

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Essays on the Heterogeneous Transmission of Macroeconomic Shocks

Permalink

<https://escholarship.org/uc/item/5f33j7s9>

ISBN

9798189225697

Author

Park, Hyunseo

Publication Date

2026-06-25

Peer reviewed|Thesis/dissertation

Essays on the Heterogeneous Transmission of Macroeconomic Shocks

By

HYUNSEO PARK
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

Òscar Jordà

Sanjay R. Singh

Xian Jiang

Committee in Charge

2026

Copyright © 2026 by

HYUNSEO PARK

All rights reserved.

Contents

Abstract	v
Acknowledgements	vi
1 How Has the Effect of Government Spending Changed? Evidence from High-Frequency Identified Shocks Since 1947	1
1.1 Introduction	2
1.2 Identification	8
1.2.1 Collecting Newspaper Articles	8
1.2.2 Quantifying News Shocks	13
1.3 Defense Spending News Shock Series	15
1.3.1 Shock Series	15
1.3.2 Diagnostics of News Shock Series	16
1.4 Econometric Methodology	21
1.4.1 Impulse Responses	21
1.4.2 Government Spending Multipliers	22
1.5 Economic Effects of Government Spending Over Time	22
1.5.1 Time-Varying Effects on GDP	22
1.5.2 Contributions of GDP Components to the Time-Varying Effects	25
1.6 Why Has the Effect of Government Spending Changed Over Time?	28
1.6.1 Procurement Share and Crowding-In Effect	28
1.6.2 Time Profile of Government Spending	33
1.7 Conclusion	36
2 Heterogeneous Effects of Public Procurement	37
2.1 Introduction	38
2.2 Data	42
2.2.1 Data Sources and Key Variables	42
2.2.2 Linking Procurement Records to Publicly Traded Firms	45
2.3 Empirical Framework	47
2.3.1 Contract Structure and the Timing of Information Arrival	47
2.3.2 Identifying Informative Procurement Contracts	49
2.3.3 Baseline Sample	52
2.3.4 Sources of Heterogeneity	52

2.3.5	Estimating Daily Stock Price Responses	54
2.3.6	Estimating Quarterly Investment Responses	55
2.4	Effects of Procurement on Stock Prices	56
2.4.1	Stock Market Responses by Contract Form	56
2.4.2	Stock Market Responses by Sector Specialization and Firm Sector	58
2.4.3	Stock Market Responses by Procurement Object	61
2.5	Effects of Procurement on Firm Investment	64
2.5.1	Investment Responses by Sector Specialization and Firm Sector	64
2.5.2	Investment Responses by Procurement Object	67
2.6	Conclusion	69
3	Demographics and Housing Booms	71
3.1	Introduction	72
3.2	Effects of Demographics on Housing Booms	76
3.2.1	Data	76
3.2.2	Large Housing Booms	76
3.2.3	Young-to-Old Population Ratio	78
3.2.4	Descriptive Evidence	79
3.2.5	Demographics and Large Housing Booms	79
3.2.6	Young Applicants and Mortgage-Market Participation During Booms	83
3.3	A Housing OLG Model	87
3.3.1	Model environment	88
3.3.2	Simulation design	90
3.3.3	Simulation results	92
3.4	Conclusion	94
	Appendices to Chapter 1	104
1.A	Evaluations of the defense news shock identification	105
1.A.1	Sentence-based text analysis model	105
1.A.2	Placebo test for the news dates	107
1.A.3	Background Noise	111
1.A.4	Details on the Topic Model Results	112
1.B	IRFs with confidence intervals	114
1.B.1	Government spending	114
1.B.2	GDP	115
1.B.3	Private Investment	116
1.B.4	Private Consumption	117
1.C	Robustness of the time-varying effects of government spending	118
1.C.1	Control Variables	118
1.C.2	Number of Lags	123
1.D	Robustness of the effects by type of government spending	128
1.E	Extra Empirical Analysis	130
1.E.1	Crowding-out Effects on Private Investment	130
1.E.2	Procurement Composition and Private Investment Crowding-In	135

1.E.3	Rolling Window Multipliers	137
1.E.4	Weak Effect of Government Spending in the Korean War Sample	138
1.E.5	State Dependent Multipliers	140
1.F	Theory: New Keynesian Model with Government Spending News	144
1.F.1	Model Description	144
1.F.2	Regime Selection	150
1.F.3	Role of the Time Profile of Government Spending	154
1.F.4	IRF Matching: Private Investment	155
1.F.5	IRF Matching: Private Consumption	160

Appendices to Chapter 2 162

2.A	Summary Statistics of Procurement Contracts	163
2.B	Balance Tables	165
2.C	Sensitivity analysis	169
2.C.1	Procurement Shock Denominator	169
2.C.2	Sector Specialization Criterion	170
2.D	Quarterly Investment Responses by Contract Form	171
2.E	Stock Price Responses to IDV Framework Awards	172

Appendices to Chapter 3 173

3.A	Additional Descriptive Evidence	174
3.B	Supplementary Logit Specifications	175
3.C	HMDA Additional Evidence	176
3.C.1	Pretrend	176
3.C.2	Alternative Boom Definitions and Sample Restrictions	178
3.C.3	Risk-Taking Outcomes Among Young Borrowers	179

Abstract

The essays show that the effects of macroeconomic shocks depend on the design of macroeconomic policy, demographic composition, and the margins along which firms and households adjust.

Chapter 1 studies how the effects of U.S. government spending have changed over time. I construct a defense spending news measure for 1947–2019 by combining sentence-level newspaper text analysis with daily stock returns of major defense contractors. Using this measure as an instrument for aggregate government spending, I show that output multipliers vary from about 0.8 to 1.5 across postwar episodes. This variation is driven mainly by private investment: procurement-intensive, persistent, and back-loaded spending plans crowd in investment more strongly and generate larger multipliers.

Chapter 2 examines the firm-level effects of U.S. federal procurement by linking transaction-level procurement contract records from 2000–2024 with daily stock prices and quarterly firm financial statements. We show that procurement awards differ in informational content. Awards that resolve uncertainty about both the winning firm and the scope of work generate significant stock price responses, while awards with a largely predetermined recipient elicit little market reaction. Among informative awards, contract content shapes both valuation and real adjustment: non-core, product, and R&D service contracts generate stronger valuation effects in stock market, while firm investments response depend on the productive capacity required for execution.

Chapter 3 studies demographics and housing-boom vulnerability. Using OECD country data and U.S. county data, we show that locations with higher young-to-old population ratios are more likely to experience large upward deviations of real house prices from trend. HMDA mortgage data further show that young applicants increase mortgage-market activity disproportionately during boom county-years. We rationalize these facts through a six-generation housing OLG model. Younger economies place more households near the active tenure-adjustment margin and therefore generate larger house-price responses to common housing-preference shocks.

Acknowledgements

I am deeply grateful to my advisor and committee chair, James Cloyne, for his guidance, support, and encouragement throughout my graduate studies. His advice has shaped not only this dissertation, but also the way I think about economic questions, empirical identification, and academic research more broadly. I am especially thankful for his patience, generosity, and constant encouragement during every stage of this project.

I am also indebted to my committee members, Òscar Jordà, Sanjay R. Singh, and Xian Jiang, for their thoughtful feedback and guidance. Their comments pushed me to clarify my arguments, strengthen the empirical analysis, and think more carefully about the broader contribution of my work. I have benefited greatly from their insights and from the intellectual environment they helped create.

I thank my coauthors and collaborators for making the research process both productive and enjoyable. I am grateful to James Cloyne for collaboration on the public procurement chapter, and to Yun Gu Cho, Jinho Kim, and Jaeyoung Yoo for collaboration on the demographics and housing booms chapter. Working with them taught me a great deal about how to develop ideas, solve empirical problems, and improve a paper through discussion.

Graduate school would have been much more difficult without the friends who shared this journey with me. I am thankful for the conversations, encouragement, and small moments of humor that made the long process of research and writing more manageable. Their support reminded me that academic life is not only about papers and deadlines, but also about community.

Finally, I owe my deepest gratitude to my family. Their love, trust, and support have been the foundation that made this dissertation possible. They encouraged me through moments of uncertainty and celebrated each small step forward. This dissertation is dedicated to them.

Chapter 1

How Has the Effect of Government Spending Changed? Evidence from High-Frequency Identified Shocks Since 1947

This paper investigates the time-varying effects of government spending on the U.S. economy. Using a novel identification strategy that combines text analysis with daily stock market data, I construct a new measure of defense spending news spanning 1947–2019, which serves as an instrument for aggregate government spending. The new measure provides richer variation than existing measures, allowing for more detailed examination of how the effects of government spending evolve over time. Empirical analysis shows that the output multiplier fluctuates substantially over time, *ranging from 0.8 to 1.5*, largely driven by differences in *private investment* responses. The key mechanism lies in the intrinsic nature of the spending—its *composition* and *time profile*. When a policy announcement signals a persistent and back-loaded expansion (time profile) of procurement spending (composition), private investment rises strongly, leading to larger output multipliers. The results highlight an important policy implication that the effectiveness of government spending is not static, and policymakers need to consider the underlying nature of spending plans, not just their size.

1.1 Introduction

As one of the primary tools of macroeconomic policy, government spending plays a central role in stabilizing the economy. However, the effectiveness of this tool is not uniform; its macroeconomic impact can vary across time and circumstances. This raises a key question: *How and why do the effects of government spending on the macroeconomy vary over time?* Understanding this issue has become increasingly important in recent years, particularly as public debt reaches historically high levels. With policymakers facing tighter fiscal constraints, it has become essential to identify under what conditions government spending has been most effective in the past. Nevertheless, empirical evidence on the magnitude and underlying sources of such time-varying effects remains limited.

A major difficulty in studying time-varying effects of government spending is finding an appropriate instrument for accurate estimation of the causal effects. A key advancement in the literature is the use of announcements (*news*) about *future* changes in defense spending (Ramey, 2011; Ben Zeev and Pappa, 2017; Ramey and Zubairy, 2018). This approach captures the anticipatory responses of economic agents to government spending plans announced well before implementation.¹ However, existing measures are not well suited for analyzing time variation: they heavily rely on announcements during a few rare episodes (e.g., World War II) and offer sparse coverage of peacetime policy changes. As a result, they lack the sufficient variation required to examine how fiscal effects evolve throughout the postwar period.

To fill this gap, I develop a new identification strategy that combines text analysis with daily stock market data. This approach yields a new measure of defense spending news for 1947–2019 that features rich variation in both wartime and peacetime. Using the measure as an instrument for aggregate government spending, I show that the effects of government spending have changed markedly over time, with output multipliers ranging from 0.8 to 1.5. I further demonstrate that this time variation is driven primarily by heterogeneous investment responses to differences in the

¹Because defense spending is largely driven by geopolitical events, it helps isolate policy variations that are exogenous to contemporaneous economic conditions.

intrinsic nature of government spending—its *composition* and *time profile*. Together, these findings highlight my central contribution: the effectiveness of government spending is not fixed; it depends on how the government plans to spend, not merely on how much it spends.

One challenge in measuring fiscal news is the complexity and irregularity of the policy-making process. Unlike monetary policy, which is announced on a predictable schedule following FOMC meetings and communicated in a relatively unified voice by the Federal Reserve, fiscal policy news is released only sporadically through multiple channels. In particular, defense spending news can originate from the presidential speeches, the annual budget process, or unexpected military events. As a result, assembling a comprehensive dataset of such news and pinpointing the exact moment when such information becomes public remains a challenging task.

I address this challenge by collecting comprehensive defense spending news items from a massive dataset of approximately 13 million newspaper articles. The data cover a wide range of defense spending announcements disseminated through various channels. Moreover, by exploiting the publication date of each article, I can precisely identify the precise timing of public disclosure.² To systematically analyze this large corpus, I employ a sentence-based text analysis model that divides the main text of each article into individual sentences and detects those related to defense spending. The model is computationally as efficient as a simple bag-of-words approach, yet achieves classification performance comparable to more sophisticated methods such as topic models and large language models (LLMs).

Another important challenge is to accurately quantify the intensity of defense spending news. Because of the complexity of the policy-making process, it is difficult to disentangle the genuinely new information contained in each announcement. For example, some announcements may merely reaffirm previously released budget plans rather than convey new information. Hence, rigorously assessing the extent to which each news item signals a meaningful change in expected future government spending is essential. To this end, I measure the intensity of defense spending news

²The news collection procedure employed in this study is highly flexible and can be extended to capture other types of policy news, including non-defense spending.

using the excess stock returns of major defense contractors within a narrow window around the announcement date. By capturing the financial market's immediate response, this measure provides a clean estimate of the expected change in defense spending implied by the news. If a policy announcement does not materially deviate from the status quo or has a low likelihood of implementation, the market reaction will likely be muted.

In the second part of the paper, I analyze the time-varying effects of government spending and investigate the sources of this variation.³ I find that the effects of government spending have changed markedly over time, with GDP multipliers ranging from 0.8 to 1.5. Among the expenditure components, time-varying responses of private investment primarily account for this variation. In response to positive government spending news, private investment exhibits both crowding-in and crowding-out, with their relative strength varying across periods. These findings are robust to alternative estimation methods.

This paper highlights the *intrinsic nature* of government spending as a key mechanism behind its time-varying effects. Specifically, I show that the crowding-in effect on private investment emerges only when a substantial portion of government spending is implemented through *procurement* from the private sector. By contrast, the effect is negligible or entirely absent during periods dominated by non-procurement spending (*composition*). Furthermore, this procurement-driven investment crowding-in channel is amplified when government spending plans are *persistent* and *back-loaded* (*time profile*). Taken together, when government spending features a high share of procurement and a persistent, back-loaded profile, private investment responds more strongly, leading to higher output multipliers.

The key policy implication of this study is that the macroeconomic impact of government spending is not a fixed constant, but varies significantly depending on its *composition* and *time profile*. In an environment of tightening fiscal constraints, effective “policy design,” characterized by

³The new measure of defense spending news is used as an instrument for future actual government spending. The measure captures exogenous variation in future spending, as they reflect information about the government's future defense plans. Moreover, given that defense spending accounts for a large share of total expenditure, they also have strong explanatory power for overall government spending.

close linkages with the private sector and persistent implementation from a long-term perspective, is more crucial than merely expanding the scale of spending. Such a design can more effectively stimulate private investment and overall output. These findings highlight the need to incorporate the timing and compositional design of fiscal policy as central variables in future policy frameworks.

Related Literature. This study is related to the empirical literature that examines the effects and transmission mechanisms of government spending, which can be broadly categorized into two streams: the Structural Vector Autoregression (SVAR) approach and the instrumental variable approach. The SVAR approach identifies government spending shocks by placing government spending as the first variable in a VAR model (Fatás and Mihov, 2001; Blanchard and Perotti, 2002; Galí et al., 2007; Monacelli and Perotti, 2008; Caldara and Kamps, 2008; Perotti et al., 2007). The instrumental variable approach uses an instrumental variable that is exogenous to GDP, such as defense-related news (Ramey and Shapiro, 1998; Edelberg et al., 1999; Burnside et al., 2004; Fisher and Peters, 2010; Ramey, 2011; Barro and Redlick, 2011; Ben Zeev and Pappa, 2017; Ramey and Zubairy, 2018; Ben Zeev, Ramey, et al., 2023; Brunet, 2022; Briganti, 2023).

This paper follows the instrumental approach and, in particular, contributes to substantially increasing the number of shocks while retaining the timing advantages of Ramey (2011)'s method. By identifying a sufficient number of shocks, this study addresses the weak instrumental variable (IV) issue related to post-Korean War period samples in existing research. Recent studies introduce alternative instrumental variables to address the weak IV issue. For example, Briganti (2023) proposes the use of series of quarterly defense contracts and Brunet (2022) suggests the use of annual defense budget authority as instrumental variables. Compared to existing instruments, the new measure developed in this paper offers an advantage in precisely identifying the timing and magnitude of defense spending news. The novel methodology of this study uses daily newspaper articles and financial market data to accurately identify only newly updated information related to policy variables at precise timing.

Several studies employ applied microeconomic identification methods to estimate government

spending multipliers using subnational-level data from recent samples. These studies provide valuable insights into the effects of government spending at the subnational level (Nakamura and Steinsson, 2014; Farhi and Werning, 2016; Chodorow-Reich, 2019). However, as noted in Ramey (2019), subnational multipliers are not equivalent to aggregate multipliers. The relationship between subnational and aggregate multipliers depends on factors such as the financing method and regional spillovers. This study contributes by proposing a new method to estimate the aggregate government spending multiplier.

This study is also related to the literature on state-dependent fiscal multipliers, as it investigates the heterogeneity in the effects of government spending. Empirical studies in this area show that fiscal multipliers are influenced by factors such as economic slack or monetary policy (Auerbach and Gorodnichenko, 2012; Owyang et al., 2013; Ramey and Zubairy, 2018; Cloyne et al., 2023), the source of business cycle fluctuations (Ghassibe and Zanetti, 2022), and the sign of government spending shocks (Hooker and Knetter, 1997; Davis et al., 1997; Barnichon et al., 2022; Ben Zeev, Ramey, et al., 2023). This study differs from prior work in that, rather than comparing the effects of government spending across states, it compares the transmission mechanisms themselves across periods. In addition, this paper contributes to the literature by interpreting the variation in fiscal multipliers through another lens of the intrinsic nature of government spending.⁴

The methodology of this study is closely related to the high-frequency identification approach, which was initially developed in the context of monetary policy (Cook and Hahn, 1989; Kuttner, 2001; Cochrane and Piazzesi, 2002; Gürkaynak et al., 2004; Nakamura and Steinsson, 2018). More recently, this approach has been extended to other contexts, such as the global oil market (Känzig, 2021), financial shocks (Ottonello and Song, 2025), and carbon policy (Känzig, 2023). Several studies have used daily data to construct instrumental variables for defense spending (Auerbach and Gorodnichenko, 2016; Choi et al., 2022; Dissanayake, 2022), but applications of high-frequency identification to fiscal policy remain limited. This paper fills this gap by introducing a novel

⁴Appendix 1.E.5 also presents comparative results for state-dependent multipliers without relying on World War I, World War II, or the Korean War.

sentence-based text analysis model within a high-frequency identification framework. The approach is broadly applicable, requiring no prior knowledge of the exact timing of policy announcements, as it infers such information directly from newspaper publications. This feature makes high-frequency identification readily applicable to a wider range of policy contexts.

A growing body of research applies high-frequency identification to the fiscal context. The study most methodologically related to this paper is McClure and Yding (2025), who construct defense spending news shocks using high-frequency identification by collecting policy announcement dates from official congressional records related to the defense budget process. In addition, Wiegand (2025) identify high-frequency fiscal deficit shocks by tracking changes in forward-looking deficit targets announced during the congressional budget resolution process, thereby determining the timing and magnitude of deficit news received by investors.

Despite the methodological similarity in identification, this study has a contribution in the way of collecting news. Rather than relying solely on a limited set of official documents, this paper employs a comprehensive review of newspaper articles. This approach allows for capturing a wider range of news, including military events, congressional budget approvals, speeches, and other defense-related news. Furthermore, Newspapers have an advantage in evaluating market reactions to news. Because newspaper articles are widely distributed, they attract significantly higher public attention. As demonstrated by Barber and Odean (2008), public attention plays a crucial role in shaping stock buying behavior, suggesting that policy-related information reported in newspaper articles is more likely to be reflected in stock prices.

Roadmap. The remainder of the paper proceeds as follows. Section 1.2 details the identification of defense spending news shocks. Section 1.3 introduces and evaluates the resulting defense spending news shock series. Section 1.4 outlines the econometric methodology used in the empirical analysis. Section 1.5 presents the macroeconomic effects of government spending over time. Section 1.6 explores why the effects of government spending change over time. Section 1.7 concludes.

1.2 Identification

Following the fiscal literature employing an instrumental-variables approach, this study focuses on defense spending as a key component of discretionary outlays. Defense spending serves as an appealing proxy for several reasons. First, its dynamics are largely driven by geopolitical events, making it plausibly exogenous to economic conditions. Second, it accounts for the largest share of discretionary federal spending and therefore captures much of the variation in overall government spending. Third, defense programs are typically large in scale and highly salient, making them more likely to be covered in major newspapers, which facilitates the identification and collection of policy announcement dates.

This study identifies defense spending news shocks using a high-frequency identification (HFI) approach guided by two key questions: (i) When is policy news released? and (ii) How does the market respond to each announcement? The next section details how defense-related news is collected and converted into a time series of news shocks.

1.2.1 Collecting Newspaper Articles

To collect defense spending news, it is important to understand how the defense budget is determined. In general, the U.S. federal budget process begins with the President's budget request for the next fiscal year in January–February. Congress then reviews, negotiates, and amends the budget, with the goal of passing appropriations bills by September 30. The specific figures proposed by the President and ultimately enacted by Congress are publicly available through official sources such as the Budget of the U.S. Government and congressional records. However, even before the President's budget request is formally submitted, internal processes within administration may be revealed through various unofficial channels. Moreover, the defense budget is shaped not only by internal administrative procedures but also by external forces such as geopolitical developments, which may trigger abrupt changes in budget planning. In light of this, it is crucial to monitor a wide range of sources to comprehensively capture all relevant developments.

In high-frequency identification, it is more important to collect all relevant news as comprehensively as possible from both official and unofficial sources. In particular, missing the earliest news can be very costly. For example, suppose the President proposes a budget and Congress later approves it without major changes. In this case, on the day of congressional approval, only little new information is revealed because the key details were already disclosed in the President’s proposal. Thus, if we miss the President’s announcement, we lose a significant portion of the information relevant to that budget decision.

Considering this importance of comprehensive news collection, this study collects defense spending news articles using an extensive ProQuest newspaper article database, consisting of approximately 13 million entries. The database includes all newspaper articles published by The New York Times, The Wall Street Journal, and The Washington Post from 1947 to 2019. To systematically review large-scale newspaper articles, I conduct sentence-based text analysis using TDM Studio. TDM Studio is a text and data mining platform that enables researchers to perform text analysis using large volumes of data from sources such as newspapers, journals, and dissertations in ProQuest’s database.

Sentence-based text analysis model for news identification

I develop a sentence-based text analysis model to collect news articles containing information about changes in defense spending plans. Compared to alternative approaches such as Bag-of-Words (BoW) frequency analysis, topic modeling, and large language models (LLMs), the sentence-based model achieves higher classification accuracy and greater computational efficiency. While BoW frequency analysis and topic modeling are effective at capturing broad themes such as “defense,”⁵ they fall short in distinguishing articles that report on future changes in defense spending from those that merely discuss defense-related issues. In contrast, the sentence-based approach searches for sentences that simultaneously satisfy conditions related to both “defense” and “changes,” thereby effectively identifying articles that signal future changes in defense spending. Although large

⁵For example, Baker et al. (2016) employs a BoW approach to capture policy uncertainty.

language models (LLMs) could achieve similar objectives, their computational cost is prohibitively high for this study, which analyzes the entire corpus of daily newspaper articles published since 1947—over 13 million in total. The sentence-based model, by contrast, employs a simple algorithm that processes each article independently, allowing for highly efficient large-scale text analysis.

The sentence-based text analysis model is designed to detect sentences that simultaneously contain defense-related keywords and references to future directions or specific numerical figures. By aggregating these sentence-level signals at the document level, the model effectively identifies articles discussing changes in defense spending while minimizing computational costs. The detailed procedure of the model is as follows.

First, I check whether the title of each article contains at least one word potentially related to defense spending from the set {“*defense*”, “*military*”, “*pentagon*”, “*arms*”, “*weapon*”, “*war*”, “*budget*”}.⁶ Including “*budget*” or “*war*” may pull in articles that are not directly related to defense budgets. However, since general budget pieces or articles covering geopolitical events such as wars may also mention defense spending, I keep the title search criteria broad but filter based on the main text, retaining only those articles genuinely related to defense budgets.

Next, I check whether the main text of each article contains any sentences related to changes in defense spending. Specifically, I split the main text into individual sentences,⁷ and perform a preprocessing procedure on each sentence.⁸ Then, I count the number of sentences in each article that satisfy both of the following conditions:

Condition 1: The sentence contains at least one defense-related word from the set {“*defense*”, “*military*”, “*pentagon*”, “*arms*”, “*weapon*”}.⁹

Condition 2: The sentence or the following sentence contains at least one quantity- or direction-

⁶I split each title into individual words, convert them to lowercase, and then apply stemming.

⁷I use the spaCy library with the pre-trained English model *en_core_web_sm* and the *sentencizer* pipeline to perform sentence segmentation.

⁸I remove all special characters except for the dollar sign (\$), convert all letters to lowercase, and apply stemming to each word.

⁹Since I use broad topics potentially related to defense spending in the article titles, I restrict the set of relevant words in the main text to those referring to the spending agents or spending targets of defense expenditure.

related word from the set {“billion”, “million”, “trillion”, “dollars”, “\$”, “increase”, “decrease”, “rise”, “cut”, “veto”}.

Through this procedure, the sentence-based model systematically identifies news articles that discuss changes in defense spending within a vast corpus of newspaper data. By distinguishing between articles that merely cover general “defense” topics and those that explicitly address changes in defense expenditure, the model substantially improves classification accuracy. Moreover, because each document is processed independently, the model can efficiently analyze the entire dataset within a reasonable amount of time. This combination of efficiency and accuracy provides a crucial foundation for constructing a comprehensive list of news release dates used in high-frequency identification.

Refining the sample articles through manual validation

Using the sentence-based text analysis model, the total number of news articles was reduced from 13,363,573 to 59,811. However, since the model simply checks whether each article contains at least one sentence that satisfies predefined conditions, an additional verification step is essential to ensure the validity of the selected articles. I therefore manually reviewed all articles identified through text analysis and selected 606 that could plausibly influence expectations regarding future defense spending. The excluded articles generally fall into two main categories.

The first category includes articles that are not related to defense spending at all. For example, the following article is about a fundraising campaign led by a lawyer for a defense appeal, so it has nothing to do with defense spending. However, this article is selected by text analysis because its title includes the word “defense”, and its text contains a sentence with both “defense” and “\$”, meeting both condition 1 and 2.

*“Lawyer heads drive for **defense** appeal,” The New York Times, April 8, 1956:*

*The Joint **Defense** Appeal will start a fund-raising campaign for \$5,600,000 at dinner in the Waldorf-Astoria Hotel Thursday night. (. . .)*

The second category includes cases where identical information is reported multiple times across different articles. For example, several articles may cover the same defense budget amount approved by the same entity.¹⁰ Because such information is already incorporated into market expectations at the time of the initial report, market reactions to subsequent articles are more likely driven by unrelated background factors rather than by the news itself. Therefore, I select only the first article that contains genuinely new information.

Supplementary search for missing articles

Text analysis is a powerful tool for systematically collecting a broad range of defense-related articles over extended periods, as it can include diverse defense-related terminology. However, articles dealing with topics that are relevant only during specific periods can be omitted, leading to false negatives in which relevant articles are mistakenly excluded during the analysis. For example, the following article discusses U.S. military support during the Korean War, a topic that could substantially influence expectations for future defense spending. Nevertheless, this article is excluded because it does not contain expressions that meet the criteria defined in the text analysis.

“Legislators hail action by Truman: A solemn trio on the way to the white house,” The New York Times, June 28, 1950:

President Truman’s announcement today that United States air and sea power would be employed to expel the Communist invaders from South Korea evoked almost unanimous support in Congress.
(...)

To mitigate these errors, I conduct an additional search process that includes key military-related events, defense programs, and the names of presidents specific to each period. As a result, I collect an additional 253 articles that are highly relevant to the defense budget but were overlooked in the text analysis, increasing the total number of selected articles to 859.

¹⁰I treat approvals by defense budget committees as equivalent to approvals by the full chamber. For instance, if a defense budget already approved by the House or Senate Armed Services Committee is subsequently approved by the full House or Senate, articles on the latter approval are excluded unless the approved amount has changed.

Properties of the selected news dates

To ensure the reliability of estimates within narrow subsamples, it is important to collect a sufficient number of news articles across the full sample period. Figure 1.1 shows the annual counts of publication dates for defense-related articles. By applying text analysis to a large dataset, a sufficient number of newspaper articles are collected, ensuring comprehensive coverage across the entire time series. Each year consistently includes at least four news dates, with an average of 12 dates annually. Furthermore, the number of dates significantly increases following major military events such as the Korean War, the Vietnam War, and 9/11.

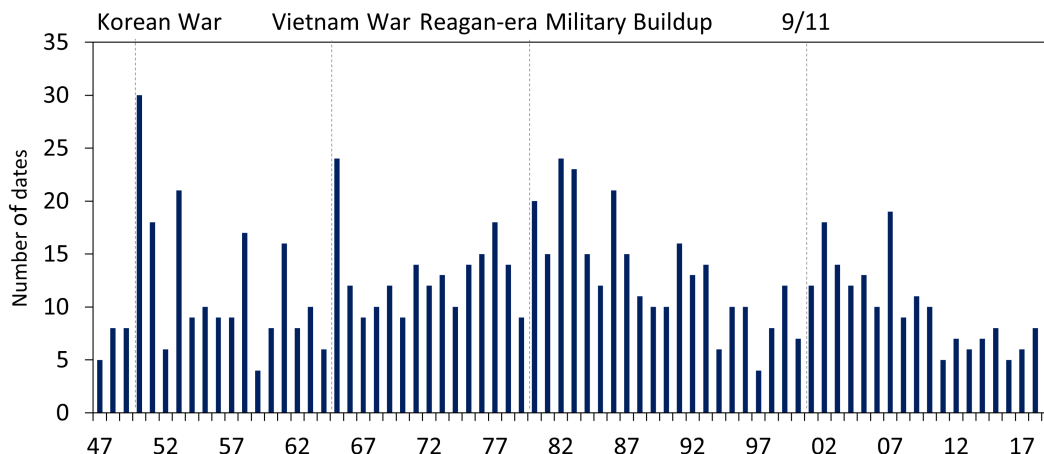


Figure 1.1: Annual count of defense spending news dates

1.2.2 Quantifying News Shocks

Excess returns of major defense contractors

The next step involves converting the collected news articles into measurable news shocks. Building on Fisher and Peters (2010), I use the excess returns of four major defense contractors relative to the overall market return.¹¹ As shown in Equation 1.1, Defense Spending News Shock (hereafter, *DNS*) is defined as the excess return on news dates and zero otherwise. The excess return is computed as a

¹¹Fisher and Peters (2010) show that the sales and excess returns of major defense contractors are highly correlated with defense spending. While they also use the excess returns of the top three defense contractors, they do not apply a high-frequency identification approach, which distinguishes the identification strategy from their paper.

market capitalization-weighted average ($\omega_{k,t}$) of each firm k 's return in excess of the market. The four primary defense contractors are Lockheed Martin, RTX Corporation, Northrop Grumman¹², and General Dynamics.¹³ The overall market return is calculated as the value-weighted sum of the total returns of all publicly traded firms. Stock return data are obtained from the Center for Research in Security Prices (CRSP).

$$DNS_t = I_t^{\text{news}} \sum_{k=1}^4 \omega_{k,t} \times (\text{Stock Return}_{k,t} - \text{Market Return}_t) \quad (1.1)$$

$$I_t^{\text{news}} = \begin{cases} 1, & \text{if defense spending news on } t, \\ 0, & \text{otherwise.} \end{cases}$$

The relationship between defense spending news shocks and expected future defense spending can be expressed as shown in Equation 1.2. Specifically, the DNS is functionally related to the expected change in defense spending implied by information contained in exogenous news. It is important to note that this requires an additional assumption that the function F is linear. Unlike high-frequency identification (HFI) approaches in other contexts, such as monetary policy, which utilize market indicators that directly represent policy variables, there is uncertainty about the functional relationship between stock excess returns and defense spending.¹⁴

$$E \left(\sum_{i=0}^H \frac{\Delta d_{t+i}}{(1+i_t)^i} \middle| \Omega_t^{\text{News}} \cap \Omega_t^{\text{Exogenous}} \right) = F(DNS_t) + u_t \quad (1.2)$$

Where Ω is the information set, u_t is the measurement error, i_t is the interest rate, d_t is defense spending, and H is the forecast horizon.

¹²Northrop Grumman is not included for the period before 1952, as its stock price data is only available from 1952 onward.

¹³As of 2024, the defense revenue of these firms is \$185.0 billion, accounting for 28.0% of the top 100 contractors' defense revenue and 17.1% of the total national defense expenditure. Moreover, these companies have consistently ranked among the top five defense contractors since 1958, when the list of the top 100 defense contractors was first published, continuing to hold these positions to the present day.

¹⁴The potential nonlinearity of this relationship will be explored more thoroughly in future work.

Window size

Another critical choice in high-frequency identification involves determining the size of the window around the news dates. In the literature, the window size typically ranges from 30 minutes to several days. I select a two-day window that includes both the publication date and the preceding day. This choice is justified, as most of the collected articles report news that occurred on the day prior to publication. Consequently, the information can be disclosed to the market at any time between the occurrence of the news and the publication of the article, allowing the stock market to respond on both days.¹⁵

1.3 Defense Spending News Shock Series

1.3.1 Shock Series

To facilitate analyses with various macroeconomic datasets, I aggregate the DNS to a quarterly frequency. Figure 1.2 shows that the identified defense spending news shocks are sufficiently frequent across the entire sample period. The largest positive shock occurs during the Korean War, and a substantial number of shocks are also captured around major episodes such as the Vietnam War, Reagan-era military buildup, and the 9/11 attacks. This enables a comparative analysis of the effects of government spending across a range of subsamples, which has been difficult in previous studies.

¹⁵Even if the actual announcement date of the news is a day before the publication date, it can still affect the stock price on the following day because it may take time for the new information to be reflected in the stock prices. As a reference, Bianchi et al. (2024) show that it takes several days for social media news to be reflected in the stock market.

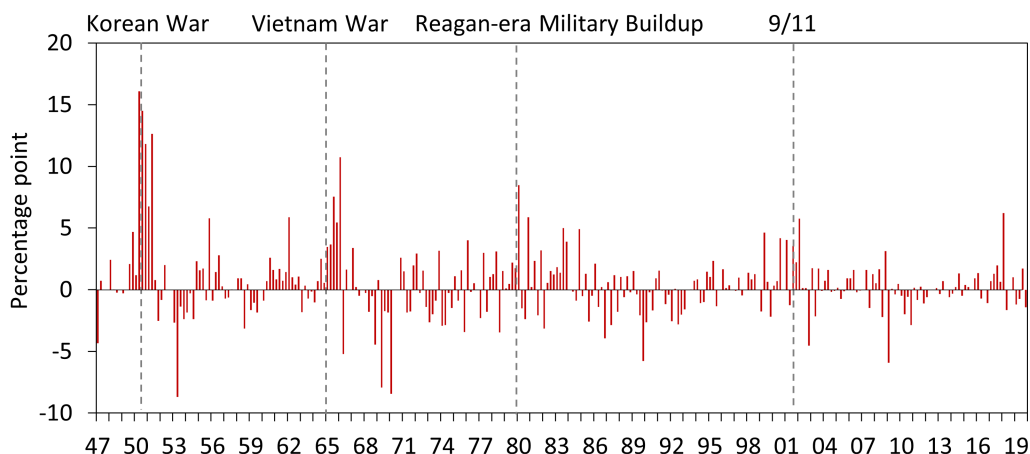


Figure 1.2: Defense Spending News Shock Series

1.3.2 Diagnostics of News Shock Series

Explanatory Power

When using DNS as an instrument to study the time-varying effects of government spending, a key consideration is whether DNS provides sufficient explanatory power for government spending across different subsamples. To assess this, I examine whether government spending responses significantly in response to DNS across various subsamples. In particular, to check whether the results are overly driven by a few major episodes, I progressively exclude key episodes one by one and evaluate whether the model maintains its explanatory power. Specifically, I analyze the full sample (1947–2019), the Post-Korean War period (1954–2019), the Post-Vietnam War period (1975–2019), and the Post-Reagan era period (1990–2019). As shown in Figure 1.3, the responses of government spending to DNS are significantly positive across all sample periods.

Another way to assess the explanatory power of DNS is to examine the first-stage F-values from the instrumental-variable (IV) estimation of the cumulative multiplier. The cumulative multiplier quantifies the cumulative response of output to a cumulative change in government spending over a given horizon. In this framework, DNS serves as the instrument for government spending, and the F-value at each horizon indicates how well the news shock at time t explains the cumulative

change in government spending up to time $t + h$. Given the potential serial correlation inherent in the local projections method, I use the effective F-statistics proposed by Olea and Pflueger (2013), which are robust to the serial correlation.¹⁶ As shown in Figure 1.4, the peak F-values in all four samples exceed the 5 percent critical threshold of 23.1, indicating that DNS has sufficiently strong explanatory power for government spending.

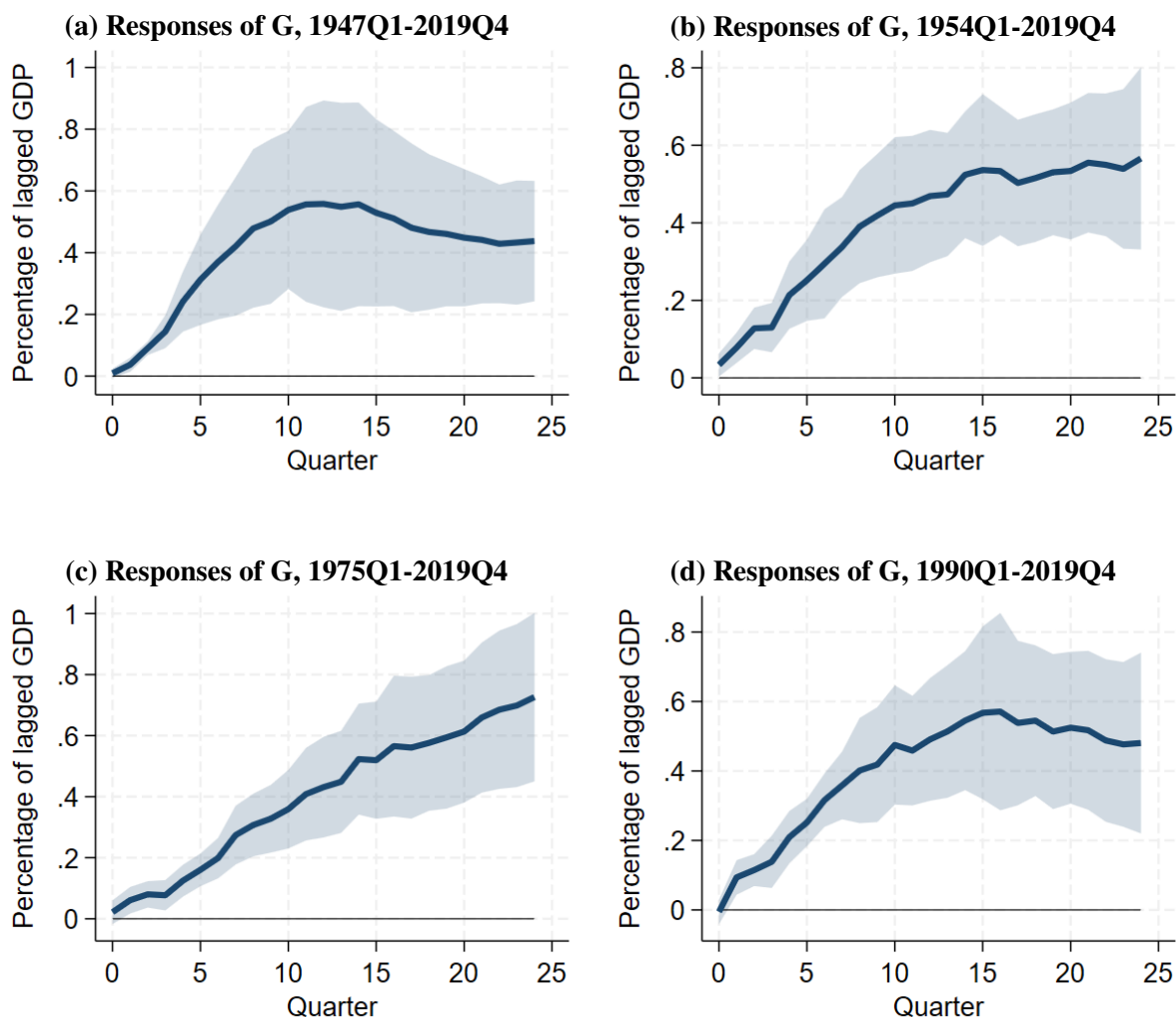


Figure 1.3: Responses of Government Spending

Notes: The figures show the response of government spending to DNS. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded area represent the 90% confidence intervals.

