ϵ_t is normalized so that a *rise* in ϵ_t means a *fall* in interest rate. The negative sign of e_t reflects the effects of the systematic response of monetary policy rule in the face of supply shocks. That is, in response to a negative supply shock $e_t > 0$, the central bank raises interest rates to stabilize inflation, which leads to higher inflation and lower output.

The main difference between my simulation design and Barnichon and Mesters (2020) lies in the assumption on the cost-push shock e_t. Instead of following a simple i.i.d. normal distribution, e_t follows an ARMA(1,1) process as in Smets and Wouters (2007):

$$e_t = \rho_e e_{t-1} + \epsilon_t^p - \rho_p \epsilon_{t-1}^p \tag{3.10}$$

where ϵ_t^p is a price markup shock which follows an i.i.d. normal distribution. One important implication of the ARMA process is that e_t now generally depends on the entire history of price markup shock ϵ_t^p. In particular, by inverting equation (3.10) we have

$$e_t = \epsilon_t^p + \sum_{j=1}^{\infty} \rho_e^{j-1}(\rho_e - \rho_p)\epsilon_{t-j}^p \tag{3.11}$$

As long as $\rho_e \neq 0$ and $\rho_e \neq \rho_p$, then endogenous variables π_t and x_t are functions of current and all lagged values of ϵ_t^p. This further implies that lags of endogenous variables are no long valid instruments as they are also functions of lagged shocks including ϵ_{t-j}^p.

Let ξ_t denote the proxy for monetary shocks that researchers construct from the data. In the baseline results, I assume $\xi_t = \epsilon_t$, i.e., we observe the true structural monetary shock. I will alter this assumption in the discussion of the violation of the exogeneity condition. The parameters are configured as follows: $\gamma_b = 0.6, \gamma_f = 0.39, \lambda = 0.3, \rho_1 = 1.2$, and $\rho_2 = -0.4$.[8] The variation in σ_ϵ relative to the standard deviation of price markup shock (normalized to one) serves as an experiment to change the importance of monetary shocks in explaining the variance observed in output and inflation. $\sigma_\epsilon = 1$ in the baseline to minimize weak instruments problem. I vary σ_ϵ in the discussion of weak instruments problem.

For each of the following experiments, I simulate $M_s = 2000$ datasets and for each dataset, I estimate the NKPC using different methods. The mean estimates are reported and compared to evaluate their performances. The set of estimators considered in this paper is as follows: (1) OLS estimator: Ordinary Least Squares estimator; (2) GIV estimator: use predetermined variables as instruments following Galí and Gertler (1999); (3) ARIV estimator: Almon-restricted IV estimator following Barnichon and Mesters (2020); (4) 2SLS estimator: use the lag sequence of monetary shocks as instruments; (5) two-step estimator: estimate IRFs using local projections as in Jordà (2005*b*) without any controls, then estimate the Phillips curve using the estimated IRFs; (6) two-step-C estimator: estimate IRFs using local projections as in Jordà (2005*b*) with extra controls, then estimate the Phillips curve using the estimated IRFs.

[8]The values of $\gamma_b, \gamma_f, \lambda$ are close to the baseline estimates of the U.S. Phillips curve instrumented by narrative monetary shocks in Barnichon and Mesters (2020). Different from them, $\gamma_f = 0.39$ instead of 0.3 in their simulation study, nor 0.4 in their empirics. $\gamma_b + \gamma_f$ is slightly less than 1 to ensure that monetary shocks do not have permanent effects on inflation. The output gap process follows their simulation study.

3.3.2 Violation of Exogeneity Condition: Predetermined Variables

As shown above, the traditional approach using predetermined variables as instruments is lag exogeneous only when $\rho_e = 0$ or $\rho_e = \rho_p$. In the default case, the GIV estimator uses $z_t = (\pi_{t-2}, \pi_{t-3}, \cdots, x_{t-1}, x_{t-3}, \cdots)$ as instruments. If $\rho_e = 0$ but $\rho_p \neq 0$, i.e., e_t is an MA(1) process, we should lag the endogenous variables by one more period as instruments, i.e., $z_t = (\pi_{t-3}, \pi_{t-4}, \cdots, x_{t-2}, x_{t-3}, \cdots)$, the resulting estimator is denoted as GIV* estimator. Table 3.1 shows the mean estimates under four combinations of ρ_e and ρ_p. The sample size is set to be $T = 4000$ to reduce small sample bias.

Table 3.1. Mean parameter estimates: lag exogeneity condition

(a) $\rho_e = \rho_p = 0$

Estimator	γ_b	γ_f	λ
OLS	0.566	0.433	0.119
GIV	0.598	0.393	0.296
GIV*	0.596	0.394	0.295
ARIV	0.604	0.386	0.308
2SLS	0.599	0.391	0.297
two-step	0.599	0.390	0.298
two-step-C	0.599	0.391	0.298

(b) $\rho_e = 0.25, \rho_p = 0$

Estimator	γ_b	γ_f	λ
OLS	0.547	0.455	0.077
GIV	0.809	0.161	0.569
GIV*	0.664	0.320	0.374
ARIV	0.615	0.374	0.328
2SLS	0.598	0.392	0.296
two-step	0.599	0.391	0.297
two-step-C	0.598	0.392	0.297

(c) $\rho_e = 0, \rho_p = 0.4$

Estimator	γ_b	γ_f	λ
OLS	0.554	0.446	0.104
GIV	0.304	0.715	-0.097
GIV*	0.593	0.398	0.290
ARIV	0.597	0.393	0.295
2SLS	0.599	0.392	0.297
two-step	0.599	0.391	0.298
two-step-C	0.599	0.392	0.297

(d) $\rho_e = 0.25, \rho_p = 0.4$

Estimator	γ_b	γ_f	λ
OLS	0.570	0.429	0.130
GIV	0.494	0.506	0.167
GIV*	0.554	0.441	0.248
ARIV	0.600	0.390	0.300
2SLS	0.599	0.391	0.298
two-step	0.599	0.390	0.299
two-step-C	0.599	0.391	0.298

Notes: Mean estimates across 2000 Monte Carlo samples of size $T = 4000$ are reported. "GIV" is the IV estimator using $z_t = (\pi_{t-2}, \pi_{t-3}, \pi_{t-4}, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4})$ as instruments. "GIV*" differs from "GIV" in that $z_t = (\pi_{t-3}, \pi_{t-4}, \pi_{t-5}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5})$. "ARIV" is the Almon-restricted IV estimator. 2SLS estimator uses a lag sequence of monetary shocks $\epsilon_{t:t-H}$ as instruments. "two-step" and "two-step-C" stand for two-step estimator with and without controlling for four lags of inflation and output gap when estimating IRFs with local projections. Different panels correspond to different combinations of ρ_e and ρ_p. The true parameter values are $\gamma_b = 0.6, \gamma_f = 0.39, \lambda = 0.3$. $H = 20$.

As expected, OLS estimates are biased towards zero due to endogeneity issues in all four panels. On the contrary, the estimates that use monetary shocks as instruments all perform quite well. What's interesting is the performance of the two estimators that use predetermined variables as instruments. In panel (a) where $\rho_e = \rho_p = 0$, i.e., cost-push shock e_t is i.i.d., all IV estimators perform quite well and the mean estimates of all three coefficients are very close to their true values. In panel (b) where $\rho_e > \rho_p = 0$, the GIV estimator overestimates λ and γ_b, but underestimates γ_f. Taking one more lag of endogenous variables as instruments help mitigate this problem. The GIV* estimates are much closer to the true values than the GIV estimates. In panel (c) where $\rho_e = 0 < \rho_p$, e_t is autocorrelated, the GIV estimator is biased if we don't take more lags. After taking one more lag in constructing instruments from endogenous variables, bias is greatly reduced, and λ goes from -0.097 to 0.290. In panel (d) where $0 < \rho_e < \rho_p$, we observe the opposite of the results in panel (b). GIV estimator underestimates γ_b and λ but overestimates γ_f, and GIV* estimates cut the biases by about half.

3.3.3 Violation of Exogeneity Condition: Monetary Shocks

In this part, I will illustrate that the two-step approach with controls is superior to the 2SLS and Almon-restricted IV approaches when the exogeneity condition is only satisfied after controlling for extra information. I consider the following two sets of experiments.

Experiment I: Contaminated Monetary Shocks In the same spirit as Lewis and Mertens (2022), the key of the first set of experiments lies in the construction of monetary shock series. To verify that my simple simulation environment is capable of replicating their main results, and put the ARIV estimator in the same comparison, I simulate the empirically relevant "Romer-Romer" type monetary shocks by augmenting the pure structural monetary shocks with a linear combination of inflation and output gap over the past four quarters:

$$\xi_t = \epsilon_t + \sum_{p=1}^{4} b_p \pi_{t-p} + \sum_{p=1}^{4} d_p x_{t-p} \tag{3.12}$$

where b_p and d_p are estimated by regressing the Romer and Romer (2004) shocks on four lags of inflation and output gap over the sample period 1969-2007. Table 3.9 in Appendix

3.5.1 shows the regression results.[9] It is noteworthy that none of the coefficient estimates is statistically significant at the 95% confidence level. Even the richest regression model explains the shock series poorly with an R-square of 0.04.

Despite the weak relevance of lagged inflation and/or output gap, the constructed monetary shocks no longer satisfy lag exogeneity condition when e_t is autocorrelated, i.e., $\rho_e \neq 0$ and $\rho_e \neq \rho_p$. Table 3.2 shows the mean parameter estimates using the contaminated monetary shocks as instruments.

Table 3.2. Mean parameter estimates: contaminated monetary shocks

(a) $\rho_e = \rho_p = 0$

Estimator	γ_b	γ_f	λ
ARIV	1.174	-0.293	1.314
ARIV*	0.587	0.404	0.275
2SLS	0.598	0.393	0.295
two-step	0.560	0.455	0.245
two-step-C	0.599	0.391	0.298

(b) $\rho_e = 0.25, \rho_p = 0$

Estimator	γ_b	γ_f	λ
ARIV	0.640	0.341	0.361
ARIV*	0.602	0.383	0.289
2SLS	0.616	0.371	0.320
two-step	0.566	0.453	0.258
two-step-C	0.598	0.392	0.297

(c) $\rho_e = 0, \rho_p = 0.4$

Estimator	γ_b	γ_f	λ
ARIV	0.570	0.425	0.253
ARIV*	0.574	0.419	0.261
2SLS	0.574	0.420	0.259
two-step	0.548	0.463	0.224
two-step-C	0.599	0.392	0.297

(d) $\rho_e = 0.25, \rho_p = 0.4$

Estimator	γ_b	γ_f	λ
ARIV	0.558	0.440	0.229
ARIV*	0.581	0.411	0.269
2SLS	0.588	0.404	0.282
two-step	0.556	0.457	0.238
two-step-C	0.599	0.391	0.298

Notes: Mean estimates across 2000 Monte Carlo samples of size $T = 4000$ are reported. The simulated monetary shocks are correlated with both inflation and output gap over the past four quarters. Different panels correspond to different combinations of ρ_e and ρ_p. "ARIV" is the Almon-restricted IV estimator. Row "ARIV*" reports the *median* estimates of the ARIV estimator. 2SLS estimator uses a lag sequence of monetary shocks $\epsilon_{t:t-H}$ as instruments. "two-step" and "two-step-C" stand for two-step estimator with and without controlling for four lags of inflation and output gap when estimating IRFs with local projections. Different panels correspond to different combinations of ρ_e and ρ_p. The true parameter values are $\gamma_b = 0.6, \gamma_f = 0.39, \lambda = 0.3$. $H = 20$.

Panel (a) serves as a benchmark where the contaminated monetary shocks should still satisfy the lag exogeneity condition since cost-push shock e_t is i.i.d. normal. The 2SLS esti-

[9] Data are obtained from the replication package of Barnichon and Mesters (2020).

mates are close to the true parameters as expected. However, it turns out that the ARIV estimator under $H = 20$ is very sensitive to the measure of monetary shocks as it puts high weights on the lagged shocks by construction. The estimates are so severely skewed that their means are very far away from their medians. Take the slope estimate, for example, the median is 0.275 but the mean exceeds 1! The two-step estimator without any controls in the IRFs estimation is biased as well, and controlling for the lagged endogenous variables helps eliminate the bias. The mean estimates of the two-step approach with controls are exactly the same as those in panel (a) of Table 3.1.

Things start to change as we introduce autocorrelation into the cost-push shock e_t. In panel (b) where $\rho_e = 0.25, \rho_p = 0$, both the ARIV estimator and the 2SLS estimator tend to overestimate the slope coefficient on average, while the two-step approach without controls underestimates the slope coefficient. Only the two-step approach with controls performs the best. In panel (c) where e_t is an MA(1) process, all these estimators tend to underestimate the slope coefficient except for the two-step estimator with controls. Results in panel (d) are similar to those in panel (c).