¹⁶The effective F-statistics are estimated using the method of Pflueger and Wang (2015), implemented via the Stata command “weakivtest.” The 5 percent critical value for testing the null hypothesis that the two-stage least squares (2SLS) bias exceeds 10 percent of the ordinary least squares (OLS) bias is 23.1.

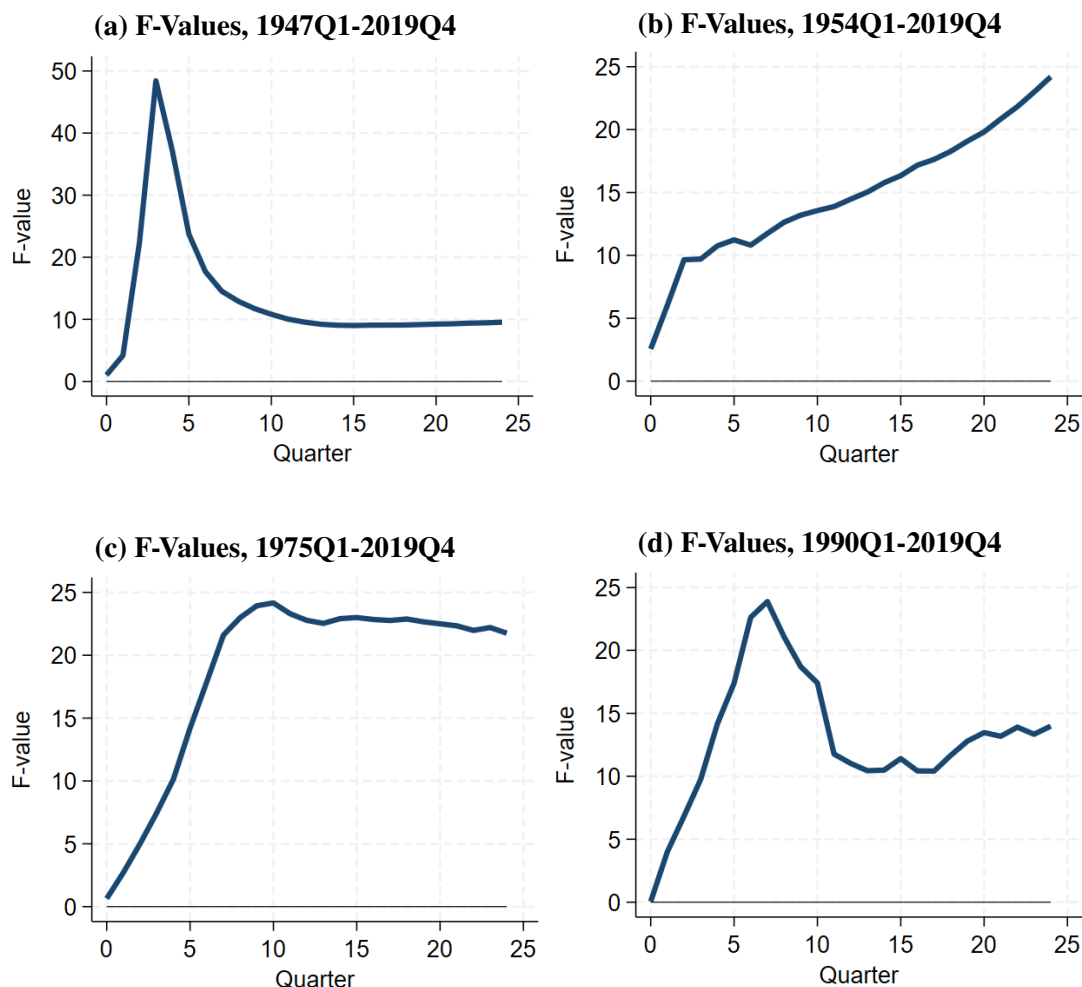


Figure 1.4: First stage F-values

Notes: The figures show the first stage F-values when estimating the government spending multiplier using DNS as an instrumental variable. The 5 percent critical value for testing the null hypothesis that the two-stage least squares bias exceeds 10 percent of the ordinary least squares bias is 23.1.

Exogeneity

It is widely recognized that HFI ensures the exogeneity of shocks by focusing exclusively on market reactions within a very narrow window around each policy event, effectively isolating the impact of these events from other concurrent economic conditions. DNS is also expected to be exogenous to other macroeconomic variables. To support this claim, I conduct Granger causality tests using four macroeconomic variables across various samples. For variables related to the business cycle, I use the total hours worked, GDP, and CPI inflation. For the variable related to financing, I use

deficit-to-GDP. As illustrated in Table 1.1, none of these four variables Granger-cause the defense spending news shocks in various samples.

Table 1.1: Granger Causality Tests (P-value)

Predictor	Predicted	1947Q1-2019Q4	1954Q1-2019Q4	1975Q1-2019Q4	1990Q1-2019Q4
GDP	DNS	0.34	0.50	0.22	0.48
Total hours worked	DNS	0.22	0.31	0.61	0.59
CPI inflation	DNS	0.85	0.57	0.62	0.90
Deficit/GDP	DNS	0.88	0.74	0.33	0.16
Joint	DNS	0.58	0.52	0.38	0.54

Notes: The table reports p-values from Granger causality tests, where the null hypothesis is that a given macroeconomic variable does not Granger-cause DNS. The number of lags is set to 4. The Global Financial Crisis period (2008–2010) is excluded. The early phase of the Korean War (Q2 1950 to Q3 1951) is also excluded to account for the effects of extreme shocks.

Timing of the shocks

As Ramey (2011) emphasizes, accurately capturing the timing of news is crucial in fiscal policy research to avoid anticipation effects. Because government budgets are typically announced several quarters before actual spending occurs, precisely identifying the timing of news is essential to capture both the announcement effects and the subsequent implementation effects of government spending. This study's identification strategy addresses this concern by pinpointing news at a daily frequency.

Figure 1.5 illustrates the correlations between DNS at time t and government spending at time $t+k$. For both defense spending and government spending, the correlation increase with k , peaking at $k = 4$. This pattern suggests that DNS effectively anticipates changes in both defense and government spending. To further evaluate the timing of DNS, I conduct Granger-causality tests to assess its predictive power for government spending. Table 1.2 shows that DNS Granger-causes both defense and government spending across various samples.

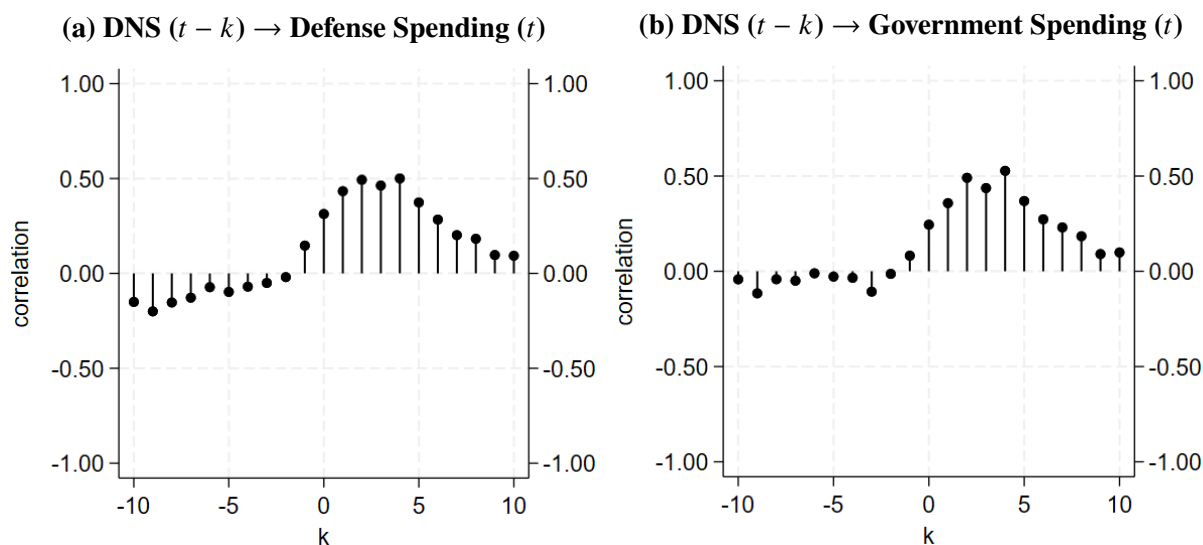


Figure 1.5: Correlation Between DNS and Government Spending

Notes: The left figure shows the correlations between DNS at time t and defense spending at time $t+k$. The right figure displays the correlations between DNS at time t and government spending at time $t+k$.

Table 1.2: Granger Causality Test (P-value)

Predictor	Perdicted	1947Q1-2019Q4	1954Q1-2019Q4	1975Q1-2019Q4	1990Q1-2019Q4
DNS	Defense spending	0.00	0.00	0.00	0.02
DNS	Government spending	0.00	0.00	0.01	0.01

Notes: The table reports p-values from Granger causality tests, where the null hypothesis is that DNS does not Granger-cause government or defense spending. The number of lags is set to 8 for the 1975Q1-2019Q4 sample, and 4 for the other samples.

1.4 Econometric Methodology

1.4.1 Impulse Responses

To estimate the impulse response functions, I employ the long-differences local projection (LP) method (Jordà (2005), Jordà and Taylor (2025)). LP has the advantage of analyzing the effects of news shocks. Depending on its nature, news can signal changes in government spending several years into the future, making the precise estimation of long-term responses essential. LP estimates responses for each horizon individually, allowing for a direct assessment of the effects at specific points in time. Also, it enables the efficient analysis of long-term effects without the need to include unnecessary control variables. In contrast, VAR requires sufficiently long lags for all variables in the system to estimate long-term relationships. This can lead to overparameterization of the model, which reduces degrees of freedom and lowers estimation efficiency.

Following Hall (2009) and Barro and Redlick (2011), I divide changes in real GDP, real government spending, and other national income and product accounts (NIPA) variables by lagged real GDP. This transformation allows the responses of government spending and GDP to be expressed in the same units, thereby allowing a direct calculation of the multiplier.¹⁷¹⁸ The baseline model looks as follows:

$$x_{t+h} - x_{t-1} = \alpha_h + Z'_{t-1} \phi_h + \beta_h DNS_t + \epsilon_{t+h} \quad (1.3)$$
$$h = 0, 1, \dots, H$$

where x denotes variable of interest, α_h is a constant term, Z represents a vector of control variables, and h is the projection horizon. Control variables include four lags of the differenced values of

¹⁷In contrast, when we use logarithmic transformations, the ratio of estimated responses corresponds to elasticities. Consequently, converting these elasticities into a multiplier requires an ad hoc conversion factor based on the sample average of the GDP-to-government-spending ratio. This widely used method can lead to biases in multiplier estimates. (Owyang et al., 2013)

¹⁸As an alternative, Ramey and Zubairy (2018) employed the Gordon and Krenn (2010) transformation, dividing the level variables by the GDP trend. However, I found that this approach is not suitable for the purpose of this study because multipliers estimated over short samples are sensitive to the choice of the sample period used to estimate the GDP trend.

variable of interest, real government spending, real GDP. Four lags of DNS are also included as a control variable to account for the possibility of autocorrelation. The Newey-West correction is used for standard errors. All NIPA variables are normalized using real GDP from the previous quarter.

1.4.2 Government Spending Multipliers

Since defense spending news is primarily about future changes in government spending, the multiplier should be estimated on a cumulative basis: it is calculated as the ratio of cumulative changes in real GDP to cumulative changes in real government spending. Cumulative multipliers can be estimated using the LP model in Equation 1.4, where DNS is used as an instrumental variable for the cumulative government spending term:

$$\sum_{j=0}^h \frac{x_{t+j} - x_{t-1}}{y_{t-1}} = \gamma_h + Z'_{t-1} \phi_h + m_h \sum_{j=0}^h \frac{g_{t+j} - g_{t-1}}{y_{t-1}} + \omega_{t+h} \quad (1.4)$$

$$h = 0, 1, \dots, H$$

where x denotes variable of interest, y is real GDP, g is real government spending, γ_h is a constant term, Z is a vector of control variables. h denotes the cumulative horizon, that is, the number of periods over which the responses are accumulated. Control variables include four lags of the differenced values of variable of interest, real government spending and real GDP. Four lags of DNS are also included as a control variable to account for the possibility of autocorrelation. The Newey-West correction is used for standard errors. All NIPA variables are normalized using real GDP from the previous quarter.

1.5 Economic Effects of Government Spending Over Time

1.5.1 Time-Varying Effects on GDP

To examine how the effects of government spending differ across periods, I divide the full sample into four subsamples, each corresponding to a major episode illustrated in Figure 1.2: the Korean

War (1947–1964), the Vietnam War (1954–1978), Reagan-era (1975–1999), and the War on Terror (1995–2019). Each subsample is constructed to maintain a sufficient length of 25 years, while avoiding overlap between major news shocks from different episodes. The only exception is the Korean War sample, which is shorter than 25 years to prevent overlap with news shocks related to the Vietnam War.¹⁹ In addition, to account for the sharp fluctuations in interest rates during the Paul Volcker period (1979–1987), lags of the interest rates are included as additional control variables.

To assess the effects of government spending, I first estimate the impulse response functions (IRFs) for government spending and GDP using Equation 1.3. One key consideration is how to make a one-unit news shock comparable across different samples. To address this, I rescale the shock so that the size of the government spending response predicted by a one-unit news shock is the same in all samples. Specifically, in each sample, a one-unit positive news shock is scaled to predict a cumulative increase in government spending over six years equivalent to 10% of GDP in period $t-1$.²⁰

Figure 1.6 shows that the effects of government spending on GDP vary considerably across samples. In the left panel, a positive news shock leads to a sustained increase in government spending in all cases. Although the degree of persistence differs somewhat, government spending rises for at least 12 quarters in each sample. In contrast, the right panel reveals substantial differences in the GDP response across samples, both in magnitude and in pattern. In the 1995–2019 sample, GDP reacts only modestly at first but increases steadily over time. In the 1975–1999 sample, GDP surges sharply within the first six quarters and rises again notably after 12 quarters. The other two samples exhibit more gradual increases, though the overall levels of response vary.

An alternative way to compare the effects of government spending is to estimate the GDP multiplier. Because spending news can predict changes in government spending well into the future, I estimate the cumulative multiplier (m_h) at each horizon h by running the IV regression

¹⁹The Appendix 1.E.3 shows that, when using a 25-year rolling window for the regression, the multiplier changes substantially across the four episodes. Excluding certain periods affected by the Global Financial Crisis, the multiplier remains at a stable level within the periods encompassing each episode.

²⁰Given that government spending news typically affects spending over a fairly long horizon, I consider a sufficiently long horizon of six years.

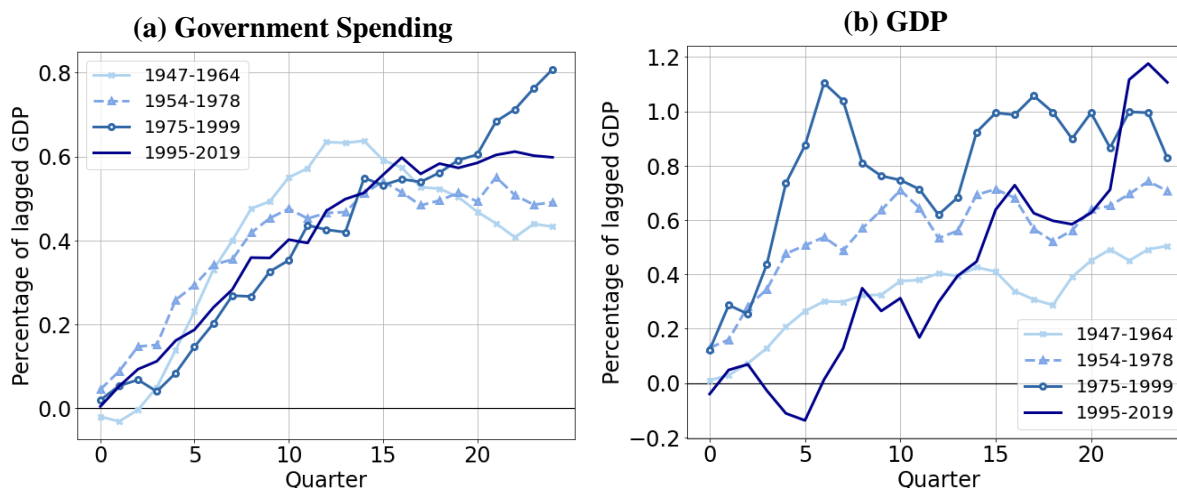


Figure 1.6: Responses of Government Spending and GDP

Notes: The figures show the response of macro variables to DNS by sample. The scale of news shock is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$.

in Equation 1.4, using DNS as an instrument for the cumulative change in government spending. Conceptually, the cumulative multiplier equals the ratio of the area under the GDP IRFs to that under the government spending IRFs in Figure 1.6.

Table 1.3 demonstrates substantial variation in the GDP multipliers across samples. The Reagan-era sample (1975–1999) exhibits the largest six-year cumulative multiplier(1.51), followed by the Vietnam War(1954–1978), the Korean War(1947–1964), and the War on Terror(1995–2019) periods, with values of 1.33, 0.82, and 0.81, respectively. Notably, the Reagan-era sample displays a much higher short-horizon multiplier, indicating a particularly strong early response. By contrast, the War on Terror period shows virtually no initial GDP reaction (a three-year multiplier of -0.00) and the lowest six-year multiplier overall. Moreover, GDP multipliers in this sample are statistically insignificant and thus unlikely to reflect an economically meaningful effect. These differences in the effects on GDP underscore the need to further investigate the underlying transmission mechanisms in each period.

Table 1.3: Cumulative GDP Multipliers Across Samples

	1947-1964	1954-1978	1975-1999	1995-2019
3-year horizon	0.87*** (0.06)	1.48*** (0.23)	2.31** (1.17)	-0.00 (2.03)
4-year horizon	0.79*** (0.05)	1.44*** (0.22)	1.79** (0.78)	0.33 (1.08)
5-year horizon	0.78*** (0.04)	1.40*** (0.23)	1.71*** (0.61)	0.45 (0.99)
6-year horizon	0.82*** (0.05)	1.33*** (0.15)	1.51*** (0.56)	0.81 (0.84)

Notes: This table shows the 3-year, 4-year, 5-year, and 6-year cumulative government spending multipliers for four sample periods: 1947Q1–1964Q4, 1954Q1–1978Q4, 1975Q1–1999Q4, and 1995Q1–2019Q4. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Standard errors are reported in parentheses.

1.5.2 Contributions of GDP Components to the Time-Varying Effects

How do different expenditure components of GDP contribute to the time-varying effects of government spending? To answer this, I first review the theoretical channels through which private consumption and investment are expected to respond. Private consumption is predicted to respond in opposite directions under the neoclassical and New Keynesian (NK) approaches. In the neoclassical framework, an increase in government spending can lead to a decline in private consumption through the wealth effect channel (Aiyagari et al. (1992); Baxter and King (1993)). In contrast, NK models with rule-of-thumb consumers suggest that government spending can increase consumption, owing to price stickiness and the presence of non-Ricardian households (Galí et al. (2007); Bilbiie (2020); Auclert et al. (2024)). On the other hand, most theoretical models predict a crowding-out effect on private investment due to rising real interest rates. Nonetheless, investment responds positively in some model specifications, depending on the policy regime. For example, Leeper et al. (2017) show that private investment can rise in a New Keynesian model under a regime characterized by active fiscal policy and passive monetary policy (Regime F).

In Figure 1.7, the responses of private investment exhibit considerable volatility, comparable to those of private consumption. In the left panel, the response of private fixed investment

differs markedly across periods. Contrary to the predictions of most theories that emphasize the crowding-out effect, the direction of the investment response is mixed. In the War on Terror sample (1995–2019), investment declines sharply over the 10 quarters following the news.²¹ In contrast, government spending has a positive impact on investment during the Vietnam War (1954–1978) and Reagan-era samples (1975–1999). Notably, the positive response of private investment during the Reagan-era sample is somewhat puzzling. This period coincides with the Volcker era, which is typically characterized as a regime of active monetary policy. An increase in private investment under such a regime is particularly difficult to reconcile with standard theoretical predictions.²²

In the right panel, the responses of private consumption also exhibit cross-sample variation. In all four samples, consumption rises, consistent with the standard predictions of New Keynesian (NK) models. In light of the NK framework, these differences are plausibly attributable to variation in the share of hand-to-mouth households across samples.

Figure 1.8 shows that cross-sample variations in private investment multipliers account for a substantial portion of the variation in the GDP multiplier. For each sample, I compute the deviation of the cumulative multiplier from the average across the four samples. The result is that the range of deviations in cumulative investment multipliers exceeds that of the GDP multipliers. This pattern is largely driven by the sizable negative values in the 1995–2019 subsample. Even when this period is excluded, the cross-sample variation in investment multipliers remains greater than or at least comparable to that of consumption multipliers through approximately year five.

Another finding is that the variation in private investment multipliers declines over longer horizons, whereas the variation of the private consumption multiplier rises. Figure 1.7 also supports this pattern, illustrating that consumption responds progressively as government spending increases, while investment reacts quickly at shorter horizons. These results indicate that the forward-looking

²¹In the War on Terror sample (1995–2019), the negative response of private investment is statistically significant, unlike the GDP response. Specifically, it is broadly significant at the 90% confidence level through the eleventh quarter. By contrast, the positive response of private consumption in this sample exhibits wide confidence intervals and is generally not statistically significant.

²²Even in Leeper et al. (2017), government spending predicts a clearly negative investment response under a regime characterized by active monetary policy and passive fiscal policy (Regime M).

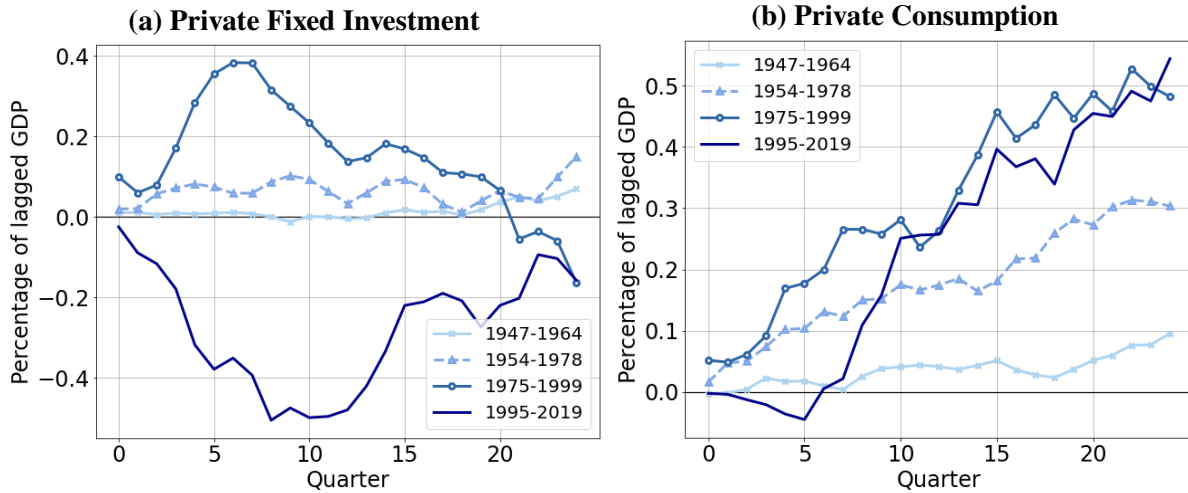


Figure 1.7: Responses of Macro Variables

Notes: The figures show the response of to DNS by sample. The left panel shows the response of private investment, while the right panel shows the response of private consumption (non-durables + services). The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$.

nature of investment, relative to consumption, led to greater variation in the initial responses following the news.

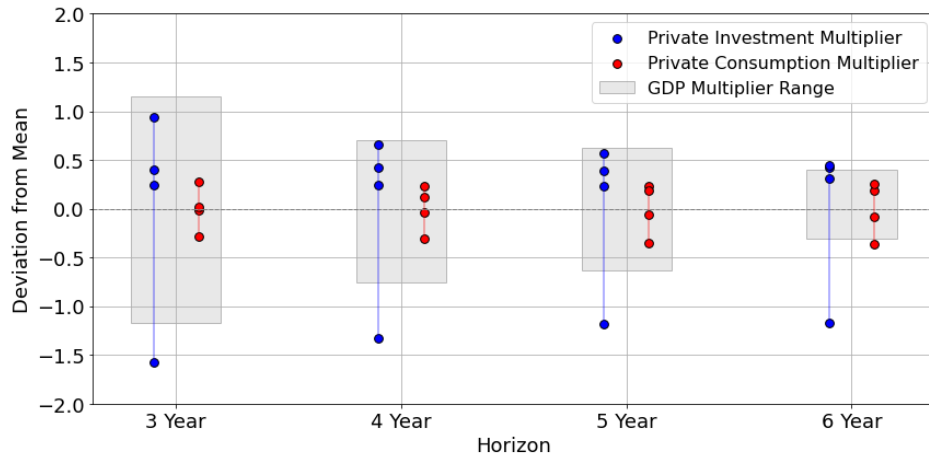


Figure 1.8: Variation in Multipliers: GDP Components

Notes: This figure shows the deviations of the cumulative multipliers from the average across the four samples. The x-axis represents the horizon of the cumulative multipliers. The blue dots indicate the sample-specific deviations of the private investment multiplier, while the red dots represent those of the private consumption multiplier. The shaded area shows the range of sample-specific deviations for the GDP multiplier.

1.6 Why Has the Effect of Government Spending Changed Over Time?

As shown in Section 1.5, private investment plays a central role in explaining the changes in the GDP multiplier over time. Government spending can either crowd in or crowd out private investment, and the relative balance between these opposing effects has shifted markedly across periods. Accordingly, understanding the time-varying effects of government spending requires examining both mechanisms. While prior studies have mainly linked changes in the crowding-out effect to movements in real interest rates, the drivers of crowding-in remain underexplored.²³ In this section, I investigate the determinants of the private investment crowding-in effect, which serves as a key mechanism for understanding the time-varying effects of government spending.

To understand the time variation in the crowding-in effects of private investment, I focus on the *intrinsic nature* of government spending. Even for the same magnitude of spending, its impact on private investment can vary depending on its *composition* and *time profile*. Specifically, the effects differ depending on whether the government executes spending through procurement or non-procurement channels, how persistent the increase in spending is, and how back-loaded its implementation becomes. This section examines how these intrinsic characteristics of government spending have evolved over time and discusses how such changes account for the time variation in the crowding-in effects.

1.6.1 Procurement Share and Crowding-In Effect

Whether government spending stimulates private investment largely depends on how and where the funds are allocated. For example, when the government procures goods or services from private firms, it can incentivize them to invest in expanding their production capacity. In contrast, if spending

²³Appendix 1.E.1 provides a detailed discussion of the crowding-out effects. In addition, the Korean War sample (1947–1964), which exhibits little response of private investment, is excluded from the main analysis. The weak responses of private sector variables during this period are discussed in Appendix 1.E.4.

is primarily directed toward non-procurement categories such as operation and maintenance or compensation for public employees, it is less likely to generate a strong stimulative effect on the private sector. Therefore, analyzing the share of procurement spending offers important insight into the heterogeneity of private investment crowding-in effects.

The main challenge in comparing the effects of government spending by whether it relates to procurement is the difficulty of classifying government spending news shocks by type of expenditure. Although the National Income and Product Accounts (NIPA) provide detailed spending categories, there is a potential bias arising from timing discrepancies between government spending news and actual expenditures. To mitigate this issue, this study uses defense budget authority (BA) as an alternative measure. BA refers to the legal authorization granted by Congress that allows agencies to commit or spend funds up to a specified amount within a given period, typically a fiscal year. Because the allocation of BA precedes actual outlays, it is more closely aligned in timing with government spending news than the spending data reported in the NIPA.²⁴

Defense BA is provided by spending type²⁵, which allows for comparing the effects of various types of government spending. Defense BA by type is available starting in fiscal year (FY) 1959. From FY1976 onward, the data are provided as a time series by the Office of Management and Budget (OMB). For the period from FY1959 to FY1975, I construct the data using the annual budget publications titled "*The Budget of the U.S. Government*."

The types of defense BA include procurement, operation and maintenance (O&M), military personnel, military construction, and research and development (R&D). The procurement finances the acquisition and upgrade of aircraft, missiles, and other weapon systems. This involves multi-year contracts with defense contractors, prompting firms to expand production capacity and invest in new plants and equipment. In contrast, non-procurement categories such as O&M and military personnel are largely absorbed within the public sector in the form of wages, administrative services,

²⁴BA still reflects a later stage than the defense spending news, which often arises before the budget approval. See Brunet (2022) for a detailed discussion of timing of BA.

²⁵The official title in the budget document is "Budget Authority by Function and Subfunction."

fuel, and routine upkeep.²⁶ Given these characteristics, increases in procurement spending are likely to generate crowd-in effects on private investment, whereas comparable expansions in O&M or military personnel budgets are unlikely to elicit much response from the private sector.²⁷

Figure 1.9 shows that the share of procurement within the defense BA has declined in recent decades. In the left panel, the average procurement share drops markedly from 26.6% during FY1960–FY1989 to 18.7% during FY1990–FY2019. This contrast becomes even more pronounced when focusing on periods marked by substantial increases in defense BA. Specifically, during the main surge years of the Vietnam War (FY1966–FY1967) and Reagan-era military buildup (FY1980–FY1985), procurement accounted for 30.0% and 29.9% of the defense BA, respectively. By contrast, during the corresponding years of the War on Terror (FY2001–FY2003), the procurement share was only 17.3%. In line with this pattern, the right panel shows that the contribution of procurement to overall defense BA growth has also weakened significantly in recent years, including during the War on Terror episode following the 9/11 terrorist attacks in 2001.²⁸

One way to estimate the effects of government spending by type is to use defense BA as an instrumental variable. The instrument is defined as the annual change in real defense BA divided by the previous year’s real GDP ($(BA_t - BA_{t-1})/GDP_{t-1}$). I construct separate instruments for procurement and non-procurement defense BA. Each instrument is then used separately to instrument for total government spending in estimating Equation 1.4, allowing for an estimation of the effects driven by each type of BA.^{29,30} Following Brunet (2022), quarterly NIPA variables are converted

²⁶For more details, see Towell (2021).

²⁷Hebous and Zimmermann (2021) and Gabriel (2024) provide micro-level evidence that procurement raises capital investment by contracting firms.

²⁸For example, in the emergency supplemental appropriations for defense and for the reconstruction of Iraq and Afghanistan for FY2004(Public Law 108-106), the budget allocated \$5.5 billion for procurement, \$17.8 billion for military personnel, and \$37.2 billion for operation and maintenance.

²⁹It is important to note that BA does not perfectly align with the timing of defense spending news. While it provides a closer approximation than the NIPA data, various forms of defense spending news often emerge prior to the official budget allocation, implying that some timing discrepancies remain. Therefore, it is necessary to assume that any measurement errors arising from such discrepancies are similar across categories and do not systematically bias the comparison.

³⁰Several recent studies have also used BA as an instrument for government spending. For example, Brunet (2022) estimates the fiscal multiplier of defense spending using BA, while Fieldhouse and Mertens (2023) analyzes the effects of government R&D spending based on the same measure.

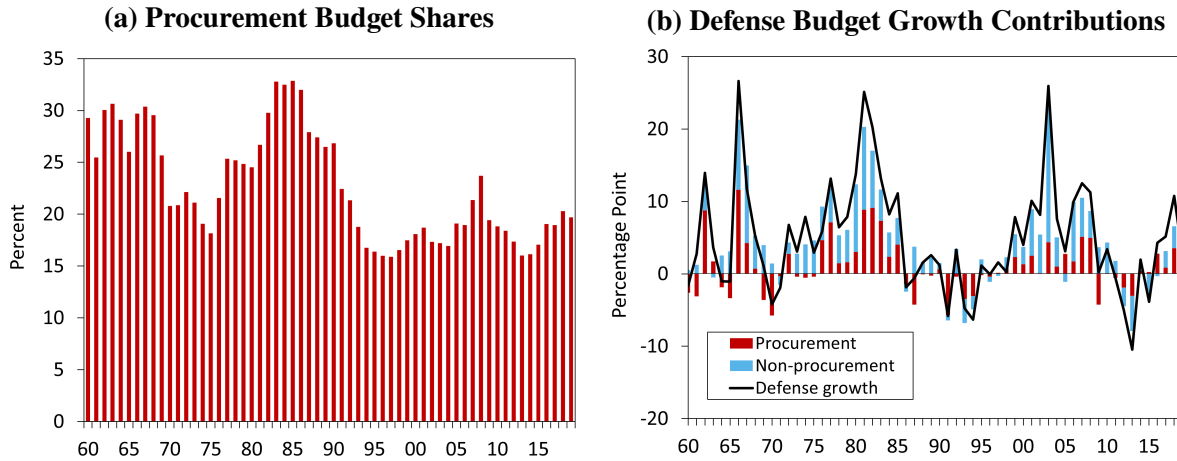


Figure 1.9: Defense Budget Authority by Type

Notes: The left figure shows the share of procurement within the defense budget authority for each fiscal year. In the right figure, the navy line represents the growth rate of total defense budget authority, while each bar indicates the contributions of procurement and non-procurement components.

to an annual frequency based on the fiscal year. The results, shown in Figure 1.10, indicate that government spending identified by the procurement BA instrument has a much stronger effect on private investment than spending identified by the non-procurement BA instrument. The estimated cumulative private investment multipliers over a three-year horizon are 0.27 for spending driven by procurement BA and -0.01 for spending driven by non-procurement BA.³¹

Another approach to estimating the effects of government spending by type is a hybrid method that combines the DNS with the procurement share of defense budget authority (BA). Specifically, Equation 1.5 extends the baseline local projection in Equation 1.3 by including an interaction term between the DNS and the procurement share. The impulse responses at each procurement share are estimated as $\beta_{i,h} + \beta_{i,h}^o \times \text{Procurement Share}$. This hybrid specification allows for a more precise identification of the timing of news shocks and enables analysis at the quarterly frequency, offering an advantage over the approach that relies solely on defense BA by type.³²

³¹Appendix 1.D presents results from two alternative specifications: (i) the defense budget authority by type is used as an instrument for the NIPA-based proxy of defense spending by type, and (ii) the defense budget authority is used directly to explain private investment. In both cases, procurement spending induces a more pronounced increase in private investment.

³²However, the procurement share of defense BA may still lag behind the DNS, making this approach a second-best alternative to directly classifying the DNS by type. It therefore relies on the assumption that any bias arising from

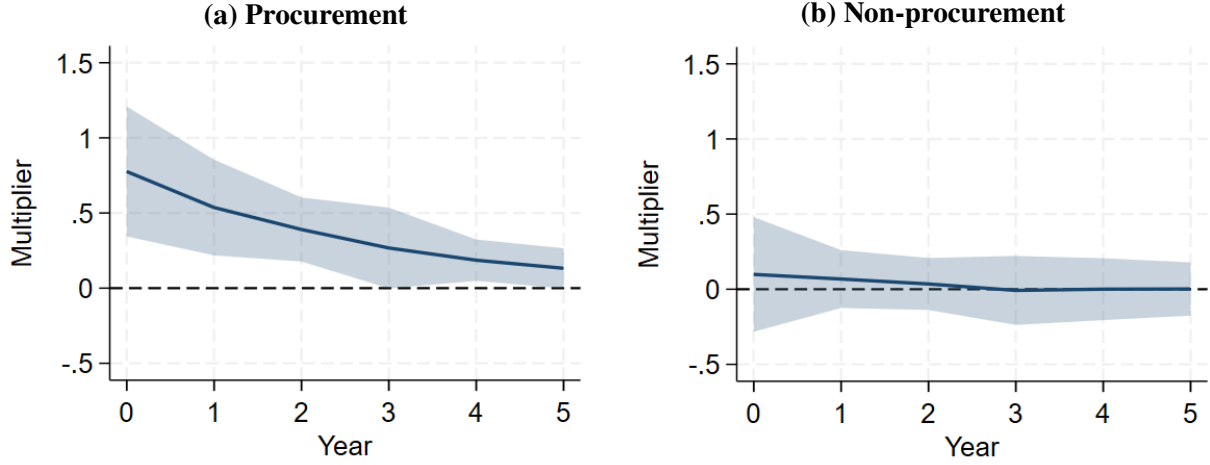


Figure 1.10: Private Investment Multipliers: Defense Budget Authority Type

Notes: The figures show the private fixed investment cumulative multipliers by defense budget type. Non-procurement budget authority is the combined total of operation and maintenance (O&M) and military personnel. The shaded area represent the 90% confidence intervals.

$$x_{i,t+h} - x_{i,t-1} = \alpha_{i,h} + \phi_{i,h}(L)z_{t-1} + \beta_{i,h}DNS_t + \beta_{i,h}^o \frac{BA_{proc,t}}{BA_t} DNS_t + \epsilon_{i,t+h} \quad (1.5)$$

$$h = 0, 1, \dots, H$$

Figure 1.11 shows that as the procurement share increases from 20% to 30%, the responses of private fixed investment shift from negative to positive. This pattern implies that non-procurement spending may not directly stimulate production by contracting firms. Instead, increased government demand creates a general equilibrium environment in which overall private investment is crowded out.

Taken together, these results indicate that changes in the procurement composition of government spending have played an important role in driving the time variation in private investment responses to government spending. As shown in Figure 1.9, the share of procurement within defense budget authority has declined markedly since the mid-1990s. In earlier periods, government procurement contracts provided strong incentives for private firms to expand production capacity and undertake

timing discrepancies between the DNS and procurement shares does not systematically distort comparisons across procurement shares.

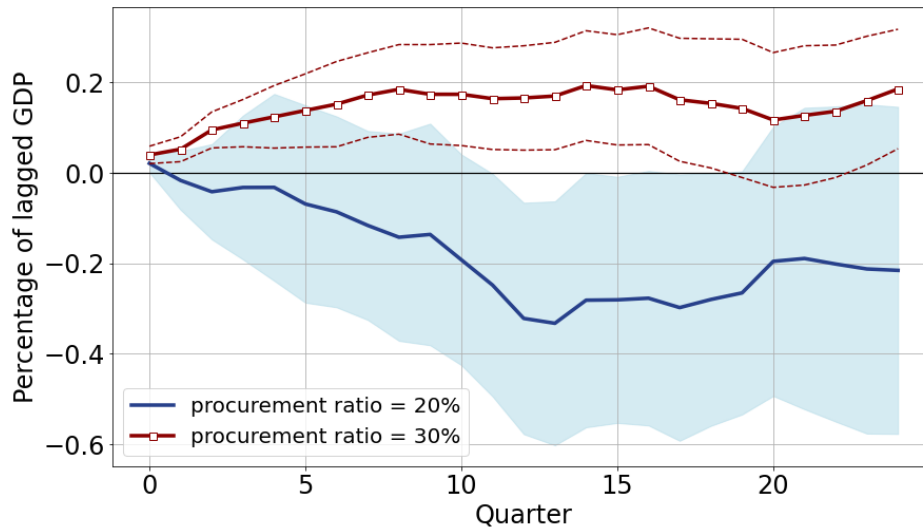


Figure 1.11: Responses of Private Fixed Investment by Share of Procurement

Notes: This figure shows the response of private fixed investment to DNS by share of procurement budget. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

new investment. In contrast, as non-procurement budgets have come to account for a larger share of defense spending in recent decades, these types of expenditures are less likely to directly stimulate private investment and may instead intensify competition for resources, thereby amplifying general-equilibrium crowding-out effects.

1.6.2 Time Profile of Government Spending

As documented in the section 1.6.1, the private investment crowding-in effects during both the Vietnam War and the Reagan era are attributable to a large procurement share in government spending. However, the difference in the magnitude of these effects between the two periods requires further explanation. Although the share of procurement in total defense budget authority is similar in both periods (around 30%), the crowding-in effect is markedly stronger during the Reagan era. This finding is puzzling, especially considering that the Reagan-era sample overlaps with the Volcker disinflation, a period of significantly tighter monetary policy than the Vietnam War era. I argue that the persistent and back-loaded nature of government spending during the Reagan era plays a key role in amplifying the private investment crowding-in effect.

Figure 1.12 shows that in response to defense spending news, government spending during the Reagan-era sample (1975–1999) is more persistent and back-loaded than during the Vietnam War sample (1954–1978). In the Vietnam period, the response begins to decline after roughly 15 quarters, whereas in the Reagan sample, it continues to rise steadily for up to 24 quarters. This contrast reflects the different historical contexts. During the 1980s, the Reagan administration pursued sustained increases in defense spending as part of a long-term strategic response to the Soviet Union, aiming to modernize the military, strengthen nuclear deterrence, and develop advanced weapons systems. In contrast, defense spending news during the Vietnam War reflected the immediate, short-term demands of an ongoing conflict. In the early phase (1965–1968), spending surged to meet urgent needs, but the pace slowed markedly after 1969, resulting in a less persistent response.³³

A more persistent increase in government spending is expected to generate a stronger crowding-in effect. This occurs because forward-looking firms incorporate the entire anticipated spending path when valuing future cash flows. When the expansion is perceived to be long-lasting, Tobin's q rises immediately, prompting firms to undertake substantial fixed investment well before the actual outlays take place. Such "anticipatory behavior" structurally amplifies the multiplier measured in empirical analysis. In Equation 1.4, I define the multiplier as the cumulative increase in private investment over a finite horizon (e.g., six years) divided by the realized increase in government spending within the same period. Under this definition, the numerator (investment) captures the response to expected long-run demand, whereas the denominator (spending) reflects only the outlays actually implemented during the window. Accordingly, the longer the expected duration of spending, the larger the finite-horizon multiplier becomes.

Monetary policy helps explain why back-loaded government spending tends to generate a stronger crowding-in effect in the early phase. When actual expenditures are concentrated in the later part of the spending horizon, inflationary pressures initially remain modest. Consequently, the central bank has little need to raise interest rates sharply, and monetary tightening proceeds

³³Appendix 1.E.2 provides further insight into this historical context by explaining the time profile of spending through the composition of procurement in each period.

gradually. In this environment, the rise in real interest rates is limited, and firms, anticipating future demand, expand investment well before the spending is implemented. In contrast, front-loaded spending expansions involve a large disbursement of funds at the outset, which quickly heightens inflationary pressures. To maintain price stability, the central bank responds by increasing interest rates rapidly and aggressively. As a result, the cost of capital surges, investment demand contracts, and the initial crowding-in effect is substantially weakened or even offset.

Overall, government spending during the Reagan era is more persistent and back-loaded than during the Vietnam War period, and this characteristic plays a central role in amplifying the crowding-in effect on private investment. When spending is expected to be sustained over a longer horizon, firms anticipate future demand growth and increase investment preemptively, generating stimulative effects even before the actual outlays occur. At the same time, when expenditures are concentrated in the later phase, initial inflationary pressures remain moderate, allowing the central bank to tighten monetary policy more gradually. As a result, the rise in real interest rates is limited, improving the investment environment.

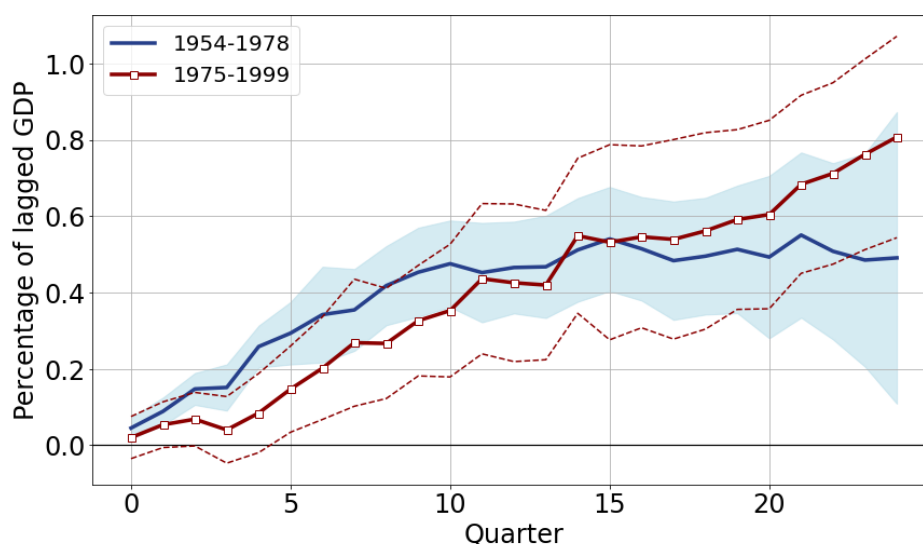


Figure 1.12: Responses of Government Spending by Sample Period

Notes: This figure shows the response of government spending to DNS. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

1.7 Conclusion

This study investigated how and why the effects of government spending have changed over time. To this end, I constructed a new series of defense spending news shocks spanning 1947–2019 by combining sentence-level text analysis of approximately 13 million newspaper articles with daily stock market data. This new series provides rich variation in government spending news across the entire sample period, allowing for more detailed estimation of the time-varying causal effects of government spending.

The empirical analysis found that the output multiplier has fluctuated substantially over time, ranging from about 0.8 to 1.5. This time variation was mainly accounted for by differences in private investment responses. In particular, the crowding-in effect on private investment depends on the intrinsic nature of the spending—its composition and time profile. When spending takes the form of procurement, it stimulates firms to expand production capacity and invest more. In contrast, when the increase is concentrated in non-procurement spending, the stimulative effect is negligible. Moreover, the time profile of spending further amplifies the crowding-in channel through procurement. When spending plans are persistent and back-loaded, the crowding-in effect on private investment becomes stronger, leading to larger output multipliers.

This study offers clear policy implications for today’s fiscally constrained environment. Expanding the overall scale of spending alone is insufficient to enhance its impact on the macroeconomy. Instead, policymakers need to consider the design of fiscal programs, specifically strengthening the linkage between public outlays and private-sector activity and ensuring a well-structured time profile of implementation. The findings indicate that a sustained, long-term spending plan, particularly one centered on procurement, can more effectively stimulate private investment and capacity expansion, thereby generating stronger macroeconomic outcomes.

Chapter 2

Heterogeneous Effects of Public Procurement

1

This paper studies the heterogeneous effects of the U.S. federal procurement on firm value and real investment by linking transaction-level contract records from 2000–2024 with daily stock prices and quarterly firm financial statements. We first show that contracts differ in how much new information they convey to financial markets. Awards that resolve uncertainty about both the winning firm and the scope of work generate significant stock price responses, while awards with a largely predetermined recipient elicit little market reaction. Focusing on these informative events, we document substantial heterogeneity in the effects of procurement by contract content. Stock market responses are stronger for contracts outside a firm’s core industry, suggesting that investors view such awards not only as current revenues but also as signals of new market opportunities or future growth options. Real investment responses follow a different pattern. Capital expenditures rise mainly after contracts aligned with the firm’s existing business and after product procurement, whereas R&D investment responds more strongly to non-core contracts and service procurement. These findings suggest that the translation of new demand into real investment depends not only on market-value-based investment incentives, such as Tobin’s q , but also on firms’ production and technological capabilities to meet that demand.

¹This chapter was co-authored with James Cloyne.

2.1 Introduction

Public procurement is one of the largest channels of government spending in the modern economy. In OECD countries, government procurement accounts for approximately 12–13% of gross domestic product (GDP), and in many countries the public sector functions as the largest purchaser of goods and services (OECD, 2025; World Bank, 2020).

Existing macroeconomic research primarily treats government purchases as an aggregate component of demand. However, procurement spending is highly heterogeneous in its object and mode of execution, and its economic effects may therefore differ substantially. For example, even a dollar of procurement spending may generate different real transmission channels depending on whether it is used for routine repeat purchases of existing goods, for the provision of new products or services, or for activities that extend beyond a firm’s existing business domain. Therefore, to fully understand the economic effects of public procurement, it is necessary to consider not only the aggregate volume of spending but also the heterogeneous nature embedded within it.

Although some recent studies have attempted to analyze the heterogeneous effects of government procurement, they have primarily focused on the characteristics of firms participating in procurement or on inter-industry production networks (Hebous and Zimmermann, 2021; Barattieri et al., 2023). However, these approaches largely treat government purchases as demand shocks at the firm or industry level, leaving relatively little known about how the content and structure of individual procurement contracts themselves shape the heterogeneity of procurement effects.

To fill this gap, this paper constructs transaction-level data on U.S. federal procurement contracts and links them to daily stock price data and quarterly firm financial statements. We examine how the effects of government procurement vary depending on contract characteristics. Specifically, we analyze how heterogeneously stock price responses around contract award dates, as well as subsequent capital expenditures (CAPX) and research and development (R&D) spending, react to the nature of the procurement itself. To this end, this paper examines heterogeneity in procurement

contracts along two key dimensions: the institutional form of the contract and the detailed content of the procurement.

First, we examine how the institutional characteristics of a contract dictate its information content, which in turn drives heterogeneous stock market responses to contract announcements. For example, a "definitive contract" is a standalone agreement not based on a pre-existing parent contract; thus, it is likely to contain relatively more new information about the winning firm, the contract amount, and the procurement object. By contrast, for "delivery orders" issued under a parent contract, some information regarding potential awardees or the scope of future orders may already be incorporated into market expectations. However, when the parent contract pre-specifies multiple potential awardees and the actual recipient is determined only at the individual order stage, the subsequent order may still retain substantial information content.

Consistent with this hypothesis, our empirical analysis shows clear stock price responses for definitive contracts and delivery orders under multiple-award parent contracts, whereas delivery orders under single-award parent contracts do not generate statistically significant responses. This suggests that how government demand shocks are recognized as 'new information' and transmitted to the market relies heavily on the structural design of the procurement contracts.

Second, we show that the economic effects of procurement vary fundamentally with contract content. Stock market valuations respond more strongly to contracts in industries outside the firm's existing core business than to those aligned with it. This aligns with the interpretation that the market views such contracts not merely as transient revenue boosts, but as signals of latent capabilities or valuable growth options in new markets. Similarly, product contracts and R&D service contracts generate significantly larger stock price responses than non-R&D service contracts, indicating that the market assigns divergent economic values to the same dollar of government spending depending on the procurement object.

Crucially, however, we demonstrate that these ex ante valuation effects do not correspond one-to-one with subsequent real investment adjustments. The direction of real adjustment depends less on

expected profitability itself and more on the specific type of productive capacity required for contract execution. For example, capital expenditure increases most following contracts aligned with the firm's existing business and for tangible product procurement. In contrast, R&D investment responds more strongly to contracts outside the firm's core business and to service procurement. These findings ultimately suggest that comprehensively understanding the effects of public procurement requires disentangling two distinct transmission mechanisms: a valuation channel, through which markets reprice firms at the moment new information arrives, and a capital adjustment channel, through which firms subsequently expand physical or knowledge capital depending on the nature of the contract.

These results carry significant implications for traditional macroeconomic investment models, particularly Tobin's Q theory. Under standard Q theory, an increase in stock market valuation (a rise in Q) should translate proportionally into an expansion of real investment. However, our results demonstrate that an increase in valuation metrics does not uniformly lead to a rise in physical capital expenditures (CAPX). This suggests that to accurately predict the multiplier effects of government purchases, macroeconomic models must explicitly incorporate the micro-level structure of demand shocks and the heterogeneity of capital (i.e., physical versus knowledge capital), rather than relying on aggregate Q metrics or the assumption of homogeneous capital.

Related literature. First, this paper relates to the literature analyzing the heterogeneous effects of government spending. Recent studies have shown that the effects of government procurement vary depending on micro-level characteristics of the recipient firms, such as financial constraints, net worth, and size (Hebous and Zimmermann, 2021; Giovanni et al., 2022; Gabriel, 2024), or cross-sectoral production networks (Muratori et al., 2023; Barattieri et al., 2023; Bouakez et al., 2025). While the existing literature primarily seeks the sources of heterogeneity in firms' financial conditions or macroeconomic industrial structures, this study focuses on the attributes of the individual procurement contracts themselves.

Second, this paper contributes to the literature on government procurement and firm dynamics.

The existing literature has primarily tracked firms' ex-post real responses, such as employment, investment, and growth, based on accounting and administrative data following the award of a procurement contract (Ferraz et al., 2015; Hebous and Zimmermann, 2021; Goldman, 2020). In contrast, this paper takes a forward-looking perspective and measures the ex-ante expected value that the stock market assigns to a procurement contract at the time of award. Furthermore, we show that the response of this expected value does not necessarily coincide with firms' ex-post capital adjustment. This suggests that distinguishing between the valuation channel at the time information arrives and the subsequent capital adjustment channel is essential for understanding the effects of procurement.

Third, this paper directly contributes to the recent finance literature on government procurement and capital market valuation. Recent studies have examined abnormal returns around announcements of federal procurement contracts (Ferris et al., 2025), estimated the private value of R&D procurement (Arora et al., 2025), and investigated how government procurement affects the incorporation of firm-specific information into stock prices (Wang et al., 2025). Going beyond prior studies that estimate the average firm-value effects of procurement contracts, this paper analyzes heterogeneous stock price responses according to the institutional and structural characteristics of procurement contracts. Furthermore, by comparing these market responses with subsequent real investment effects, the paper provides broader and more comprehensive economic implications.

Finally, this paper extends the literature on identifying government spending news shocks to the micro level. Existing studies emphasize the discrepancy between the timing of actual spending and the timing at which economic agents become informed, identifying news shocks using defense-related narrative measures (Ramey and Shapiro, 1998; Ramey, 2011), defense contractors' stock prices (Fisher and Peters, 2010), and time-series models (Ben Zeev and Pappa, 2017). This paper applies that insight to individual procurement contracts. We show that the timing of when a procurement contract is perceived by the market as new information depends on when uncertainty about the recipient firm and the scope of work is resolved, and that contract form plays a central role in shaping this timing.

The remainder of the paper is organized as follows. Section 2.2 introduces the transaction-level procurement data and details our procedure for linking these records to publicly traded firms. Section 2.3 establishes the empirical framework; it discusses the institutional structure of federal contracts to identify the timing of information arrival, outlines our heterogeneity classifications, and sets out the empirical methodology. Section 2.4 presents the main empirical results regarding the valuation channel, focusing on daily stock price responses across different contract forms and characteristics. Section 2.5 then examines the investment adjustment channel, tracking the subsequent firm-level responses in capital expenditure and R&D. Finally, Section 2.6 concludes.

2.2 Data

2.2.1 Data Sources and Key Variables

Procurement Data: FPDS-NG

For the empirical analysis, we construct our procurement dataset using transaction-level data from the Federal Procurement Data System-Next Generation (FPDS-NG), accessed via [USAspending.gov](https://www.usaspending.gov). The FPDS-NG is a comprehensive database that records all federal procurement activities at the individual contract action level. In this context, a “contract action” encompasses not only the initial contract award but also modifications to existing contracts, as well as follow-on transactions such as delivery orders and purchase orders. We collect data from this platform covering the period from 2000 to 2024.

To conduct a rigorous empirical analysis, we extract several key variables from this database, each serving a specific functional role in our identification strategy.

First, *Action date* serves as the event date ($t = 0$) in our analysis. This date represents the exact time when the federal agency officially executed the contract action. *Federal obligation* indicates the amount the federal government has committed to pay in the respective contract action, which allows us to measure the magnitude of the public procurement shock. For merging with stock price

and financial statement data, we use the *parent company name* and *parent company DUNS numbers* as our primary identifiers. *Competition status* denotes whether the contract was awarded through open competition or on a sole-source basis, while the number of offers received is also recorded to control for the intensity of competition and the novelty of the information.

Contract type categorizes awards into definitive contracts, Indefinite Delivery Vehicles (IDVs), delivery orders, and other types. We utilize this classification to distinguish the ex ante predictability and cash flow clarity of each contract. Specifically, a definitive contract is a standalone agreement with a clearly defined scope and legally binding commitments, thereby providing a highly transparent cash flow at the time of the award. In contrast, an IDV acts as an overarching "parent contract" that establishes fundamental terms and potential vendors in advance but does not guarantee any specific, immediate orders. The actual cash flows associated with an IDV are realized only when subsequent delivery orders are issued under it.

To analyze the content of the procurement, we use *Product Service Codes* (PSCs) to classify the procurement items into three broad categories: tangible products, R&D services, and non-R&D services. Furthermore, we leverage *NAICS codes* to determine whether the industry of the procurement contract matches the recipient firm's primary core business domain or constitutes a "mismatch." Such cross-industry mismatches can be interpreted by the market as a signal that the firm is pursuing new growth opportunities or validating latent capabilities.

In addition in our empirical analysis, an "event" is not defined as just any contract action. Rather, the estimation sample is strictly restricted to initial award actions—representing the first issuance of an individual contract or order—while simple follow-on modifications to existing contracts are excluded. Furthermore, among these initial award actions, we retain only observations with strictly positive federal action obligations. Therefore, unless specified otherwise, the terms "award," "procurement event," and "contract shock" in this study refer exclusively to these newly executed contracts with positive financial commitments that satisfy our rigorous filtering criteria.

Stock Return Data: CRSP

Daily stock returns are collected from the Center for Research in Security Prices (CRSP), which provides comprehensive data on stock prices, returns, and market capitalizations for firms listed on major U.S. exchanges. We utilize daily stock market data primarily because its high-frequency nature allows for a precise identification of "news shocks" that is unattainable with lower-frequency accounting data. By observing price movements in a narrow window around contract awards, we can isolate the immediate impact of procurement events from other confounding firm-level or macroeconomic factors. This approach is essential for capturing the market's real-time reassessment of a firm's future value at the exact moment information becomes available.

The key variables constructed from CRSP include daily stock returns and excess returns, the latter of which is calculated by subtracting the market return to isolate firm-specific responses from aggregate market movements. We also use market capitalization (shares outstanding multiplied by price) to measure firm size. To assess the relative economic importance of each award, we normalize the federal obligation by the firm's average market capitalization over the 10 trading days preceding the event. This normalization ensures that the impact of a given contract amount is comparable across firms of varying sizes, effectively scaling the demand shock by the firm's total equity value.

Corporate Financial Data: COMPUSTAT

Quarterly financial statements and R&D-related accounting variables are collected from COMPUSTAT. While stock price responses immediately reflect market expectations, accounting data serve to measure the adjustments actually undertaken by the firm. To fulfill new procurement demand, firms may expand production facilities or allocate additional resources to R&D projects and technological capabilities. However, unlike market valuations, these real adjustments materialize with significant lags due to budgeting, capital goods ordering, project execution, and the process of technological accumulation. Thus, utilizing quarterly data is appropriate for analyzing these structural responses.

The key variables drawn from COMPUSTAT are as follows. CAPX represents funds spent

on acquiring or improving physical assets, directly reflecting the firm's commitment to expanding productive capacity. R&D expenditure is a key variable used to capture the impact of contract awards on the accumulation of knowledge capital and the strengthening of technological capabilities, which this study interprets as the R&D investment response. We use total assets (atq) to measure firm size. In addition, we construct several control variables, including the market-to-book ratio (MTB), leverage, return on assets (ROA), and the cash ratio.

Table 2.1: Data Sources and Key Variables

Variable	Definition	Unit	Frequency	Source
Event date	Action date of the procurement contract	Date	Daily	FPDS-NG
Federal obligation	Obligated amount for the action	Dollars	Per action	FPDS-NG
Contract type	Definitive, IDV, Delivery/Task Order, etc.	Categorical	Per action	FPDS-NG
Competition	Competition status and number of offers	Categorical	Per action	FPDS-NG
Procurement object	Product, R&D service, non-R&D service	Categorical	Per action	FPDS-NG
Contract industry	NAICS/PSC-based industry classification	Categorical	Per action	FPDS-NG
Daily stock price	Firm's daily stock price	Dollars	Daily	CRSP
Stock market price	S&P 500 index	Index	Daily	CRSP
Market capitalization	Firm market capitalization	Dollars	Daily	CRSP
Tangible Assets	Property Plant and Equipment (PPENT)	Dollars	Quarterly	COMPUSTAT
Total Assets	Total assets (AT)	Dollars	Quarterly	COMPUSTAT
Capital Investment	Capital expenditure (CAPX)	Dollars	Quarterly	COMPUSTAT
R&D Investment	R&D Expenditure (XRD)	Dollars	Quarterly	COMPUSTAT
Firm characteristics	MTB, leverage, ROA, cash ratio, etc.	Various	Quarterly	COMPUSTAT

Notes: Accounting variables other than market capitalization are attached using the most recently available fiscal quarter-end value prior to the event date.

2.2.2 Linking Procurement Records to Publicly Traded Firms

A critical empirical challenge in this study is accurately linking government procurement contracts to the correct publicly traded companies. This difficulty primarily stems from the use of disparate firm identifiers across different databases. While procurement data rely on administrative identifiers

such as DUNS (Data Universal Numbering System) numbers, financial databases employ database-specific identifiers, such as PERMCO for CRSP and GVKEY for Compustat. To resolve this issue, we adopt a supplementary matching strategy by combining the existing crosswalk with our own matching process based on parent firm names.

Coverage Gaps in the Existing Matching

As one of the most widely recognized benchmarks, Hebous and Zimmermann (2021) (hereafter HZ) construct their crosswalk by uploading DUNS numbers to Bureau van Dijk's Orbis database to match firm records, and then linking them to the Osiris database via ticker symbols to obtain public firm identifiers. However, relying solely on the HZ crosswalk is insufficient for our comprehensive sample due to two major limitations.

First, it suffers from temporal constraints. The HZ crosswalk relies entirely on an Orbis-Osiris snapshot valid only up to 2012, assuming that the matching established in 2012 remain static in subsequent years. Consequently, firms that went public or newly entered the federal procurement market after 2012 are systematically excluded.

Second, the HZ crosswalk contains substantial matching gaps even in the pre-2012 period. As shown in Figure 2.1, our extended matching procedure captures a considerably larger number of additional firms than HZ throughout the 2000–2011 period. This appears to reflect various forms of information loss that can arise in the ticker-based linking process, such as the absence of ticker information for some observations in each dataset. Accordingly, our matching procedure not only extends temporal coverage beyond 2012, but also helps fill structural matching gaps that existed in the earlier period.

Extended Multi-Stage Matching Procedure

This study takes the HZ crosswalk as a starting point but implements a three-step matching procedure based on parent company names to extend coverage through 2024.

Step 1 (Existing crosswalk matching). We first apply the DUNS–GVKEY/PERMCO crosswalk

provided by Hebous and Zimmermann (2021).

Step 2 (Conservative name matching). For observations unmatched in Step 1, we perform name matching between the *parent company name* in FPDS-NG and firm names in COMPUSTAT, applying only minimal normalization to limit false positives. Specifically, we (i) standardize letter case, (ii) remove special characters, and (iii) standardize legal suffix notation. Because legal suffixes themselves are retained at this stage, “Apple Inc” and “Apple Corp” are treated as distinct entities.

Step 3 (Aggressive name matching). For observations still unmatched after Step 2, we remove all legal suffixes and compress multiple whitespace characters into a single space before re-matching. Because this step carries a relatively higher risk of false positives, all matches generated in Step 3 were manually verified to confirm their accuracy.

Matching Results

Figure 2.1 compares the number of matched publicly listed firms by year under two approaches: using only the HZ crosswalk versus applying our extended procedure. The temporal extension effect beyond 2012 is clear. Relying solely on the HZ crosswalk, the number of matched firms declines progressively after 2012, whereas our procedure maintains relatively stable matching coverage throughout the sample period. Moreover, our procedure generates a substantial number of additional matches even in the 2000–2011 interval. As discussed above, this suggests that the HZ crosswalk contained structural gaps arising from the Orbis-Osiris linking process, beyond its temporal limitations, and that our name-based supplementary procedure partially fills these gaps.

2.3 Empirical Framework

2.3.1 Contract Structure and the Timing of Information Arrival

Federal procurement contracts differ in the timing and intensity of the information they convey to market participants based on their contractual structure. While some spending events capture

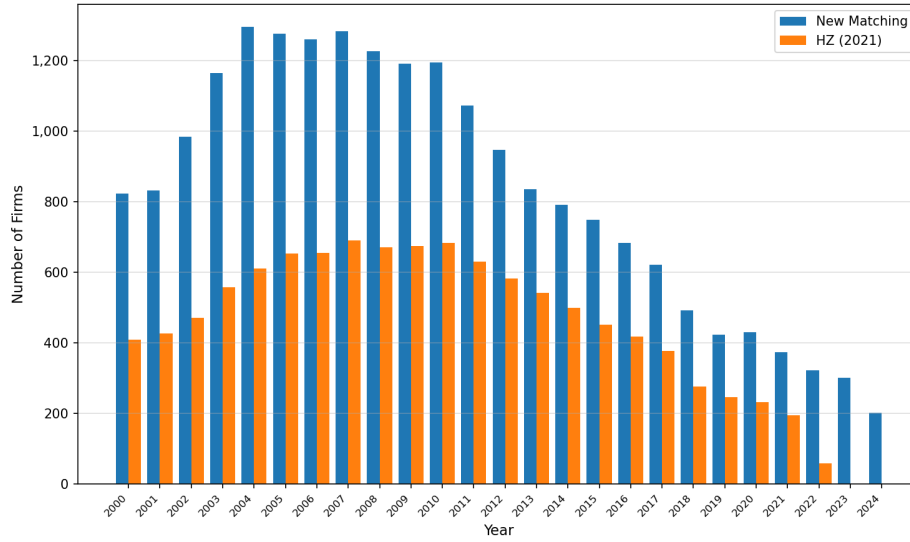


Figure 2.1: Annual Matching Coverage

Notes: This figure illustrates the results of the multi-stage matching procedure. The panel displays the number of matched publicly listed firms by year, comparing the coverage of the existing crosswalk by Hebous and Zimmermann (2021) with our new matching results.

market attention because the winning contractor was unexpected, others generate new information when specific project details and cash flows are finalized, even if the recipient's identity was already somewhat anticipated. Therefore, the differences between contract types can be understood in terms of when two key uncertainties—recipient uncertainty and scope uncertainty—are resolved. Figure 2.2 illustrates these differences. In a definitive contract (panel a), both recipient uncertainty and scope uncertainty are resolved simultaneously at a single point: the contractor is selected and the price and scope are specified at the time of award. As a result, the award represents a single, concentrated information event.

Under Indefinite-Delivery Vehicles (IDVs), by contrast, these two uncertainties are resolved in stages—first at the framework award and then at the subsequent Delivery Order (DO)—and the pattern of resolution varies between single-award and multiple-award structures. In a single-award IDV (panel c), recipient uncertainty is fully resolved at the framework stage, since the contractor relationship is established at that point. Scope uncertainty, however, is only partially resolved at the framework stage and is fully settled only when specific task scopes and cash flows are determined at the DO stage. In a multiple-award IDV (panel b), the framework award determines only a pool of

candidate firms, leaving recipient uncertainty only partially resolved. Scope uncertainty remains entirely unresolved at the framework stage. Both dimensions are fully resolved only at the DO stage, meaning that each individual DO conveys more new information to the market compared to DOs under a single-award structure.

Consequently, the events that trigger significant stock price or real-sector responses are not simply those with large dollar amounts, but those that resolve recipient and scope uncertainties more sharply at a particular point in time.

2.3.2 Identifying Informative Procurement Contracts

The discussion in the previous section on contract structure and the timing of information arrival also has important implications for identifying procurement news shocks. To interpret the obligation amount of a procurement contract as closely approximating an unanticipated news shock, it is necessary to identify contract forms that contain relatively more new information. From this perspective, definitive contracts and delivery orders issued under multiple-award IDVs are particularly important. These contracts are more likely to resolve uncertainty about the awardee, the contract amount, or whether an actual order will be placed at the time of contract award or order issuance. Conversely, if delivery orders issued under single-award IDVs, where most of the relevant information may already have been incorporated into market expectations, are included, the estimated market response may be attenuated, leading to a downward bias in the estimated procurement effect.

This discussion is important because it provides an additional criterion for identifying informative procurement news shocks beyond the “full and open competition with at least two bidders” criterion commonly used in the existing literature. (Hebous and Zimmermann (2021) and Barattieri et al. (2023)) To be sure, competition codes are an important starting point for identifying firm-level procurement news shocks, since they help restrict attention to firms selected through a competitive process rather than to pre-determined recipients. However, in federal procurement data, the competition identifier for delivery orders is often inherited from the parent framework contract.

Therefore, even if a delivery order is recorded as a follow-on order under a competitively awarded IDV, when that order is issued under a single-award IDV, the actual recipient is likely to have already been effectively determined at the parent contract award stage. In this case, an individual delivery order may be classified as competitive based on the competition code, but the amount of new information it conveys to the market may be limited. Thus, identifying informative procurement shocks requires taking into account not only competition codes, but also the structure of the contract.

Taken together, this institutional discussion suggests that definitive contracts and DOs issued under multiple-award IDVs are the procurement events relatively more likely to convey new information to the market. This classification yields a clear empirical prediction: if differences in the timing of information arrival across contract structures matter, daily stock price responses should be stronger for definitive contracts and DOs under multiple-award IDVs than for DOs under single-award IDVs. In Section 2.4, we verify this prediction directly using stock market responses.

Table 2.2: Informativeness of Procurement by Contract Type

Event	Recipient	Scope	Interpretation
Definitive contract	High	High	Both the recipient and scope are determined simultaneously at award
Multiple-award IDV	Medium	Low	The candidate pool is disclosed, but the final recipient and specific scope remain undetermined
DO under multiple IDV	Medium	High	The final recipient is selected from the pool and the specific scope of work is determined
Single-award IDV	High	Low	The recipient is fixed, but the scope of follow-on orders remains largely unresolved
DO under single IDV	Low	High	The recipient is already known; new information pertains primarily to the task scope and cash flows

Notes: DO denotes Delivery Order. IDV denotes Indefinite-Delivery Vehicles. “Recipient” and “Scope” columns indicate the degree to which new information about each dimension is conveyed at the respective event. High, Medium, and Low reflect the amount of uncertainty resolved at that stage.

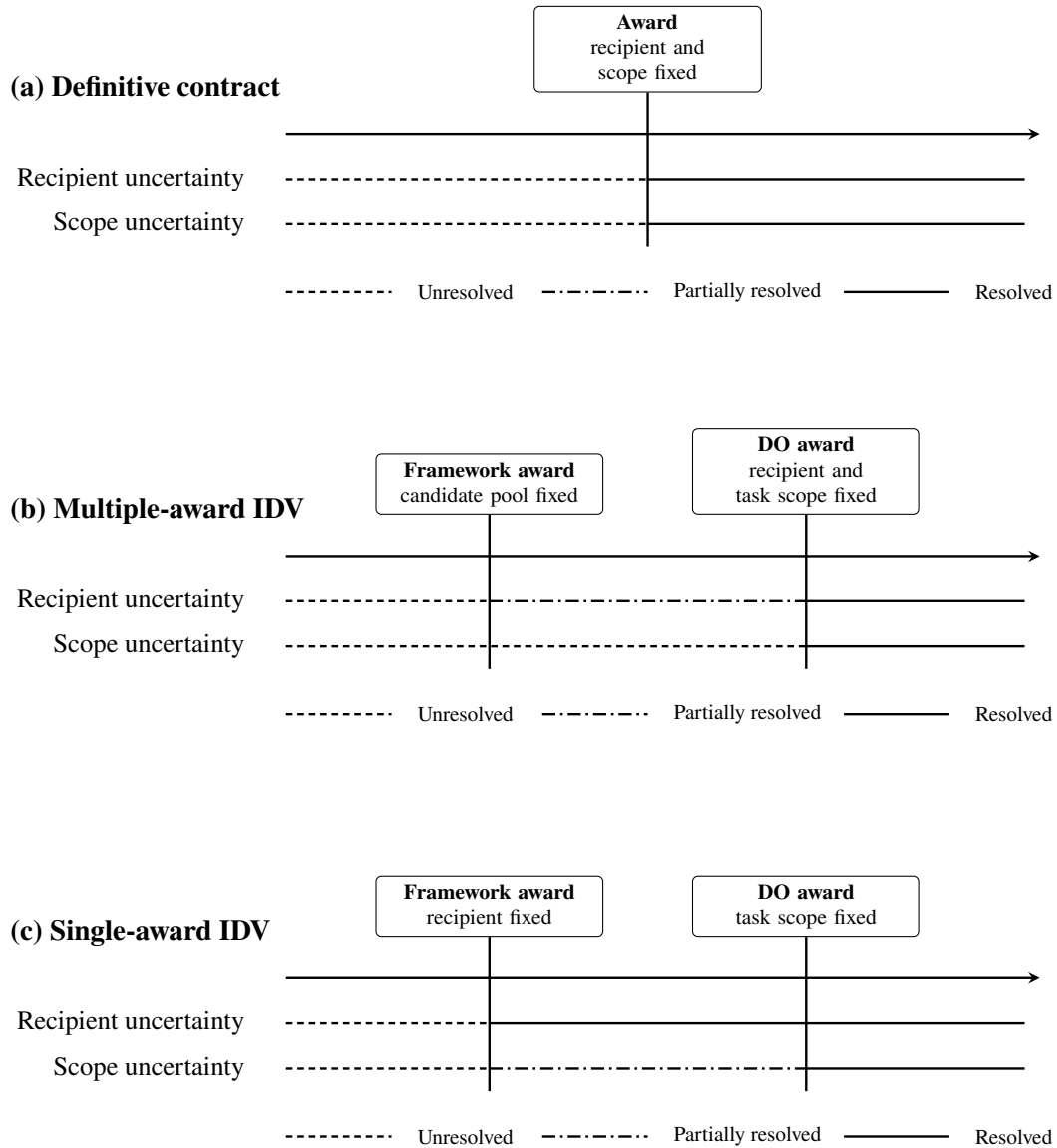


Figure 2.2: Timing of Recipient and Scope Resolution across Contract Forms

Notes: Dashed, dash-dot, and solid lines represent unresolved, partially resolved, and resolved uncertainty, respectively. (a) For definitive contracts, both the recipient and scope are fixed at the time of the award. (b) Under multiple-award IDVs, the framework agreement fixes the candidate pool, while the final recipient and specific task scope are determined only at the delivery-order (DO) stage. (c) Under single-award IDVs, the framework agreement fixes the recipient, but the task-specific scope is resolved gradually at the DO stage

2.3.3 Baseline Sample

Based on the institutional discussion in section 2.3.2, our empirical analysis focuses on definitive contracts and delivery orders under multi-award IDVs. These contract types are characterized by relatively low ex-ante predictability of the recipient. At the same time, they likely possess higher information content because the scope of work and cash flows are more concretely defined at the time of issuance.

It is important to note that this selection does not entirely eliminate concerns regarding selection bias. Even in competitive bidding environments, participating firms and eventual winners may systematically differ in unobservable characteristics, and the award itself may reflect a firm's growth opportunities or pre-existing government relationships. In unconditional comparisons, treated firms are substantially larger, more profitable, and more liquid than control firms (see Appendix Table 2.B.2).

Our identification strategy relies on two-way fixed effects—firm fixed effects and industry×time fixed effects—so the relevant notion of balance is conditional on this structure. Table 2.3 reports treatment–control comparisons after partialling out both sets of fixed effects. All normalized differences fall well below the conventional threshold of 0.10 (Austin, 2009; Austin, 2011), indicating that, within the same firm and conditional on industry-specific time shocks, the observable characteristics of firm-periods that coincide with a contract award are virtually indistinguishable from those that do not. The largest normalized difference is 0.067 for $\log(\text{Assets})$ in the quarterly panel, corresponding to less than a one-percentage-point deviation from the within-firm mean. These results suggest that, after conditioning on the fixed-effect structure employed in our regressions, residual selection on observables is negligible.

2.3.4 Sources of Heterogeneity

We examine heterogeneity along two dimensions: the procurement object and industry alignment. The procurement object is related not only to which production inputs the contract requires the

Table 2.3: Firm Characteristics: Treatment vs. Control (FE-Adjusted)

Variable	Treatment Mean	Control Mean	Difference	SMD	Obs.
<i>Panel A: Daily</i>					$N_t=100,510; N_c=5,477,994$
Excess return	0.004	0.000	0.004	0.002	5,578,504
log(Market cap)	-0.001	0.000	-0.001	0.003	5,567,389
<i>Panel B: Quarterly</i>					$N_t=18,893; N_c=78,694$
log(Assets)	0.025	-0.006	0.031	0.067	97,587
Market-to-book ratio	-0.003	0.001	-0.003	0.003	93,930
Leverage	0.013	-0.003	0.016	0.005	95,066
ROA	0.000	0.000	0.000	0.011	92,457
Cash ratio	0.006	-0.002	0.008	0.005	90,207

Notes: All variables are residualized using firm fixed effects and industry (4-digit NAICS)×time fixed effects. In Panel A, excess return is the one-day lagged excess return and log(Market cap) is the average of log daily market capitalization over the prior 10 trading days. In Panel B, all variables are lagged one quarter; time fixed effects are at the quarterly level; the control group consists of non-event firm-quarters from firms with at least one contract. |SMD| is the absolute Standardized Mean Difference (SMD), defined as $|\bar{X}_t - \bar{X}_c|/\sqrt{(s_t^2 + s_c^2)/2}$, where absolute values above 0.25 is a standard threshold for assessing imbalance (Stuart and Rubin, 2008; Imbens and Rubin, 2015).

firm to adjust, but also to the economic meaning of a given contract value for firm value. Industry alignment is related to whether the market interprets the contract as an extension of the firm's existing business or as a signal of new capabilities or entry into a new market.

Heterogeneity by Procurement Object Procurement of physical products is more likely to require adjustments in physical inputs such as equipment, inventories, and production capacity, and may therefore be more closely linked to CAPX responses. Service procurement, by contrast, more directly requires inputs of labor, organization, and specialized knowledge, and may therefore be more closely related to R&D investment and the accumulation of knowledge capital rather than the immediate expansion of physical capital. At the same time, the market may attach different economic values to the same contract amount depending on the procurement object, because input structures, expected margins, implementation difficulty, relationship persistence, and follow-on contracting opportunities may differ across categories. R&D services, in particular, may be linked not only to short-run revenue effects but also to knowledge accumulation, the possibility of follow-on project awards, and the expansion of future growth opportunities, and are therefore analyzed separately

from non-R&D services. The distinction by procurement object is thus useful for understanding not only real responses such as CAPX and R&D investment, but also heterogeneity in the magnitude of stock market reactions and in the valuation of contracts.

Heterogeneity by Industry Alignment The relationship between contract content and the firm's existing core business is another important source of heterogeneity. Contracts that align with the firm's existing industry activities are likely to be interpreted as extensions of its established production capacity or commercial relationships. By contrast, contracts characterized by industry mismatch may be viewed as evidence of latent capabilities, signals of entry into new markets, or expansion into non-core activities. Such contracts may therefore generate larger valuation responses, but market reactions may also be dampened by execution risk and uncertainty about profitability. The empirical analysis focuses on determining which effect dominates.

2.3.5 Estimating Daily Stock Price Responses

Daily stock price responses are estimated using separate local projections for each horizon h . The dependent variable is defined as the market-adjusted cumulative log return from the day before the event through h days after the event:

$$CAR_{i,t+h} \equiv [\log P_{i,t+h} - \log P_{m,t+h}] - [\log P_{i,t-1} - \log P_{m,t-1}]. \quad (2.1)$$

Here, $P_{i,t}$ denotes the stock price of firm i , and $P_{m,t}$ denotes the market index. This definition allows the dependent variable to capture the total valuation revision over the event window without requiring the separate estimation of an expected-return model.

The estimating equation is

$$CAR_{i,t+h} = \alpha_i^h + \gamma_{j \times \tau}^h + \sum_{k \in \mathcal{K}} \beta_h^k \left(\frac{\text{Obligation}_{i,t}^k}{\overline{\text{MktCap}}_{i,t-10:t-1}} \right) + \Gamma_h' X_{i,t-1} + \varepsilon_{i,t+h}. \quad (2.2)$$

Here, $\text{Obligation}_{i,t}^k$ is the total obligation received by firm i on date t under contract type k , and the denominator is the firm's average market capitalization over the preceding 10 trading days. This normalization allows contract size to be interpreted as a relative shock to the firm's pre-existing market value. α_i^h denotes firm fixed effects, and $\gamma_{j \times \tau}^h$ denotes 4-digit industry $\times$ calendar-time fixed effects. The vector $X_{i,t-1}$ includes predetermined controls such as pre-event excess returns and pre-event government procurement exposure. Standard errors are computed using the Driscoll–Kraay method, which is robust to both cross-sectional and serial correlation.