Experiment II: Other Confounding Channels Another experiment that makes the two-step approach stand out relative to the ARIV and 2SLS methods is particularly relevant in the cross-country studies of monetary policy effects.

Consider the following open economy NKPC as in Razin and Binyamini (2007):

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \lambda x_t + \delta x_t^* + e_t \tag{3.13}$$

where x_t and x_t^* denote the domestic and foreign country's output gap, respectively. δ measures the sensitivity of domestic inflation with respect to the foreign demand x_t^*.

The simulated domestic output gap x_t follows the same AR(2) process as in the baseline model, i.e., equation (3.9). More importantly, the foreign country's output gap x_t^* is affected by both local and base country's monetary policy changes ϵ_t^* and ϵ_t:

$$x_t^* = \rho_1^* x_{t-1}^* + \rho_2^* x_{t-2}^* + \epsilon_t^* + \alpha \epsilon_t - e_t^* \tag{3.14}$$

where α measures the sensitivity of the foreign country's aggregate demand to the base country's monetary shocks. Take open pegs, for example, the exogenous base country's interest rate movement will spill into the peg country's interest rates. In the following experiment, I set $\rho_1^* = \rho_1$ and $\rho_2^* = \rho_2$, the same as the AR(2) process for x_t for simplicity. I set $\delta = 0.2$, $\alpha = 0.4$ so that domestic inflation is more sensitive to domestic demand than foreign demand, and the foreign country's demand is less sensitive to the base country's interest rate change than its own. $e_t, e_t^*, \epsilon_t, \epsilon_t^*$ are four i.i.d. standard normal structural shocks.

What will happen if researchers directly estimate the closed economy Phillips curve without taking x_t^* into account? This is probably because they are only interested in the slope coefficient λ but not δ. To illustrate, let's rewrite equation (3.13) as

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f \pi_{t+1} + \lambda x_t + \underbrace{\delta x_t^* + e_t - \gamma_f(\pi_{t+1} - E_t \pi_{t+1})}_{u_t} \tag{3.15}$$

Since ϵ_t is correlated with x_t^*, monetary shocks in the base country are no longer valid instruments due to the violation of the exogeneity condition, i.e., $E[\epsilon_t u_t] \neq 0$. All the estimators we've considered so far will be biased. In terms of the magnitude, for open pegs, the expansionary monetary policy in the base country forces the local interest rate to decline, otherwise, the capital inflows will put appreciation pressure on the exchange rate.

Table 3.3 presents the mean estimates of the closed economy Phillips curve. As expected, OLS estimates are still biased for all three parameters. The GIV estimates that use lagged inflation and output gap as instruments are also severely biased. The lagged output gap x_{t-j} is now correlated with the omitted foreign output gap x_t^* in that x_t^* is a function of all current and lags of ϵ_t. The ARIV and 2SLS estimators overestimate λ, so do the two-step estimators without any controls and with traditional control list, i.e., $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}$. But γ_b and γ_f are close to their true values.

The ability of the two-step approach to allow for extra controls opens the door for solving or at least mitigating the above endogeneity issue. However, it is important to add the key and a sufficient number of controls to reduce the bias. If we add four lags of x_t^* in addition to the traditional control list, the results essentially stay the same. In contrast, if we add the contemporaneous value of x_t^* in addition to the traditional control list, the mean estimate of

Table 3.3. Mean parameter estimates: confounding channels

Estimator	γ_b	γ_f	λ
True	0.6	0.39	0.3
OLS	0.526	0.478	0.072
GIV	0.416	0.594	0.008

(a) $H = 20$

Estimator	γ_b	γ_f	λ
ARIV	0.653	0.331	0.509
ARIV*	0.594	0.397	0.364
2SLS	0.595	0.396	0.368
two-step	0.596	0.394	0.371
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}$)	0.596	0.395	0.370
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_{t-1}, \cdots, x^*_{t-4}$)	0.596	0.395	0.370
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_t$)	0.607	0.382	0.324
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_t, x^*_{t-1}, \cdots, x^*_{t-4}$)	0.594	0.396	0.290

(b) $H = 8$

Estimator	γ_b	γ_f	λ
ARIV	0.601	0.389	0.384
ARIV*	0.599	0.391	0.379
2SLS	0.600	0.390	0.380
two-step	0.600	0.389	0.381
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}$)	0.600	0.390	0.380
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_{t-1}, \cdots, x^*_{t-4}$)	0.599	0.391	0.380
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_t$)	0.614	0.374	0.338
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_t, x^*_{t-1}, \cdots, x^*_{t-4}$)	0.599	0.391	0.300

Notes: Mean estimates across 2000 Monte Carlo samples of size $T = 4000$ are reported. “ARIV” is the Almon-restricted IV estimator. Row “ARIV*” reports the *median* estimates of the ARIV estimator. 2SLS estimator uses a lag sequence of monetary shocks $\epsilon_{t:t-H}$ as instruments. “two-step” and “two-step-C” stand for two-step estimator without and with controls when estimating IRFs with local projections. See the text for the information on different control lists. Different panels correspond to different choices of H.

the slope coefficient becomes 0.324 under $H = 20$ (or 0.338 under $H = 8$), the bias is cut by almost two thirds, but this introduces small biases in the estimates of γ_b and γ_f. Finally, if we add both contemporaneous and four lags of x_t^* in the control set, the mean estimates of all three parameters get extremely close to their true values, especially under $H = 8$. The observed difference in point estimates under $H = 20$ versus $H = 8$ is at least partially due to the many weak instruments problem discussed in Section 3.3.6.

Remarks: In practice, we seldomly include contemporaneous variables in the control sets when estimating impulse responses by LPs to avoid the absorption of the on-impact effects by contemporaneous variables. However, in the context at hand, we include x_t^* in order to absorb the effects through the alternative channel that domestic monetary shocks spill into the foreign output gap and thus inflation through imports. Since λ is the *partial correlation coefficient* between IRFs of π_t and x_t, it is necessary to control for this confounding channel when estimating the IRFs in the first step, especially the IRFs of inflation in this example. In Appendix 3.5.2, I show further details about the magnitude of the omitted variable bias using the traditional control list, and how adding additional controls $x_{t-j}^*(j \geq 0)$ affects the estimated IRFs and thus helps reduce the bias.

3.3.4 Failure of Rank Condition

Consider the general hybrid New Keynesian Phillips curve (3.8) and assume that the output gap follows an AR(1) process:

$$x_t = \rho x_{t-1} + \epsilon_t + \nu e_t \tag{3.16}$$

where ϵ_t and e_t are i.i.d. normal monetary shock and cost-push shock, respectively. $\nu = -1$.

Proposition 5. *The model characterized by (3.8) and (3.16) generally can not be identified using the sequence of shocks $z_t = \epsilon_{t:t-2}$ as instruments due to the failure of rank conditions. That is, the impulse response functions $\mathscr{R}_{h-1}^{\pi}, \mathscr{R}_{h+1}^{\pi}$, and $\mathscr{R}_h^x$ are linearly dependent.*

In particular,

$$\mathscr{R}_{h+1}^{\pi} = \delta_1^2 \mathscr{R}_{h-1}^{\pi} + (\delta_1 + \rho)\kappa \mathscr{R}_h^x \tag{3.17}$$

where $\kappa = \frac{\lambda}{\gamma_f(\delta_2 - \rho)}$, δ_1, δ_2 are the stable and unstable roots of second-order difference equation $\pi_t = \gamma_b \pi_{t-1} + \gamma_f \boldsymbol{E}_t \pi_{t+1}$.

Proof. See Appendix 3.5.3. □

By substituting (3.17) into the second-step regression where $\mathscr{R}_h^\pi = \gamma_b \mathscr{R}_{h-1}^\pi + \gamma_f \mathscr{R}_{h+1}^\pi + \lambda \mathscr{R}_h^x$, we have three extreme cases depending on which term is omitted.

Proposition 6. *If the forward-looking term is omitted, then the second-step regression yields*

$$\mathscr{R}_h^\pi = \delta_1 \mathscr{R}_{h-1}^\pi + \kappa \mathscr{R}_h^x \tag{3.18}$$

If the backward-looking term is omitted, then the second-step regression yields

$$\mathscr{R}_h^\pi = \frac{1}{\delta_1} \mathscr{R}_{h+1}^\pi - \kappa \frac{\rho}{\delta_1} \mathscr{R}_h^x \tag{3.19}$$

Otherwise, if the output gap term is omitted, then the second-step regression yields

$$\mathscr{R}_h^\pi = \frac{\rho\delta_1}{\delta_1 + \rho} \mathscr{R}_{h-1}^\pi + \frac{1}{\delta_1 + \rho} \mathscr{R}_{h+1}^\pi \tag{3.20}$$

Proof. See Appendix 3.5.4. □

Proposition 2 implies that the failure of the rank condition renders uncertainty in point estimates. If the forward-looking term is omitted, the slope estimate will be inflated as $\kappa > \lambda$. However, if the backward-looking term is omitted, the slope estimate will have the wrong sign. What's more, the absolute values of slope in both cases are increasing in ρ. The results lie in between if the output gap term is omitted. The actual point estimates should lie in between two extreme cases. That is, γ_f ranges between 0 and $\frac{1}{\delta_1}(> 1)$, γ_b ranges between 0 and $\delta_1(< 1)$, and λ ranges between $-\kappa\frac{\rho}{\delta_1}(\leq 0)$ and $\kappa(> \lambda)$.

Table 3.4 shows the mean estimates of $\boldsymbol{\beta} = [\gamma_b, \gamma_f, \lambda]$ across 2000 Monte Carlo samples for four parameter choices of ρ. Under the true parameter values of $\gamma_b = 0.6, \gamma_f = 0.39, \lambda = 0.3$, the theoretical ranges for $\hat{\gamma}_b$ and $\hat{\gamma}_f$ only depend on γ_b and γ_f, but the range for $\hat{\lambda}$ also depends on ρ, they are calculated accordingly in the table.

As expected, OLS estimates are severely biased in all four cases due to endogeneity issues. The GIV estimates using predetermined variables as instruments are very close to the 2SLS

Table 3.4. Mean parameter estimates: AR(1) output gap

(a) $\rho = 0$

Estimator	γ_b	γ_f	λ
OLS	0.535	0.461	-0.178
GIV	0.537	0.459	-0.198
ARIV	0.017	1.023	0.022
ARIV*	0.426	0.579	0.212
2SLS	0.364	0.648	0.178
two-step	0.361	0.651	0.177
two-step-C	0.363	0.649	0.178
γ_f omitted	0.958	-	0.479
γ_b omitted	-	1.044	0
λ omitted	0	1.044	-

(b) $\rho = 0.25$

Estimator	γ_b	γ_f	λ
OLS	0.501	0.494	-0.166
GIV	0.348	0.665	0.108
ARIV	-0.047	1.088	-0.188
ARIV*	0.418	0.589	0.163
2SLS	0.347	0.667	0.107
two-step	0.343	0.670	0.105
two-step-C	0.345	0.669	0.106
γ_f omitted	0.958	-	0.567
γ_b omitted	-	1.044	-0.148
λ omitted	0.198	0.828	-

(c) $\rho = 0.5$

Estimator	γ_b	γ_f	λ
OLS	0.453	0.553	-0.147
GIV	0.335	0.679	0.007
ARIV	-0.096	1.146	-0.485
ARIV*	0.409	0.600	0.089
2SLS	0.330	0.685	-0.001
two-step	0.329	0.686	-0.002
two-step-C	0.329	0.686	-0.002
γ_f omitted	0.958	-	0.695
γ_b omitted	-	1.044	-0.363
λ omitted	0.328	0.686	-

(d) $\rho = 0.75$

Estimator	γ_b	γ_f	λ
OLS	0.407	0.607	-0.187
GIV	0.334	0.681	-0.146
ARIV	0.977	-0.017	0.913
ARIV*	0.407	0.601	-0.024
2SLS	0.331	0.683	-0.151
two-step	0.333	0.681	-0.147
two-step-C	0.330	0.684	-0.152
γ_f omitted	0.958	-	0.898
γ_b omitted	-	1.044	-0.703
λ omitted	0.421	0.586	-

Notes: Mean estimates across 2000 Monte Carlo samples of size $T = 4000$ are reported except for ARIV* which reports the median of ARIV estimates. "GIV" is the IV estimator using $z_t = (\pi_{t-2}, \pi_{t-3}, \pi_{t-4}, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4})$ as instruments. "ARIV" is the Almon-restricted IV estimator. 2SLS estimator uses a lag sequence of monetary shocks as instruments. "two-step" and "two-step-C" stand for two-step estimator with and without controlling for lags of endogenous variables when estimating IRFs in the first step. Different panels correspond to different parameter values of ρ. The true parameter values are $\gamma_b = 0.6, \gamma_f = 0.39, \lambda = 0.3$. $H = 20$.

and two-step estimators except when $\rho = 0$.[10] As ρ increases, the slope estimates bias towards the negative territory, while the estimates for γ_b and γ_f are quite stable, and stay close to the average of the three possibilities. This implies that the data can't distinguish all three coefficients, and there's an approximately equal probability of ending up with any one of the three cases.