The key coefficient, β_h^k , measures the average valuation revision over horizon h associated with a one-unit increase in relative contract size for contract type k . Differences in β_h^k therefore directly compare how the market values government purchases of the same relative size across different contractual forms.

As a note, we do not use a standard difference-in-differences design, such as Callaway and Sant'Anna (2021), because procurement contracts are repeated and non-absorbing treatments rather than one-time adoption events. The same firm may receive multiple contracts on different dates, of different sizes, and under different contract types. This makes it difficult to assign firms to a single treatment cohort or to define a clean control group. A first-treatment DID design would ignore subsequent contract awards and would not directly exploit variation in contract size and type. The local projection specification instead treats each contract award as a firm-date shock and estimates the market response to the relative size of that shock.

2.3.6 Estimating Quarterly Investment Responses

For the analysis of real responses, daily procurement events are linked to accounting data at the firm-quarter level. The quarter in which the event occurs is taken as period t , and the most recent accounting period prior to the event is defined as $t - 1$. When the same firm experiences multiple procurement events within the same accounting period, obligations are aggregated by contract type to form a single period-level shock.

Investment responses are estimated using the following local projection:

$$\frac{Y_{i,t+h} - Y_{i,t-1}}{N_{i,t-1}} = \alpha_i^h + \gamma_{j \times \tau}^h + \sum_{k \in \mathcal{K}} \theta_h^k \left(\frac{\text{Obligation}_{i,t}^k}{N_{i,t-1}} \right) + \Pi_h' Z_{i,t-1} + u_{i,t+h}. \quad (2.3)$$

Where $Y_{i,t+h}$ represents capital or R&D investment at horizon h , and $N_{i,t-1}$ is a predetermined scaling factor measured pre-event (lagged PP&E for CAPX and lagged total assets for R&D). Under this scaling, CAPX reflects investment adjustments relative to physical capital, while R&D captures knowledge capital adjustments relative to the asset base. Crucially, since the procurement shock is also normalized by the same denominator ($N_{i,t-1}$), the coefficient θ_h^k can be interpreted as a dollar-for-dollar sensitivity—measuring the dollar change in investment or R&D for every one-dollar change in the procurement shock. Standard errors are computed using the Driscoll–Kraay method, which is robust to both cross-sectional and serial correlation.

2.4 Effects of Procurement on Stock Prices

This section analyzes when government procurement contracts are recognized as new market information and how such news affects financial market valuations. First, we identify contract types with high informativeness by examining daily stock price reactions according to contract structures. Next, we investigate how market valuations vary based on industrial congruence and the nature of the procurement target.

2.4.1 Stock Market Responses by Contract Form

Figure 2.3 and Table 2.4 show that market reactions differ significantly depending on the institutional form of the procurement contract. For definitive contracts and delivery orders under multi-award IDVs, excess returns increase significantly after the event date. For example, the coefficients for definitive contracts are 0.109 at $h = 1$, 0.168 at $h = 2$, and 0.292 at $h = 10$. Similarly, the coefficients for multi-award IDV delivery orders are 0.221, 0.421, and 0.833 at horizons $h = 1$, 2, and 10,

respectively. In the baseline sample that combines these two types, the coefficients also remain positive and statistically significant: 0.119 at $h = 1$, 0.240 at $h = 2$, and 0.638 at $h = 10$. In contrast, delivery orders under single-award IDVs yield small and statistically unstable coefficients across all horizons.

These results suggest that the market responds more strongly when contract awards resolve recipient uncertainty and contemporaneous scope uncertainty at the time of issuance, rather than to contract size alone. In particular, these patterns align with the previous discussion suggesting that the timing and intensity of information transmission to the market vary depending on the institutional structure. In other words, the market interprets a contract as a new government demand shock more strongly when it is difficult to predict the beneficiary ex-ante and when the scope of work and cash flows are relatively well-defined at the time of signing. Conversely, under a single-award structure, the beneficiary of subsequent delivery orders is largely predetermined, limiting the extent to which the orders themselves can act as "new" news.

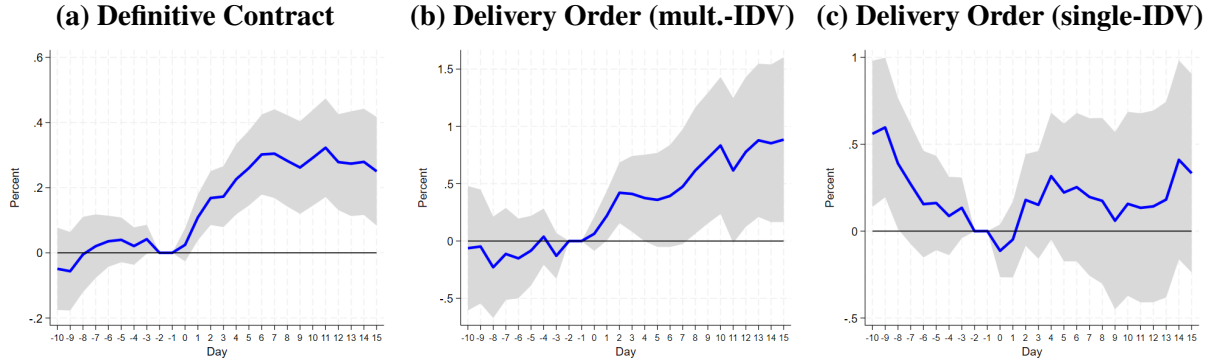


Figure 2.3: Daily Stock Price Responses by Contract Form

Notes: This figure plots the estimated coefficients from local projections of excess returns on relative procurement shocks. The x-axis represents the trading days relative to the action date ($t = 0$). Panels (a) and (b) show responses for definitive contracts and delivery orders (DO) under multiple-award IDVs, respectively. Panel (c) displays the response for DOs under single-award IDVs. Shaded areas denote 90% confidence intervals. Driscoll–Kraay standard errors with bandwidth 12.

Table 2.4: Daily Stock Price Responses by Contract Form

	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 7$	$h = 10$
Definitive	0.024 (0.031) [4,491,143]	0.109** (0.043) [4,490,124]	0.168*** (0.050) [4,489,141]	0.172*** (0.057) [4,488,161]	0.304*** (0.083) [4,484,264]	0.292*** (0.089) [4,481,358]
DO under mult-IDV	0.063 (0.090) [3,350,344]	0.221* (0.131) [3,349,586]	0.421*** (0.161) [3,348,848]	0.411** (0.201) [3,348,112]	0.475 (0.304) [3,345,191]	0.833** (0.363) [3,343,014]
DO under single-IDV	-0.114 (0.092) [2,906,511]	-0.048 (0.133) [2,905,889]	0.179 (0.160) [2,905,280]	0.151 (0.189) [2,904,674]	0.196 (0.275) [2,902,252]	0.157 (0.322) [2,900,440]
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry $\times$ Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each row is estimated from a separate regression for the respective contract type. Coefficients are interpreted as the stock price response to obligation/lagged market capitalization. Standard errors in parentheses; observations in brackets. Driscoll–Kraay standard errors with bandwidth 12. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.4.2 Stock Market Responses by Sector Specialization and Firm Sector

This section analyzes how stock price responses to procurement shocks vary depending on whether the contract falls within the firm’s sector of specialization. Throughout this section and the analyses that follow, we use definitive contracts and multiple-award IDV delivery orders as the baseline sample, based on the evidence in Section 2.4.1 that these contract forms are relatively more informative.

Our results show that stock price responses are more pronounced when the contract falls outside the firm’s specialized sector. According to Figure 2.4 and Table 2.5, the response coefficients for specialized-sector contracts remain small and statistically insignificant across most horizons, whereas those for non-specialized-sector contracts increase rapidly. Specifically, they are 0.190 at $h = 1$, 0.319 at $h = 2$, 0.420 at $h = 3$, 0.713 at $h = 7$, and 0.917 at $h = 10$. Notably, the difference between the two groups becomes statistically significant starting at $h = 3$, with values of 0.353, 0.595, and 0.604 at $h = 3, 7$, and 10, respectively. This suggests that the strong reaction to non-specialized-sector contracts is not an immediate jump on the event date, but rather a gradual accumulation as the market digests the contract details.

By contrast, a coarser classification based only on the firm’s broad industry, such as goods

versus service sectors, provides little explanatory power for differences in stock price responses. In Figure 2.5 and Table 2.6, the coefficients for both groups are positive at longer horizons, but the difference between the two groups is not statistically significant at any horizon. These results suggest that the market responds more sensitively to the extent to which an individual contract overlaps with, or diverges from, the firm's existing business portfolio than to broad industry categories.

These results are consistent with at least two interpretations. First, non-specialized-sector contracts may contain a larger surprise component because they are less likely to have been anticipated given the existing information set. Second, such contracts may serve as signals of hidden capabilities or potential entry into new markets. For example, if a firm that has historically supplied a specific type of military component wins a contract outside its existing procurement, such as communications equipment or R&D testing and evaluation services, the market may interpret this not merely as an increase in current revenue but as a signal that the firm's technological scope or business opportunities have expanded. By contrast, contracts within the firm's existing sector of specialization are more likely to be viewed as an extension of already known capabilities and business relationships.

Table 2.5: Stock Price Responses by Sector Specialization

	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 7$	$h = 10$
Specialized sector	0.007 (0.055)	0.028 (0.078)	0.132 (0.105)	0.067 (0.115)	0.118 (0.151)	0.313 (0.193)
Non-specialized sector	-0.014 (0.072)	0.190** (0.096)	0.319*** (0.122)	0.420*** (0.143)	0.713*** (0.203)	0.917*** (0.228)
Difference	-0.021	0.162	0.187	0.353**	0.595**	0.604**
p -value	[0.817]	[0.181]	[0.229]	[0.044]	[0.016]	[0.039]
Observations	5,578,504	5,577,160	5,575,865	5,574,575	5,569,438	5,565,605
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry $\times$ Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Specialized-sector and non-specialized-sector contracts are classified by whether the contract's 4-digit NAICS code matches the firm's primary industry. Both categories are included simultaneously in the same regression. Difference = non-specialized – specialized. Driscoll–Kraay standard errors in parentheses (bandwidth 12). F -test p -values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

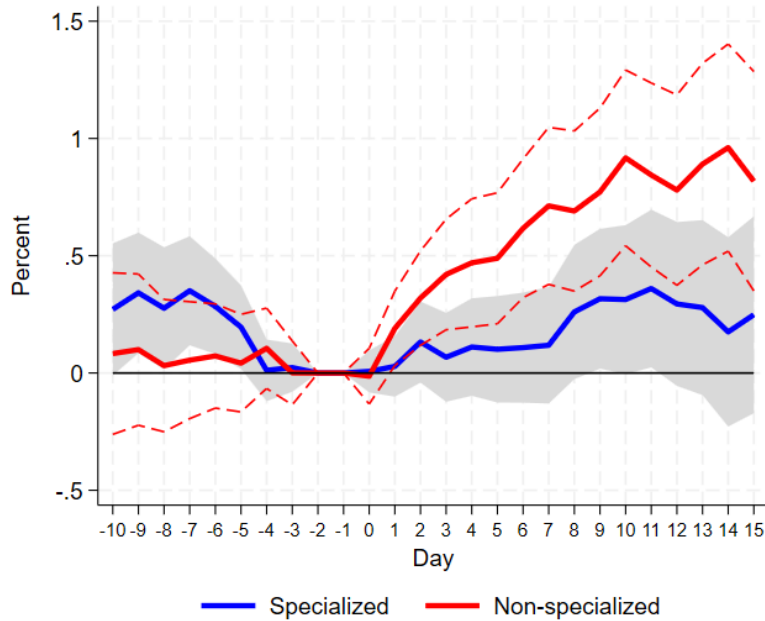


Figure 2.4: Daily Stock Price Responses by Sector Specialization

Notes: This figure compares stock price responses for specialized-sector contracts, where the contract's 4-digit NAICS code matches the firm's primary industry, with non-specialized-sector contracts. Shaded areas denote 90% confidence intervals. Driscoll–Kraay standard errors with bandwidth 12.

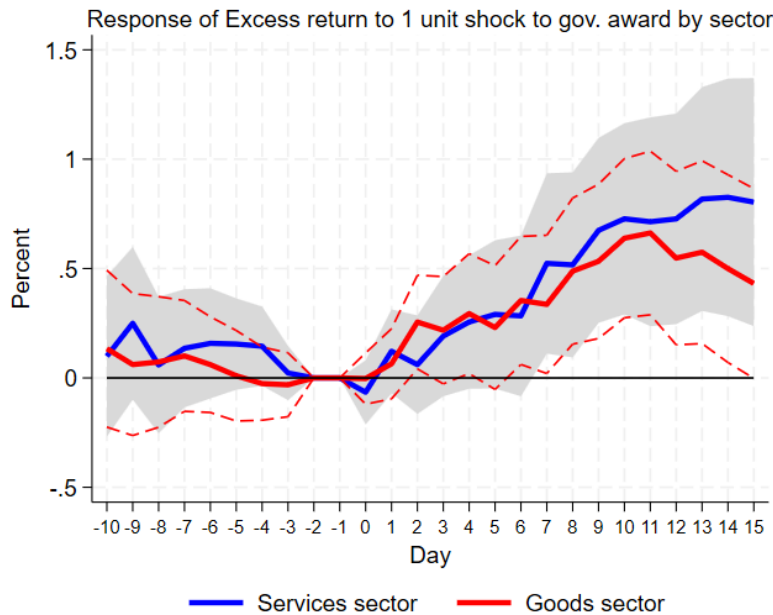


Figure 2.5: Daily Stock Price Responses by Firm Sector

Notes: This figure compares stock price responses for firms primarily operating in the Goods sector versus the Service sector, classified by their primary SIC codes. Shaded areas denote 90% confidence intervals. Driscoll–Kraay standard errors with bandwidth 12.

Table 2.6: Stock Price Responses by Firm Sector

	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 7$	$h = 10$
Goods sector	−0.003 (0.071)	0.065 (0.098)	0.255** (0.130)	0.218 (0.149)	0.336* (0.192)	0.639*** (0.221)
Service sector	−0.065 (0.089)	0.122 (0.117)	0.061 (0.137)	0.191 (0.167)	0.524** (0.251)	0.727*** (0.266)
Difference	−0.062	0.057	−0.194	−0.027	0.188	0.088
p -value	[0.592]	[0.711]	[0.306]	[0.906]	[0.553]	[0.809]
Observations	4,137,467	4,136,440	4,135,449	4,134,463	4,130,540	4,127,627
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry $\times$ Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Goods and service sector effects are estimated via interaction with a service sector dummy. Firms are classified by their primary SIC code. Difference = service – goods. Driscoll–Kraay standard errors in parentheses (bandwidth 12). F -test p -values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.4.3 Stock Market Responses by Procurement Object

This section examines whether stock price responses to procurement shocks differ by the object of procurement. Specifically, I classify procurement contracts into product contracts, R&D service contracts, and non-R&D service contracts. Product contracts include contracts for physical goods, such as aircraft components, military vehicles, and communication equipment. R&D service contracts include research-and-development-related activities, such as technology development, prototype design, testing, and experimental work. Non-R&D service contracts refer to services that are not directly related to research and development, such as facility management, logistics support, administrative services, and routine maintenance.

Figure 2.6 and Table 2.7 show that market responses to procurement shocks differ substantially by procurement object. For product contracts, the short-run response is not particularly large, but the effect accumulates over time, with the cumulative excess return rising to 1.677 at the 10-day horizon and becoming statistically significant at the 1% level. R&D service contracts exhibit a more immediate response. A significant positive response of 0.660 already appears at $h = 1$, and the positive effect persists through $h = 10$. At the 10-day horizon, the cumulative effect of R&D service contracts is estimated to be 1.419.

By contrast, although non-R&D service contracts also generate statistically significant positive responses at some horizons, their economic magnitude is relatively limited. At the 10-day horizon, the cumulative effect of non-R&D service contracts is only 0.234, much smaller than those for product contracts and R&D service contracts. These differences are also supported statistically. The difference between product contracts and non-R&D service contracts is 1.443 at the 10-day horizon and is significant at the 5% level. The difference between R&D service contracts and non-R&D service contracts is also significant at the 10% level at the 10-day horizon, with an estimated difference of 1.185. In particular, the difference between R&D service contracts and non-R&D service contracts is already significant at the 5% level at $h = 1$, suggesting that R&D-related contracts are incorporated into market valuations more quickly.

These results are consistent with the view that investors interpret the economic meaning of procurement contracts differently depending on the object of procurement. Product procurement involves tangible deliverables and relatively well-defined outputs, allowing investors to assess more directly how the contract may translate into future revenues and cash flows. R&D service contracts may provide information beyond immediate revenue effects, signaling a firm's technological capabilities, innovative capacity, and potential for future follow-on contracts. As a result, R&D service contracts appear to generate strong stock price responses even in the short run. By contrast, non-R&D service contracts often have a more operational-support nature, such as facility management, logistics support, and administrative services, and may therefore provide relatively less new information about the firm's growth prospects or technological advantage. Hence, the larger market responses to product and R&D service contracts suggest that investors view these contracts not simply as increases in government spending, but as signals of concrete profit opportunities or future growth potential.

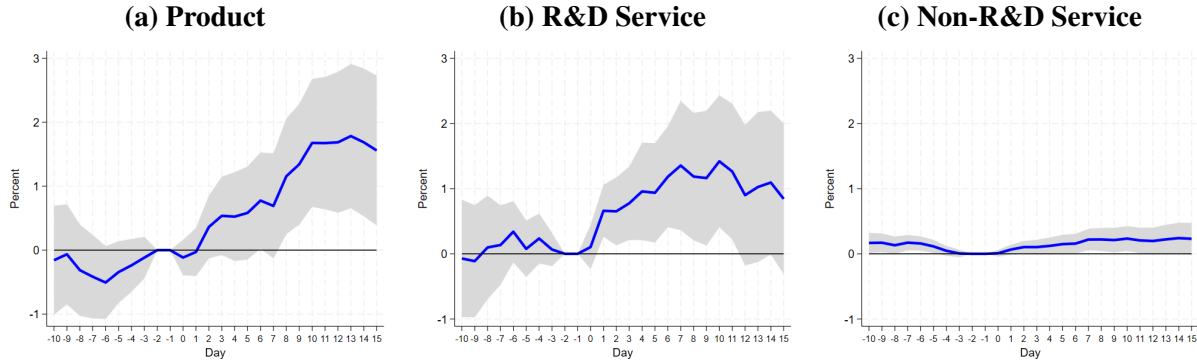


Figure 2.6: Daily Stock Price Responses by Procurement Object

Notes: This figure illustrates how the market valuation of a contract varies by the nature of the procurement target. Panel (a) shows the responses for Products. Panel (b) shows the responses for R&D service. Panel (c) shows the responses for non-R&D service. Contracts are classified using Product Service Codes (PSC). Shaded areas denote 90% confidence intervals. Driscoll–Kraay standard errors with bandwidth 12.

Table 2.7: Daily Stock Price Responses by Procurement Object

	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 7$	$h = 10$
Product [A]	-0.116 (0.169)	-0.029 (0.229)	0.362 (0.299)	0.537 (0.372)	0.692 (0.499)	1.677*** (0.609)
R&D service [B]	0.104 (0.206)	0.660*** (0.243)	0.652** (0.317)	0.775** (0.342)	1.355** (0.603)	1.419** (0.615)
Non-R&D service [C]	0.008 (0.032)	0.066 (0.046)	0.103* (0.057)	0.103 (0.068)	0.219** (0.100)	0.234** (0.116)
Difference [A – C]	-0.124 [0.473]	-0.095 [0.688]	0.259 [0.393]	0.434 [0.251]	0.473 [0.359]	1.443** [0.021]
p -value						
Difference [B – C]	0.096 [0.642]	0.594** [0.016]	0.549* [0.086]	0.672* [0.054]	1.136* [0.065]	1.185* [0.062]
p -value						
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry $\times$ Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,578,504	5,577,160	5,575,865	5,574,575	5,569,438	5,565,605

Notes: Product, R&D service, and Non-R&D service are classified by PSC and included simultaneously in the same regression. Driscoll–Kraay standard errors in parentheses (bandwidth 12). F -test p -values in brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5 Effects of Procurement on Firm Investment

We now examine how procurement contracts lead to actual corporate adjustments in capital expenditure and R&D. Based on the results in Section 2.4.1, this analysis focuses on the baseline sample of definitive contracts and DOs under multiple-award IDVs, which are more informative than DOs under single-award IDVs. Appendix 2.D shows that these informational differences across contract forms are also reflected in the results for quarterly real variables.

Our key finding is that financial market valuations and real adjustments do not correspond one-to-one. Non-specialized-sector contracts generate stronger stock price reactions, yet it is specialized-sector contracts that drive CAPX expansion—while R&D investment moves in the opposite direction. Similarly, product contracts and R&D service contracts both elicit large market responses, yet they trigger fundamentally different types of real adjustment. These patterns suggest that the corporate impact of government procurement cannot be summarized by a single real response; instead, it is necessary to distinguish between a valuation channel, where the re-evaluation of expectations is reflected in stock prices, and an investment adjustment channel, where physical and knowledge capital are expanded along different dimensions depending on the nature of the contract.

2.5.1 Investment Responses by Sector Specialization and Firm Sector

The contrast between CAPX and R&D is most striking along the sector specialization dimension. According to Figure 2.7 and Table 2.8, CAPX increases markedly only for specialized-sector contracts, with coefficients of 0.075, 0.071, and 0.085 at $h = 4, 5$, and 6, respectively. By contrast, the corresponding coefficients for non-specialized-sector contracts are stagnant at 0.013, 0.013, and 0.010, and the difference between the two groups is statistically significant at all three horizons. R&D investment exhibits a mirror-image pattern: non-specialized-sector contracts generate positive and significant R&D coefficients of 0.143, 0.146, and 0.179 at $h = 4, 5$, and 6, while specialized-sector contracts yield negative coefficients of -0.121 , -0.083 , and -0.047 . The difference is again

statistically significant throughout. These findings are consistent with the interpretation that contracts aligned with the firm’s existing business primarily involve the expansion of existing production lines, where additional physical capital is required to meet new demand, whereas contracts outside the firm’s specialized sector are accompanied by the accumulation of new capabilities and knowledge capital.

A coarser classification by firm sector does not generate meaningful differences in either CAPX or R&D responses. As shown in Figure 2.8 and Table 2.9, both goods-sector and service-sector firms exhibit modestly positive CAPX coefficients and positive R&D coefficients at longer horizons, but the differences between the two groups are not statistically significant at any horizon for either outcome. This confirms that what matters for real adjustment is the contract-firm alignment at the granular industry level, rather than the firm’s broad sectoral classification.

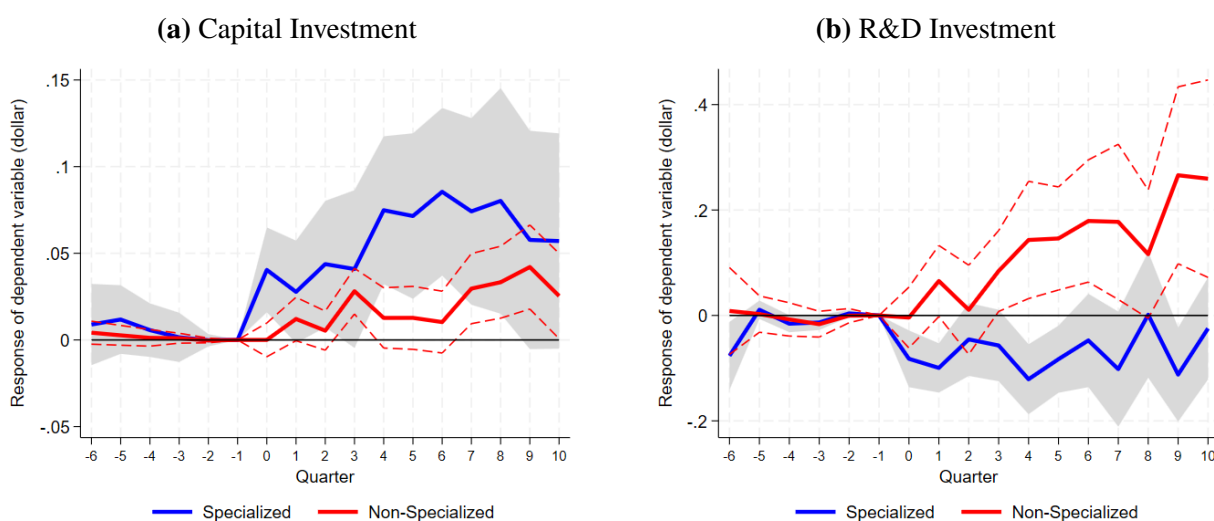


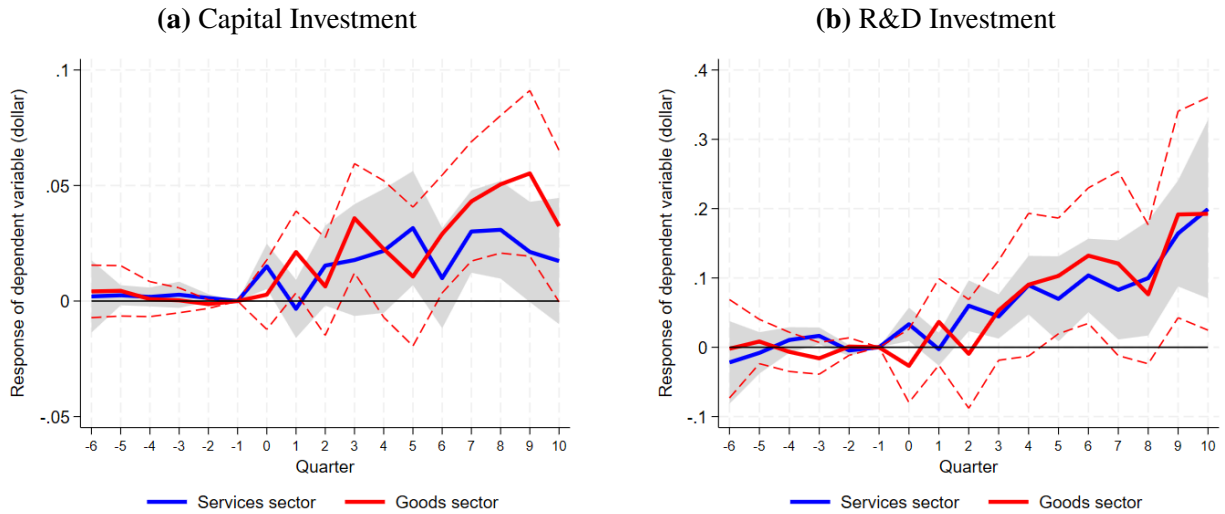
Figure 2.7: Quarterly Real Responses by Sector Specialization

Notes: This figure plots the quarterly cumulative responses of physical and knowledge capital adjustments by sector specialization. Panel (a) shows the response of capital investment (CAPX normalized by lagged PP&E), and Panel (b) shows the response of R&D investment (XRD normalized by lagged total assets). Specialized-sector contracts are those where the contract’s 4-digit NAICS code matches the firm’s primary industry. Shaded areas represent 90% confidence intervals. Driscoll–Kraay standard errors with bandwidth 4.

Table 2.8: Capital Investment and R&D Responses by Sector Specialization

	Capital Investment			R&D Investment		
	$h = 4$	$h = 5$	$h = 6$	$h = 4$	$h = 5$	$h = 6$
Specialized sector	0.075*** (0.026)	0.071** (0.029)	0.085*** (0.029)	-0.121*** (0.040)	-0.083** (0.039)	-0.047 (0.054)
Non-specialized sector	0.013 (0.011)	0.013 (0.011)	0.010 (0.011)	0.143** (0.068)	0.146** (0.059)	0.179** (0.070)
Difference	-0.062** [0.027]	-0.059* [0.063]	-0.075** [0.018]	0.264*** [0.003]	0.229*** [0.008]	0.227** [0.038]
Observations	74,728	73,527	72,328	32,321	31,696	31,072
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry $\times$ Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Horizons are quarterly. Capital investment is capital expenditure divided by lagged PP&E; R&D investment is R&D expenditure divided by lagged total assets. Specialized-sector and non-specialized-sector contracts are classified by whether the contract's 4-digit NAICS code matches the firm's primary industry. Both categories are included simultaneously. Difference = non-specialized – specialized. Standard errors in parentheses (Driscoll–Kraay). F -test p -values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Figure 2.8:** Quarterly Real Responses by Firm Sector

Notes: This figure plots the quarterly cumulative responses of physical and knowledge capital adjustments by firm sector. Panel (a) shows the response of capital investment (CAPX normalized by lagged PP&E), and Panel (b) shows the response of R&D investment (XRD normalized by lagged total assets). Firms are classified into Goods and Service sectors by their primary SIC codes. Shaded areas represent 90% confidence intervals. Driscoll–Kraay standard errors with bandwidth 4.

Table 2.9: Capital Investment and R&D Responses by Firm Sector

	Capital Investment			R&D Investment		
	$h = 4$	$h = 5$	$h = 6$	$h = 4$	$h = 5$	$h = 6$
Goods sector	0.023 (0.018)	0.011 (0.018)	0.029* (0.015)	0.090 (0.063)	0.103** (0.051)	0.132** (0.059)
Service sector	0.022 (0.016)	0.032** (0.015)	0.010 (0.013)	0.090*** (0.026)	0.070* (0.037)	0.104*** (0.032)
Difference	-0.001	0.021	-0.019	0.000	-0.033	-0.029
p -value	[0.975]	[0.384]	[0.316]	[0.992]	[0.602]	[0.671]
Observations	58,144	57,195	56,250	30,182	29,610	29,035
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry $\times$ Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Horizons are quarterly. Capital investment is capital expenditure divided by lagged PP&E; R&D investment is R&D expenditure divided by lagged total assets. Goods and service sector effects are estimated via interaction with a service sector dummy. Firms are classified by their primary SIC code. Difference = service – goods. Standard errors in parentheses (Driscoll–Kraay). F -test p -values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5.2 Investment Responses by Procurement Object

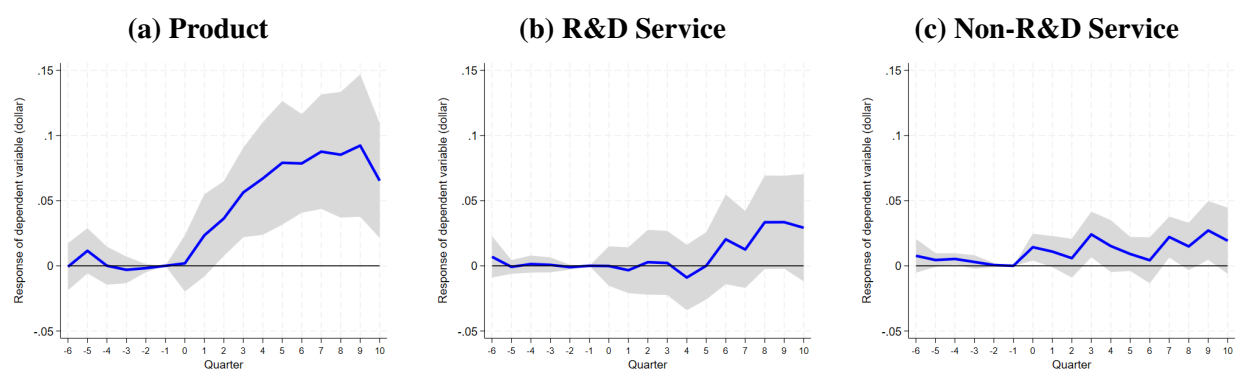
Real effects by procurement object exhibit a comparable pattern. According to Figure 2.9 and Table 2.10, CAPX increases are most pronounced for product contracts, with coefficients of 0.067, 0.079, and 0.079 at $h = 4, 5$, and 6 , respectively. Over the same horizons, the CAPX coefficients for non-R&D services are only 0.015, 0.009, and 0.004, with the difference between the two categories being statistically significant at $h = 5$ and $h = 6$. R&D service contracts show no meaningful CAPX response. This pattern is intuitive: product procurement involves tangible deliverables that are more likely to require the expansion of physical production capacity, whereas service contracts place demands primarily on labor and organizational inputs.

R&D investment, by contrast, reacts more strongly to R&D service contracts. The coefficients for R&D services are positive and significant at 0.158 and 0.156 at $h = 5$ and $h = 6$, whereas the R&D response to product contracts remains near zero throughout. However, since the difference between R&D services and non-R&D services is not consistently significant, these results are best interpreted as suggesting that service contracts in general—rather than products—are more likely to

be associated with knowledge capital adjustments.

Taken together with the sector specialization results, the evidence in this section reveals a clear pattern: product procurement and specialized-sector contracts primarily expand physical capital, while non-specialized-sector contracts and service-type contracts expand knowledge capital through R&D investment. Government procurement therefore operates through at least two distinct channels: a valuation channel, through which markets reprice firms when new information arrives, and an investment adjustment channel, through which firms expand either physical capital or knowledge capital depending on the nature of the contract.

I. Capital Investment



II. R&D Investment

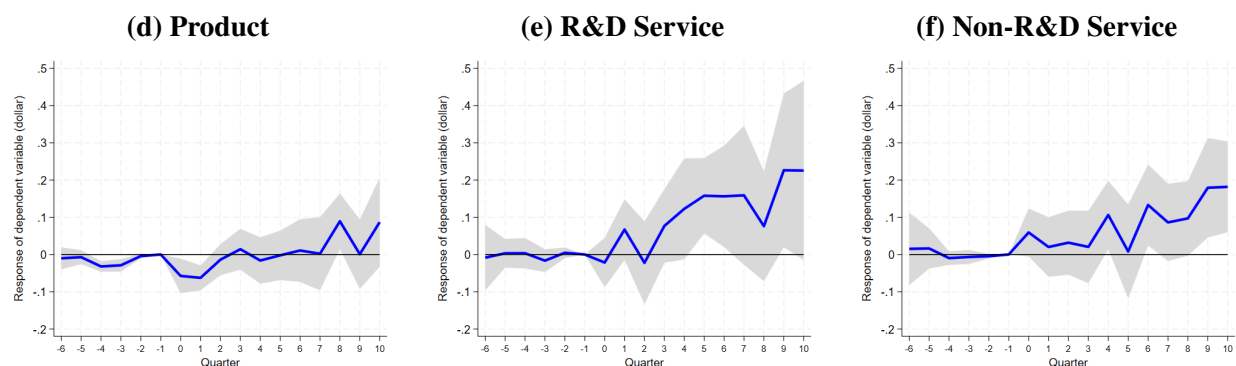


Figure 2.9: Investment Responses by Procurement Object

Notes: This figure shows how capital investment (Panel I) and R&D investment (Panel II) responses vary by the nature of the procurement target. In each panel, subfigure (a) shows the responses for Products, subfigure (b) for R&D service, and subfigure (c) for non-R&D service. Contracts are classified using Product Service Codes (PSC). Shaded areas denote 90% confidence intervals. Driscoll–Kraay standard errors with bandwidth 4.

Table 2.10: Investment and R&D Responses by Procurement Object

	Capital Investment			R&D Investment		
	$h = 4$	$h = 5$	$h = 6$	$h = 4$	$h = 5$	$h = 6$
Product [A]	0.067** (0.026)	0.079*** (0.029)	0.079*** (0.023)	-0.016 (0.038)	-0.002 (0.040)	0.011 (0.051)
R&D service [B]	-0.009 (0.015)	0.000 (0.016)	0.020 (0.021)	0.123 (0.082)	0.158** (0.061)	0.156* (0.083)
Non-R&D service [C]	0.015 (0.012)	0.009 (0.008)	0.004 (0.011)	0.106* (0.056)	0.008 (0.077)	0.133** (0.066)
Difference [A – C]	0.052 [0.108]	0.070** [0.034]	0.074*** [0.008]	-0.122* [0.090]	-0.010 [0.912]	-0.122 [0.211]
p -value						
Difference [B – C]	-0.024 [0.257]	-0.009 [0.637]	0.016 [0.459]	0.017 [0.868]	0.150 [0.126]	0.023 [0.843]
p -value						
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry $\times$ Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	74,728	73,527	72,328	32,321	31,696	31,072

Notes: Horizons are quarterly. Capital investment is capital expenditure divided by lagged PP&E; R&D investment is R&D expenditure divided by lagged total asset. Product, R&D service, and Non-R&D service are classified by PSC and included simultaneously. Standard errors in parentheses (Driscoll–Kraay). F -test p -values in brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.6 Conclusion

This study has analyzed how the institutional structure and content of public procurement contracts determine the quality of information transmitted to financial markets and subsequently lead to real investment adjustments by firms. To this end, we tracked the transmission channels of procurement shocks by linking individual transaction records from U.S. federal procurement data (FPDS-NG) for the period 2000–2024 with daily stock prices (CRSP) and quarterly financial data (COMPUSTAT). The key findings are as follows:

First, this paper shows that the market effect of procurement contracts varies depending on contract structure. Contract structure determines the extent to which uncertainty about the winning firm, contract size, and scope of work is newly resolved at the time of the award, and differences in this informational content lead to heterogeneous stock market responses. Indeed, the stock market responds clearly to definitive contracts and delivery orders issued under multiple-award

IDVs, whereas it does not respond significantly to delivery orders issued under single-award IDVs. This suggests that procurement contracts should not be treated as uniform government demand shocks. Instead, it is important to distinguish contracts that convey new information to the market at the time of award.

Second, financial market valuations and actual investment adjustments do not necessarily move in tandem. The stock market reacts more strongly to contracts in non-core sectors, suggesting that investors interpret these contracts as signals of firms' latent capabilities or potential entry into new markets. In contrast, real investment responses are driven less by expected profitability than by the type of capabilities required to fulfill the contract. Physical investment tends to expand mainly in response to contracts related to firms' core sectors or product procurement, whereas R&D investment increases in response to non-core sector contracts or R&D service contracts that require new technological capabilities. These results show that the translation of government demand into real investment is not determined solely by market-value-based investment incentives, such as Tobin's q , but also depends on the nature of the production and knowledge assets required for contract fulfillment.

In short, government procurement operates through two distinct channels: a *valuation channel*, where markets immediately reprice firms upon the arrival of new information, and an *investment adjustment channel*, where firms expand either physical or knowledge capital depending on the nature of the contract. These findings imply that the transmission of procurement shocks cannot be understood through a single average multiplier or a simple real-side response alone. The specific impact of a government procurement shock on the macroeconomy depends fundamentally on the contract form, informational novelty, industrial alignment, and the type of productive capacity required for execution.

Chapter 3

Demographics and Housing Booms

1

This paper studies whether demographic age structure is associated with housing-boom vulnerability. In an OECD panel of 29 countries over 1970–2021 and a U.S. county panel over 1975–2024, locations with higher young-to-old population ratios are more likely to experience large housing booms. In the country and time fixed-effects specification, a one-unit increase in the young-to-old ratio is associated with an 18.6 percentage-point higher boom probability; in the corresponding U.S. county specification, the estimate is 4.3 percentage points. HMDA mortgage data further show that young applicants increase mortgage-market participation disproportionately in boom county-years, with similar results for loan originations. To rationalize these facts, we study a six-generation housing OLG model with tenure choice, gradual price adjustment, adaptive house-price-growth beliefs, and age-specific income profiles. The model shows that a young economy can generate a larger house-price response to the same housing-preference shock.

¹This chapter was co-authored with Yun Gu Cho, Jinho Kim, and Jaeyoung Yoo.

3.1 Introduction

The relationship between demographic structure and housing markets has long been a central topic in macroeconomics and housing economics. A substantial body of existing research has shown that the age composition of households is an important determinant of the level and long-run trend of housing demand. For example, it is well known that demand for housing space and homeownership rises sharply at particular stages of the life cycle, while housing consumption and asset composition change after retirement (Mankiw and Weil, 1989; Bakshi and Chen, 1994; Takáts, 2012; Nishimura and Takáts, 2012).

However, much less is known about whether demographic structure affects not only these long-run trends, but also the likelihood and magnitude of house price booms. Since young households are more likely to be located at the margin between homeownership and renting, actual home purchases may increase more strongly in response to the same demand shock. In other words, demographic composition may be an important state variable governing both the likelihood and the magnitude of house price booms. Understanding this relationship is important for identifying the sources of housing-market fluctuations and for designing the timing and intensity of policy responses. If demographic structure amplifies or dampens the effects of a given shock, demographic composition can be interpreted as a key state variable for predicting the risk of house price booms and financial instability.

This paper shows that demographic age structure is systematically associated with the incidence of housing booms and this pattern is consistent with a life-cycle housing-demand mechanism. Using data at both the OECD country level and the U.S. regional level, we show that locations with relatively younger populations are more likely to experience large upward deviations of house prices from trend. We then examine whether this empirical pattern can be reproduced using a simple overlapping-generations (OLG) model.

Our empirical analysis begins by documenting the relationship between age structure and the

incidence of housing booms. We define a large housing boom as an episode in which the cyclical component of real house prices is more than one standard deviation above its region-specific trend. Using both a panel of 29 OECD countries from 1970 to 2021 and a U.S. county-year panel from 1975 to 2024, we show that locations with a higher ratio of young adults to older adults are more likely to experience large upward deviations of house prices from trend.

In the linear probability model using the country panel, a one-unit increase in the young-to-old ratio, defined as the population aged 20–39 relative to the population aged 50–69, is associated with a statistically significant 18.6 percentage point increase in the probability of a large housing boom. The corresponding estimate from the U.S. county panel implies a 4.3 percentage point increase in the probability of a housing boom. Thus, across different datasets and specifications, the evidence robustly indicates that a higher young-to-old ratio is associated with a greater incidence of housing booms.²

Next, we examine the household-level demand margins underlying this aggregate pattern. The empirical evidence based on the young-to-old ratio shows that housing booms occur more frequently in younger regions, but in principle this could reflect other regional factors correlated with age structure. A natural implication of the life-cycle housing-demand view is that, when a local housing boom occurs, younger households should increase home purchases more actively than older households.

We test this implication using mortgage application data from the Home Mortgage Disclosure Act (HMDA). Specifically, we examine whether young applicants increase home-purchase activity more strongly when local house prices are in a boom, thereby contributing disproportionately to mortgage-market activity. The results show that, in counties experiencing a housing boom, the application intensity of young applicants is 4.3–6.6 percent higher when normalized by the boom-period average of the older comparison group. These findings are robust across alternative

²This relationship does not fully rule out the possibility of reverse causality, whereby young households move into regions experiencing rapid house-price growth. However, the fact that the results are observed not only in U.S. county-level data, where migration can operate relatively quickly, but also in country-level panels is more directly consistent with a life-cycle housing-demand mechanism than with a migration explanation.

definitions of age groups, application-based outcome variables, and more stringent boom thresholds.

In the second part of the paper, we examine whether these empirical patterns can be reproduced within a theoretical model. To this end, we construct an OLG model with six generations that differ in age, income, asset holdings, housing tenure, and survival probabilities. Households choose consumption and savings, and decide whether to own or rent housing. A housing preference shock raises the utility value of homeownership, while house prices adjust gradually in response to excess demand. Income follows a hump-shaped life-cycle profile: young households have relatively low income, income peaks in middle age, and then declines in old age.

In simulations that apply the same housing preference shock, house prices rise by 15.6 percent at the peak in the young economy, compared with 13.7 percent in the old economy. Thus, a life-cycle housing OLG model that incorporates both demographic structure and age-specific income differences can qualitatively reproduce the key empirical pattern. When a housing preference shock raises the value of homeownership, the response is strongest among households that have sufficient income and assets to transition into homeownership but whose housing demand is not yet fully saturated. Because the young economy has a larger share of households located on this margin, it exhibits a stronger house-price response than the old economy.

Related literature. This paper is related to three strands of literature. First, the paper contributes to the literature on demographics and housing markets. From a life-cycle perspective, age structure shapes household formation, saving behavior, portfolio allocation, and the demand for owner-occupied housing (Bakshi and Chen, 1994). The classic “baby boom” hypothesis of Mankiw and Weil (1989) argues that changes in cohort size generate large shifts in housing demand and house prices. Subsequent work shows that population aging can affect house prices and asset valuations over long horizons (Takáts, 2010; Takáts, 2012; Nishimura and Takáts, 2012; Gevorgyan, 2019). This literature establishes that demographics are a powerful determinant of long-run housing demand and valuation. Our paper shifts the focus from long-run price levels to boom vulnerability. We show that demographic composition also predicts the likelihood of large upward deviations of

house prices from trend.

Second, the paper contributes to the literature on housing booms, credit constraints, and macro-financial amplification. A large literature studies housing booms as unusually large deviations of real house prices from trend and emphasizes the macro-financial risks associated with asset-price booms (Borio and Lowe, 2002; Detken and Smets, 2004; Goodhart and Hofmann, 2008; Jordà, Schularick, et al., 2015). Closely related work studies how credit conditions, household leverage, housing finance, and beliefs amplify housing boom-bust cycles (Mian and Sufi, 2011; Favilukis et al., 2017; Justiniano et al., 2019; Kaplan et al., 2020). Our analysis builds on this literature by asking why similar housing-market shocks generate larger booms in some locations than in others. While urban research emphasizes housing-supply elasticity as a key source of local boom sensitivity (Glaeser et al., 2008; Saiz, 2010), we highlight a complementary demand-side state variable: the local young-to-old population ratio. Younger demographic structures place more households near the owner-renter margin, amplifying mortgage-market participation and upward deviations of house prices from trend.

Third, the paper contributes to quantitative life-cycle and OLG approaches to housing demand. The quantitative housing-finance models discussed above provide rich accounts of credit, belief, and leverage channels in housing boom-bust episodes. Our model is deliberately more parsimonious and isolates a different mechanism: how demographic weights interact with tenure choice and age-specific income profiles. By embedding tenure choice, demographic weights, gradual price adjustment, and age-specific income profiles in a six-generation housing model, we show that the effect of demographics on boom amplification depends on more than the mechanical share of young households. The model clarifies why intergenerational income differences are important for reproducing the empirical pattern: they determine which cohorts are close to the owner-renter margin and therefore how strongly aggregate demand responds to a common preference shock.

The rest of the paper proceeds as follows. Section 3.2 presents the empirical evidence on demographics and large housing booms. Section 3.3 develops the housing OLG model and uses it to

assess whether the empirical pattern can be rationalized by a life-cycle housing-demand mechanism. Section 3.4 concludes.

3.2 Effects of Demographics on Housing Booms

3.2.1 Data

Our empirical analysis combines several datasets at both the country and regional levels. For the cross-country analysis, we use OECD house price indicators, OECD population data by age group, and real GDP per capita for 29 countries over the period 1970–2021. For the U.S. county analysis, we use annual FHFA county house price indexes over 1975–2024, together with county population data and housing supply elasticity estimates from Saiz (2010). The number of county-year observations differs across specifications because housing-price availability and the availability of the supply-elasticity control are not identical across counties.

Table 3.1: Summary of Datasets

Analysis	Source	Unit	Sample	Variables
Cross-country	OECD	Country $\times$ year	1970–2021	Real house prices, real GDP per capita, population by age group
U.S. county	FHFA, U.S. Census	County $\times$ year	1975–2024	Real house prices, population by age group, supply elasticity

Notes: For the U.S. regional analysis, housing supply elasticity is from Saiz (2010).

3.2.2 Large Housing Booms

We define a large housing boom as an episode in which real house prices rise substantially above their long-run trend. This definition builds on the literature on housing booms and bubbles. For example, Borio and Lowe (2002), Detken and Smets (2004), and Goodhart and Hofmann (2008) define housing booms as periods in which real house prices rise sufficiently above their HP-filtered trend under relatively high smoothing parameters. Jordà, Schularick, et al. (2015) identify housing bubble episodes using two conditions: the cyclical component of real house prices must exceed one

standard deviation, and house prices must subsequently decline by at least 15 percent within three years.

In this study, we follow Jordà, Schularick, et al. (2015) in using the HP filter (Hodrick and Prescott, 1997) to extract the cyclical component of real house prices, but we omit the subsequent price-decline condition. By dropping this ex-post crash requirement, we retain episodes in which house prices rise sharply above trend, regardless of whether a large decline materializes within a specific window.

This modification is motivated by our different research objective. Whereas the definition in Jordà, Schularick, et al. (2015) is designed to identify boom-bust episodes and study the macro-financial consequences of bursting housing bubbles, our focus is on the incidence of large upward house-price deviations associated with demographic structure. Requiring an ex-post price collapse would condition the boom definition on a subsequent outcome that may depend on factors outside the demographic mechanism, such as credit conditions, policy responses, housing supply, or monetary tightening. We therefore retain the price-elevation criterion but omit the subsequent crash requirement, so that our boom indicator captures large upward deviations in real house prices rather than realized boom-bust episodes.

We decompose real house prices into trend and cyclical components using the HP filter:

$$p_{i,t} = \tau_{i,t} + c_{i,t}, \quad (3.1)$$

where $p_{i,t}$ denotes the log real house price in location i at time t , $\tau_{i,t}$ is the trend component, and $c_{i,t}$ is the cyclical component. Because both the country and county panels are annual, we set the HP smoothing parameter to $\lambda = 100$.

We then define the large housing boom indicator as

$$\text{LHB}_{i,t} = \begin{cases} 1, & \text{if } c_{i,t} > \sigma_{c_i}, \\ 0, & \text{otherwise,} \end{cases} \quad (3.2)$$

where σ_{c_i} is the standard deviation of $c_{i,t}$ for location i over the sample period. Thus, a large housing boom is identified whenever the cyclical component of real house prices exceeds one standard deviation of the location-specific cyclical component. Because the HP trend and σ_{c_i} are constructed from the available sample, the indicator should be interpreted as an ex-post classification of large upward deviations rather than as a real-time forecast signal.

3.2.3 Young-to-Old Population Ratio

The literature does not adopt a single age cutoff for distinguishing young, middle-aged, and old cohorts. The relevant boundaries vary across studies depending on the question of interest. In the housing literature, age profiles are typically motivated by life-cycle patterns in household formation, housing demand, and asset accumulation. The demographic literature similarly uses alternative cohort definitions depending on the mechanism under consideration.

We define the young cohort as individuals aged 20–39 and the older cohort as individuals aged 50–69. The numerator is intended to capture younger adults who are entering the housing market and forming housing demand, whereas the denominator captures older households who are, on average, more likely to have accumulated housing wealth and to adjust housing consumption or rebalance their portfolios. Accordingly, our measure should be interpreted not as a conventional old-age dependency ratio, but as a cohort-balance measure designed to capture life-cycle patterns in housing demand and asset holdings.

We define the young-to-old population ratio as

$$\text{YO}_{i,t} \equiv \frac{\text{Population}_{i,t}^{(20-39)}}{\text{Population}_{i,t}^{(50-69)}}, \quad (3.3)$$

where a higher value of $YO_{i,t}$ indicates a relatively larger young cohort compared with the older cohort. In the Appendix, we examine the robustness of the results to alternative age-group definitions.

3.2.4 Descriptive Evidence

Before turning to regression analysis, we present descriptive evidence on the relationship between demographic structure and large housing booms. Figure 3.1 plots the probability of a large housing boom against the young-to-old population ratio in the country panel and the U.S. county panel. Panel (a) displays the country-level relationship, while Panel (b) shows the analogous relationship at the U.S. county level.³

The country-level plot reveals a clear positive association: countries with a higher young-to-old ratio tend to experience large housing booms more frequently. The U.S. county-level plot shows a similar positive relationship, although the association is weaker than in the country panel. This difference is not surprising given the finer geographic resolution and greater heterogeneity of local housing markets. These patterns are descriptive, but they are consistent with the hypothesis that younger demographic structures are associated with more frequent large housing booms and motivate the regression analysis that follows.

3.2.5 Demographics and Large Housing Booms

We estimate a sequence of linear probability model (LPM) specifications to examine whether locations with a higher young-to-old population ratio are more vulnerable to large housing booms. In both the cross-country and U.S. county analyses, we begin with parsimonious specifications and then progressively introduce additional controls and fixed effects in order to assess the robustness of

³The country-level plot is reported without population weights because each country is treated as a national housing-market regime, with its own macroeconomic, institutional, and mortgage-market environment. The U.S. county-level plot, by contrast, uses county population weights. Counties differ substantially in market size, and treating very small and very large counties symmetrically can give disproportionate influence to idiosyncratic demographic variation in small local markets. Because the proposed mechanism operates through household demand and local housing-market participation, population provides a natural proxy for the size of the relevant housing market. Appendix Figure 3.A.1 reports the corresponding unweighted county-level bin-scatter plot; the relationship is weaker but remains positive.

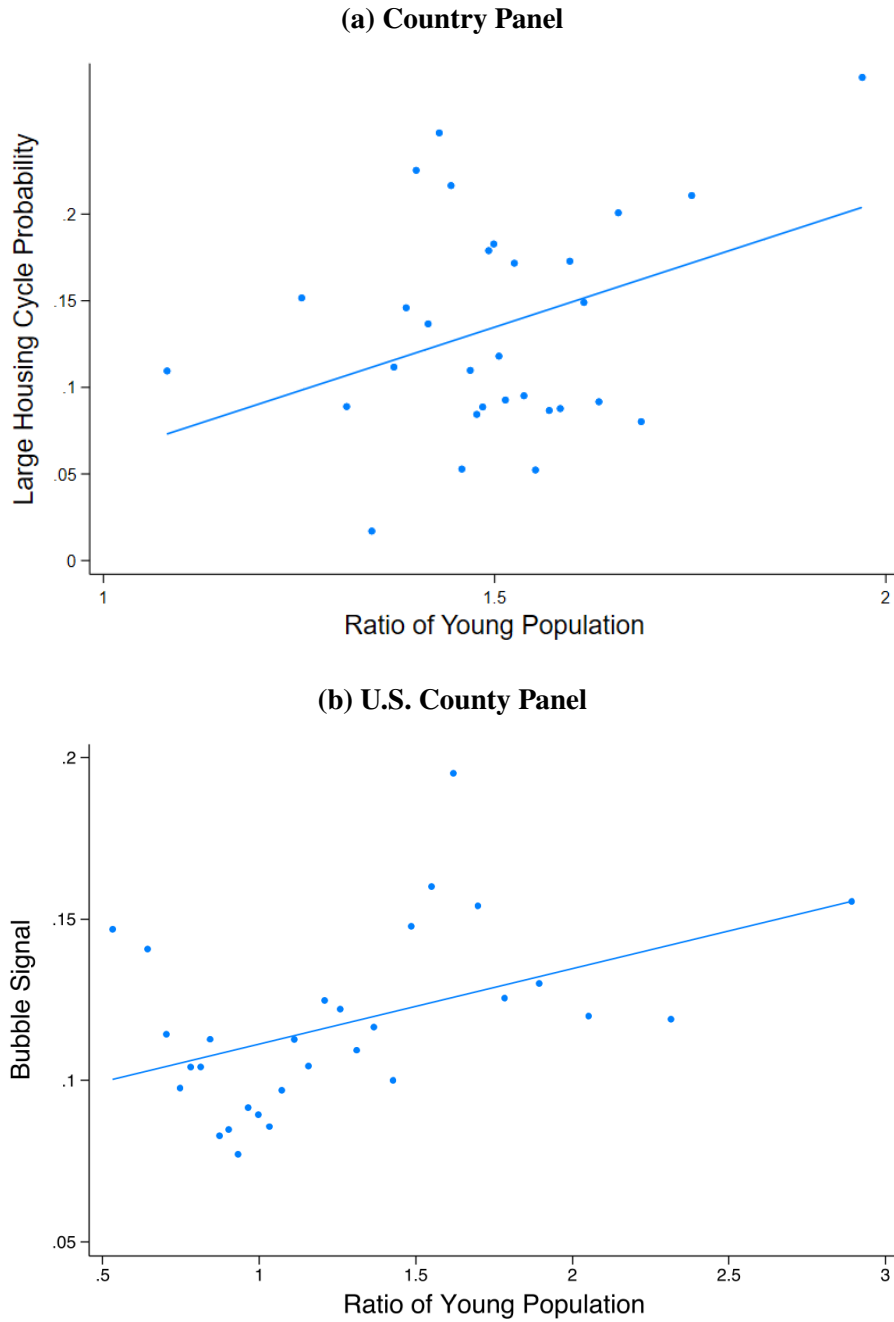


Figure 3.1: Bin Scatter Plot of Large Housing Booms and Young-to-Old Population Ratio

Notes: This figure presents bin-scatter plots of the large housing boom probability against the young-to-old population ratio. Panel (a) shows the country-level result and Panel (b) shows the U.S. county-level result. Observations are grouped into 30 equal-sized bins based on the young-to-old ratio, and each plot displays the mean large housing boom probability within each bin along with a linear fit.

the estimated relationship.

Our baseline specification is given by

$$\text{LHB}_{i,t} = \beta_0 + \beta_1 \text{YO}_{i,t} + Z'_{i,t} \boldsymbol{\phi} + \alpha_i + \gamma_t + \varepsilon_{i,t}, \quad (3.4)$$

where $\text{LHB}_{i,t}$ is an indicator equal to one if location i experiences a large housing boom at time t , $\text{YO}_{i,t}$ ⁴ is the young-to-old population ratio, $Z_{i,t}$ is a vector of control variables, α_i denotes location-specific fixed effects when included, and γ_t denotes time fixed effects when included. Standard errors are clustered at the location level throughout.

In the cross-country analysis, the control vector includes log real GDP per capita, while in the U.S. county analysis it includes housing supply elasticity. Across specifications, we progressively add controls, location fixed effects, and time fixed effects. Because the LPM is linear, the coefficient β_1 is directly interpretable as the change in the probability of a large housing boom associated with a one-unit increase in the young-to-old ratio, holding other covariates constant. This approach allows us to assess whether the relationship between demographic structure and large housing booms is robust to various specifications without the complications arising from nonlinear models.

Tables 3.2 and 3.3 report the results. Across both the cross-country and U.S. county samples, the coefficient on the young-to-old ratio is positive and statistically significant in all specifications, indicating that locations with younger demographic structures have higher incidence of large upward deviations of house prices from trend.

Table 3.2 shows a strong and robust relationship in the country panel. The coefficient on the young-to-old ratio is positive in all five specifications, ranging from 0.045 in the most parsimonious specification to 0.231 in the specification with country fixed effects and GDP per capita. In the most demanding specification with both country and time fixed effects (column 5), a one-unit increase

⁴We use the contemporaneous young-to-old ratio because it is a slow-moving demographic state variable. Since the ratio is constructed from broad twenty-year age bins, its one-year lag is mechanically very close to the contemporaneous value, especially in the cross-country panel where annual changes in age composition are limited. Hence, we interpret β_1 as a conditional demographic gradient in housing-boom vulnerability, not as the causal effect of an exogenous demographic shock.

in the young-to-old ratio is associated with an 18.6 percentage-point increase in the probability of a large housing boom. Because this variable is a ratio rather than a standardized regressor, the one-unit coefficient is best interpreted as the slope of the boom-probability relationship. The estimated relationship remains statistically significant after introducing country and time fixed effects, suggesting that the result is not driven solely by time-invariant cross-country heterogeneity or common global shocks. GDP per capita enters positively and is statistically significant at the 5% level in both specifications in which it is included, indicating that higher income levels are also associated with a greater likelihood of large housing booms, though the demographic composition variable remains the more robust predictor.

Table 3.2: Linear Probability Model: Country Panel

	(1)	(2)	(3)	(4)	(5)
YO ratio	0.045*	0.160***	0.231***	0.147**	0.186**
	(0.023)	(0.030)	(0.052)	(0.064)	(0.077)
GDP per Cap.			0.083**		0.323**
			(0.039)		(0.118)
Observations	1,262	1,262	1,225	1,262	1,225
Country FE	No	Yes	Yes	Yes	Yes
Time FE	No	No	No	Yes	Yes
R^2	0.0045	0.0252	0.0298	0.1895	0.2058

Notes: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Standard errors are clustered at the country level.

Table 3.3 shows that the same qualitative relationship holds in the U.S. county panel. The coefficient on the young-to-old ratio is positive and highly significant across all specifications, ranging from 0.013 in the time-FE-only specification (column 3) to 0.075 in the pooled specification (column 1). In the specification with both county and time fixed effects (column 5), a one-unit increase in the young-to-old ratio is associated with a 4.3 percentage-point increase in the probability of a large housing boom. Supply elasticity enters with a negative coefficient in column (4), consistent with the view that more elastic housing supply dampens the translation of demand shifts into house price fluctuations, though the estimate is not statistically significant. The substantial rise in R^2 in specifications with time fixed effects suggests that common temporal shocks explain an important

share of boom incidence in U.S. counties, although the demographic effect remains statistically meaningful.

The difference in the estimated magnitude of the young-to-old population ratio effect across the country and U.S. county panels may reflect several factors. One possible explanation is that the country panel reflects slow-moving differences in national demographic regimes, housing-finance institutions, and macroeconomic conditions. In contrast, the county panel exploits within-country variation across local housing markets, where population mobility, commuting patterns, and idiosyncratic shocks can weaken the mapping from age structure to local housing booms. More broadly, the comparison suggests that while the magnitude of the coefficients differs across samples, the positive relationship between the young-to-old ratio and the likelihood of a large housing boom remains robust across very different empirical settings.

Table 3.3: Linear Probability Model: U.S. County Panel

	(1)	(2)	(3)	(4)	(5)
YO ratio	0.075*** (0.001)	0.044*** (0.006)	0.013*** (0.004)	0.022*** (0.006)	0.043** (0.019)
Supply Elasticity				-0.002 (0.002)	
Observations	101,567	101,567	101,567	37,205	101,567
County FE	No	Yes	No	No	Yes
Time FE	No	No	Yes	Yes	Yes
R^2	0.1178	0.0159	0.2887	0.3198	0.3046

Notes: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. All regressions are weighted by county population. Standard errors are clustered at the county level. Column (4) includes housing supply elasticity; column (5) reports the county and year fixed-effects specification without the elasticity control.

3.2.6 Young Applicants and Mortgage-Market Participation During Booms

The previous regression results connect the two ends of the proposed mechanism: locations with younger demographic structures have higher boom incidence. They do not, however, identify the intermediate demand margin through which demographics may translate into larger housing booms.

To examine this margin directly, we turn to mortgage application data from the Home Mortgage Disclosure Act (HMDA). HMDA provides loan-level data on mortgage applications and their outcomes for covered financial institutions. The data span a broad segment of the U.S. mortgage market and include information on loan purpose, property location, applicant characteristics, and applicant age in decade-based bins.⁵ This makes HMDA a natural source for measuring the extensive-margin demand for housing finance: an application reflects a household’s attempt to obtain mortgage credit, while originated loans provide a narrower measure of completed housing-finance transactions.

We construct a county–age-group–year panel over the 2018–2024 HMDA sample period, which is the period over which applicant age is reported in the public HMDA files. The outcome is mortgage applications per 1,000 age-specific population in county c , age group a , and year t . We identify housing boom county-years using the same large-boom definition as in the country and county analyses above: the cyclical component of real house prices must exceed one standard deviation above its location-specific trend. We then estimate the following triple-difference specification:

$$Y_{c,a,t} = \beta (\text{Boom}_{c,t} \times \text{Young}_a) + \alpha_{c,a} + \gamma_{t,a} + \delta_{c,t} + \varepsilon_{c,a,t}, \quad (3.5)$$

where $Y_{c,a,t}$ denotes applications per 1,000 age-specific population. The fixed effects are county-by-age effects $\alpha_{c,a}$, year-by-age effects $\gamma_{t,a}$, and county-by-year effects $\delta_{c,t}$. The specification is highly saturated: the county-by-year fixed effects absorb all county-year shocks and any level effect of the boom indicator itself, including local credit conditions, local house-price growth, and other time-varying county-level factors common across age groups. Identification therefore comes from whether, within the same county-year, young applicants increase mortgage-market activity more than older applicants when that county is in a boom. The coefficient β measures the differential increase in young applicants’ application intensity in boom county-years relative to the older comparison

⁵Estimates from the Consumer Financial Protection Bureau (CFPB) indicate that HMDA captures about 92 percent of mortgage originations nationally (2017). Estimates from the Federal Reserve Bank of Minneapolis suggest coverage above 90 percent in metropolitan counties and about 80 percent in remote nonmetropolitan counties (2015).

group. We interpret this specification as evidence on within-county-year age-group differences in mortgage-market participation, rather than as a fully causal estimate of the effect of boom onset on young households.

Table 3.4 reports the results. Across columns (1)–(4), the estimated boom-period young-minus-old differential ranges from 1.00 to 1.69 additional mortgage applications per 1,000 age-specific population. The natural denominator for interpreting these coefficients is the comparison group’s application intensity in boom county-years, because the regression coefficient is the incremental young-versus-old gap during boom episodes. Relative to the older comparison groups’ mean boom-period application intensity, these estimates correspond to roughly 4.3% to 6.6% higher application intensity for young groups. Using the older groups’ full-sample mean application intensity instead yields a somewhat larger range of about 5.3% to 8.0%. These results indicate that young applicants disproportionately increase their mortgage-market activity during housing booms relative to older applicants. The finding complements the LPM evidence by showing not only that younger areas have higher boom incidence, but also that young cohorts are themselves more active participants in the mortgage market precisely when booms occur. This pattern is consistent with the life-cycle interpretation that younger households, who are at the stage of household formation and initial home purchase, respond more strongly to boom-period conditions.

One potential concern is that the increase in mortgage applications by younger applicants does not necessarily reflect actual housing demand. For example, younger households may submit more applications per property by shopping across multiple lenders, or their applications may be less likely to translate into completed home purchases. While HMDA records only applications that enter the lender’s formal underwriting process—that is, oral or written requests submitted according to the institution’s application procedures, rather than informal information inquiries—it is nonetheless useful to examine whether our results are robust to alternative outcomes and comparison groups.

To address this concern, we re-estimate our baseline specification using alternative outcome measures: originated loans and denied applications. Table 3.5 reports the results. When restricting

Table 3.4: HMDA Baseline: Application Intensity in Boom County-Years

	(1)	(2)	(3)	(4)
Young group	25–34	25–44	25–34	25–44
Comparison group	45–74	45–74	55–74	55–74
Boom _{c,t} × Young _a	1.002*** (0.179)	1.346*** (0.227)	1.335*** (0.235)	1.686*** (0.264)
Observations	76,328	95,587	57,161	76,420
Counties	2,779	2,779	2,779	2,779
County-year cells	19,099	19,099	19,099	19,099
County-age cells	11,113	13,892	8,349	11,113
R ²	0.8896	0.8755	0.8991	0.8798

Notes: The outcome is mortgage applications per 1,000 age-specific population. Each column estimates equation (3.5) using the indicated young and comparison groups. Columns (3)–(4) use an older comparison group as a robustness check. Regressions are weighted by age-specific population. Standard errors, clustered at the county level, are reported in parentheses. *** indicates significance at the 1% level.

the outcome to originated loans, the estimated coefficients decline in magnitude but remain large and statistically significant. This stricter outcome indicates that the increased application activity among young applicants during booms translates into higher realized borrowing, rather than reflecting only multiple applications or unsuccessful credit searches. Denied applications also rise significantly for young applicants during boom periods, suggesting that part of the increased activity reflects a broader expansion in credit-seeking behavior. Quantitatively, the origination margin accounts for the bulk of the baseline application response, suggesting that the result largely reflects completed mortgage transactions rather than unsuccessful credit searches. This pattern supports the interpretation that effective housing demand by young households rises disproportionately during boom periods. Additional robustness checks are reported in Appendix Section 3.C.2.

Beyond young households' extensive-margin response, Appendix 3.C.3 examines whether they also amplify housing booms through an intensive-margin channel by taking on higher leverage during boom periods. The estimated effects are modest in magnitude but statistically significant: relative to older borrowers, young borrowers exhibit higher loan-to-value and debt-to-income ratios in boom county-years.

Table 3.5: HMDA Robustness: Mortgage Application Outcomes

	(1)	(2)	(3)	(4)
Young group	25–34	25–44	25–34	25–44
Outcome	Originations	Originations	Denials	Denials
Comparison group	45–74	45–74	45–74	45–74
$\text{Boom}_{c,t} \times \text{Young}_a$	0.820*** (0.123)	0.968*** (0.148)	0.071*** (0.021)	0.105*** (0.024)
Observations	76,328	95,587	76,328	95,587
Counties	2,779	2,779	2,779	2,779
County-year cells	19,099	19,099	19,099	19,099
County-age cells	11,113	13,892	11,113	13,892

Notes: Outcomes are measured per 1,000 age-specific population. Each column estimates equation (3.5) using the indicated outcome and age-group definition. Regressions are weighted by age-specific population. Standard errors, clustered at the county level, are reported in parentheses. *** indicates significance at the 1% level.

3.3 A Housing OLG Model

We next examine whether our empirical pattern can be reproduced in a stylized life-cycle housing environment. To this end, we use a structural overlapping-generations (OLG) model to analyze whether, in response to the same housing-preference shock, a young economy can generate a larger house-price boom response than an older economy. Crucially, we show that in an OLG model with a standard age-specific income profile, young households are more likely to be close to the tenure-adjustment margin. As a result, the same home-purchase preference shock can generate a larger increase in aggregate housing demand and a larger peak response of house prices in an economy with a larger share of young households.

Our model is deliberately stylized and is not intended to replicate a full quantitative housing model. It builds on the quantitative life-cycle housing literature that studies tenure choice and housing adjustment over the life cycle (Li and Yao, 2007; Chambers et al., 2009). In these models, age, income, assets, and housing tenure jointly determine whether households are renters, first-time buyers, or existing owners. Our key mechanism is also related to Ortalo-Magne and Rady (2006), who emphasize that the ability of young households to move into homeownership can be an important

driver of housing-market dynamics. Relative to this literature, we strip the model down to the elements needed for our question: demographic weights, age-specific income profiles, tenure choice, transaction costs, adaptive house-price beliefs, and gradual price adjustment.

3.3.1 Model environment

The economy contains six overlapping generations, indexed by $g = 0, \dots, 5$, corresponding to the age groups 21–30, 31–40, 41–50, 51–60, 61–70, and 71–80. Demographic composition is governed by generation-specific survival probabilities. Let ρ_g denote the probability that a member of generation g survives into generation $g + 1$. Population masses are constructed as

$$N_0 = 1, \quad N_g = N_{g-1}\rho_{g-1}, \quad (3.6)$$

and normalized as

$$\pi_g = \frac{N_g}{\sum_{h=0}^5 N_h}. \quad (3.7)$$

These normalized weights determine aggregate demand. We use survival probabilities as a parsimonious device for generating two stationary demographic distributions. The young demographic scenario places more mass on younger and middle-aged cohorts, while the old scenario shifts population mass toward older cohorts.

In each period, households choose consumption, savings, and tenure status. Tenure is denoted by $s \in \{0, 1\}$, where $s = 1$ is owner-occupation and $s = 0$ is renting. Flow utility is

$$u(c_t, s_t; \Omega_t) = \frac{c_t^{1-\sigma} - 1}{1 - \sigma} + \Omega_t s_t, \quad (3.8)$$

where Ω_t is the housing-preference shifter. We use Ω_t for this shifter to avoid confusing it with the young-population share used elsewhere in the paper. The terminal generation consumes and leaves

a warm-glow bequest, with bequest utility

$$v(B) = \phi \frac{(B + \underline{B})^{1-\sigma} - 1}{1 - \sigma}. \quad (3.9)$$

Households face transaction costs when buying and selling housing, pay rent when they do not own, and earn a risk-free return on financial assets. Income differs both across and within generations. The deterministic component follows a hump-shaped life-cycle profile based on Guvenen et al. (2015), normalized across the six generations as

$$\bar{y}_g = (0.705, 1.027, 1.245, 1.240, 1.000, 0.969), \quad (3.10)$$

which captures lower income among young households, peak income in middle age, and lower income among older households. This intergenerational income profile is central to the model exercise because it determines which cohorts are near the owner-renter margin. Within each generation, households also draw an idiosyncratic income shock from a grid of nine equally-spaced points. The joint distribution of assets, tenure status, and income type is tracked explicitly for each cohort.

Households form expectations adaptively. Let $y_t = \log p_t - \log p_{t-1}$ denote observed house-price growth. All cohorts share a common belief $\hat{b}_t$ that follows a constant-gain Kalman updating rule,

$$\hat{b}_t = (1 - K)\hat{b}_{t-1} + Ky_t. \quad (3.11)$$

Each household projects future house prices according to

$$E_t p_{t+k} = p_t \exp\left(\hat{b}_t \frac{1 - \delta^k}{1 - \delta}\right), \quad \delta = 0.85. \quad (3.12)$$

The decay parameter δ prevents long-horizon beliefs from generating explosive price projections in the finite life-cycle problem.

For each generation, the household problem is solved by backward induction. Each household chooses tenure status $s \in \{0, 1\}$ and next-period assets by maximizing lifetime utility; tenure is selected deterministically as the argmax of the value of owning versus renting. Aggregate housing demand is the population-weighted sum of cohort-level ownership shares:

$$D_t = \sum_{g=0}^5 \pi_g \tilde{s}_{g,t}, \quad (3.13)$$

where $\tilde{s}_{g,t} = \int \mathbf{1}\{s^*(a, s, i; g, t) = 1\} d\mu_{g,t}(a, s, i)$ is the ownership share in cohort g , integrated over the joint distribution $\mu_{g,t}$ of assets, current tenure status, and income. House prices adjust gradually to excess demand,

$$\log p_{t+1} = \log p_t + \lambda(D_t - D^{ss}), \quad \lambda = 1.0, \quad (3.14)$$

where D^{ss} is the scenario-specific steady-state demand. Thus, prices are not re-cleared instantaneously each period; instead, the model generates a dynamic price-adjustment process in which demand pressure feeds into future prices and then back into beliefs.

The numerical analysis uses the parameter values summarized in Table 3.6. Several elements are disciplined by existing work. The life-cycle and demographic structure follows the long tradition linking age composition to housing demand (Mankiw and Weil, 1989; Takáts, 2012). The deterministic income profile is based on Guvenen et al. (2015). The housing block follows the macro-housing tradition in which housing demand, house prices, and household balance-sheet positions interact dynamically. The belief-updating rule for house prices is a constant-gain implementation of standard linear filtering logic (Kalman, 1960; Harvey, 1989; Durbin and Koopman, 2012).

3.3.2 Simulation design

The simulation compares two demographic scenarios: a young society and an old society. The scenarios differ only in the normalized population weights generated by equation (3.7). For each

Table 3.6: Calibration of the Housing OLG Model

Object	Value	Description
Life cycle	$G = 6$	Six age groups: 21–30, 31–40, 41–50, 51–60, 61–70, and 71–80.
Preferences	$\beta = 0.96, \sigma = 2.0$	Discount factor and coefficient of relative risk aversion.
Bequest motive	$\phi = 0.40, \underline{B} = 0.30$	Warm-glow bequest weight and bequest floor for the terminal generation.
Financial return	$r_a = 0.02$	Risk-free return on financial assets.
Rental cost	Rent = 0.06	Per-period cost paid by renter households.
Housing transaction costs	$\kappa_b = \kappa_s = 0.02$	Proportional buying and selling costs.
Housing supply	$\bar{H} = 0.55$	Fixed aggregate housing supply used to pin down the steady-state house price.
Housing preference	$\Omega^{ss} = 0.10$	Steady-state value of the owner-occupation preference shifter.
Price adjustment	$\lambda = 1.0$	Speed in the law of motion $\log p_{t+1} = \log p_t + \lambda(D_t - D^{ss})$.
Belief updating	$K = 0.618, \delta = 0.85$	Constant Kalman gain and decay parameter in projected future house prices.
Income profile	(0.705, 1.027, 1.245, 1.240, 1.000, 0.969)	Deterministic generation-specific income levels from equation (3.10).
Idiosyncratic income risk	$N_i = 9, \eta_i \in [-0.45, 0.45]$	Nine-point iid income shock grid with normal weights.
Asset and savings grids	$N_a = 24, a \in [0, 12]; N_f = 13, f \in [0.05, 0.95]$	Asset grid and candidate savings fractions used in the household problem.
Demographic scenarios	Young: (0.995, 0.990, 0.950, 0.880, 0.700, 0) Old: (0.998, 0.995, 0.990, 0.975, 0.940, 0)	Survival probabilities used to construct the young and old population distributions.
Preference shock	$\rho_\omega = 0.85, \varepsilon_{t^*} = 0.15, t^* = 20, T = 50$	AR(1) housing-preference shock and simulation horizon.

Notes: Values are the parameter values used in the OLG model in this paper. The Kalman gain is implied by signal and noise standard deviations equal to one. Survival probabilities are used to generate the normalized demographic weights in equation (3.7); all other parameters are held fixed across the young and old demographic scenarios.

scenario, we first compute the steady-state house price under $\Omega^{ss} = 0.10$ and zero house-price-growth beliefs. We then start each economy from its own steady state and apply the same housing-preference shock. Because the empirical object is a large upward deviation of house prices from trend, the main model statistic is the peak log deviation of house prices from steady state.

The shock follows an AR(1) process: $\Omega_t = \Omega^{ss} + \hat{\omega}_t$ where $\hat{\omega}_t = \rho_\omega \hat{\omega}_{t-1} + \varepsilon_t$, with persistence $\rho_\omega = 0.85$ and a single innovation $\varepsilon_{t^*} = 0.15$ at period $t^* = 20$ (all other $\varepsilon_t = 0$). After the innovation, the preference shifter decays back to its steady-state value at rate ρ_ω . The simulation horizon is 50 model periods, and the reported price response is the log deviation from each scenario's own steady-state price:

$$\text{dev}_t = 100 \times (\log p_t - \log p^{ss}). \quad (3.15)$$

Panel (b) of Figure 3.2 shows the resulting demographic compositions. The young society assigns survival probabilities (0.995, 0.990, 0.950, 0.880, 0.700, 0), placing 56.0 percent of the population in the first three age groups. The old society uses (0.998, 0.995, 0.990, 0.975, 0.940, 0), shifting mass toward the 61–80 age range (31.0 percent of population). The shock, preference process, belief updating rule, and income profile are otherwise identical across the two economies. Differences in the house-price response therefore come from the interaction between demographic weights π_g and the life-cycle income and tenure-choice structure.

3.3.3 Simulation results

Figure 3.2 reports the results. Panel (a) shows the house-price response in each demographic scenario. The young economy generates the larger boom: house prices rise by 15.6 percent above steady state at the peak, compared with 13.7 percent in the old economy. The peak response is therefore about 1.9 percentage points larger in the young economy in response to the same housing-preference shock. This is the model counterpart of the empirical finding that younger locations are more likely to experience large positive deviations of house prices from trend.

The mechanism operates through the interaction of income, tenure choice, and demographic

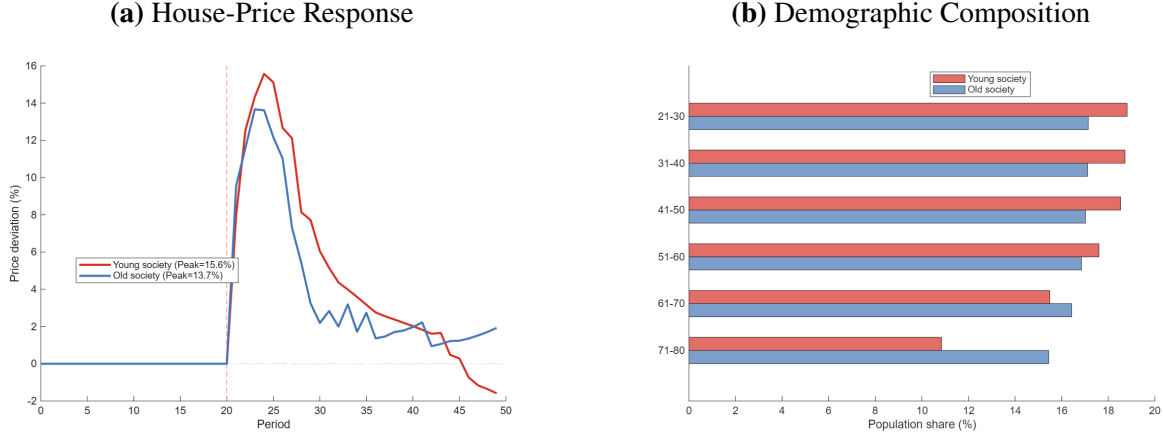


Figure 3.2: Structural OLG Simulation: Young vs. Old Economy

Notes: The figure reports the six-generation structural OLG simulation for a young society (red, solid) and an old society (blue, dashed). Panel (a) plots house prices as percent log deviations from each scenario's own steady state in response to an identical housing-preference shock (AR(1) with $\rho_\omega = 0.85$, single innovation $\varepsilon_{t^*} = 0.15$ at $t = 20$). Panel (b) shows the demographic composition implied by the survival probabilities for each scenario. The simulation uses the hump-shaped intergenerational income profile in equation (3.10). The common Kalman gain is $K = 0.618$; the price-adjustment speed is $\lambda = 1.0$.

composition. Following the shock, the preference value of ownership rises for all cohorts. The income profile then determines which households can translate this higher preference for ownership into actual housing demand. Very young households have relatively low income and limited assets, while older households are more likely to have already made their tenure choice. Middle-aged households, by contrast, have higher income and are closer to the active tenure-adjustment margin. A young demographic structure places more population mass in and around these margin-sensitive cohorts, so the same preference shock generates a larger aggregate demand response.