It is noteworthy that the ARIV approach produces unreasonably large or small point estimates. They are so severely skewed that the mean estimates lie far away from the medians, which are comparable to other IV estimates using structural monetary shocks as instruments but less downward biased.

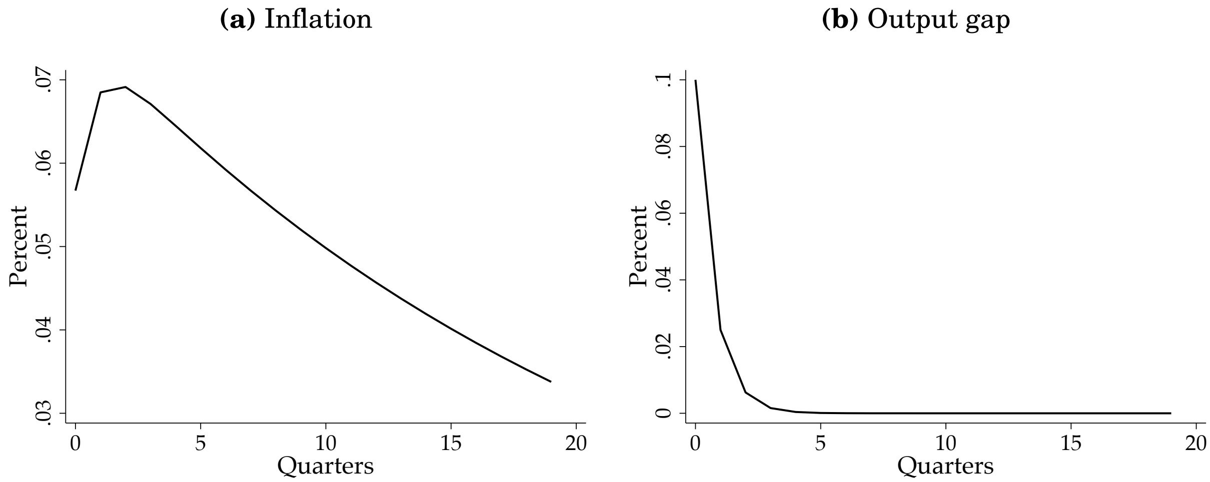
One way to improve the estimation results is to impose range inequality constraints, i.e., $0 < \gamma_b, \gamma_f, \lambda < 1$, together with the long-run restriction, i.e., $\gamma_b + \gamma_f = 1$. Proposition 2 suggests that omitting any one of the terms will introduce bias, it thus makes sense to restrict the coefficients to be positive. Besides, the long-run restriction is often imposed in the empirical estimation of the hybrid New Keynesian Phillips curve and seems to hold approximately in the data as well, see e.g., Barnichon and Mesters (2020).

To illustrate, I extract the theoretical IRFs for 20 periods under $\rho = 0.25$ and compare the two-step estimates with versus without imposing various constraints. In Figure 3.1, following an expansionary monetary shock, the impulse responses of inflation exhibit a hump-shaped pattern in the short run and then die out slowly over time. The IRFs of the output gap, however, jump up high initially and die out very quickly. It takes no longer than five years before the output gap completely recovers. For the periods $h \geq 5$, the output gap stays close to zero while inflation keeps declining.

Table 3.5 lists the second-step regression results using the theoretical IRFs. As expected, the unconstrained OLS estimates suffer from the failure of rank condition. When $H = 20$, γ_f is omitted in estimation, resulting in a much inflated λ. On the other hand, when $H = 8$, γ_b is omitted in estimation, resulting in $\gamma_f > 1$ and a negative λ. Imposing only the long-

[10] When $\rho = 0$, x_t is a function of contemporaneous shocks only. Lagged values of endogenous variables are no longer relevant since both shocks are i.i.d. normal. When $\rho > 0$, they are valid instruments.

Figure 3.1. Theoretical impulse responses



Notes: Theoretical impulse response functions of inflation (Panel a) and output gap (Panel b) from the simulation where output gap follows an AR(1) process with $\rho = 0.25$.

run restriction basically shifts the estimates to the third extreme case where λ is omitted. The Nonlinear Least Squares (NLS) estimates after imposing range inequalities depend on the initial values that one specifies for the optimization algorithm. It's very unlikely that researchers have the right priori of initial values, thus hard to evaluate its performance. With both sets of constraints imposed, the biases are greatly reduced. In addition, if we impose the true long-run restriction, then we are able to recover the true parameter values.

Table 3.5. Two-step approach: second-step regression results, $\rho = 0.25$

Method	Constraints	γ_b	γ_f	λ
OLS, H = 20	–	0.958	(omitted)	0.567
OLS, H = 8	–	(omitted)	1.044	-0.148
OLS, H = 20	$\gamma_b + \gamma_f = 1$	0.187	0.813	(omitted)
OLS, H = 8	$\gamma_b + \gamma_f = 1$	0.177	0.823	(omitted)
NLS, H = 8 or 20	$0 < \gamma_b, \gamma_f, \lambda < 1$	(depends)	(depends)	(0,0.567]
NLS, H = 8 or 20	$0 < \gamma_b, \gamma_f, \lambda < 1; \gamma_b + \gamma_f = 1$	0.489	0.511	0.217
NLS, H = 8 or 20	$0 < \gamma_b, \gamma_f, \lambda < 1; \gamma_b + \gamma_f = 0.99$	0.6	0.39	0.3

Notes: The first two rows report the unconstrained OLS estimates. The next two rows report the OLS estimates under the long-run constraint. H is the number of periods of IRFs used in the second-step regression. The fifth row reports the NLS estimates after imposing the range inequalities. The coefficient estimates for γ_b and γ_f depend on the initial value specified in the NLS estimation. The last two rows show the NLS estimates with both sets of constraints imposed.

Other econometric methods that deal with multicollinearity problems, such as ridge regression, lasso regression, and principal component analysis, aim to refine the set of regressors or use a weighted average of the regressors, which does not meet our purpose of recovering all three underlying parameters. It still remains a challenge in the empirical estimation of the Phillips curve when the rank condition fails in the data.

3.3.5 Weak Instruments Problem

One of the main concerns with the identification of the Phillips curve using monetary shocks as instruments is the weak instruments problem. That is, monetary shocks do not necessarily explain a large fraction of the variations in inflation and output gap (e.g., Gorodnichenko and Lee, 2020; Plagborg-Møller and Wolf, 2022).

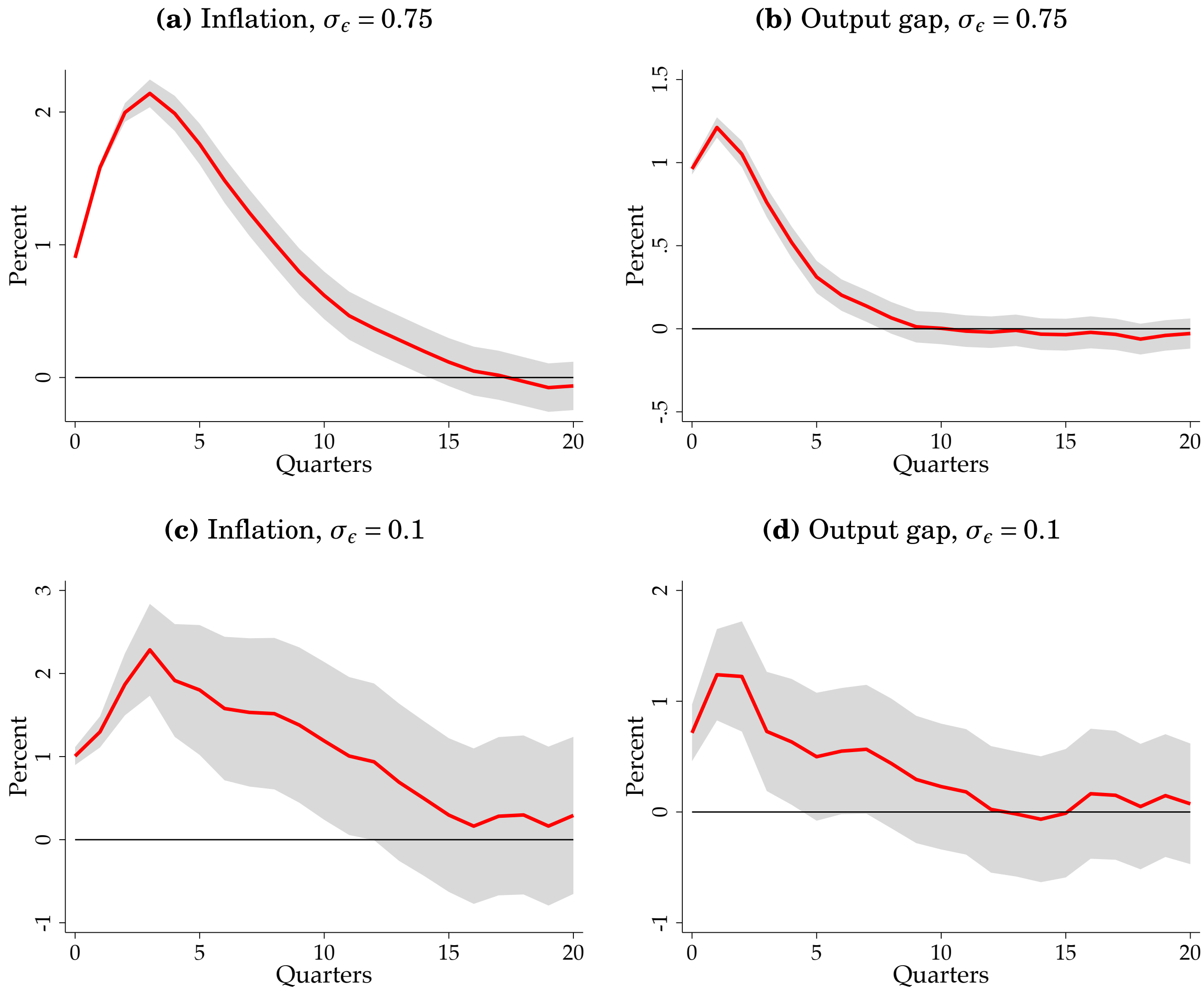
So far this has not been an issue since I set $\sigma_\epsilon = 1$, which is the same as the standard deviation of price markup shock ϵ_t^p or cost-push shock e_t when it's i.i.d. normal. In the following experiment, output gap follows an AR(2) process, and $\rho_e = \rho_p = 0, \sigma_e = 1$, i.e., e_t is i.i.d. standard normal. I will vary the standard deviation of monetary shocks σ_ϵ to change the importance of monetary shocks as one of the drivers of economic fluctuations.

In the context of the two-step approach, the presence of cost-push shock overshadows the estimation of the impulse responses in the first step. The smaller the σ_ϵ relative to σ_e, the larger the noise, and the more difficult it becomes to identify the impulse responses to monetary shocks. To illustrate, Figure 3.2 plots the estimated IRFs of inflation and output gap under $\sigma_\epsilon = 0.75$ versus $\sigma_\epsilon = 0.1$. The estimated IRFs under $\sigma_\epsilon = 0.1$ are very imprecise and less smooth compared to those under $\sigma_\epsilon = 0.75$.

Table 3.6 presents the mean estimates of the Phillips curve under different values of σ_ϵ. As expected, OLS estimates are severely biased in all cases due to endogeneity issues. In particular, the bias decreases with σ_ϵ as the endogeneity issues alleviate. On the other hand, the GIV estimates are close to the true parameter values since lags of inflation and output gap satisfy the lag exogeneity condition when cost-push shock e_t is i.i.d.

Among the IV estimators using monetary shocks as instruments, the ARIV estimator is less biased under small σ_ϵ, however, it's right skewed such that its mean is much larger than

Figure 3.2. Impulse responses when monetary shocks are strong vs. weak IVs



Notes: Data are simulated under $\sigma_\epsilon = 0.75$ versus $\sigma_\epsilon = 0.1$. The local projection is specified as follows:

$$y_{t+h} - y_{t-1} = \alpha^h + \beta^h \epsilon_t + \sum_{p=1}^{P} w'_{t-p} \delta_p^h + v_{t+h}, \quad h = 0, \cdots, H$$

where $y \in \{\pi, x\}, w = (\pi\ x)', P = 4, H = 20$. The grey areas are 90% Newey-West confidence bands.

its median. Take $\sigma_\epsilon = 0.5$, for example, the mean ARIV estimate of the slope coefficient is almost 0.4 while the median is 0.29. The 2SLS and two-step estimators are severely biased as well under small σ_ϵ. In the extreme case under $\sigma_\epsilon = 0.1$, their estimates even have the wrong sign. The biases are greatly reduced as σ_ϵ increases, and their estimates become quite close to the true values once σ_ϵ exceeds 0.5.