The result should be interpreted as a model-based consistency check for the empirical pattern rather than as a full quantitative explanation of boom incidence. The model shows that a stylized housing OLG environment can produce a stronger peak boom response in a younger economy once it includes intergenerational income differences. The mechanism does not require young households to have stronger extrapolative beliefs or a larger preference shock. Instead, demographic structure changes the distribution of households across life-cycle income and tenure positions, and that distribution governs the aggregate price response.

3.4 Conclusion

This paper studies whether demographic structure is associated with housing-boom vulnerability. Using both an OECD country panel and a U.S. county panel, we document that locations with a higher young-to-old population ratio are more likely to experience large upward deviations of real house prices from trend. This relationship is robust to location and time fixed effects, income controls, housing supply elasticity controls, and alternative specifications. We also provide micro-level evidence from HMDA mortgage application data showing that young applicants increase their mortgage-market activity disproportionately during boom episodes. Taken together, these findings suggest that demographics matter not only for long-run housing demand, but also for housing-boom vulnerability.

The model exercise asks whether this empirical pattern can arise in a stylized life-cycle housing environment. We study a six-generation housing OLG model in which households differ by age, assets, tenure status, survival probabilities, and income. The key model ingredient is the intergenerational income profile. When income differs across generations in a hump-shaped life-cycle pattern, a young economy places more households near the active tenure-adjustment margin. In response to the same housing-preference shock, this young economy generates a larger peak house-price response than the old economy.

The central message is that demographic composition can help explain housing-boom vulnerability through standard life-cycle forces. Young populations do not mechanically create larger booms simply because young households receive more weight. They create larger boom responses when age composition interacts with income differences and tenure-adjustment margins. This interaction implies that demographic structure should be considered when assessing housing-market risk, designing macroprudential policy, and evaluating the transmission of housing and preference shocks.

Bibliography

- Aiyagari, S. R. et al. (1992). “The output, employment, and interest rate effects of government consumption”. *Journal of Monetary Economics* 30.1, pp. 73–86.
- Arora, A. et al. (2025). *The private value of innovating for the government*. Tech. rep. National Bureau of Economic Research.
- Ascari, G. et al. (2023). “Fiscal foresight and the effects of government spending: It’s all in the monetary-fiscal mix”. *Journal of Monetary Economics* 134, pp. 1–15.
- Auclert, A. et al. (2024). “The intertemporal keynesian cross”. *Journal of Political Economy* 132.12, pp. 4068–4121.
- Auerbach, A. J., Y. Gorodnichenko, et al. (2011). *Fiscal multipliers in recession and expansion*. National Bureau of Economic Research Cambridge, MA.
- Auerbach, A. J. and Y. Gorodnichenko (2012). “Measuring the output responses to fiscal policy”. *American Economic Journal: Economic Policy* 4.2, pp. 1–27.
- (2016). “Effects of fiscal shocks in a globalized world”. *IMF Economic Review* 64.1, pp. 177–215.
- Austin, P. C. (2009). “Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples”. *Statistics in medicine* 28.25, pp. 3083–3107.
- (2011). “An introduction to propensity score methods for reducing the effects of confounding in observational studies”. *Multivariate behavioral research* 46.3, pp. 399–424.

- Baker, S. R. et al. (2016). “Measuring economic policy uncertainty”. *The quarterly journal of economics* 131.4, pp. 1593–1636.
- Bakshi, G. S. and Z. Chen (1994). “Baby Boom, Population Aging, and Capital Markets”. *The Journal of Business* 67.2, pp. 165–202. DOI: 10.1086/296629.
- Barattieri, A. et al. (2023). *Estimating the effects of government spending through the production network*. Tech. rep. National Bureau of Economic Research.
- Barber, B. M. and T. Odean (2008). “All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors”. *The review of financial studies* 21.2, pp. 785–818.
- Barnichon, R. et al. (2022). “Understanding the size of the government spending multiplier: It’s in the sign”. *The Review of Economic Studies* 89.1, pp. 87–117.
- Barro, R. J. and C. J. Redlick (2011). “Macroeconomic effects from government purchases and taxes”. *The Quarterly Journal of Economics* 126.1, pp. 51–102.
- Baxter, M. and R. G. King (1993). “Fiscal policy in general equilibrium”. *The American Economic Review*, pp. 315–334.
- Ben Zeev, N. and E. Pappa (2017). “Chronicle of a war foretold: The macroeconomic effects of anticipated defence spending shocks”. *The Economic Journal* 127.603, pp. 1568–1597.
- Ben Zeev, N., V. A. Ramey, et al. (2023). “Do government spending multipliers depend on the sign of the shock?” *AEA Papers and Proceedings*. Vol. 113. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 382–387.
- Bianchi, F. et al. (2024). “Using social media to identify the effects of congressional viewpoints on asset prices”. *The Review of Financial Studies* 37.7, pp. 2244–2272.
- Bilbiie, F. O. (2020). “The new Keynesian cross”. *Journal of Monetary Economics* 114, pp. 90–108.
- Blanchard, O. and R. Perotti (2002). “An empirical characterization of the dynamic effects of changes in government spending and taxes on output”. *the Quarterly Journal of economics* 117.4, pp. 1329–1368.

- Boehm, C. E. and N. Pandalai-Nayar (2022). “Convex supply curves”. *American Economic Review* 112.12, pp. 3941–3969.
- Borio, C. E. and P. W. Lowe (2002). “Asset prices, financial and monetary stability: exploring the nexus”.
- Bouakez, H. et al. (2025). “The sectoral origins of heterogeneous spending multipliers”. *Journal of Public Economics* 248, p. 105404.
- Briganti, E. (2023). “On the Effects of Government Purchases and Their Transmission Mechanism”. *Unpublished*. https://edoardobriganti.github.io/research/edoardo_briganti_JMP.pdf.
- Brunet, G. (2022). “When Does Government Spending Matter? Evidence from a New Measure of Spending”. *en. In: Working Paper (May 2020)*, p. 41.
- Burnside, C. et al. (2004). “Fiscal shocks and their consequences”. *Journal of Economic theory* 115.1, pp. 89–117.
- Caldara, D. and C. Kamps (2008). “What are the effects of fiscal policy shocks? A VAR-based comparative analysis”. *ECB working paper*.
- Callaway, B. and P. H. Sant’Anna (2021). “Difference-in-differences with multiple time periods”. *Journal of econometrics* 225.2, pp. 200–230.
- Chambers, M. et al. (2009). “Accounting for changes in the homeownership rate”. *International Economic Review* 50.3, pp. 677–726.
- Chodorow-Reich, G. (2019). “Geographic cross-sectional fiscal spending multipliers: What have we learned?” *American Economic Journal: Economic Policy* 11.2, pp. 1–34.
- Choi, S. et al. (2022). “Are government spending shocks inflationary at the zero lower bound? New evidence from daily data”. *Journal of Economic Dynamics and Control* 139, p. 104423.
- Cloyne, J. et al. (2023). *State-dependent local projections: Understanding impulse response heterogeneity*. National Bureau of Economic Research.
- Cochrane, J. H. (2022). “Fiscal histories”. *Journal of Economic Perspectives* 36.4, pp. 125–146.
- Cochrane, J. H. and M. Piazzesi (2002). “The fed and interest rates—a high-frequency identification”. *American economic review* 92.2, pp. 90–95.

- Cook, T. and T. Hahn (1989). “The effect of changes in the federal funds rate target on market interest rates in the 1970s”. *Journal of monetary economics* 24.3, pp. 331–351.
- Cox, L. et al. (2024). “Big g”. *Journal of Political Economy* 132.10, pp. 3260–3297.
- Davis, S. J. et al. (1997). “Regional labor fluctuations: oil shocks, military spending, and other driving forces”. *Board of Governors of the Federal Reserve System International Finance Disc. Papers* 578.
- Detken, C. and F. Smets (2004). “Asset price booms and monetary policy”. *Macroeconomic policies in the world economy* 329, p. 189.
- Dissanayake, R. (2022). “Using Stock Returns to Identify Government Spending Shocks: New Insights”. *Available at SSRN 4006335*.
- Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. 2nd ed. Oxford University Press.
- Edelberg, W. et al. (1999). “Understanding the effects of a shock to government purchases”. *Review of Economic Dynamics* 2.1, pp. 166–206.
- Farhi, E. and I. Werning (2016). “A theory of macroprudential policies in the presence of nominal rigidities”. *Econometrica* 84.5, pp. 1645–1704.
- Fatás, A. and I. Mihov (2001). “The effects of fiscal policy on consumption and employment: theory and evidence”. *Available at SSRN 267281*.
- Favilukis, J. et al. (2017). “The macroeconomic effects of housing wealth, housing finance, and limited risk sharing in general equilibrium”. *Journal of Political Economy* 125.1, pp. 140–223.
- Ferraz, C. et al. (2015). “Procuring Firm Growth: The Effects of Government Purchases on Firm Dynamics”. *Working Paper*. NBER Working Paper No. 21219. DOI: 10.3386/w21219.
- Ferris, S. P. et al. (2025). “Corporate Value Creation and the Award of Procurement Contracts”. *Journal of Corporate Accounting & Finance* 36.3, pp. 119–136.
- Fieldhouse, A. J. and K. Mertens (2023). *The returns to government R&D: Evidence from US appropriations shocks*. Federal Reserve Bank of Dallas, Research Department.

- Fisher, J. D. and R. Peters (2010). “Using stock returns to identify government spending shocks”. *The Economic Journal* 120.544, pp. 414–436.
- Gabriel, R. D. (2024). “The credit channel of public procurement”. *Journal of Monetary Economics* 147, p. 103601.
- Galí, J. et al. (2007). “Understanding the effects of government spending on consumption”. *Journal of the european economic association* 5.1, pp. 227–270.
- Gevorgyan, K. (2019). “Do Demographic Changes Affect House Prices?” *Journal of Demographic Economics* 85.4, pp. 305–320. DOI: 10.1017/dem.2019.9.
- Ghassibe, M. and F. Zanetti (2022). “State dependence of fiscal multipliers: the source of fluctuations matters”. *Journal of Monetary Economics* 132, pp. 1–23.
- Giovanni, J. di et al. (2022). “Buy Big or Buy Small? Procurement Policies, Firm Size, and Aggregate Productivity”. *Working Paper*. Available at SSRN.
- Glaeser, E. L. et al. (2008). “Housing supply and housing bubbles”. *Journal of urban Economics* 64.2, pp. 198–217.
- Goldman, J. (2020). “Government as Customer of Last Resort: The Stabilizing Effects of Government Purchases on Firms”. *Review of Financial Studies* 33.2, pp. 610–643. DOI: 10.1093/rfs/hhz059.
- Goodhart, C. and B. Hofmann (2008). “House prices, money, credit, and the macroeconomy”. *Oxford review of economic policy* 24.1, pp. 180–205.
- Gordon, R. J. and R. Krenn (2010). *The end of the great depression 1939-41: Policy contributions and fiscal multipliers*. National Bureau of Economic Research.
- Gürkaynak, R. S. et al. (2004). “Do actions speak louder than words? The response of asset prices to monetary policy actions and statements”. *The Response of Asset Prices to Monetary Policy Actions and Statements (November 2004)*.
- Güvenen, F. et al. (2015). *What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?* NBER Working Paper 20913. National Bureau of Economic Research.

- Hall, R. E. (2009). *By how much does GDP rise if the government buys more output?* National Bureau of Economic Research.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- Hebous, S. and T. Zimmermann (2021). “Can government demand stimulate private investment? Evidence from US federal procurement”. *Journal of Monetary Economics* 118, pp. 178–194.
- Hodrick, R. J. and E. C. Prescott (1997). “Postwar U.S. Business Cycles: An Empirical Investigation”. *Journal of Money, Credit and Banking* 29.1, pp. 1–16. DOI: 10.2307/2953682.
- Hooker, M. A. and M. M. Knetter (1997). “The effects of military spending on economic activity: Evidence from state procurement spending”. *Journal of Money, Credit, and Banking*, pp. 400–421.
- Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press.
- Jordà, Ò. (2005). “Estimation and inference of impulse responses by local projections”. *American economic review* 95.1, pp. 161–182.
- Jordà, Ò., M. Schularick, et al. (2015). “Leveraged bubbles”. *Journal of Monetary Economics* 76, S1–S20.
- Jordà, Ò. and A. M. Taylor (2025). “Local projections”. *Journal of Economic Literature* 63.1, pp. 59–110.
- Justiniano, A. et al. (2019). “Credit supply and the housing boom”. *Journal of political economy* 127.3, pp. 1317–1350.
- Kalman, R. E. (1960). “A New Approach to Linear Filtering and Prediction Problems”. *Journal of Basic Engineering* 82.1, pp. 35–45.
- Känzig, D. R. (2021). “The macroeconomic effects of oil supply news: Evidence from OPEC announcements”. *American Economic Review* 111.4, pp. 1092–1125.
- (2023). *The unequal economic consequences of carbon pricing*. National Bureau of Economic Research.

- Kaplan, G. et al. (2020). “The housing boom and bust: Model meets evidence”. *Journal of Political Economy* 128.9, pp. 3285–3345.
- Kuttner, K. N. (2001). “Monetary policy surprises and interest rates: Evidence from the Fed funds futures market”. *Journal of monetary economics* 47.3, pp. 523–544.
- Leeper, E. M. et al. (2017). “Clearing up the fiscal multiplier morass”. *American Economic Review* 107.8, pp. 2409–2454.
- Li, W. and R. Yao (2007). “The life-cycle effects of house price changes”. *Journal of Money, Credit and banking* 39.6, pp. 1375–1409.
- Mankiw, N. G. and D. N. Weil (1989). “The Baby Boom, the Baby Bust, and the Housing Market”. *Regional Science and Urban Economics* 19.2, pp. 235–258. doi: 10.1016/0166-0462(89)90005-7.
- McClure, E. and A. Yding (2025). “The Macroeconomic Effects of Defense Spending News”. *Working paper*.
- Mian, A. and A. Sufi (2011). “House prices, home equity-based borrowing, and the us household leverage crisis”. *American economic review* 101.5, pp. 2132–2156.
- Monacelli, T. and R. Perotti (2008). *Fiscal policy, wealth effects, and markups*. National Bureau of Economic Research.
- Muratori, U. et al. (2023). *Heterogeneous Spending, Heterogeneous Multipliers*. International Monetary Fund.
- Nakamura, E. and J. Steinsson (2014). “Fiscal stimulus in a monetary union: Evidence from US regions”. *American Economic Review* 104.3, pp. 753–792.
- (2018). “High-frequency identification of monetary non-neutrality: the information effect”. *The Quarterly Journal of Economics* 133.3, pp. 1283–1330.
- Nishimura, K. G. and E. Takáts (2012). *Ageing, Property Prices and Money Demand*. BIS Working Papers 385. Bank for International Settlements.

- OECD (2025). *Government at a Glance 2025: Public Procurement*. Tech. rep. Paris: Organisation for Economic Co-operation and Development. URL: <https://www.oecd.org/en/topics/public-procurement.html>.
- Olea, J. L. M. and C. Pflueger (2013). “A robust test for weak instruments”. *Journal of Business & Economic Statistics* 31.3, pp. 358–369.
- Ortalo-Magne, F. and S. Rady (2006). “Housing market dynamics: On the contribution of income shocks and credit constraints”. *The Review of Economic Studies* 73.2, pp. 459–485.
- Ottonello, P. and W. Song (2025). “Financial intermediaries and the macroeconomy: Evidence from a high-frequency identification”. *The Economic Journal*, ueaf119.
- Owyang, M. T. et al. (2013). “Are government spending multipliers greater during periods of slack? Evidence from twentieth-century historical data”. *American Economic Review* 103.3, pp. 129–134.
- Perotti, R. et al. (2007). “In search of the transmission mechanism of fiscal policy [with comments and discussion]”. *NBER macroeconomics Annual* 22, pp. 169–249.
- Pflueger, C. E. and S. Wang (2015). “A robust test for weak instruments in Stata”. *The Stata Journal* 15.1, pp. 216–225.
- Ramey, V. A. (2011). “Identifying government spending shocks: It’s all in the timing”. *The Quarterly Journal of Economics* 126.1, pp. 1–50.
- (2019). “Ten years after the financial crisis: What have we learned from the renaissance in fiscal research?” *Journal of Economic Perspectives* 33.2, pp. 89–114.
- Ramey, V. A. and M. D. Shapiro (1998). “Costly capital reallocation and the effects of government spending”. *Carnegie-Rochester conference series on public policy*. Vol. 48. Elsevier, pp. 145–194.
- Ramey, V. A. and S. Zubairy (2018). “Government spending multipliers in good times and in bad: evidence from US historical data”. *Journal of political economy* 126.2, pp. 850–901.
- Ramey, V. A. (2009). *Identifying Government Spending Shocks: It’s All in the Timing*. 15464. National Bureau of Economic Research.

- Saiz, A. (2010). “The Geographic Determinants of Housing Supply”. *The Quarterly Journal of Economics* 125.3, pp. 1253–1296. doi: 10.1162/qjec.2010.125.3.1253.
- Stuart, E. A. and D. B. Rubin (2008). “Best practices in quasi-experimental designs”. *Best practices in quantitative methods*, pp. 155–176.
- Takáts, E. (2010). *Ageing and Asset Prices*. BIS Working Papers 318. Bank for International Settlements.
- (2012). “Aging and House Prices”. *Journal of Housing Economics* 21.2, pp. 131–141. doi: 10.1016/j.jhe.2012.04.001.
- Towell, P. (Nov. 2021). *The Department of Defense (DOD) Budget: An Orientation*. Tech. rep. Congressional Research Service, Library of Congress.
- Wang, N. et al. (2025). “Government procurement and stock price synchronicity”. *Pacific-Basin Finance Journal*, p. 102896.
- Wiegand, C. (2025). “The Effect of Fiscal Policy Shocks on Asset Prices”. *Working paper*.
- World Bank (2020). *Global Public Procurement Database: Share of Public Procurement in GDP*. Tech. rep. World Bank Group. URL: <https://www.worldbank.org/en/topic/procurement>.

Appendices to Chapter 1

1.A Evaluations of the defense news shock identification

1.A.1 Sentence-based text analysis model

To compare the performance of the sentence-based model with alternative models, I use the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC). The ROC curve visualizes the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) as the classification threshold varies, defined as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (1.A.1)$$

Here, TP denotes true positives, FN denotes false negatives, FP denotes false positives, and TN denotes true negatives. The classification threshold is based on the relevance score evaluated by the model for each document. In the sentence-based model, the threshold corresponds to the number of sentences related to changes in defense spending, and any article with more relevant sentences than the threshold is classified as positive. Raising the threshold typically reduces both the true positive rate and the false positive rate, while lowering the threshold increases the true positive rate but also raises the false positive rate. The AUC is defined as the area under the ROC curve and can be expressed mathematically as:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR}) \quad (1.A.2)$$

An AUC value closer to 1 indicates better classification performance, with a perfect classifier achieving an AUC of 1 and a random classifier approximating 0.5.

To conduct the ROC-AUC analysis, it is necessary to construct a labeled dataset that distinguishes between articles related to changes in defense spending (positive cases) and those that are not (negative cases). For this purpose, I collect a sample of 2,396 articles published in The Wall Street

Journal between January 1980 and December 1981, retrieved from the ProQuest database using the keyword “defense.”¹ After carefully reviewing the content of each article, I classify 164 as news articles related to changes in defense spending (positive cases) and the remaining 2,232 as non-defense spending news (negative cases), thereby constructing a labeled dataset for comparison.

I compare the sentence-based model with three alternative approaches. First, the keyword-based model calculates the proportion of words that match a predefined list of defense-related terms, and this proportion serves as the article’s relevance score.² Second, the topic model extracts k latent topics from the corpus and assigns each article a topic-probability vector. The share of probability mass allocated to each topic serves as the article’s relevance score.³ I set the topic model to extract five latent topics and select the topic with the highest AUC value for comparison with the sentence-based model. Lastly, I use a large language model (LLM), ChatGPT (gpt-4o-mini), to assess whether each article contains information related to changes in defense spending. This method does not provide a continuous relevance score but instead produces binary outcomes of “Yes” or “No.”⁴

Figure 1.A.1 shows that the sentence-based model outperforms both the keyword ratio model and the topic model. The sentence-based model achieves an AUC of 0.84, which is higher than that of the keyword ratio model (0.64) and the topic model (0.67). If the sentence-based model classifies an article as positive when it contains at least one sentence related to changes in defense spending, the model yields a true positive rate (TPR) of 0.95 and a false positive rate (FPR) of 0.51. If one attempts to match either of these metrics (TPR or FPR) using the alternative models, the other metric necessarily worsens. In the upper panel, maintaining a TPR of 0.95 in the keyword ratio model requires increasing the FPR from 0.51 to 0.84, while holding the FPR at 0.51 lowers the TPR

¹This period marks the onset of the Reagan administration’s military buildup and is characterized by a large number of defense spending-related news articles that are not tied to period-specific keywords, such as those associated with the Korean War or Vietnam War.

²The predefined list of defense-related terms includes: *defense*, *military*, *pentagon*, *arms*, *weapon*, and *war*.

³I use the built-in Latent Dirichlet Allocation (LDA) topic-modeling feature available in TDM Studio.

⁴The model is prompted with the following instruction: “Respond ‘Yes’ if this article contains information that could help predict changes in the U.S. defense budget; otherwise, respond ‘No’.” The model’s binary reply (Yes or No) is then used as the classification outcome.

from 0.95 to 0.71. Similarly, in the topic model, keeping the TPR at 0.95 raises the FPR from 0.51 to 0.76, and maintaining the FPR at 0.51 reduces the TPR from 0.95 to 0.89.

Since ChatGPT directly classifies each article as either defense-related or not, it does not allow for adjusting the classification threshold. As shown in Table 1.A.1, ChatGPT achieves a true positive rate (TPR) of 0.49 and a false positive rate (FPR) of 0.02, indicating that it substantially reduces false positives at the cost of a significant drop in true positives. A comparable TPR of 0.51 can be achieved by the sentence-based model when the model classifies an article as positive when it contains at least four relevant sentences. However, the corresponding FPR rises to 0.10, which is considerably higher than GPT's 0.02. While ChatGPT excels at minimizing false positives, its low TPR entails a nontrivial opportunity cost in the context of high-frequency identification strategies, which depend on capturing as many relevant news articles as possible to accurately determine the timing and magnitude of news shocks.

Table 1.A.1: True Positive Rate & False Positive Rate

	True Positive Rate (TPR)	False Positive Rate (FPR)
<i>Sentence</i> ≥ 1	0.95	0.51
<i>Sentence</i> ≥ 2	0.84	0.26
<i>Sentence</i> ≥ 3	0.66	0.16
<i>Sentence</i> ≥ 4	0.51	0.10
<i>ChatGPT = Yes</i>	0.49	0.02

Notes: The table shows the true positive rate and false positive rate of the text analysis results for each relevant sentence threshold. I use 2,396 Wall Street Journal articles containing the keyword “defense” from ProQuest, covering the period from January 1, 1980, to December 31, 1981. Based on the manual review, 164 articles were identified as positive cases and 2,232 as negative.

1.A.2 Placebo test for the news dates

To assess the validity of the identification strategy, I conduct two placebo tests using alternative government spending news shock series constructed from dates unrelated to actual news events. Specifically, first placebo test randomly selects 869 dates from the sample period. Second placebo test shifts all news dates 7 days earlier. For each test, I estimate the responses of government

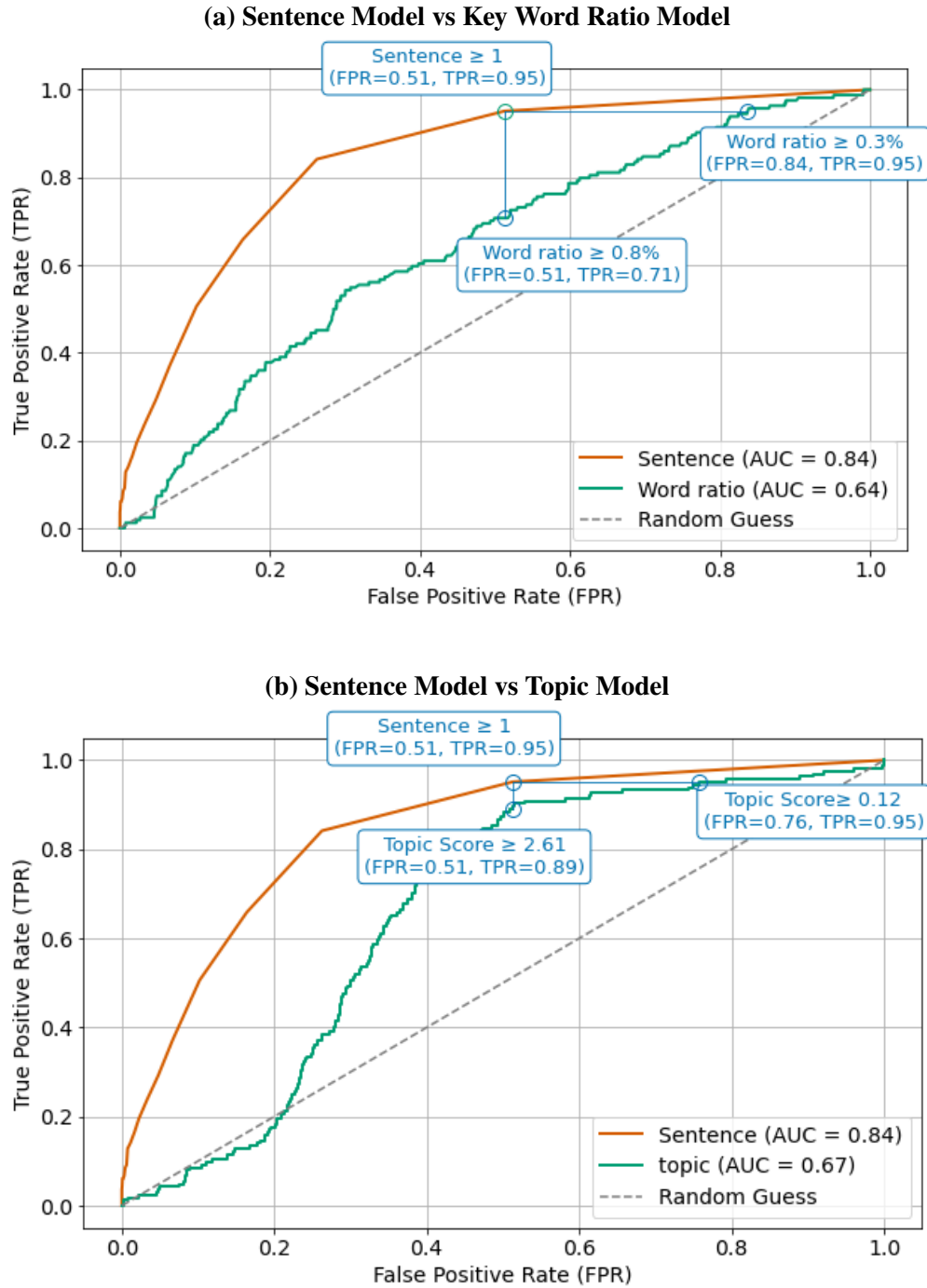


Figure 1.A.1: ROC Curves by Text Analysis Model

Notes: These figures show the receiver operating characteristic (ROC) curves for two text analysis approaches. The blue line represents the sentence-based model, while the green line shows the results from the comparison model. The top panel compares the sentence model with the keyword ratio model. The bottom panel compares the sentence model with the topic model. For the topic model, the number of topics is set to five, and the best-performing result (i.e., the one with the highest AUC among the five topics) is reported.

spending using these placebo shock series. If the identification strategy is valid, the estimated responses should be substantially weaker compared to those based on the actual news dates.

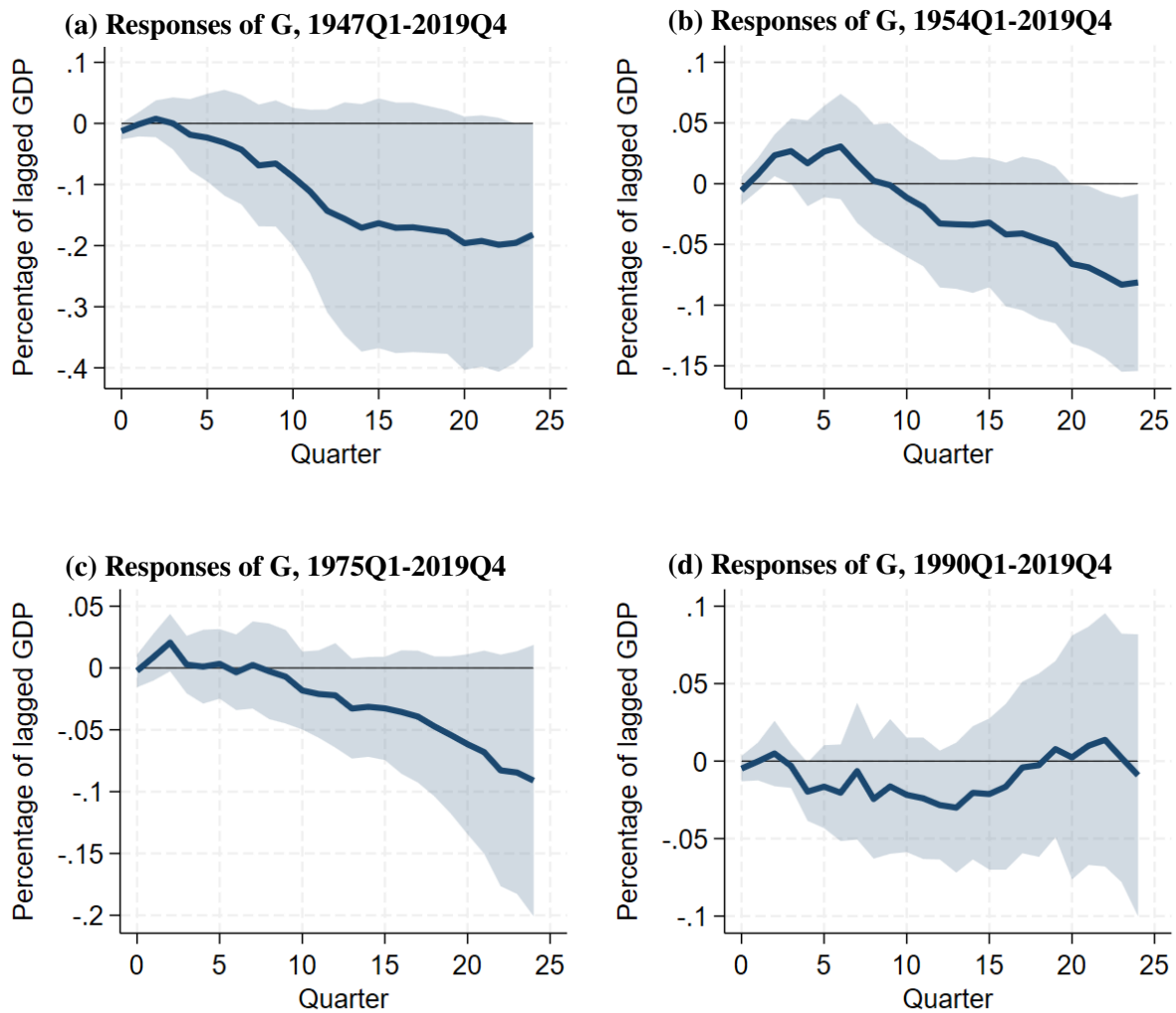


Figure 1.A.2: Placebo Test I, Random News Dates

Notes: The figures show the response of government spending to placebo shocks constructed using randomly selected dates. The shaded area represent the 90% confidence intervals.

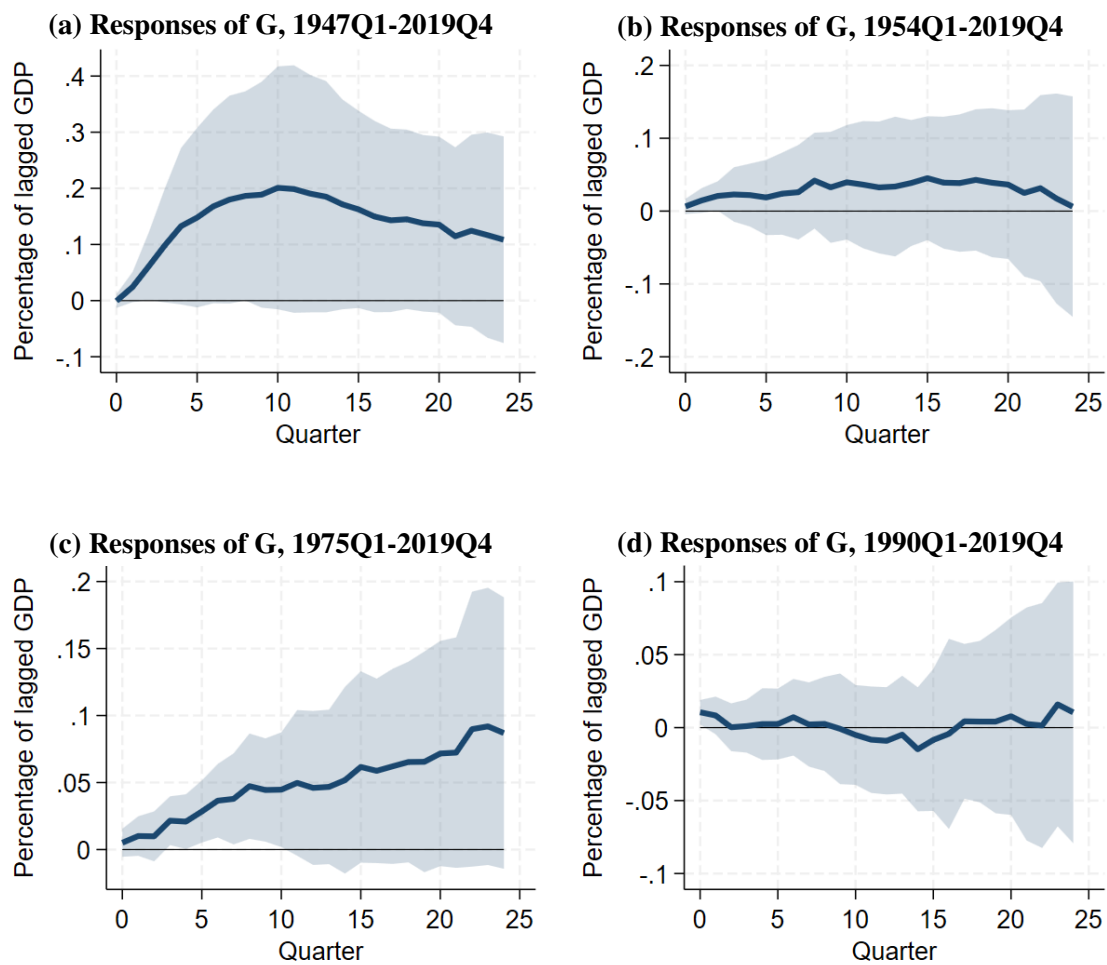


Figure 1.A.3: Placebo Test II, A Week Prior to News

Notes: The figures show the response of government spending to placebo shocks constructed using dates one week prior to the news dates. The shaded area represent the 90% confidence intervals.

1.A.3 Background Noise

One potential concern with the high-frequency identification approach is that factors unrelated to defense may significantly influence excess returns during the event window. To assess this possibility, I follow the method of Känzig (2021) and compare the excess stock return on each defense spending news date with that on a corresponding control date. The control date is defined as the business day that falls on the same weekday at least one month prior to the news date. If the variation in excess returns on defense spending news dates is significantly larger than that on control dates, it suggests that the constructed shock effectively captures information related to defense spending news. As shown in Figure 1.A.4, excess returns exhibit significantly greater volatility on defense spending news dates, with their variance being approximately 1.56 times higher than on control dates. Additionally, the density function of excess returns on defense spending news dates reveals notably fatter tails.

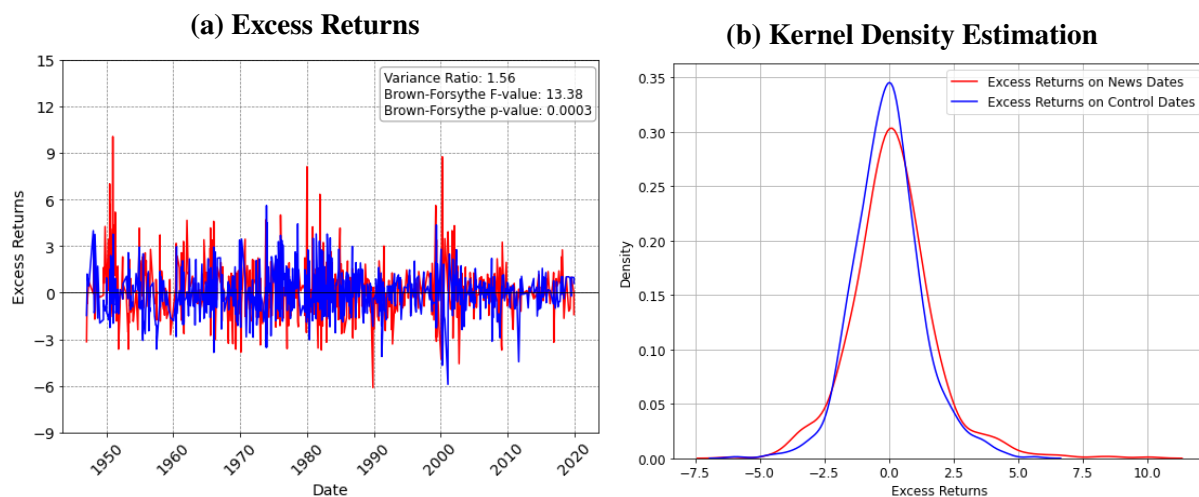


Figure 1.A.4: News Versus Control Days

Notes: The left panel shows the series of excess returns computed over a 2-day window for defense spending news dates and control dates. The right panel presents the kernel density functions estimated using these two series.

1.A.4 Details on the Topic Model Results

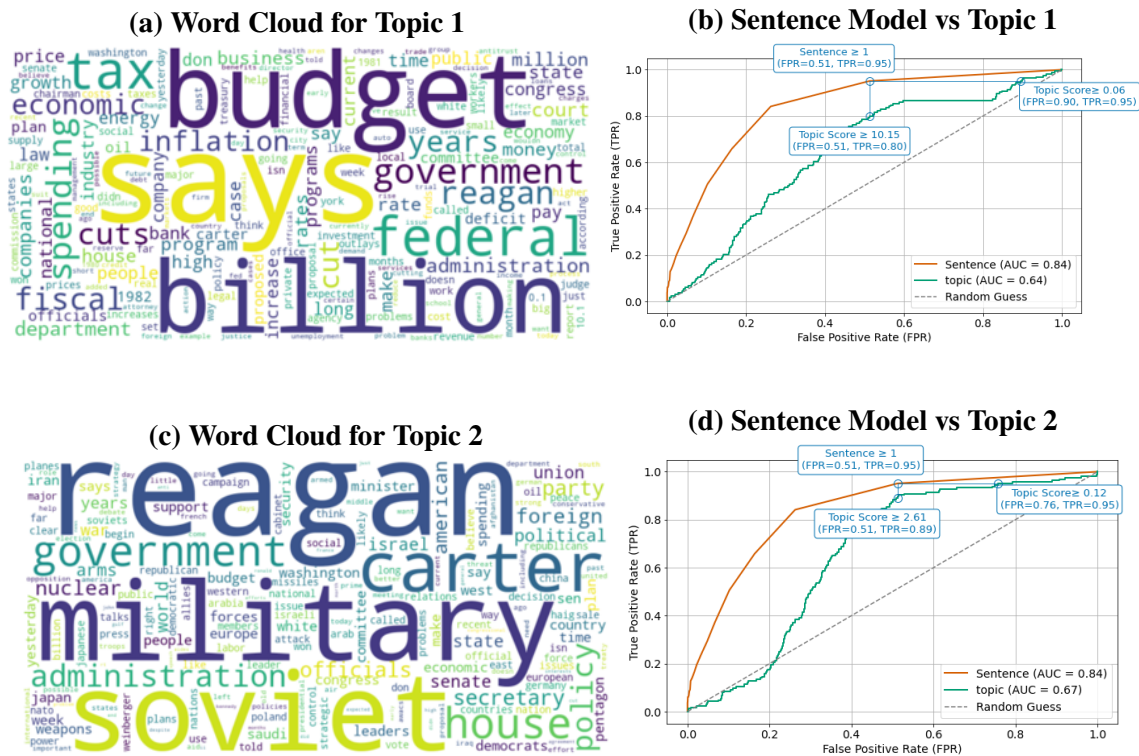


Figure 1.A.5: Word clouds and ROC Curves by Topic: Topic 1 and 2

Notes: The left panels show the word clouds for Topics 1 and 2, while the right panels compare the ROC curves and AUC values of the sentence model with those of Topics 1 and 2.