Part of the reason for which these IV estimators are so biased under small σ_ϵ also lies in the choice of H when constructing instruments. In fact, this is related to another IV

Table 3.6. Mean parameter estimates: weak instruments, $H = 20$

(a) $\sigma_\epsilon = 0.1$

Estimator	γ_b	γ_f	λ
OLS	-0.003	1.066	-0.786
GIV	0.593	0.398	0.289
ARIV	0.467	0.538	0.080
ARIV*	0.403	0.613	-0.078
2SLS	0.306	0.729	-0.306
two-step	0.337	0.706	-0.255
two-step-C	0.320	0.711	-0.280

(b) $\sigma_\epsilon = 0.25$

Estimator	γ_b	γ_f	λ
OLS	0.104	0.947	-0.635
GIV	0.594	0.397	0.290
ARIV	0.576	0.413	0.263
ARIV*	0.558	0.437	0.216
2SLS	0.531	0.469	0.159
two-step	0.534	0.465	0.165
two-step-C	0.538	0.461	0.174

(c) $\sigma_\epsilon = 0.5$

Estimator	γ_b	γ_f	λ
OLS	0.314	0.715	-0.321
GIV	0.595	0.396	0.292
ARIV	0.652	0.332	0.397
ARIV*	0.597	0.394	0.290
2SLS	0.585	0.408	0.268
two-step	0.586	0.405	0.272
two-step-C	0.586	0.406	0.272

(d) $\sigma_\epsilon = 0.75$

Estimator	γ_b	γ_f	λ
OLS	0.466	0.545	-0.067
GIV	0.596	0.394	0.294
ARIV	0.614	0.375	0.325
ARIV*	0.599	0.392	0.295
2SLS	0.595	0.396	0.290
two-step	0.596	0.394	0.292
two-step-C	0.596	0.395	0.291

Notes: Mean estimates across 2000 Monte Carlo samples of size $T = 4000$ are reported except for ARIV* which reports the median of ARIV estimates. Different panels correspond to different parameter values of σ_ϵ. "GIV" is the IV estimator using $z_t = (\pi_{t-2}, \pi_{t-3}, \pi_{t-4}, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4})$ as instruments. "ARIV" is the Almon-restricted IV estimator. 2SLS estimator uses a lag sequence of monetary shocks as instruments. "two-step" and "two-step-C" stand for two-step estimator with and without controlling for lags of endogenous variables when estimating IRFs in the first step. The true parameter values are $\gamma_b = 0.6, \gamma_f = 0.39, \lambda = 0.3$.

problem, i.e., the many instruments problem, in the sense that H is also the number of lagged monetary shocks as instruments for ARIV and 2SLS estimators. In Appendix Table 3.11, I show the mean estimates of the Phillips curve under $H = 8$. After alleviating the many instruments problem, the mean estimates have already become quite close to the true values when σ_ϵ increases to 0.25.

3.3.6 Projection Horizon and Lags of Monetary Shocks as IVs

The many instruments problem is another well-known IV problem associated with the identification of the Phillips curve using monetary shocks as instruments. To alleviate the concern of the many instruments problem, Barnichon and Mesters (2020) reduce the number of instruments to three by applying Almon (1965) parameterization on the impulse responses, which leads to the ARIV estimator.

In practice, both *many* and *weak* instruments problems are embedded in the choice of H, which is the number of lagged shocks included in the construction of instruments for the ARIV and 2SLS estimators, or the projection horizon in the estimation of IRFs for the two-step approach. It is noteworthy that there are two competing forces involved: on the one hand, as H increases, impulse responses of inflation and output gap tend to decay to zero (less relevant), which mechanically reduces the correlation between $\hat{\mathscr{R}}_h^{\pi}$ and $\hat{\mathscr{R}}_h^{x}$, i.e., $\lambda \to 0$ as H increases. On the other hand, it often takes time for macroeconomic shocks to have noticeable impacts on aggregate variables, e.g., due to the inside lag and outside lag in policymaking. As a result, H needs to be large enough to depict the full interactions between the IRFs of inflation and the output gap to help identification.

To illustrate, consider the baseline model with AR(2) output gap and i.i.d. cost-push shock. I consider two parameter choices of $\sigma_\epsilon \in \{0.25, 0.75\}$ and three choices of $H \in \{8, 20, 32\}$. As shown in Figure 3.2 under $\sigma_\epsilon = 0.75$, inflation and output gap respond immediately upon the arrival of monetary shocks, the first force is likely to dominate. That is, as H increases, the IV estimators based on monetary shocks tend to suffer more from many instruments problem. Table 3.7 shows the mean estimates of the Phillips curve for different choices of H under $\sigma_\epsilon = 0.25$ in Panel (a) and $\sigma_\epsilon = 0.75$ in Panel (b).

As expected, all the IV estimators that use monetary shocks as instruments bias towards zero as H increases, but the bias is quite small for reasonably large H, e.g., 32, under $\sigma_\epsilon = 0.75$ in Panel (a). That is, under strong identification, using a long sequence of lagged monetary shocks as instruments will not cause much bias in the point estimates, except for the ARIV estimates. However, it is noteworthy that many and weak instruments problems are intertwined under weak identification where monetary shocks cannot explain much of the variations in inflation and the output gap, e.g., $\sigma_\epsilon = 0.25$ in Panel (b). The biases in point estimates increase dramatically with the increase in H. By including more weak

Table 3.7. Mean parameter estimates: choice of H

Estimator	$\sigma_\epsilon = 0.75$			$\sigma_\epsilon = 0.25$		
	γ_b	γ_f	λ	γ_b	γ_f	λ
OLS	0.466	0.545	-0.067	0.104	0.947	-0.635
GIV	0.596	0.394	0.294	0.594	0.397	0.290

(a) $\sigma_\epsilon = 0.75$

Estimator	$H = 8$			$H = 20$			$H = 32$		
	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
ARIV	0.600	0.390	0.301	0.614	0.375	0.325	0.446	0.560	0.013
ARIV*	0.599	0.391	0.298	0.599	0.392	0.295	0.571	0.423	0.247
2SLS	0.599	0.391	0.298	0.595	0.396	0.290	0.591	0.400	0.281
two-step	0.600	0.389	0.301	0.596	0.394	0.292	0.589	0.402	0.277
two-step-C	0.600	0.390	0.300	0.596	0.395	0.291	0.592	0.400	0.282

(b) $\sigma_\epsilon = 0.25$

Estimator	$H = 8$			$H = 20$			$H = 32$		
	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
ARIV	0.614	0.374	0.330	0.576	0.413	0.263	0.666	0.317	0.409
ARIV*	0.599	0.392	0.295	0.558	0.437	0.216	0.400	0.611	-0.051
2SLS	0.576	0.419	0.248	0.531	0.469	0.159	0.492	0.512	0.080
two-step	0.586	0.407	0.267	0.534	0.465	0.165	0.490	0.517	0.074
two-step-C	0.584	0.410	0.265	0.538	0.461	0.174	0.498	0.505	0.092

Notes: Mean estimates across 2000 Monte Carlo samples of size $T = 4000$ are reported. Different panels correspond to different parameter values of σ_ϵ. "GIV" is the IV estimator using predetermined endogenous variables as instruments (see text). "ARIV" is the Almon-restricted IV estimator. ARIV* reports the medians of ARIV estimates. 2SLS estimator uses a lag sequence of monetary shocks as instruments. "two-step" and "two-step-C" stand for two-step estimator with and without controlling for lags of inflation and output gap when estimating IRFs with local projections. The true parameter values are $\gamma_b = 0.6, \gamma_f = 0.39, \lambda = 0.3$.

instruments in the estimation, the point estimates bias towards zero quickly.

Among these IV estimators based on monetary shocks, the ARIV estimator was proposed to mitigate the many instruments problem from which the 2SLS estimator suffers. By reducing the number of IVs to three, the ARIV estimator should perform better for any given H.

In fact, this is only true before H reaches a certain threshold, but after that ARIV estimator becomes very sensitive to the choice of H. In both panels, the ARIV estimates exhibit smaller biases for both $H = 8$ and $H = 20$, and the mean estimates are quite close to medians. However, we find much larger biases and skewness in ARIV estimates when $H = 32$.

To see why, by equation (3.6), two out of the three instruments are "weighted" by the number of lags h or h^2. Larger H adds more lagged monetary shocks into the instruments and at the same time shifts the weights towards the shocks that occurred far in the past. This potentially leads to more severe many weak instruments problems, and thus larger biases and uncertainty in the point estimates. Therefore, it is highly recommended to choose relatively small H when applying the ARIV approach.

3.3.7 Small Sample Bias

The main goal of the last simulation study is to investigate how these estimators compare in small sample bias. As usual, the output gap follows an AR(2) process and cost-push shock is i.i.d. normal. Table 3.8 reports the mean estimates of the Phillips curve for sample sizes ranging from $n = 200, 500, 1000$ to 4000. To minimize the weak intruments problem, $\sigma_\epsilon = 1$.

As expected, OLS is stable but severely biased due to endogeneity issues. The GIV estimator exhibits moderate bias under $T = 200$ even though the predetermined macroeconomic variables are valid instruments. Among the other three approaches, the Almon-restricted IV approach exhibits a much smaller bias in small samples under $H = 8$, but its performance significantly deteriorates under large H. Its mean slope estimate even has the wrong sign for $T = 200$ and 500 under $H = 20$. As the sample size increases, the slope estimate slowly improves. It is not until $T = 4000$ does the slope estimate become as close to the true value as the other three IV estimators.

In contrast, the 2SLS approach and the two-step approach produce relatively larger biases in small samples under $H = 8$, and their biases get widened under larger H. Unlike the ARIV estimates, the 2SLS and two-step estimates of the slope coefficient preserve the right sign in both cases, and the estimates improve quite significantly when the sample size increases from 200 to 500.

Table 3.8. Mean parameter estimates: small sample bias

Estimator	$T = 200$			$T = 500$			$T = 4000$		
	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
OLS	0.566	0.437	0.115	0.566	0.435	0.117	0.566	0.433	0.119
GIV	0.580	0.416	0.265	0.587	0.405	0.279	0.598	0.393	0.296

(a) $H = 8$

Estimator	$T = 200$			$T = 500$			$T = 4000$		
	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
ARIV	0.602	0.389	0.311	0.599	0.391	0.300	0.600	0.390	0.300
ARIV*	0.592	0.400	0.290	0.596	0.395	0.290	0.600	0.390	0.299
2SLS	0.583	0.413	0.265	0.594	0.398	0.288	0.600	0.390	0.301
two-step	0.588	0.399	0.279	0.596	0.393	0.292	0.601	0.389	0.302
two-step-C	0.582	0.414	0.267	0.593	0.399	0.288	0.600	0.390	0.301

(b) $H = 20$

Estimator	$T = 200$			$T = 500$			$T = 4000$		
	γ_b	γ_f	λ	γ_b	γ_f	λ	γ_b	γ_f	λ
ARIV	0.400	0.618	-0.084	0.432	0.580	-0.039	0.604	0.386	0.308
ARIV*	0.528	0.473	0.166	0.581	0.415	0.261	0.599	0.392	0.296
2SLS	0.566	0.433	0.218	0.585	0.409	0.265	0.599	0.391	0.297
two-step	0.566	0.427	0.227	0.585	0.405	0.268	0.599	0.390	0.298
two-step-C	0.567	0.431	0.227	0.586	0.408	0.269	0.599	0.391	0.298

Notes: Mean estimates across 2000 Monte Carlo samples of size $T = 4000$ are reported. Different panels correspond to different choices of H. "GIV" is the IV estimator using predetermined endogenous variables as instruments (see text). "ARIV" is the Almon-restricted IV estimator. ARIV* reports the medians of ARIV estimates. 2SLS estimator uses a lag sequence of monetary shocks as instruments. "two-step" and "two-step-C" stand for two-step estimator with and without controlling for lags of inflation and output gap when estimating IRFs with local projections. The true parameter values are $\gamma_b = 0.6, \gamma_f = 0.39, \lambda = 0.3$.

So far, the simulation study in this part has been based on the simplest data-generating process where all IV estimators perform quite well in a large sample, the strength of each estimator in other scenarios discussed above may distinguish one from the other in their small sample performance. The interaction between different challenges is of paramount importance in empirical analysis but is beyond the scope of this paper.