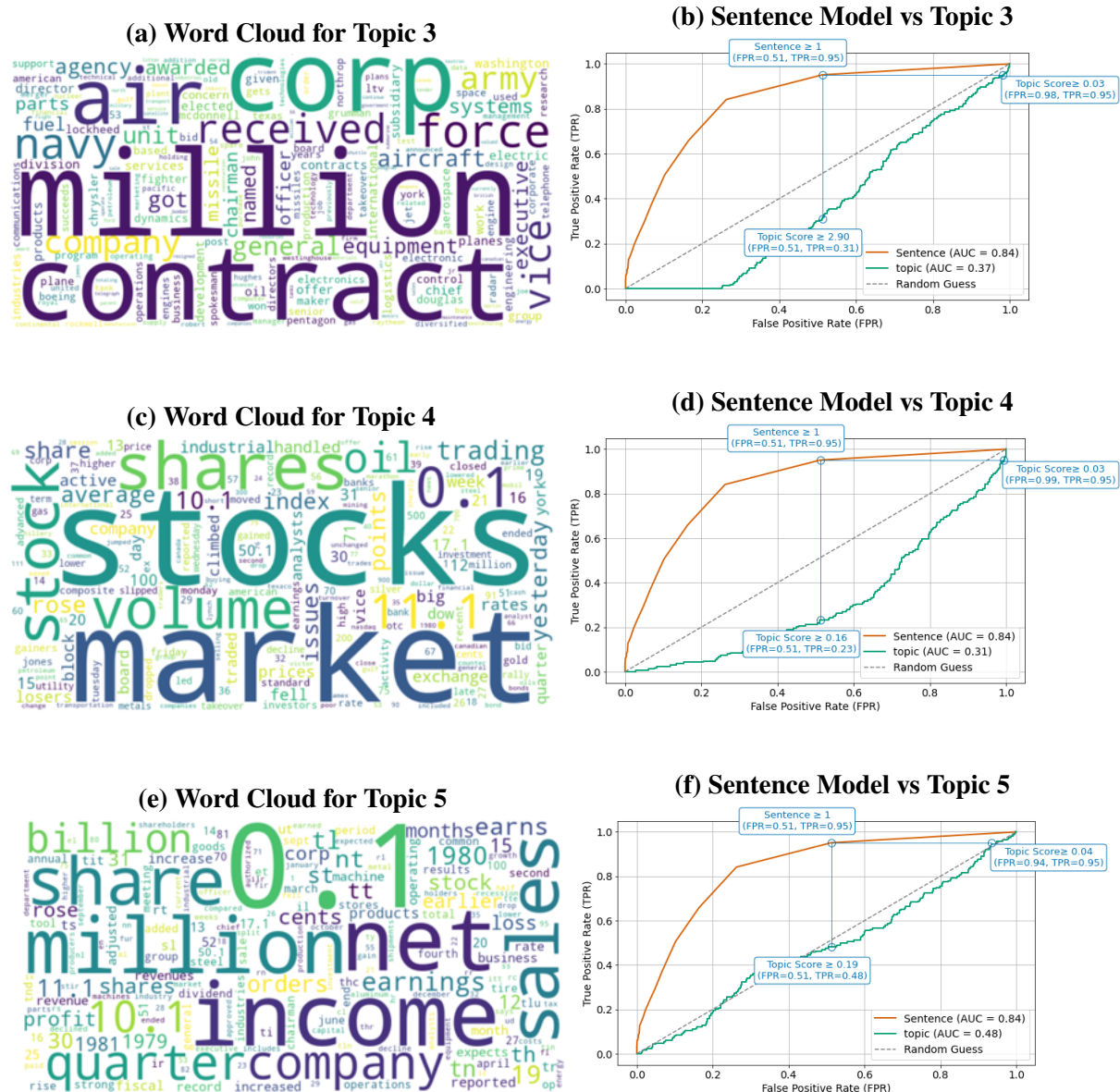


Figure 1.A.6: Word clouds and ROC Curves by Topic: Topic 3, 4 and 5

Notes: The left panels show the word clouds for Topics 3, 4, and 5, while the right panels compare the ROC curves and AUC values of the sentence model with those of Topics 3, 4, and 5.

1.B IRFs with confidence intervals

1.B.1 Government spending

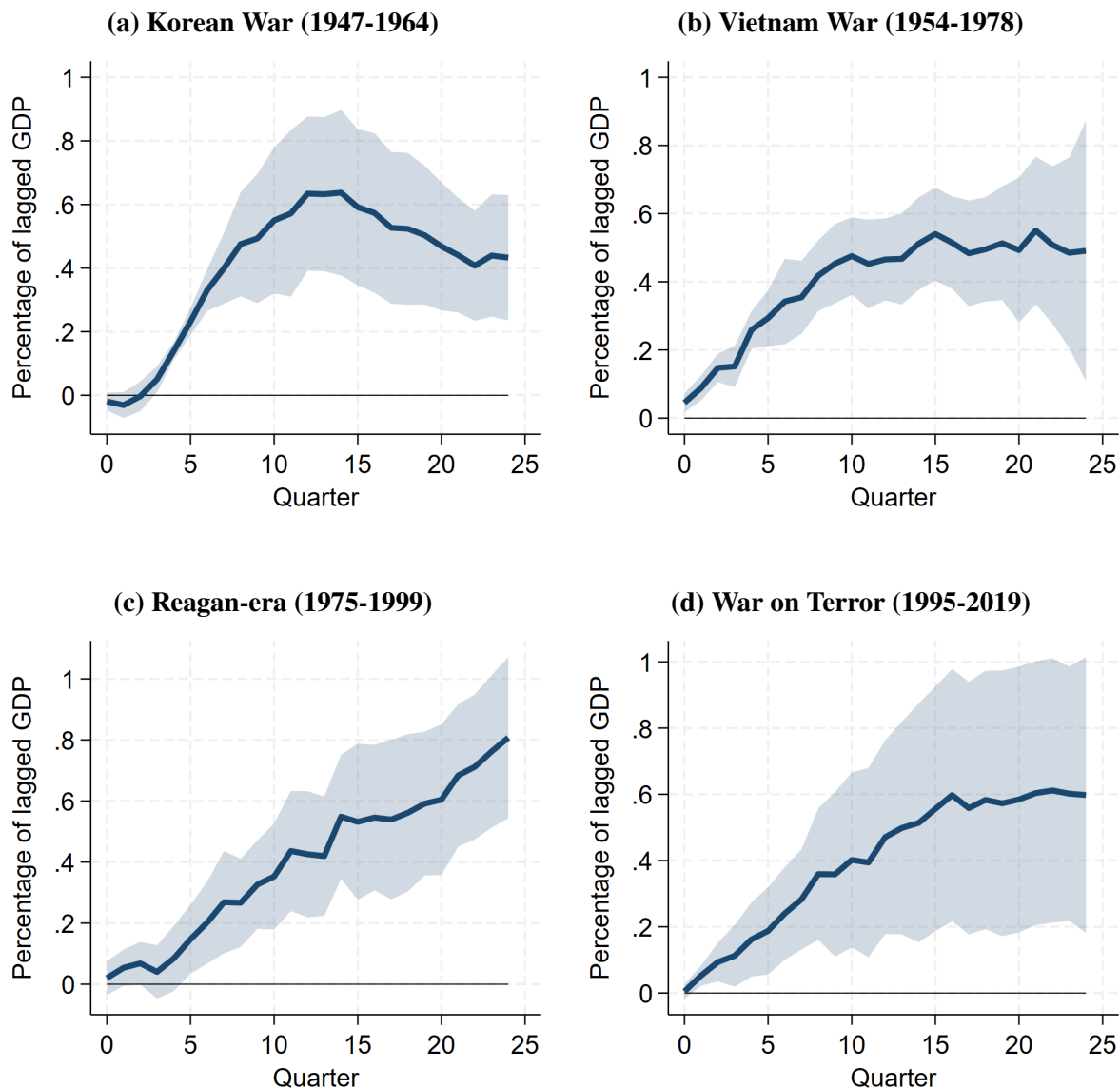


Figure 1.B.1: Responses of Government Spending by Sample

Notes: The figures show the response of government spending to DNS by sample period. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

1.B.2 GDP

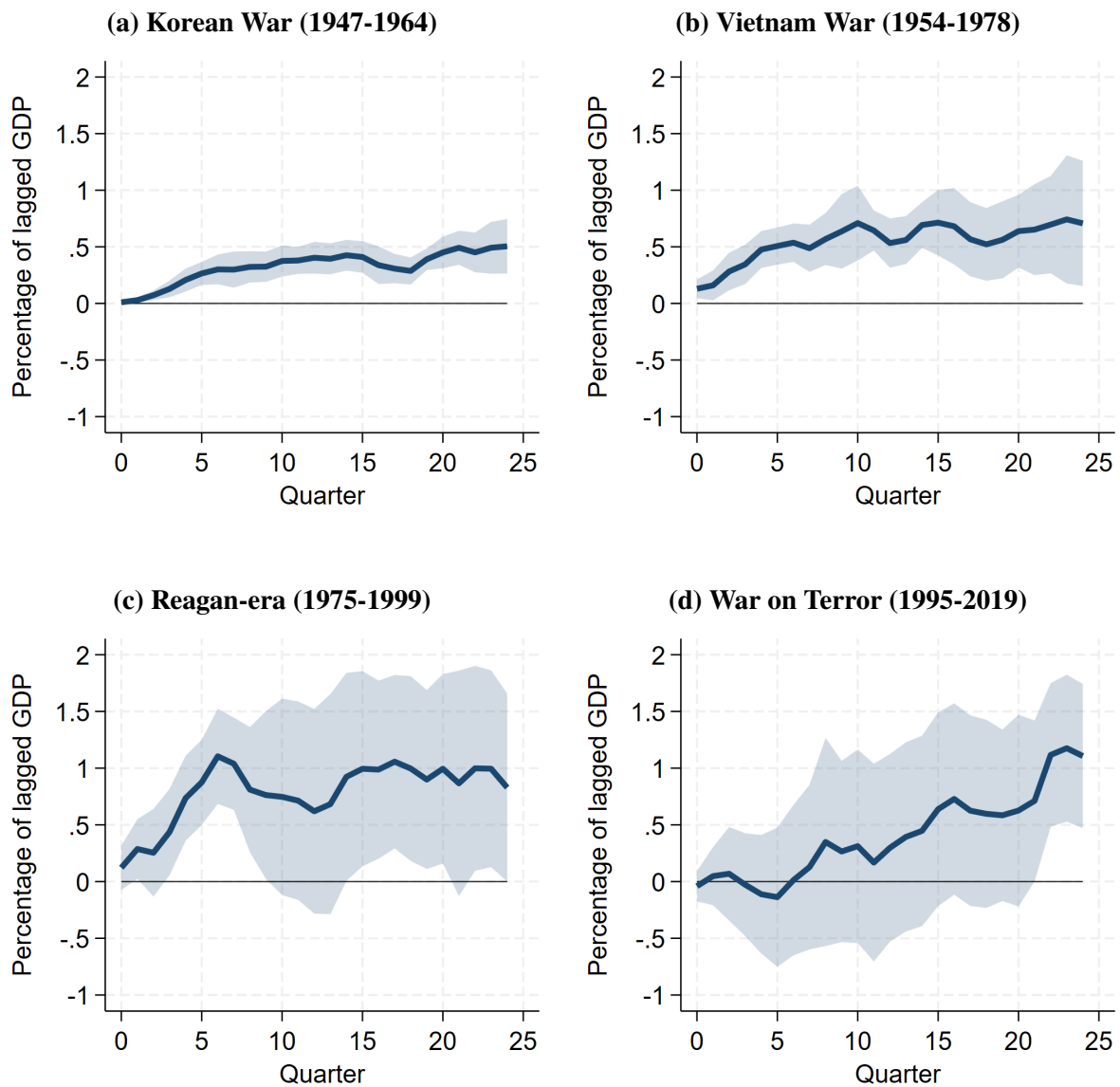


Figure 1.B.2: Responses of GDP by Sample

Notes: The figures show the response of GDP to DNS by sample period. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

1.B.3 Private Investment

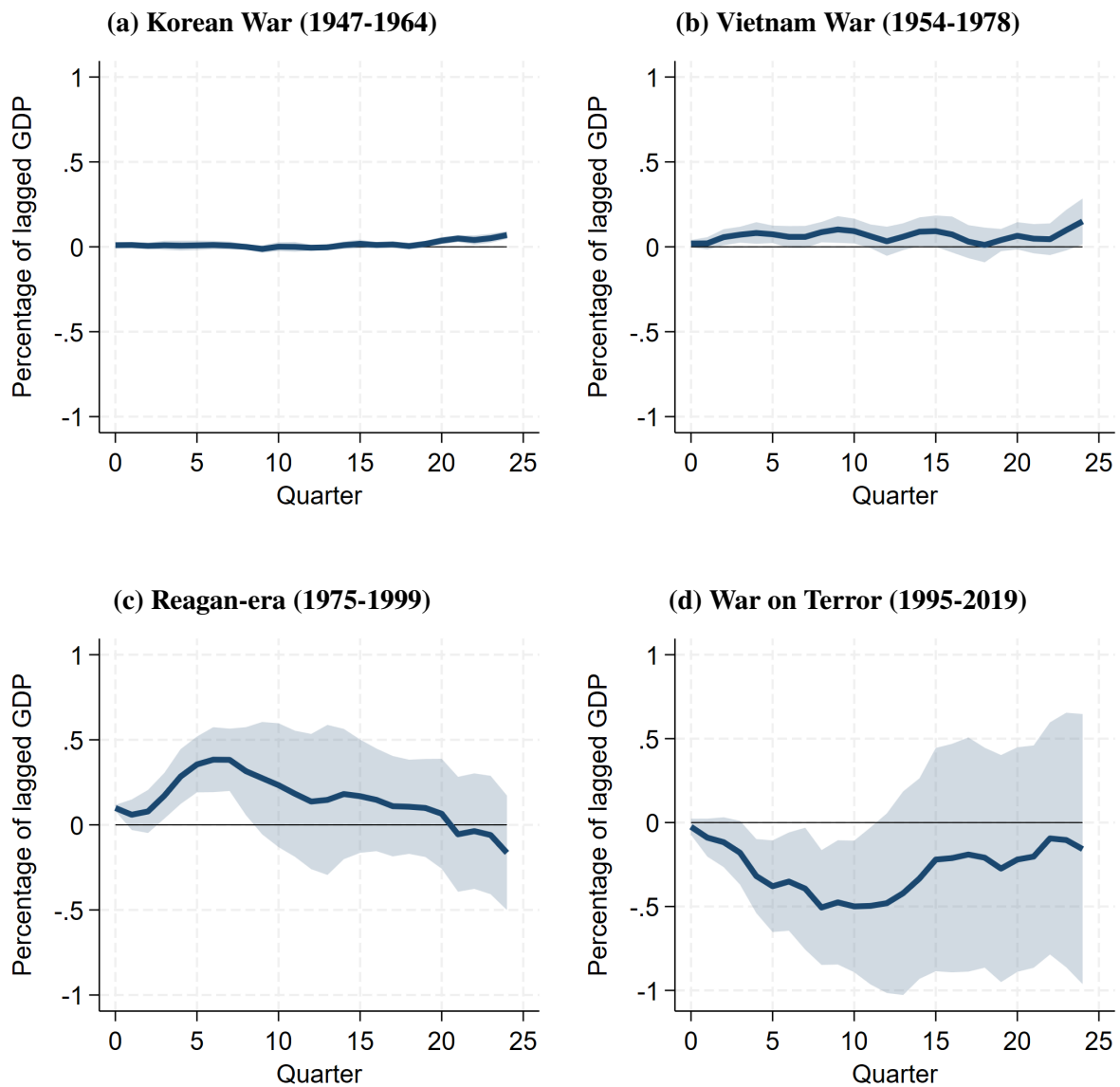


Figure 1.B.3: Responses of Private Investment by Sample

Notes: The figures show the response of private investment to DNS by sample period. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

1.B.4 Private Consumption

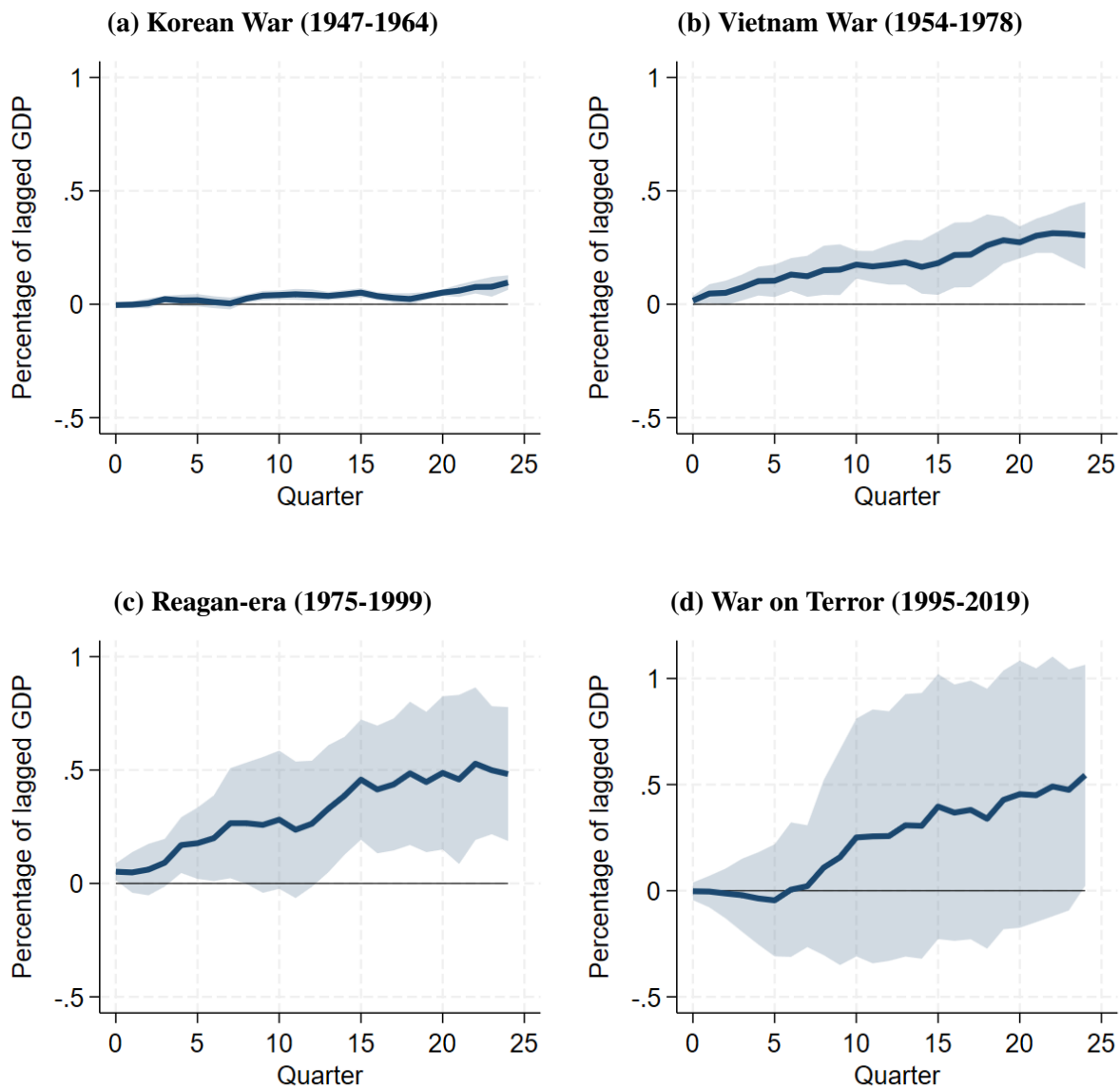


Figure 1.B.4: Responses of Private Consumption by Sample

Notes: The figures show the response of private consumption (non-durables + services) to DNS by sample period. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

1.C Robustness of the time-varying effects of government spending

1.C.1 Control Variables

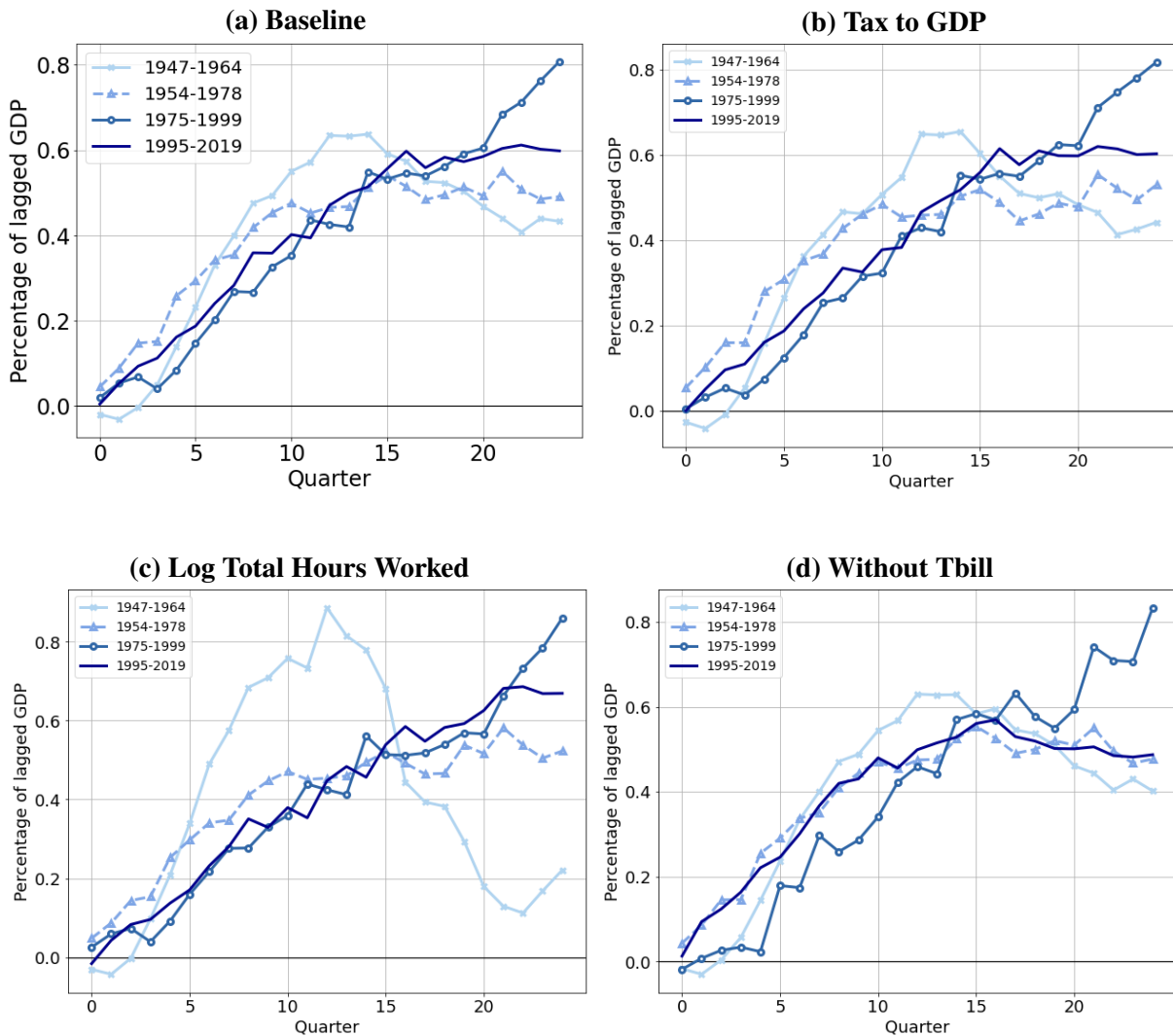


Figure 1.C.1: Sensitivity to Controls- Government Spending

Notes: The figures show the response of government spending to DNS by sample. The scale of news shock is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. For the model without T-bill, I control for the period of high interest rate volatility from 1978:Q1 to 1982:Q2.

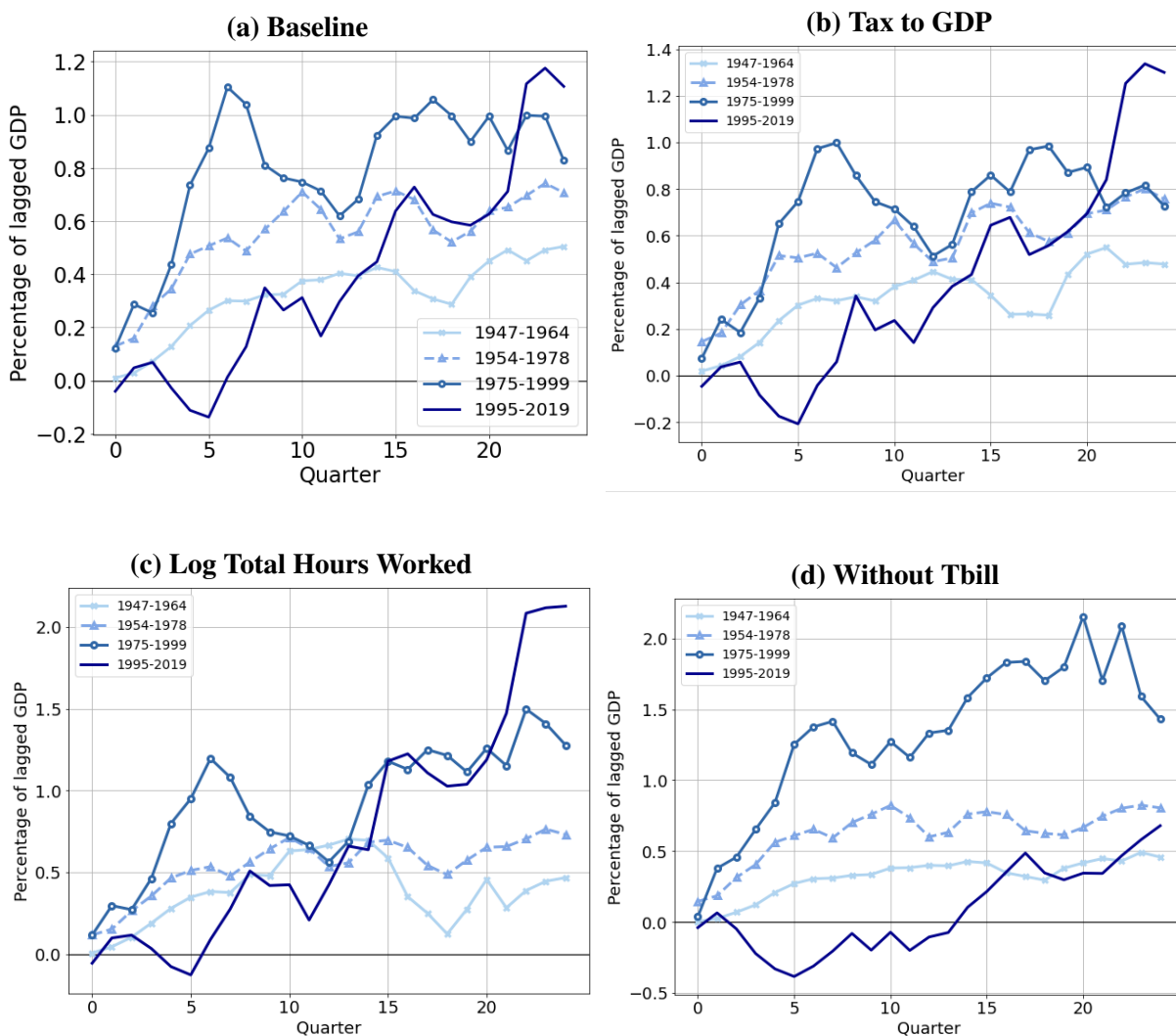


Figure 1.C.2: Sensitivity to Controls- GDP

Notes: The figures show the response of GDP to DNS by sample. The scale of news shock is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. For the model without T-bill, I control for the period of high interest rate volatility from 1978:Q1 to 1982:Q2.

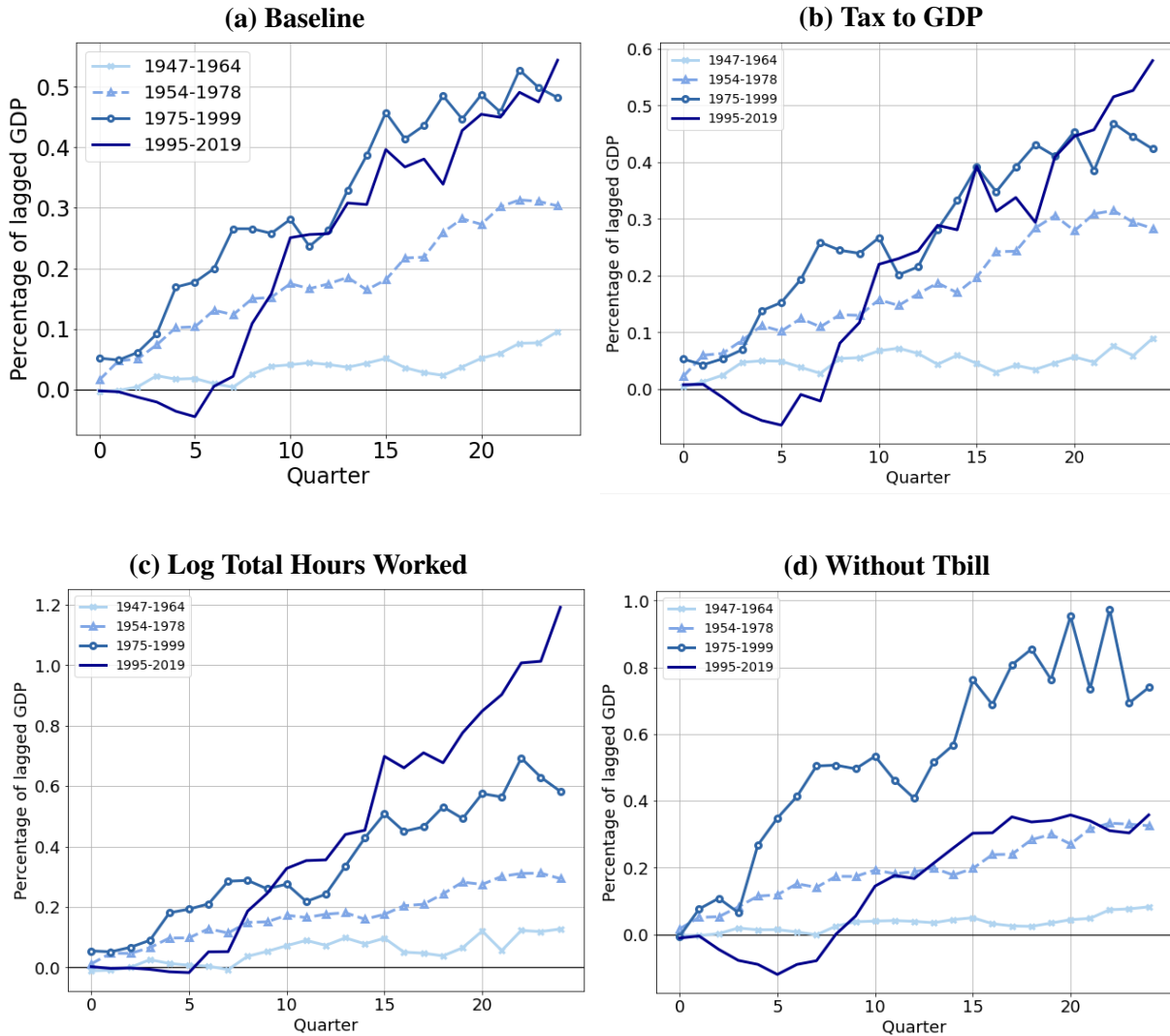


Figure 1.C.3: Sensitivity to Controls- Private Consumption

Notes: The figures show the response of private consumption (non-durable and service) to DNS by sample. The scale of news shock is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. For the model without T-bill, I control for the period of high interest rate volatility from 1978:Q1 to 1982:Q2.

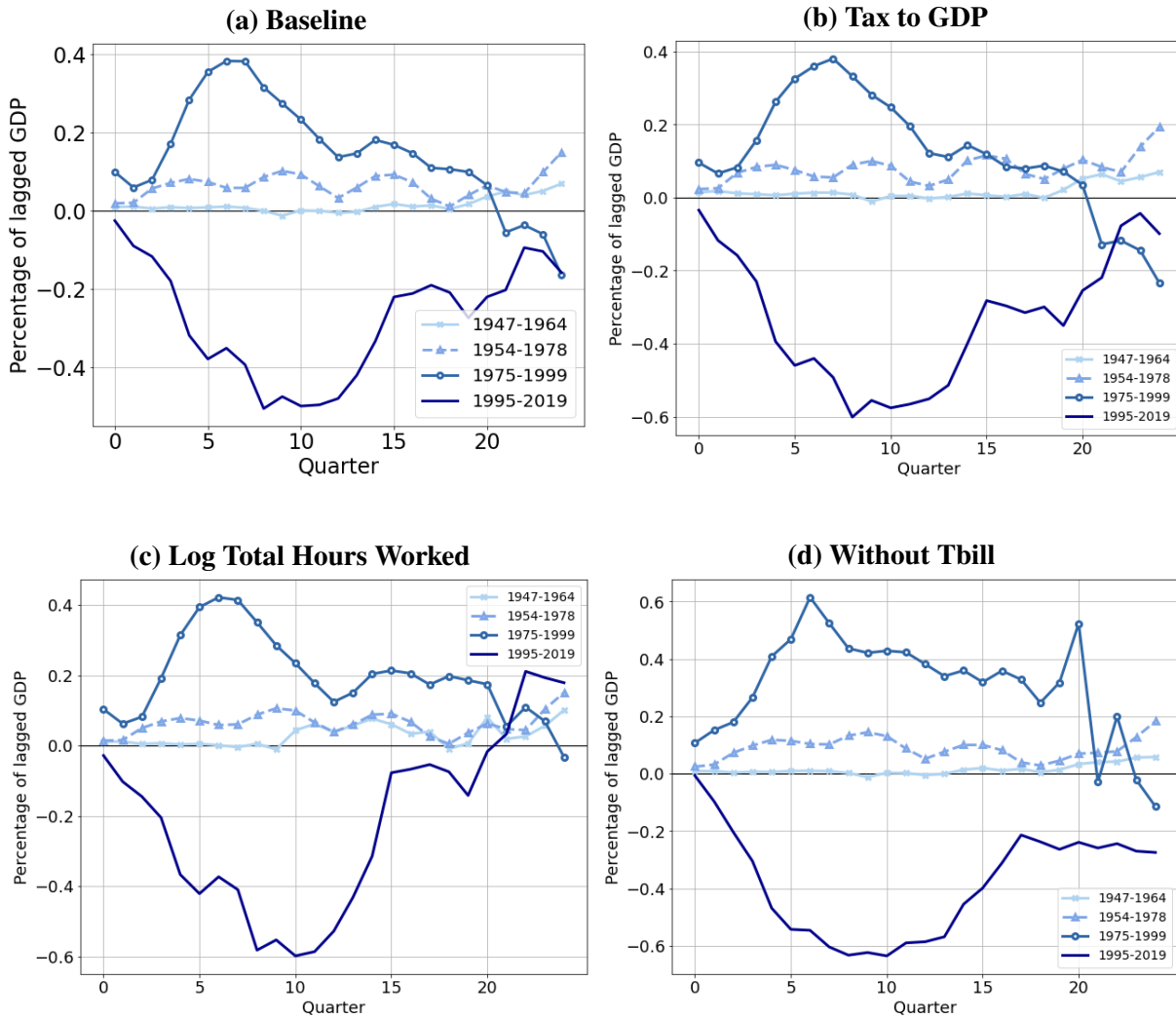


Figure 1.C.4: Sensitivity to Controls- Private Fixed Investment

Notes: The figures show the response of private fixed investment to DNS by sample. The scale of news shock is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. For the model without T-bill, I control for the period of high interest rate volatility from 1978:Q1 to 1982:Q2.

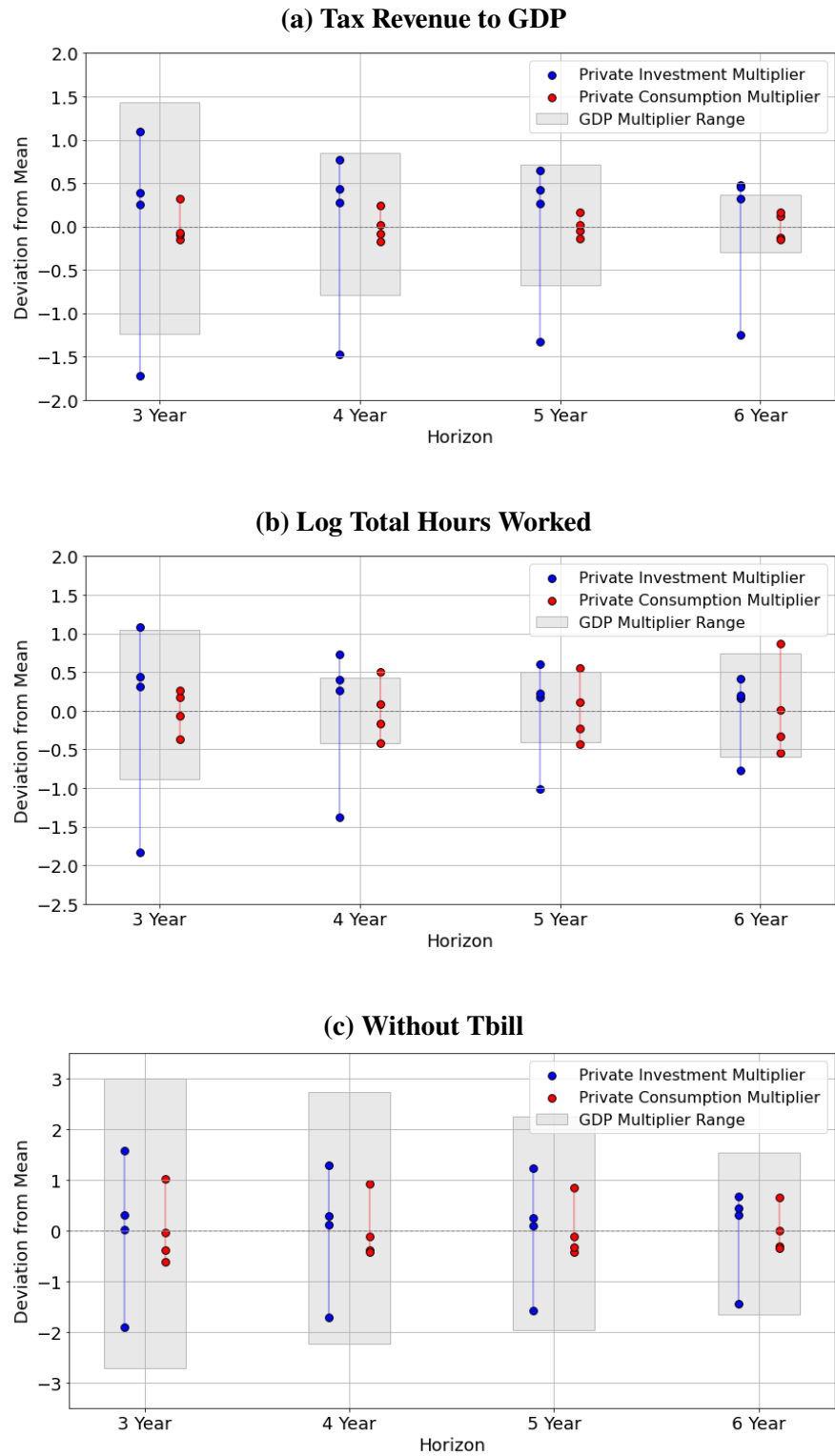


Figure 1.C.5: Sensitivity to Control Variables

Note: For the model without T-bill, I control for the period of high interest rate volatility from 1978:Q1 to 1982:Q2.