3.4 Concluding Remarks and Future Research

Recent developments in the empirical estimation of the Phillips curve deliver new sources of aggregate variations – structural demand shocks – to help resolve the endogeneity issues. In spite of the rising popularity, the IV estimators based on proxy measures of monetary shocks could suffer from similar pitfalls in empirical estimation due to instrument and data issues. Using carefully designed simulation studies, this paper evaluates the strengths and weaknesses of these estimators in the face of the following challenges: violation of the exogeneity condition, failure of the rank condition, weak instruments problem, many instruments problem, and small sample bias.

The main takeaway is as follows: first, given that the cost-push shocks are autocorrelated, the IV estimators based on true structural monetary shocks are consistent, while the GIV estimators using predetermined variables as instruments are not due to the failure of lag exogeneity condition. Second, the flexibility of the two-step approach to allow for extra controls in estimating the impulse responses gains its additional advantage over the other IV estimators under at least two scenarios: (i) proxies for monetary shocks are correlated with lags of inflation and/or output gap; (ii) monetary shocks affect domestic inflation through channels other than domestic output gap. Under both scenarios, the measured monetary shocks are valid instruments only after conditioning on extra information. Third, imposing range inequality constraints along with long-run constraint bring the point estimates close to their true values when the rank condition fails. Fourth, the many instruments and weak instruments problems are now intertwined together in that including more lags of monetary shocks or longer projection horizon provides more variations for identification, but endogenous variables become harder to predict at longer horizons. The natural tension between relevance and many weak instruments remains one of the main concerns. Fourth, among all IV estimators based on monetary shocks, the ARIV estimator exhibits smaller biases when the instruments are constructed using fewer lags of monetary shocks. However, it becomes unstable and skewed when the instruments are constructed using long sequence of monetary shocks.

Given the advantage of the two-step approach with extra controls in dealing with various endogeneity issues, it is worth further developing the inference procedure to construct confidence intervals that are robust to many weak instruments. One such attempt has been made by Lewis and Mertens (2022), they nicely transform the two-step approach into a GMM problem and develop weak instrument robust inference methods. Further research can build on their approach and develop inference methods that are also robust to many instruments. Another possibility is to rely on boostrap methods, e.g., Wang and Kaffo (2016).

The second direction for future research is to develop methods to determine the optimal projection horizon, which greatly depends on the dynamics of impulse responses and the trade-off between the relevance of instruments and many weak instruments problems. The choice of optimal projection horizon requires finer consideration, but its upper bound and lower bound are quite intuitive to figure out. As discussed in Section 3.3.6, on the one hand, after the impulse responses approach zero, extending the projection horizon only adds more zeros in the impulse response sequences used in the second-step regression, which mechanically reduces the slope estimate. On the other hand, it often takes time for macroeconomic shocks to have noticeable impacts on aggregate variables, the projection horizon needs to be large enough. Therefore, it amounts to finding out the projection horizons at which the IRFs of inflation and output gap depart from zero and flatten out. One potential approach is to estimate the IRFs by smooth LPs under some conventional choice of horizon, e.g., $H = 20$ for quarterly data. Given the choice of basis function for smoothing, we can then *predict* the horizons at which the smoothed IRFs would pick up or flatten out. Examples of basis functions for smoothing in the literature include Gaussian (Barnichon and Matthes, 2018) and B-spline (Barnichon and Brownlees, 2019).

3.5 Appendix

3.5.1 Conditional Exogeneity: Contaminated Monetary Shocks

Table 3.9 shows the auxiliary regression of Romer-Romer shocks on past inflation and/or output gap. Data are obtained from Barnichon and Mesters (2020).

Table 3.9. Auxiliary regression: Romer-Romer shocks on past inflation and output gap

	(1)	(2)	(3)
π_{t-1}	0.021		0.008
	(0.038)		(0.040)
π_{t-2}	-0.005		-0.000
	(0.043)		(0.043)
π_{t-3}	0.031		0.031
	(0.043)		(0.043)
π_{t-4}	-0.063		-0.058
	(0.038)		(0.040)
x_{t-1}		0.079	0.065
		(0.067)	(0.069)
x_{t-2}		-0.019	-0.018
		(0.097)	(0.098)
x_{t-3}		-0.111	-0.119
		(0.097)	(0.098)
x_{t-4}		0.100	0.113
		(0.067)	(0.069)
Constant	0.059	-0.006	0.075
	(0.092)	(0.048)	(0.095)
N	152	152	152
R^2	0.02	0.02	0.04

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.5.2 Conditional Exogeneity: Confounding Channels

The theoretical IRFs of the simulated model should satisfy the following relationship:

$$\mathscr{R}_h^{\pi} = \gamma_b \mathscr{R}_{h-1}^{\pi} + \gamma_f \mathscr{R}_{h+1}^{\pi} + \lambda \mathscr{R}_h^{x} + \delta \mathscr{R}_h^{x^*}, \quad h = 0, \cdots, H \tag{3.21}$$

where $\mathscr{R}_h^{y}$ is the IRFs of outcome variable y to the monetary shock ϵ_t at horizon h without conditioning on $x_{t-j}^*, j \geq 0$. To illustrate, let's rewrite equation (3.13) as follows:

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f \pi_{t+1} + \lambda x_t + \delta x_t^* + \underbrace{e_t - \gamma_f(\pi_{t+1} - E_t \pi_{t+1})}_{v_t}$$

Taking h-periods ahead,

$$\pi_{t+h} = \gamma_b \pi_{t+h-1} + \gamma_f \pi_{t+h+1} + \lambda x_{t+h} + \delta x_{t+h}^* + v_{t+h}$$

Multiplying both sides by monetary shocks ϵ_t, taking expectations, and dividing both sides by $E[\epsilon_t^2]$, we arrive at equation (3.21) where $\mathscr{R}_h^{v} = 0$ if $E[v_{t+h}\epsilon_t] = 0$.

In general, the correlation between $\mathscr{R}_h^{x^*}$ and $\mathscr{R}_h^{x}$ depends on autoregressive coefficients. However, when x_t and x_t^* have the same autoregressive coefficients, i.e., $\rho_1 = \rho_1^* = 1.2, \rho_2 = \rho_2^* = -0.4$, $\mathscr{R}_h^{x}$ and $\mathscr{R}_h^{x^*}$ are perfectly colinear. In particular, $\mathscr{R}_h^{x^*} = \alpha \mathscr{R}_h^{x}$ as their on-impact impulse responses are proportional.

Substituting $\mathscr{R}_h^{x^*} = \alpha \mathscr{R}_h^{x}$ in equation (3.21), we have

$$\mathscr{R}_h^{\pi} = \gamma_b \mathscr{R}_{h-1}^{\pi} + \gamma_f \mathscr{R}_{h+1}^{\pi} + (\lambda + \alpha\delta) \mathscr{R}_h^{x}$$

Given that $\lambda = 0.3, \alpha = 0.4, \delta = 0.2$, the slope $\lambda + \alpha\delta = 0.3 + 0.4 * 0.2 = 0.38$. The point estimates of the slope coefficient in Table 3.3 are quite close to this theoretical value, especially under $H = 8$.

To see how controlling for $x_{t-j}^*(j \geq 0)$ may help identification, instead of taking unconditional expectation, we take conditional expectation given some control set C and divide both sides by $E[\epsilon_t^2|C]$,

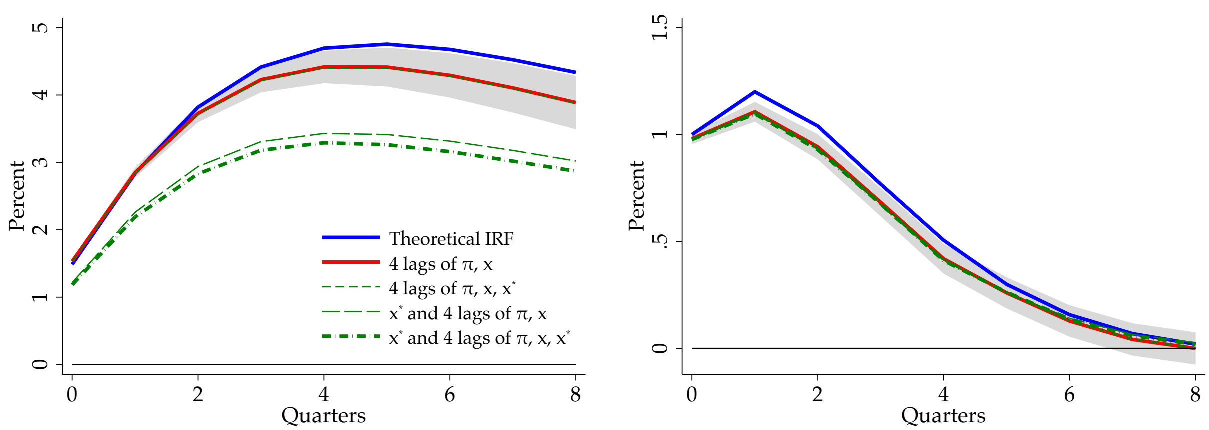
$$\mathscr{R}_h^{\pi|C} = \gamma_b \mathscr{R}_{h-1}^{\pi|C} + \gamma_f \mathscr{R}_{h+1}^{\pi|C} + \lambda \mathscr{R}_h^{x|C} + \delta \mathscr{R}_h^{x^*|C} + \mathscr{R}_h^{v|C}$$

where $\mathscr{R}_h^{y|C}$ is the projection of outcome variable y on the monetary shock ϵ_t conditional on the control set C at horizon h. Since ϵ_t affects x_{t+h}^* only through x_t^* and its autoregressive process, if C includes x_t^* as well as its lags, then $\mathscr{R}_h^{x^*|C} = 0$. Besides, the exogeneity condition only needs to hold conditional on the control set C, i.e., $E[v_{t+h}\epsilon_t|C] = 0$.

Figure 3.3 compares the IRFs to monetary shocks conditioning on different sets of controls. The main difference lies in the IRFs of inflation, while the IRFs of the domestic output gap do not differ much. At first glance, we find that IRFs estimated by LPs without controlling for x_{t-j}^* are close to the theoretical counterparts at the short horizons, but biased downward at the longer horizons. Adding four lags of the foreign output gap in the control set does not have much effect on the IRFs of inflation. In contrast, once we add the contemporaneous foreign output gap, the conditional projections on monetary shocks are much smaller than unconditional IRFs, which deflates the magnitude of the slope estimate.

Figure 3.3. Impulse responses estimated by LPs with different control sets

(a) Inflation **(b)** Domestic output gap



Notes: Impulse responses estimated by local projections are specified as follows:

$$y_{t+h} - y_{t-1} = \alpha^h + \beta^h \epsilon_t + \sum_{p=0}^{P} c'_{t-p}\delta_p^h + v_{t+h}, \quad h = 0, \cdots, H$$

where $y \in \{\pi, x\}$, control set c may also include contemporaneous foreign output gap x^*. The blue solid lines denote the theoretical IRFs consistent with equation (3.21). The red solid lines and grey shaded areas denote the IRFs and their 90 percent confidence bands estimated conditional on $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}$. The green short-dashed lines are IRFs estimated conditional on $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x_{t-1}^*, \cdots, x_{t-4}^*$. The green long-dashed lines are IRFs estimated conditional on $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}$ and x_t^*. The green dash-dotted lines are IRFs estimated conditional on $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}$ and $x_t^*, x_{t-1}^*, \cdots, x_{t-4}^*$.

To alleviate the concern that the baseline results might be a special case when x_t and x_t^*

share the same autoregressive coefficients. What if $\rho_1 \neq \rho_1^*, \rho_2 \neq \rho_2^*$? What if x_t^* follows a simple AR(1) process? To provide further checks, I repeat the practice for the following two cases: (a) keep $\rho_1 = 1.2, \rho_2 = -0.4$, but $\rho_1^* = 1.5, \rho_2^* = -0.6$; (b) $\rho_1^* = 0.25, \rho_2^* = 0$. The mean estimates are shown in Table 3.10. Informed by the baseline results, the number of lagged monetary shocks as instruments and the projection horizon $H = 8$ produces less biased results.

Under $\rho_1^* = 1.5, \rho_2^* = -0.6$, the slope estimates are biased upward by more for all the IV estimators except the two-step estimator that controls both contemporaneous and lagged values of foreign output gap when estimating the IRFs in the first step. Unlike the baseline results where $\rho_1^* = \rho_1 = 1.2$ and $\rho_2^* = \rho_2 = -0.4$, the inflation expectation terms γ_b and γ_f are also severely biased for other IV estimators. Furthermore, adding x_t^* alone does not improve much compared to 2SLS estimates, but only distorts the estimates of inflation expectation terms. Only by including x_t^* as well as its four lags in the estimation of IRFs are the mean estimates close to their true values.

In Panel (b), x_t^* follows an AR(1) process with $\rho_1 = 0.25$. Such simple shock propagation structure leads to two new observations: first, the slope estimates are biased *downward* instead, and second, the two-step approach with x_t^* included in the control set produces estimates that are already close enough to the true values.

Table 3.10. Mean parameter estimates: other channels and alternative processes for x^*

(a) $\rho_1^* = 1.5, \rho_2^* = -0.6$

Estimator	γ_b	γ_f	λ
OLS	0.511	0.494	0.044
GIV	0.442	0.565	0.039
ARIV	0.685	0.305	0.616
ARIV*	0.676	0.315	0.590
2SLS	0.634	0.362	0.466
two-step	0.628	0.371	0.436
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}$)	0.627	0.372	0.436
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_{t-1}, \cdots, x^*_{t-4}$)	0.627	0.372	0.435
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_t$)	0.659	0.336	0.441
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_t, x^*_{t-1}, \cdots, x^*_{t-4}$)	0.599	0.390	0.301

(b) $\rho_1^* = 0.25, \rho_2^* = 0$

Estimator	γ_b	γ_f	λ
OLS	0.569	0.430	0.135
GIV	0.582	0.409	0.277
ARIV	0.541	0.455	0.216
ARIV*	0.542	0.455	0.217
2SLS	0.534	0.464	0.199
two-step	0.531	0.468	0.190
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}$)	0.531	0.468	0.189
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_{t-1}, \cdots, x^*_{t-4}$)	0.531	0.468	0.189
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_t$)	0.602	0.385	0.304
two-step-C (ctr. list: $\pi_{t-1}, \cdots \pi_{t-4}, x_{t-1}, \cdots, x_{t-4}, x^*_t, x^*_{t-1}, \cdots, x^*_{t-4}$)	0.599	0.391	0.299

Notes: Mean estimates across 2000 Monte Carlo samples of size $T = 4000$ are reported. x_t^* follows an AR(2) process with autoregressive coefficients distinct from x_t in Panel (a), and an AR(1) process in Panel (b). "GIV" is the IV estimator using $z_t = (\pi_{t-2}, \pi_{t-3}, \pi_{t-4}, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4})$ as instruments. "ARIV" is the Almon-restricted IV estimator. Row "ARIV*" reports the *median* estimates of the ARIV estimator. 2SLS estimator uses a lag sequence of monetary shocks $\epsilon_{t:t-H}$ as instruments. "two-step" and "two-step-C" stand for two-step estimator without and with controls when estimating IRFs with local projections. See the text for the information on different control lists. The true parameter values are $\gamma_b = 0.6, \gamma_f = 0.39, \lambda = 0.3$. $H = 8$.

3.5.3 Proof of Proposition 1

Proof. First, apply the undetermined coefficient approach and rewrite equation (3.8) as follows:

$$\delta_2(\pi_t - \delta_1\pi_{t-1}) = E_t\pi_{t+1} - \delta_1\pi_t + \lambda' x_t + \alpha e_t$$

That is,

$$\left(1+\frac{\delta_1}{\delta_2}\right)\pi_t = \delta_1\pi_{t-1} + \frac{1}{\delta_2}E_t\pi_{t+1} + \frac{\lambda'}{\delta_2}x_t + \frac{\alpha}{\delta_2}e_t$$

Matching coefficients with equation (3.8), we have

$$\begin{aligned}
\delta_1 &= \gamma_b\left(1+\frac{\delta_1}{\delta_2}\right) &\Longrightarrow&& \delta_1\delta_2 &= \gamma_b(\delta_1+\delta_2)\\
\frac{1}{\delta_2} &= \gamma_f\left(1+\frac{\delta_1}{\delta_2}\right) &\Longrightarrow&& 1 &= \gamma_f(\delta_1+\delta_2)\\
\frac{\lambda'}{\delta_2} &= \lambda\left(1+\frac{\delta_1}{\delta_2}\right) &\Longrightarrow&& \lambda' &= \lambda(\delta_1+\delta_2)\\
\frac{\alpha}{\delta_2} &= \left(1+\frac{\delta_1}{\delta_2}\right) &\Longrightarrow&& \alpha &= \delta_1+\delta_2
\end{aligned}$$

So δ_1 and δ_2 are the solutions of the following quadratic equation

$$\gamma_f w^2 - w + \gamma_b = 0$$

which is the same as solving the homogeneous difference equation $\pi_t = \gamma_b\pi_{t-1} + \gamma_f E_t(\pi_{t+1})$. The two roots are

$$\delta_{1,2} = \frac{1 \pm \sqrt{1-4\gamma_b\gamma_f}}{2\gamma_f}$$

So

$$\pi_t - \delta_1\pi_{t-1} = \frac{1}{\delta_2}(E_t\pi_{t+1} - \delta_1\pi_t) + \frac{\lambda}{\delta_2\gamma_f}x_t + \frac{1}{\delta_2\gamma_f}e_t$$

Iterating forward and taking expectations, we have

$$E_t\pi_{t+1} - \delta_1\pi_t = \frac{1}{\delta_2}(E_t\pi_{t+2} - \delta_1 E_t\pi_{t+1}) + \frac{\lambda}{\delta_2\gamma_f}E_t x_{t+1}$$

where $E_t[E_{t+1}\pi_{t+2}] = E_t\pi_{t+2}$ by law of iterated expectation, and $E_t e_{t+1} = 0$.

Therefore,

$$\pi_t - \delta_1\pi_{t-1} = \left(\frac{1}{\delta_2}\right)^2(E_t\pi_{t+2} - \delta_1 E_t\pi_{t+1}) + \frac{1}{\delta_2}\frac{\lambda}{\delta_2\gamma_f}E_t x_{t+1} + \frac{\lambda}{\delta_2\gamma_f}x_t + \frac{1}{\delta_2\gamma_f}e_t$$

Repeating this process, we get

$$\pi_t - \delta_1 \pi_{t-1} = \frac{\lambda}{\delta_2 \gamma_f} \sum_{j=0}^{\infty} \left(\frac{1}{\delta_2}\right)^j E_t x_{t+j} + \frac{1}{\delta_2 \gamma_f} e_t$$

For this to be a stable solution, δ_2 must be the unstable root, that is,

$$\delta_1 = \frac{1-\sqrt{1-4\gamma_b\gamma_f}}{2\gamma_f}, \delta_2 = \frac{1+\sqrt{1-4\gamma_b\gamma_f}}{2\gamma_f}$$

On the other hand, we can iterate equation 3.16 backward to get

$$x_t = \sum_{j=0}^{\infty} \rho^j (\epsilon_{t-j} + \nu e_{t-j})$$

Therefore, the solution of π_t and x_t is

$$\begin{cases} \pi_t & = \delta_1 \pi_{t-1} + \frac{\lambda}{\delta_2 \gamma_f} \sum_{j=0}^{\infty} \left(\frac{1}{\delta_2}\right)^j E_t x_{t+j} + \frac{1}{\delta_2 \gamma_f} e_t \\ x_t & = \sum_{j=0}^{\infty} \rho^j (\epsilon_{t-j} + \nu e_{t-j}) \end{cases}$$

Let $w_t = \begin{bmatrix} x_t & \pi_{t-1} & \pi_{t+1} \end{bmatrix}'$, and $z_t = \begin{bmatrix} \epsilon_t & \epsilon_{t-1} & \epsilon_{t-2} \end{bmatrix}'$, then

$$\Gamma = E(w_t z_t') = \begin{bmatrix} E(x_t\epsilon_t) & E(x_t\epsilon_{t-1}) & E(x_t\epsilon_{t-2}) \\ E(\pi_{t-1}\epsilon_t) & E(\pi_{t-1}\epsilon_{t-1}) & E(\pi_{t-1}\epsilon_{t-2}) \\ E(\pi_{t+1}\epsilon_t) & E(\pi_{t+1}\epsilon_{t-1}) & E(\pi_{t+1}\epsilon_{t-2}) \end{bmatrix}$$

Since $E(\epsilon_{t-j}\epsilon_{t-j'}) = 1$ if $j = j'$, and 0, otherwise, and $E(\epsilon_{t-j} e_{t-j'}) = 0, \forall j \neq j'$, and

$$E_t x_{t+j} = \sum_{s=0}^{\infty} \rho^s (E\epsilon_{t+j-s} + \nu E_t e_{t+j-s}) = \sum_{s=j}^{\infty} \rho^s (\epsilon_{t+j-s} + \nu e_{t+j-s}) = \rho^j \sum_{s'=0}^{\infty} \rho^{s'} (\epsilon_{t-s'} + \nu e_{t-s'}) = \rho^j x_t$$

we have the following results

- $E(x_t\epsilon_{t-j}) = \rho^j, \forall j \geq 0$
- $E(\pi_{t-j}\epsilon_{t-j'}) = 0, \forall j > j'$
- $E(\pi_t\epsilon_t) = \delta_1 E(\pi_{t-1}\epsilon_t) + \frac{\lambda}{\delta_2\gamma_f} \sum_{j'=0}^{\infty} \left(\frac{1}{\delta_2}\right)^{j'} \rho^{j'} E(x_t\epsilon_t) = \frac{\lambda}{\delta_2\gamma_f} \frac{1}{1-\rho/\delta_2} \equiv \kappa$
- $E(\pi_t\epsilon_{t-j}) = \delta_1 E(\pi_{t-1}\epsilon_{t-j}) + \frac{\lambda}{\delta_2\gamma_f} \sum_{s=0}^{\infty} \left(\frac{1}{\delta_2}\right)^s \rho^s E(x_t\epsilon_{t-j}) = \delta_1 E(\pi_{t-1}\epsilon_{t-j}) + \rho^j \kappa$
- $E(\pi_t\epsilon_{t-1}) = \delta_1\kappa + \rho\kappa$

- $E(\pi_t \epsilon_{t-2}) = \delta_1 E(\pi_{t-1}\epsilon_{t-2}) + \rho^2\kappa = \delta_1(\delta_1\kappa + \rho\kappa) + \rho^2\kappa$
- $E(\pi_{t+1}\epsilon_{t-j}) = \delta_1 E(\pi_t \epsilon_{t-j}) + \rho^{j+1}\kappa$
- $E(\pi_{t+1}\epsilon_t) = \delta_1 E(\pi_t\epsilon_t) + \rho\kappa = \delta_1\kappa + \rho\kappa$
- $E(\pi_{t+1}\epsilon_{t-1}) = \delta_1 E(\pi_t \epsilon_{t-1}) + \rho^2\kappa = \delta_1(\delta_1\kappa + \rho\kappa) + \rho^2\kappa$
- $E(\pi_{t+1}\epsilon_{t-2}) = \delta_1 E(\pi_t\epsilon_{t-2}) + \rho^3\kappa = \delta_1(\delta_1(\delta_1\kappa + \rho\kappa) + \rho^2\kappa) + \rho^3\kappa$

Hence,

$$\Gamma = \begin{bmatrix} 1 & \rho & \rho^2 \\ 0 & \kappa & \delta_1\kappa + \rho\kappa \\ \delta_1\kappa + \rho\kappa & \delta_1(\delta_1\kappa + \rho\kappa) + \rho^2\kappa & \delta_1(\delta_1(\delta_1\kappa + \rho\kappa) + \rho^2\kappa) + \rho^3\kappa \end{bmatrix}$$

The third row can be obtained by a linear combination of the first two rows, or

$$\begin{aligned}\det\Gamma &= \begin{vmatrix} 1 & \rho & \rho^2 \\ 0 & \kappa & \delta_1\kappa + \rho\kappa \\ \delta_1\kappa + \rho\kappa & \delta_1(\delta_1\kappa + \rho\kappa) + \rho^2\kappa & \delta_1(\delta_1(\delta_1\kappa + \rho\kappa) + \rho^2\kappa) + \rho^3\kappa \end{vmatrix} \\ &= \begin{vmatrix} 1 & \rho & \rho^2 \\ 0 & \kappa & \delta_1\kappa + \rho\kappa \\ 0 & \delta_1^2\kappa & \delta_1^2(\delta_1\kappa + \rho\kappa) \end{vmatrix} = \begin{vmatrix} 1 & \rho & \rho^2 \\ 0 & \kappa & \delta_1\kappa + \rho\kappa \\ 0 & 0 & 0 \end{vmatrix} = 0\end{aligned}$$

That is, the rank condition always fails given that the output gap is AR(1).

Furthermore, we can show that $E(\pi_{t+1}\epsilon_{t-j}) = (\delta_1\kappa + \rho\kappa)E(x_t\epsilon_{t-j}) + \delta_1^2 E(\pi_{t-1}\epsilon_{t-j}), \forall j \geq 0$. Since $E(x_t\epsilon_{t-j}) = \rho^j, \forall j \geq 0$,

$$\begin{aligned} E(\pi_{t+1}\epsilon_{t-j}) &= \delta_1 E(\pi_t\epsilon_{t-j}) + \rho^{j+1}\kappa \\ &= \delta_1^2 E(\pi_{t-1}\epsilon_{t-j}) + \delta_1\rho^j\kappa + \rho^{j+1}\kappa \\ &= \delta_1^2 E(\pi_{t-1}\epsilon_{t-j}) + \rho^j(\delta_1\kappa + \rho\kappa) \\ &= \delta_1^2 E(\pi_{t-1}\epsilon_{t-j}) + E(x_t\epsilon_{t-j})(\delta_1\kappa + \rho\kappa) \end{aligned}$$

Dividing both sides by $E(\epsilon_{t-j}^2)$, we have equation (3.17).

□

3.5.4 Proof of Proposition 2

If the forward-looking inflation expectation term is omitted, then

$$\begin{aligned}\mathscr{R}^{\pi}_{h} &= \gamma_b \mathscr{R}^{\pi}_{h-1} + \gamma_f\left(\delta_1^2 \mathscr{R}^{\pi}_{h-1} + (\delta_1+\rho)\kappa \mathscr{R}^{x_t}_{h}\right) + \lambda \mathscr{R}^{x}_{h} \\ &= (\gamma_b + \gamma_f \delta_1^2)\mathscr{R}^{\pi}_{h-1} + \left[(\delta_1+\rho)\gamma_f\kappa + \lambda\right]\mathscr{R}^{x}_{h}\end{aligned}$$

Since $\kappa = \frac{\lambda}{\gamma_f(\delta_2-\rho)}$,

$$\mathscr{R}^{\pi}_{h} = (\gamma_b + \gamma_f \delta_1^2)\mathscr{R}^{\pi}_{h-1} + \frac{\delta_1+\delta_2}{\delta_2-\rho}\lambda \mathscr{R}^{x}_{h}$$

$\because \delta_1 + \delta_2 = \frac{1}{\gamma_f}, \delta_1\delta_2 = \frac{\gamma_b}{\gamma_f}$,

$$\therefore \gamma_b + \gamma_f\delta_1^2 = \gamma_f\left(\frac{\gamma_b}{\gamma_f} + \delta_1^2\right) = \gamma_f(\delta_1^2 + \delta_1\delta_2) = \gamma_f\delta_1(\delta_1+\delta_2) = \delta_1$$

and

$$\frac{\delta_1+\delta_2}{\delta_2-\rho}\lambda = \frac{\lambda}{\gamma_f(\delta_2-\rho)} = \kappa > \lambda$$

That is,

$$\mathscr{R}^{\pi}_{h} = \delta_1 \mathscr{R}^{\pi}_{h-1} + \kappa \mathscr{R}^{x}_{h}$$

In this case, the slope coefficient will be inflated and increasing in ρ.

If the backward-looking term is omitted, then

$$\begin{aligned}\mathscr{R}^{\pi}_{h} &= \frac{1}{\delta_1}\left[\mathscr{R}^{\pi}_{h+1} - (\delta_1+\rho)\kappa \mathscr{R}^{x}_{h}\right] + \kappa \mathscr{R}^{x}_{h} \\ &= \frac{1}{\delta_1}\mathscr{R}^{\pi}_{h+1} - \kappa\frac{\rho}{\delta_1}\mathscr{R}^{x}_{h}\end{aligned}$$

In this case, the slope coefficient becomes negative! Its absolute value is increasing in ρ.

If the output gap term is omitted, then

$$\begin{aligned}\mathscr{R}^{\pi}_{h} &= \frac{1}{\delta_1}\mathscr{R}^{\pi}_{h+1} - \frac{\rho}{\delta_1}\frac{1}{\delta_1+\rho}\left(\mathscr{R}^{\pi}_{h+1} - \delta_1^2\mathscr{R}^{\pi}_{h-1}\right) \\ &= \frac{\rho\delta_1}{\delta_1+\rho}\mathscr{R}^{\pi}_{h-1} + \frac{1}{\delta_1+\rho}\mathscr{R}^{\pi}_{h+1}\end{aligned}$$

In this case, the slope coefficient is zero, the forward-looking term is decreasing in ρ while the backward-looking term is increasing in ρ.

3.5.5 Appendix Tables

Table 3.11. Mean parameter estimates: weak instruments, $H = 8$

(a) $\sigma_\epsilon = 0.1$

Estimator	γ_b	γ_f	λ
OLS	-0.003	1.066	-0.786
GIV	0.593	0.398	0.289
ARIV	0.776	0.186	0.670
ARIV*	0.591	0.406	0.271
2SLS	0.429	0.602	-0.090
two-step	0.451	0.592	-0.063
two-step-C	0.466	0.560	-0.024

(b) $\sigma_\epsilon = 0.25$

Estimator	γ_b	γ_f	λ
OLS	0.104	0.947	-0.635
GIV	0.594	0.397	0.290
ARIV	0.614	0.374	0.330
ARIV*	0.599	0.392	0.295
2SLS	0.576	0.419	0.248
two-step	0.586	0.407	0.267
two-step-C	0.584	0.410	0.265

(c) $\sigma_\epsilon = 0.5$

Estimator	γ_b	γ_f	λ
OLS	0.314	0.715	-0.320
GIV	0.595	0.396	0.292
ARIV	0.602	0.388	0.305
ARIV*	0.598	0.392	0.298
2SLS	0.596	0.395	0.291
two-step	0.598	0.392	0.296
two-step-C	0.598	0.393	0.295

(d) $\sigma_\epsilon = 0.75$

Estimator	γ_b	γ_f	λ
OLS	0.466	0.545	-0.067
GIV	0.596	0.394	0.294
ARIV	0.600	0.390	0.301
ARIV*	0.599	0.391	0.298
2SLS	0.599	0.391	0.298
two-step	0.600	0.389	0.301
two-step-C	0.600	0.390	0.300

Notes: Mean estimates across 2000 Monte Carlo samples of size $T = 4000$ are reported except for ARIV* which reports the median of ARIV estimates. Different panels correspond to different parameter values of σ_ϵ. "GIV" is the IV estimator using $z_t = (\pi_{t-2}, \pi_{t-3}, \pi_{t-4}, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4})$ as instruments. "ARIV" is the Almon-restricted IV estimator. 2SLS estimator uses a lag sequence of monetary shocks as instruments. "two-step" and "two-step-C" stand for two-step estimator with and without controlling for lags of endogenous variables when estimating IRFs in the first step. The true parameter values are $\gamma_b = 0.6, \gamma_f = 0.39, \lambda = 0.3$.

Bibliography

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J Klenow, and Huiyu Li. 2019. "A Theory of Falling Growth and Rising Rents." NBER Working Paper 26448.

Aghion, Philippe, Ufuk Akcigit, and Peter Howitt. 2014. "What Do We Learn From Schumpeterian Growth Theory?" In *Handbook of Economic Growth*. Vol. 2, 515–563. Elsevier.

Akcigit, Ufuk, and Sina T Ates. 2019. "What Happened to US Business Dynamism?" NBER Working Paper 25756.

Akcigit, Ufuk, and Sina T. Ates. 2021. "Ten Facts on Declining Business Dynamism and Lessons From Endogenous Growth Theory." *American Economic Journal: Macroeconomics*, 13(1): 257–98.

Almon, Shirley. 1965. "The Distributed Lag Between Capital Appropriations and Expenditures." *Econometrica: Journal of the Econometric Society*, 178–196.

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

Attanasio, Orazio P, and Guglielmo Weber. 1995. "Is Consumption Growth Consistent With Intertemporal Optimization? Evidence From the Consumer Expenditure Survey." *Journal of political Economy*, 103(6): 1121–1157.

Bai, Jushan, and Pierre Perron. 1998. "Estimating and Testing Linear Models With Multiple Structural Changes." *Econometrica*, 47–78.

Bai, Jushan, and Pierre Perron. 2003. "Computation and Analysis of Multiple Structural Change Models." *Journal of Applied Econometrics*, 18(1): 1–22.

Bajgar, Matej, Giuseppe Berlingieri, Sara Calligaris, Chiara Criscuolo, and Jonathan Timmis. 2019. "Industry Concentration in Europe and North America." *OECD Productivity Working Papers, No. 18, OECD Publishing, Paris*.

Ball, Laurence, and Sandeep Mazumder. 2019. "A Phillips Curve With Anchored Expectations and Short-Term Unemployment." *Journal of Money, Credit and Banking*, 51(1): 111–137.

Baqaee, David, Emmanuel Farhi, and Kunal Sangani. 2021. "The Supply-Side Effects of Monetary Policy." NBER Working Paper 28345.

Barnichon, Regis, and Christian Brownlees. 2019. "Impulse Response Estimation by Smooth Local Projections." *Review of Economics and Statistics*, 101(3): 522–530.

Barnichon, Regis, and Christian Matthes. 2018. "Functional Approximation of Impulse Responses." *Journal of Monetary Economics*, 99: 41–55.

Barnichon, Regis, and Geert Mesters. 2020. "Identifying Modern Macro Equations With Old Shocks." *The Quarterly Journal of Economics*, 135(4): 2255–2298.

Barsky, Robert B, and Eric R Sims. 2009. "News Shocks." National Bureau of Economic Research Working Paper 15312.

Barsky, Robert B., and Eric R. Sims. 2011. "News Shocks and Business Cycles." *Journal of Monetary Economics*, 58(3): 273 – 289.

Beaudry, Paul, and Bernd Lucke. 2010. "Letting Different Views about Business Cycles Compete." *NBER Macroeconomics Annual 2009, Volume 24*, 413–455. University of Chicago Press.

Beaudry, Paul, and Franck Portier. 2004. "An Exploration into Pigou's Theory of Cycles." *Journal of Monetary Economics*, 51(6): 1183–1216.

Beaudry, Paul, and Franck Portier. 2006. "Stock Prices, News, and Economic Fluctuations." *American Economic Review*, 96(4): 1293–1307.

Bergeaud, Antonin, Gilbert Cette, and Rémy Lecat. 2016. "Productivity Trends in Advanced Countries Between 1890 and 2012." *Review of Income and Wealth*, 62(3): 420–444.

Bernanke, Ben, Mark Gertler, and Simon Gilchrist. 1996. "The Financial Accelerator and the Flight to Quality." *The Review of Economics and Statistics*, 78(1): 1–15.

Bernanke, Ben S, Mark Gertler, and Simon Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." *Handbook of macroeconomics*, 1: 1341–1393.

Bilbiie, Florin O, Fabio Ghironi, and Marc J Melitz. 2019. "Monopoly Power and Endogenous Product Variety: Distortions and Remedies." *American Economic Journal: Macroeconomics*, 11(4): 140–74.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb. 2020. "Are Ideas Getting Harder to Find?" *American Economic Review*, 110(4): 1104–44.

Borio, Claudio EV, and Andrew J Filardo. 2007. "Globalization and Inflation: New Cross-Country Evidence on the Global Determinants of Domestic Inflation."

Calligaris, Sara, Chiara Criscuolo, and Luca Marcolin. 2018. "Mark-Ups in the Digital Era." OECD Science, Technology and Industry Working Papers 2018/10, OECD Publishing, Paris.

Calvino, Flavio, Chiara Criscuolo, and Rudy Verlhac. 2020. "Declining Business Dynamism: Structural and Policy Determinants." OECD Science, Technology and Industry Policy Papers 94, OECD Publishing, Paris.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics*, 12(3): 383–398.

Cascaldi-Garcia, Danilo, and Marija Vukotić. 2022. "Patent-Based News Shocks." *The Review of Economics and Statistics*, 104(1): 51–66.

Chen, Kaiji, and Zheng Song. 2013. "Financial Frictions on Capital Allocation: A Transmission Mechanism of TFP Fluctuations." *Journal of Monetary Economics*, 60(6): 683–703.

Chen, Zhang, Zulfiqar Ali Wagan, and Hakimzadi Seelro. 2018. "New Evidence on the Robust Identification of News Shocks: Role of Revisions in Utilization-Adjusted TFP Series and Term Structure Data." *Journal of Forecasting*, 37(3): 352–370.

Christiano, Lawrence, Roberto Motto, Massimo Rostagno, and Cosmin Ilut. 2008. "Monetary Policy and Stock Market Boom-Bust Cycles." ECB Working Paper Series 955.

Cloyne, James, and Patrick Hürtgen. 2016. "The Macroeconomic Effects of Monetary Policy: A New Measure for the United Kingdom." *American Economic Journal: Macroeconomics*, 8(4): 75–102.

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

Cloyne, J., C. Ferreira, M. Froemel, and P. Surico. 2018. "Monetary Policy, Corporate Finance and Investment." NBER Working Paper 25366.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Is the Phillips Curve Alive and Well After All? Inflation Expectations and the Missing Disinflation." *American Economic Journal: Macroeconomics*, 7(1): 197–232.

Dedola, Luca, Giulia Rivolta, and Livio Stracca. 2017. "If the Fed Sneezes, Who Catches a Cold?" *Journal of International Economics*, 108: S23–S41.

Del Negro, Marco, Michele Lenza, Giorgio E Primiceri, and Andrea Tambalotti. 2020. "What's up With the Phillips Curve?" NBER Working Paper 27003.

De Loecker, Jan, and Jan Eeckhout. 2018. "Global Market Power." NBER Working Paper 24768.

Dixit, Avinash K, and Joseph E Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *The American Economic Review*, 67(3): 297–308.

Faust, Jon. 1998. “The Robustness of Identified VAR Conclusions about Money.” *Carnegie-Rochester Conference Series on Public Policy*, 49: 207 – 244.

Feenstra, Robert C. 2003. “A Homothetic Utility Function for Monopolistic Competition Models, Without Constant Price Elasticity.” *Economics Letters*, 78(1): 79–86.

Fernald, John. 2014*a*. “A Quarterly, Utilization-Adjusted Series on Total Factor Productivity.” *Federal Reserve Bank of San Francisco Working Paper*, , (2012-19).

Fernald, John G. 2014*b*. “Productivity and Potential Output Before, During, and After the Great Recession.” NBER Working Paper 20248.

Francis, Neville, Michael T. Owyang, and Jennifer E. Roush. 2007. “A Flexible Finite-Horizon Identification of Technology Shocks.” Federal Reserve Bank of St. Louis Working Paper No. 2005-024D.

Friedman, Milton. 1968. “The Role of Monetary Policy.” *American Economic Review*, 58(1): 1–17.

Friedman, Milton. 1969. “The Monetary Studies of the National Bureau, 44th Annual Report.” *The Optimal Quantity of Money and Other Essays*, 261–284.

Fujiwara, Ippei, and Kiminori Matsuyama. 2022. “Competition and the Phillips Curve.” CEPR Press Discussion Paper 17521.

Gabriel, Ricardo Duque. 2021. “Monetary Policy and the Wage Inflation-Unemployment Tradeoff.” *Available at SSRN 3689791.*

Gai, Prasanna, and Eric Tong. 2022. “Information Spillovers of US Monetary Policy.” *Journal of Macroeconomics*, 72: 103401.

Galí, Jordi. 2015. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications.* Princeton University Press.

Galí, Jordi, and Mark Gertler. 1999. “Inflation Dynamics: A Structural Econometric Analysis.” *Journal of Monetary Economics*, 44(2): 195–222.

Georgiadis, Georgios. 2016. “Determinants of Global Spillovers From US Monetary Policy.” *Journal of international Money and Finance*, 67: 41–61.

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

Global Economic Prospects. June, 2020. *Global Economic Prospects.* The World Bank.

Gorodnichenko, Yuriy, and Byoungchan Lee. 2020. “Forecast Error Variance Decompositions With Local Projections.” *Journal of Business & Economic Statistics*, 38(4): 921–933.

Görtz, Christoph, Dimitris Korobilis, John D. Tsoukalas, and Francesco Zanetti. 2019. "The Effect of News Shocks and Monetary Policy." CESifo Working Paper No. 7578.

Görtz, Christoph, John D Tsoukalas, and Francesco Zanetti. 2022. "News Shocks Under Financial Frictions." *American Economic Journal: Macroeconomics*, 14(4): 210–43.

Hadlock, Charles J., and Joshua R. Pierce. 2010. "New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index." *The Review of Financial Studies*, 23(5): 1909 – 1940.

Hansen, Lars Peter, and Kenneth J Singleton. 1982. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica: Journal of the Econometric Society*, 1269–1286.

Hazell, Jonathon, Juan Herreno, Emi Nakamura, and Jón Steinsson. 2022. "The Slope of the Phillips Curve: Evidence From Us States." *The Quarterly Journal of Economics*, 137(3): 1299–1344.

Heise, Sebastian, Fatih Karahan, and Ayşegül Şahin. 2022. "The Missing Inflation Puzzle: The Role of the Wage-Price Pass-Through." *Journal of Money, Credit and Banking*, 54(S1): 7–51.

Hooper, Peter, Frederic S Mishkin, and Amir Sufi. 2020. "Prospects for Inflation in a High Pressure Economy: Is the Phillips Curve Dead or Is It Just Hibernating?" *Research in Economics*, 74(1): 26–62.

Jaimovich, Nir, and Sergio Rebelo. 2009. "Can News about the Future Drive the Business Cycle?" *American Economic Review*, 99(4): 1097–1118.

Jeenas, Priit. 2019. "Firm Balance Sheet Liquidity, Monetary Policy Shocks, and Investment Dynamics." *Working Paper, New York University*.

Jermann, Urban, and Vincenzo Quadrini. 2012. "Macroeconomic Effects of Financial Shocks." *The American Economic Review*, 102(1): 238–271.

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

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

Jordà, Òscar, Moritz Schularick, and Alan M Taylor. 2017. "Macrofinancial History and the New Business Cycle Facts." *NBER Macroeconomics Annual*, 31(1): 213–263.

Jordà, Òscar, Moritz Schularick, and Alan M Taylor. 2020*a*. "The Effects of Quasi-Random Monetary Experiments." *Journal of Monetary Economics*, 112: 22–40.

Jordà, Òscar, Sanjay R Singh, and Alan M Taylor. 2020*b*. "The Long-Run Effects of Monetary Policy." NBER Working Paper 26666.

Jorgensen, Peter Lihn, and Kevin J Lansing. 2022. "Anchored Inflation Expectations and the Slope of the Phillips Curve." Federal Reserve Bank of San Francisco.

Kahn, James A, and Robert W Rich. 2007. "Tracking the New Economy: Using Growth Theory To Detect Changes in Trend Productivity." *Journal of Monetary Economics*, 54(6): 1670–1701.

Kamber, Gunes, Konstantinos Theodoridis, and Christoph Thoenissen. 2017. "News-Driven Business Cycles in Small Open Economies." *Journal of International Economics*, 105: 77–89.

Kaplan, Steven N, and Luigi Zingales. 1997. "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?" *The quarterly journal of economics*, 112(1): 169–215.

Kiyotaki, Nobuhiro, and John Moore. 1997. "Credit cycles." *Journal of political economy*, 105(2): 211–248.

Kobayashi, Keiichiro, Tomoyuki Nakajima, and Masaru Inaba. 2012. "Collateral Constraint and News-Driven Cycles." *Macroeconomic Dynamics*, 16(5): 752–776.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. "Technological Innovation, Resource Allocation, and Growth." *The Quarterly Journal of Economics*, 132(2): 665–712.

Koltay, Gabor, Szabolcs Lorincz, and Tommaso M Valletti. 2022. "Concentration and Competition: Evidence From Europe and Implications for Policy." CESifo Working Paper 9640.

Krusell, Per L., and Alisdair McKay. 2010. "News Shocks and Business Cycles." *FRB Richmond Economic Quarterly*, 96(4): 373–397.

Kung, Howard. 2015. "Macroeconomic Linkages Between Monetary Policy and the Term Structure of Interest Rates." *Journal of Financial Economics*, 115(1): 42–57.

Kurmann, André, and Christopher Otrok. 2013. "News Shocks and the Slope of the Term Structure of Interest Rates." *American Economic Review*, 103(6): 2612–2632.

Kurmann, André, and Eric Sims. 2021. "Revisions in Utilization-Adjusted Tfp and Robust Identification of News Shocks." *The Review of Economics and Statistics*, 103(2): 216–235.

Kuttner, Ken, and Tim Robinson. 2010. "Understanding the Flattening Phillips Curve." *The North American Journal of Economics and Finance*, 21(2): 110–125.

Kuttner, Kenneth N. 2001. "Monetary Policy Surprises and Interest Rates: Evidence From the Fed Funds Futures Market." *Journal of Monetary Economics*, 47(3): 523–544.

Lewis, Daniel J., and Karel Mertens. 2022. "Dynamic Identification Using System Projections and Instrumental Variables." CEPR Discussion Paper No. DP17153, Available at SSRN: https://ssrn.com/abstract=4121257.

Lucas, Robert Jr. 1972. "Expectations and the Neutrality of Money." *Journal of economic theory*, 4(2): 103–124.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H Stock. 2014. "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve." *Journal of Economic Literature*, 52(1): 124–88.

Miranda-Agrippino, Silvia, Sinem Hacioglu Hoke, and Kristina Bluwstein. 2019. "When Creativity Strikes: News Shocks and Business Cycle Fluctuations." Bank of England Working Paper No. 788.

Nakamura, Emi, and Jón Steinsson. 2013. "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications." *Annu. Rev. Econ.*, 5(1): 133–163.

Nakamura, Emi, and Jón Steinsson. 2018. "High-frequency identification of monetary non-neutrality: the information effect." *The Quarterly Journal of Economics*, 133(3): 1283–1330.

Oliner, Stephen D., and Glenn D. Rudebusch. 1996. "Monetary Policy and Credit Conditions: Evidence from the Composition of External Finance: Comment." *The American Economic Review*, 86(1): 300–309.

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

Phelps, Edmund S. 1967. "Phillips Curves, Expectations of Inflation and Optimal Unemployment Over Time." *Economica*, 254–281.

Phillips, Alban W. 1958. "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957." *economica*, 25(100): 283–299.

Pigou, Arthur C. 1927. *Industrial Fluctuations.* London: Macmillan.

Plagborg-Møller, Mikkel, and Christian K Wolf. 2022. "Instrumental Variable Identification of Dynamic Variance Decompositions." *Journal of Political Economy*, 130(8): 2164–2202.

Ramos-Francia, Manuel, and Alberto Torres. 2008. "Inflation Dynamics in Mexico: A Characterization Using the New Phillips Curve." *The North American Journal of Economics and Finance*, 19(3): 274–289.

Ratner, David, and Jae Sim. 2022. "Who Killed the Phillips Curve? A Murder Mystery." FEDS Working Paper No. 2022-28.

Razin, Assaf, and Alon Binyamini. 2007. "Flattened Inflation-Output Tradeoff and Enhanced Anti-Inflation Policy: Outcome of Globalization?" NBER Working Paper 13280.

Roberts, John M. 1995. "New Keynesian Economics and the Phillips Curve." *Journal of money, credit and banking*, 27(4): 975–984.

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

Rotemberg, Julio J. 1982. "Sticky Prices in the United States." *Journal of Political Economy*, 90(6): 1187–1211.

Samuelson, Paul A, and Robert M Solow. 1960. "Analytical Aspects of Anti-inflation Policy." *The American Economic Review*, 50(2): 177–194.

Silajdzic, Sabina, and Eldin Mehic. 2018. "Trade Openness and Economic Growth: Empirical Evidence From Transition Economies." In *Trade and Global Market*. IntechOpen.

Smets, Frank, and Rafael Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian Dsge Approach." *American economic review*, 97(3): 586–606.

Smulders, Sjak, and Theo Van de Klundert. 1995. "Imperfect Competition, Concentration and Growth With Firm-Specific R&D." *European Economic Review*, 39(1): 139–160.

Smulders, Sjak, and Theo Van de Klundert. 1997. "Growth, Competition and Welfare." *Scandinavian Journal of Economics*, 99(1): 99–118.

Stock, James H, and Mark W Watson. 2019. "Slack and Cyclically Sensitive Inflation." NBER Working Paper 25987.

Tenreyro, Silvana, and Gregory Thwaites. 2016. "Pushing on a String: US Monetary Policy Is Less Powerful in Recessions." *American Economic Journal: Macroeconomics*, 8(4): 43–74.

Uhlig, Harald. 2003. "What Moves GNP?" *Unpublished manuscript*, Euro Area Business Cycle Network.

Uhlig, Harald. 2004. "Do Technology Shock Lead to A Fall in Total Hours Worked?" *Journal of the European Economic Association*, 2(2-3): 361–371.

Walentin, Karl. 2014. "Expectation Driven Business Cycles with Limited Enforcement." *Economics Letters*, 124(2): 300–303.

Wang, Olivier, and Iván Werning. 2022. "Dynamic Oligopoly and Price Stickiness." *American Economic Review*, 112(8): 2815–49.

Wang, Wenjie, and Maximilien Kaffo. 2016. "Bootstrap Inference for Instrumental Variable Models With Many Weak Instruments." *Journal of Econometrics*, 192(1): 231–268.