1.C.2 Number of Lags

For the model with 5 lags, I exclude the T-bill control in the 1995–2019 sample. To reduce volatility in the consumption response.

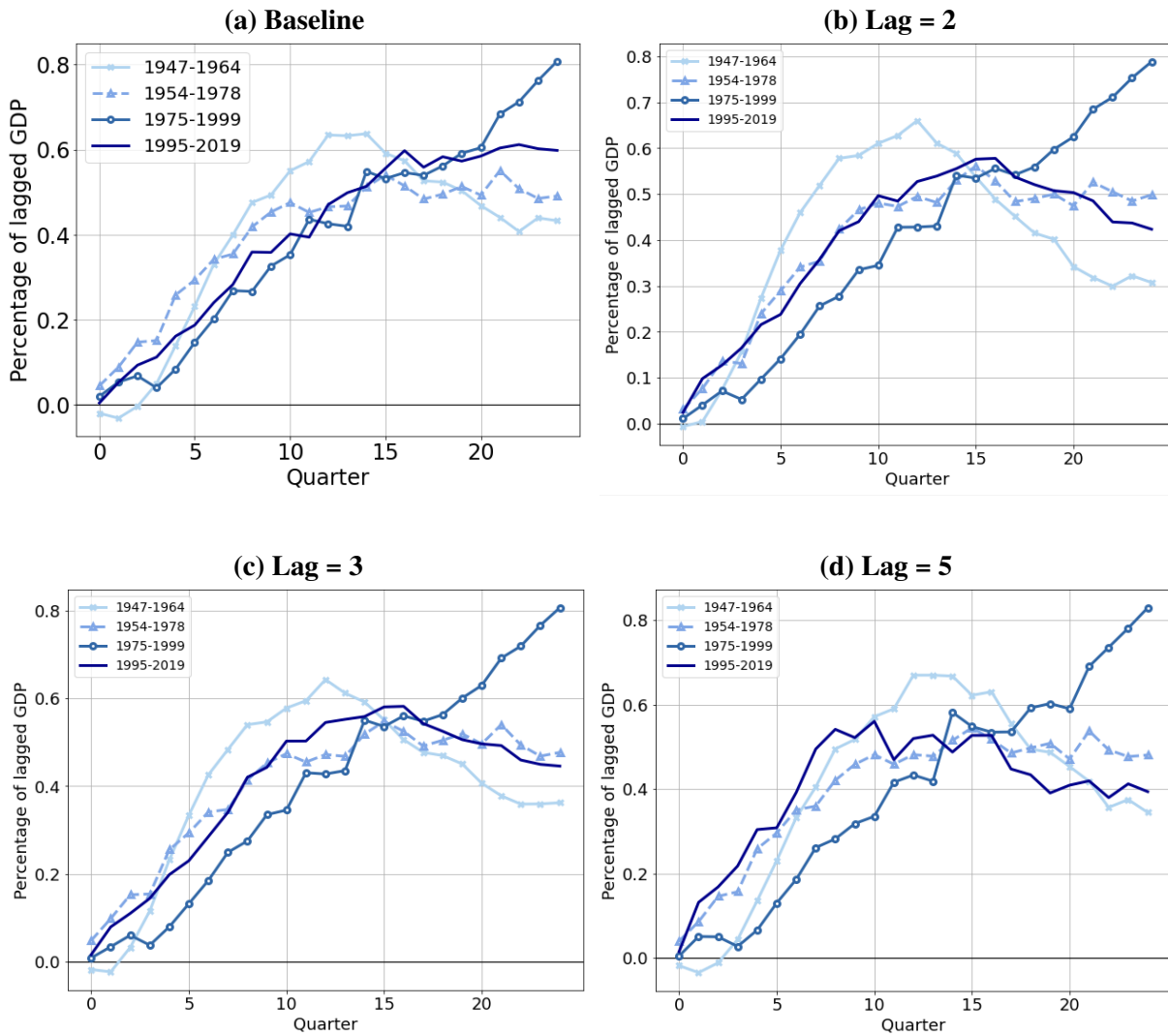


Figure 1.C.6: Sensitivity to Number of Lags- Government Spending

Notes: The figures show the response of government spending to DNS by sample. The scale of news shock is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. For the model with 5 lags, the T-bill is excluded from the set of control variables in the 1995–2019 sample.

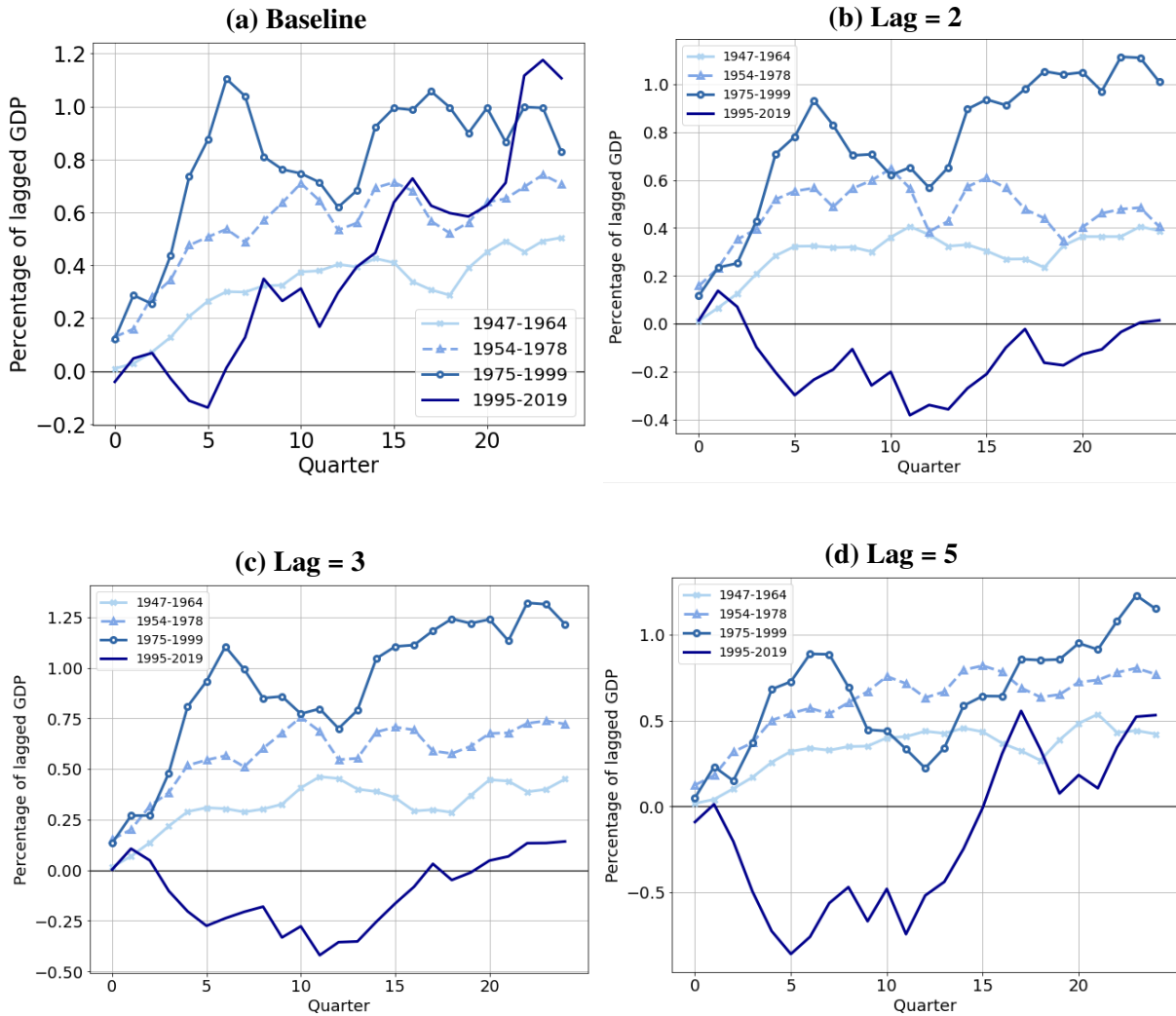


Figure 1.C.7: Sensitivity to Number of Lags- GDP

Notes: The figures show the response of GDP to DNS by sample. The scale of news shock is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. For the model with 5 lags, the T-bill is excluded from the set of control variables in the 1995–2019 sample.

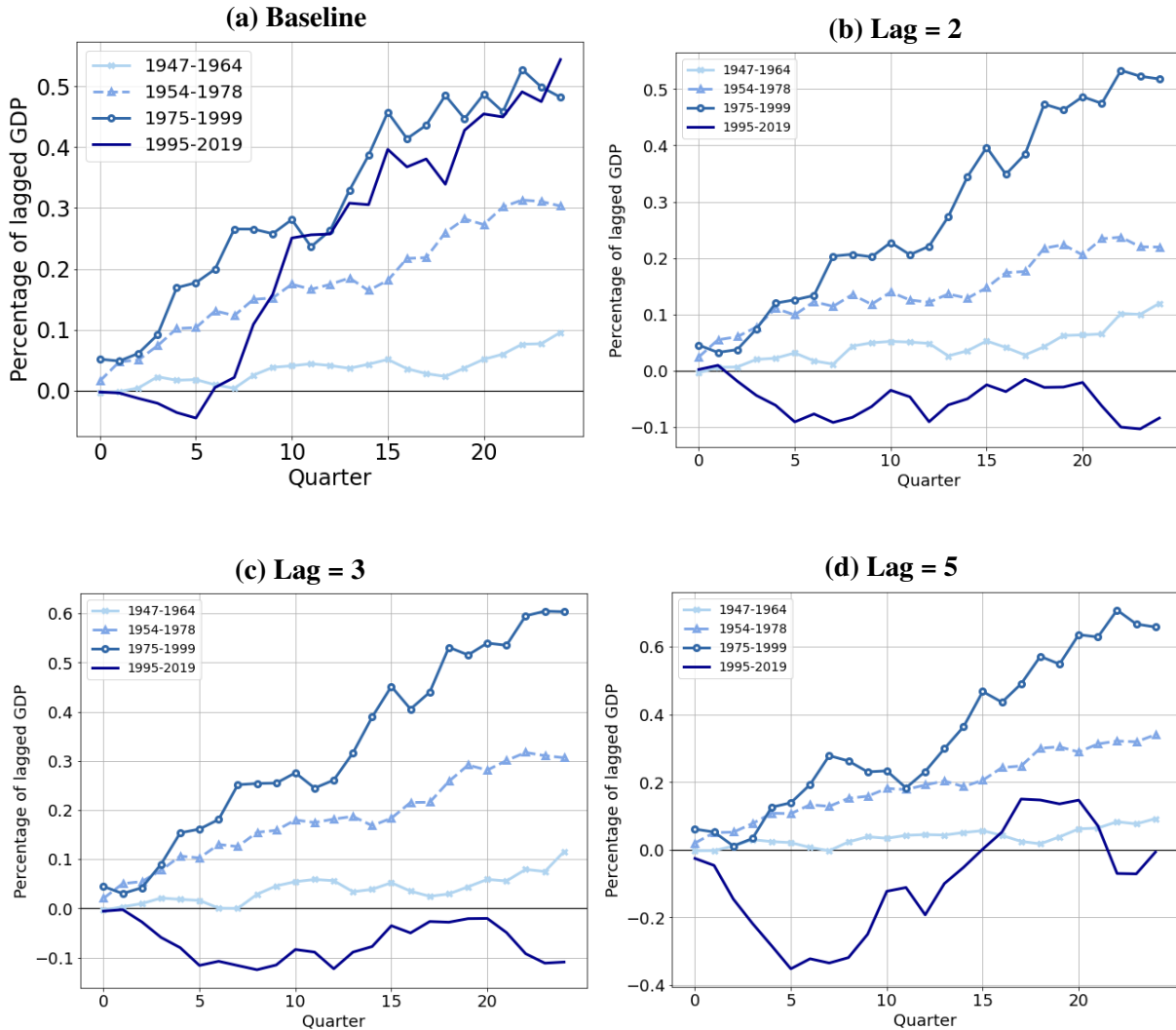


Figure 1.C.8: Sensitivity to Number of Lags- Private Consumption

Notes: The figures show the response of private consumption (non-durable and service) to DNS by sample. The scale of news shock is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. For the model with 5 lags, the T-bill is excluded from the set of control variables in the 1995–2019 sample.

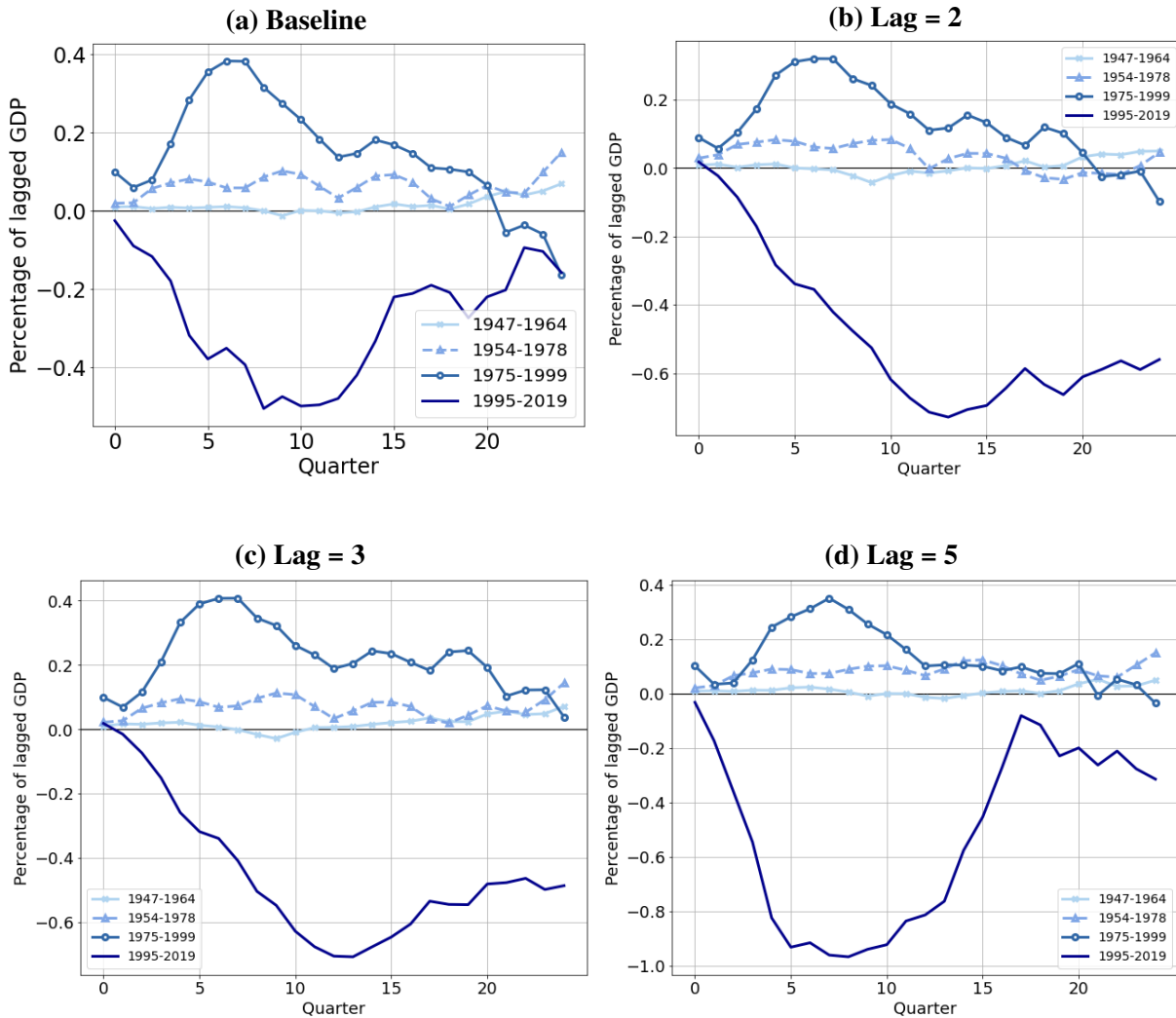


Figure 1.C.9: Sensitivity to Number of Lags- Private Fixed Investment

Notes: The figures show the response of private fixed investment to DNS by sample. The scale of news shock is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. For the model with 5 lags, the T-bill is excluded from the set of control variables in the 1995–2019 sample.

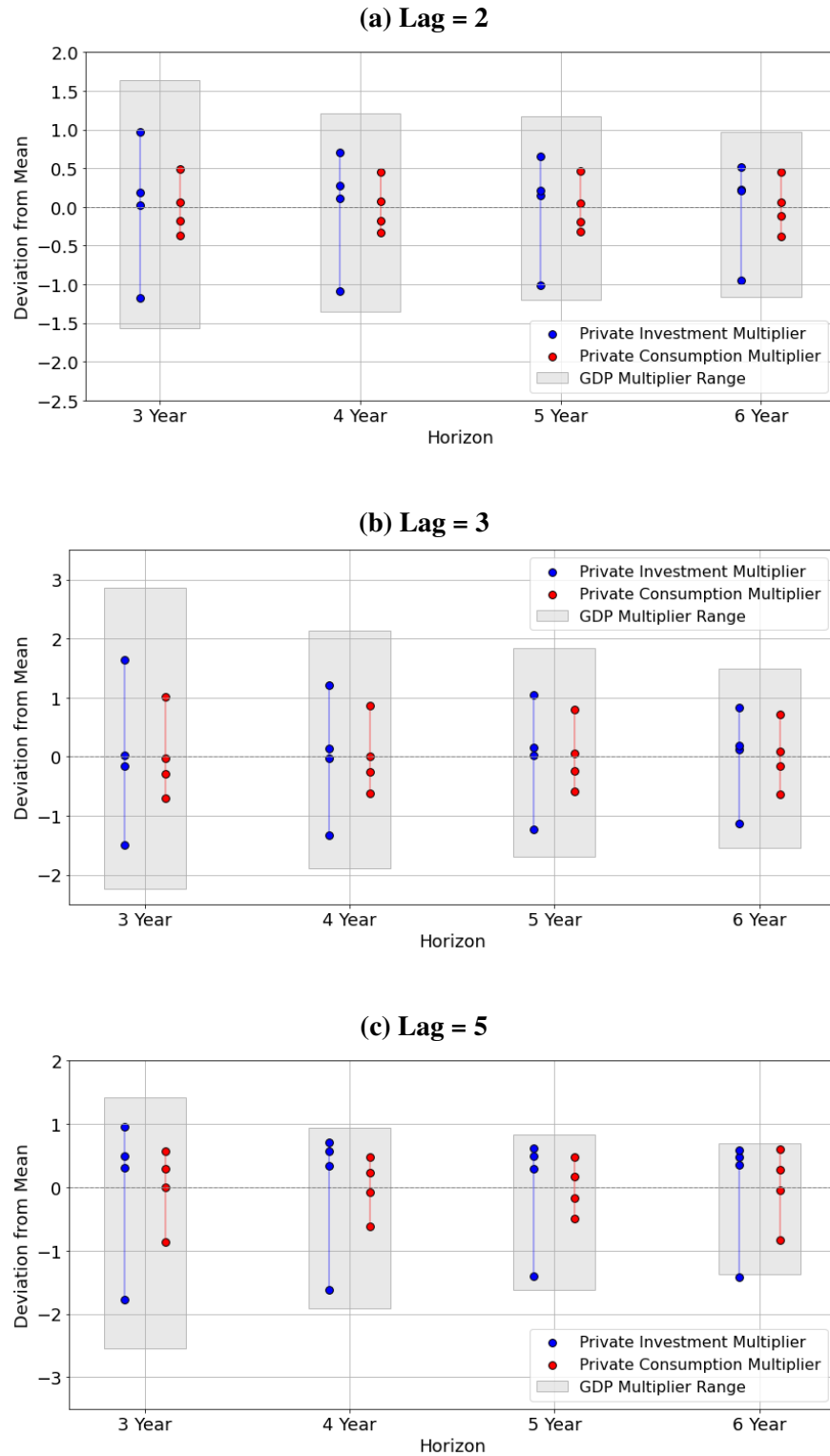


Figure 1.C.10: Sensitivity to Number of Lags

Notes: For the model with 5 lags, the T-bill is excluded from the set of control variables in the 1995–2019 sample.

1.D Robustness of the effects by type of government spending

In the baseline specification, I instrument the cumulative change in total government spending with defense budget authority by type. As an alternative specification, I construct a proxy for defense procurement spending using data from the National Income and Product Accounts (NIPA), defined as the sum of defense intermediate goods and defense investment within total defense expenditures. Non-procurement defense spending is defined as the remainder of total defense spending excluding procurement spending.⁵ I then use defense budget authority as an instrument for each type-specific cumulative change in defense spending when estimating the multiplier equation (1.D.1). Figure 1.D.1 shows, consistent with the main text, that procurement spending induces a pronounced crowding-in of private investment, whereas the effect of non-procurement spending is much weaker.

$$\sum_{j=0}^h \frac{x_{t+j} - x_{t-1}}{y_{t-1}} = \gamma_h + Z'_{t-1} \phi_h + m_h^{(\ell)} \sum_{j=0}^h \frac{g_{t+j}^{(\ell)} - g_{t-1}^{(\ell)}}{y_{t-1}} + \omega_{t+h} \quad (1.D.1)$$

$$\ell \in \{P, NP\} \text{ (P : procurement, NP : non-procurement)}$$

As another alternative specification, following Brunet (2022), I directly use the defense budget authority by type, without employing an instrumental variable. The cumulative changes in procurement and non-procurement budget authority are included in the regression as explanatory variables. In equation (1.D.2), m_h^P and m_h^{NP} denote the cumulative multipliers associated with

⁵Cox et al. (2024) include intermediate services when constructing their proxy for procurement using NIPA data. In contrast, defense procurement budget authority refers to purchases of tangible items, such as equipment and weapon systems, and therefore excludes most service-related expenditures. The “intermediate services” category in NIPA is conceptually closer to the “Operation and Maintenance (O&M)” category in the defense budget, which covers a wide range of activities such as equipment maintenance, personnel training, administrative operations, and logistics support. When the government outsources these functions to the private sector through service contracts instead of directly hiring personnel, the corresponding payments are recorded under both the O&M budget and NIPA’s intermediate services. Although such outsourcing formally involves transactions with private firms, its economic substance may be closer to an indirect purchase of labor inputs, mediated through private firms, rather than the acquisition of final goods or services produced for the market. In this sense, including intermediate services in the definition of procurement could blur the conceptual boundary between purchasing production inputs and procuring private-sector outputs. Accordingly, the defense procurement budget authority—focused on equipment and other tangible items—can be viewed as more directly aligned with the purchase of final outputs.

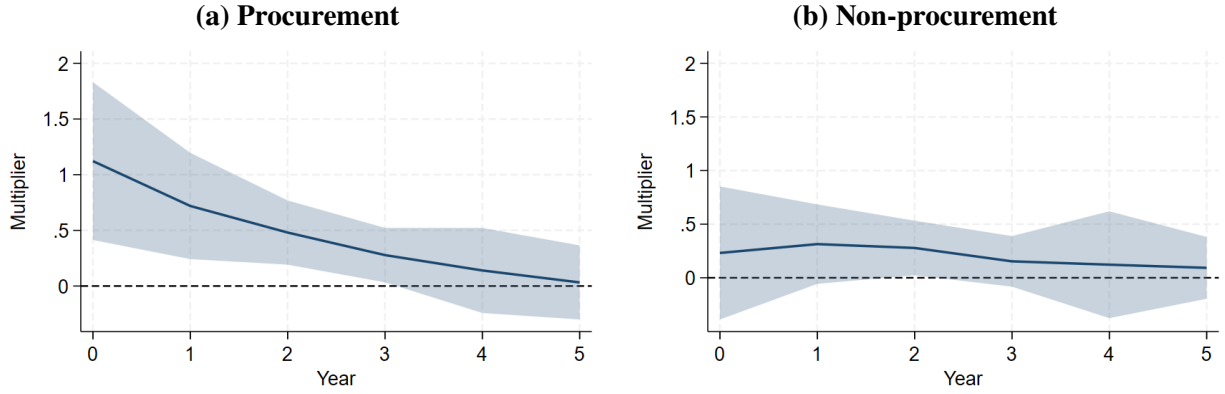


Figure 1.D.1: Private Investment Multipliers by Defense Budget Authority Type

Notes: The figures show the private fixed investment cumulative multipliers by defense budget type. Non-procurement budget authority is the combined total of operation and maintenance (O&M) and military personnel. The shaded area represent the 90% confidence intervals.

procurement and non-procurement budget authority, respectively. Figure 1.D.2 shows a clear contrast between the two components: procurement budget authority exerts a positive and significant effect on private investment, whereas non-procurement budget authority tends to crowd out private investment.

$$\sum_{j=0}^h \frac{x_{t+j} - x_{t-1}}{y_{t-1}} = \gamma_h + Z'_{t-1} \phi_h + m_h^P \sum_{j=0}^h \frac{BA_{t+j}^P - BA_{t-1}^P}{y_{t-1}} + m_h^{NP} \sum_{j=0}^h \frac{BA_{t+j}^{NP} - BA_{t-1}^{NP}}{y_{t-1}} + \omega_{t+h} \quad (1.D.2)$$

(P : procurement, NP : non-procurement)

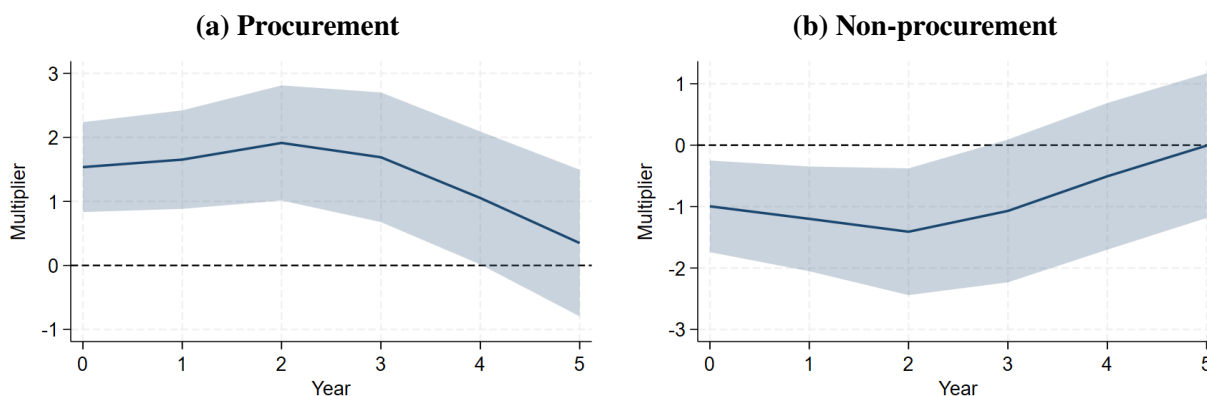


Figure 1.D.2: Private Investment Multipliers by Defense Budget Authority Type

Notes: The figures show the private fixed investment cumulative multipliers by defense budget type. Non-procurement budget authority is the combined total of operation and maintenance (O&M) and military personnel. The shaded area represent the 90% confidence intervals.

1.E Extra Empirical Analysis

1.E.1 Crowding-out Effects on Private Investment

Government spending may also reduce private investment by raising real interest rates. This mechanism, commonly referred to as the crowding-out effect, is a well-known concept in classical economics. Variation in the crowding-out effects is also an important factor behind the time-varying impact of government spending. In this subsection, I examine how the crowding-out effect differs across samples and explores its underlying drivers.

Real Corporate Bond Rates and Crowding-Out

Figure 1.E.1 shows that the crowding-out effect emerges only in the Reagan-era sample (1975–1999) and the War on Terror sample (1995–2019). The left panel shows that the real BAA corporate bond rate increases in response to government spending news only in these two periods. While both samples exhibit a rise in real rates, their patterns somewhat differ. During the Reagan-era sample, the real BAA rate gradually rises following positive government spending news and remains elevated through quarter 12. In contrast, during the War on Terror period, the rate jumps immediately after

the news shock and fluctuates at a high level until around quarter 10. The right panel shows that in both samples, increases in real interest rates are followed by declines in private investment. In the Reagan-era sample, the negative correlation between real interest rates at horizon $h-k$ and private investment at horizon h begins at $k = 2$ and grows in magnitude until $k = 6$. This delayed effect reflects the presence of an initial crowding-in response in this period, which masks the crowding-out effect in the early quarters. By contrast, in the War on Terror sample, real interest rates exhibit a strong negative correlation with private investment as early as $k = 1$, suggesting that the immediate rise in rates quickly crowds out private activity without generating any crowding-in effect.

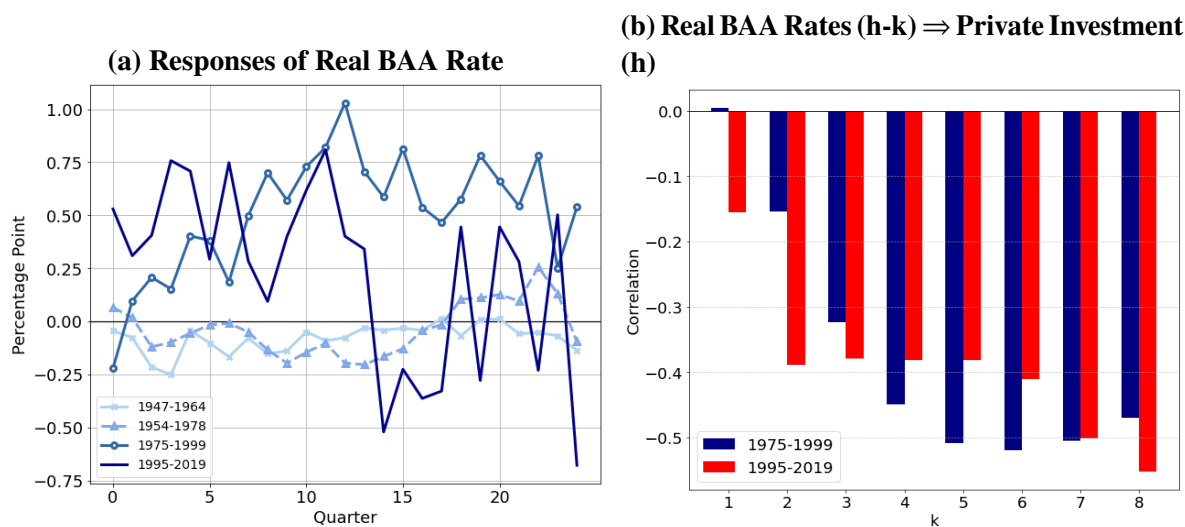


Figure 1.E.1: Responses of Real Baa Rate and Correlation with Private Investment Responses

Notes: The left panel shows the impulse response of the real BAA corporate bond rate to a DNS. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The right panel shows the correlation between the responses of the real BAA corporate bond rate at horizon $h-k$ and those of private investment at horizon h .

Why Does Real Corporate Bond Rate Rise?

Active Monetary Policy: What determines the crowding-out effect through higher real corporate bond rates? Monetary policy plays an important role in shaping the magnitude of this effect. For example, when the central bank follows an “active” Taylor rule, it responds more aggressively to increases in expected inflation and the output gap triggered by government spending news shocks.

This policy response transmits to higher real market interest rates, thereby dampening private investment. Thus, examining how the monetary policy regime differs across subsamples helps us understand the time variation in the magnitude of the crowding-out effect.

Figure 1.E.2 shows that the crowding-out effect during the Reagan-era sample can be attributed to active monetary policy, in contrast to the War on Terror period, when monetary policy appears to be passive. In the left panel, the 3-month real Treasury bill rate rises significantly in response to positive government spending news shocks. The result is related to the historical context of the period. This sample includes the Volcker disinflation episode (1979–87), during which the Federal Reserve maintained positive real interest rates to suppress high inflation. In contrast, the War on Terror period is better characterized as a “passive” monetary policy regime. In the right panel, the 3-month real Treasury bill rate does not rise but instead declines in response to positive government spending news shocks. This result is related to the historical context in which the Federal Reserve implemented preemptive monetary easing in response to the economic slowdown and financial instability following the 9/11 attacks.

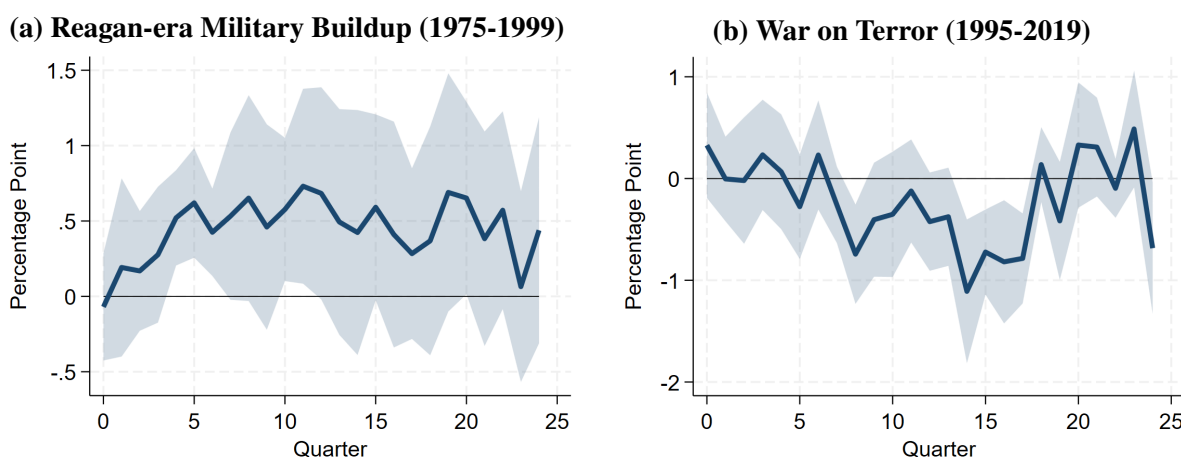


Figure 1.E.2: Response of 3 Month Real T-bill Rate

Notes: These figures show the responses of the real 3 month treasury bill rates to a DNS. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

Term Premium: How can private investment be crowded out during the War on Terror period despite the passive monetary policy stance? Given that the Baa corporate bonds in our sample have maturities of 20 years or longer, an increase in the term premium can offer a plausible explanation. For this channel to operate, the extent to which government spending raises fiscal deficits plays a critical role. Spending news that signals higher future deficits can elevate uncertainty about long-term interest rates and inflation, thereby increasing the term premium. Consistent with this mechanism, Wiegand (2025) provides empirical evidence that deficit news shocks raise the term premium.

Figure 1.E.3 shows that fiscal policy is more active during the War on Terror period, as fiscal deficit responds quickly and substantially to government spending news. The left panel, which covers the Reagan-era sample, shows that fiscal deficit responds only modestly in the short run and begins to rise noticeably after 15 quarters following a positive government spending news shock. In contrast, the right panel shows that during the War on Terror period, the same shock leads to an immediate and pronounced increase in deficits. The response continues to build through the first 15 quarters and remains elevated thereafter.

Figure 1.E.4 shows how the difference in fiscal stances between the two samples translated into distinct responses in the term structure. The left panel shows that during the Reagan-era sample, the initial responses of the term premium is modest. Although the term premium gradually increased after fiscal deficits began to rise around the 15th quarter, the response remains statistically insignificant compared to that observed in the War on Terror sample. In contrast, the right panel shows that during the War on Terror period, the term premium rises sharply and significantly over the first 10 quarters following the shock. This pattern reflects the immediate rise in fiscal deficits triggered by government spending news during this period.

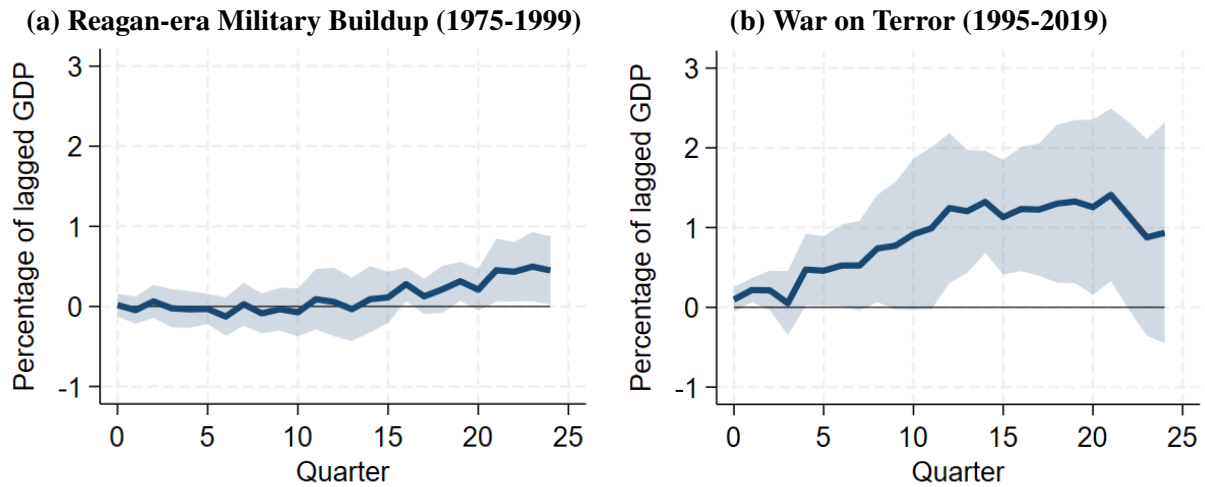


Figure 1.E.3: Response of Real Federal Deficit

Notes: The figures show the responses of the real federal deficit to a DNS. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

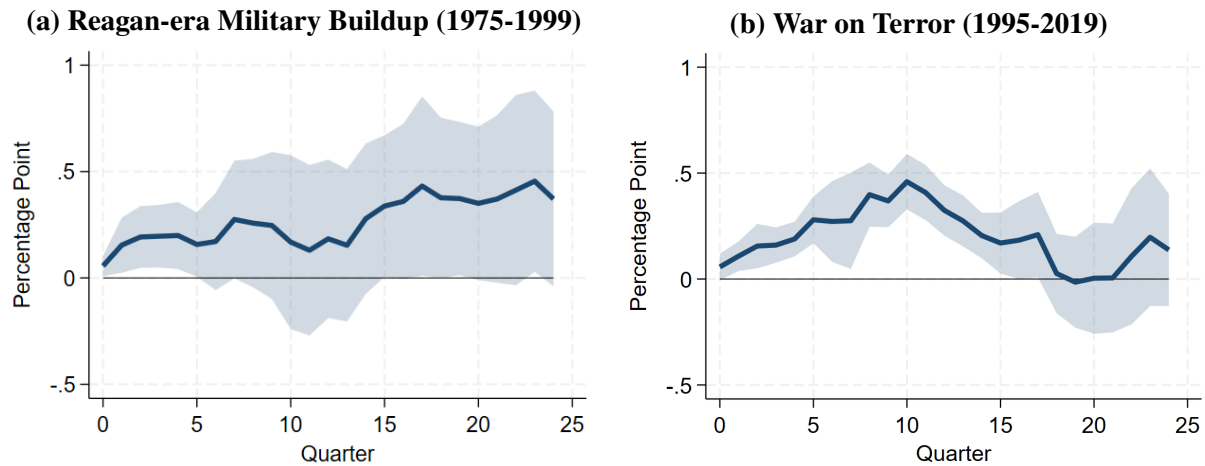


Figure 1.E.4: Responses of the 10 Year Term Premium

Notes: The figures show the responses of the 10 year term premium to a DNS. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

1.E.2 Procurement Composition and Private Investment Crowding-In

Section 1.6.2 has shown that the time profile of government spending plans is a key factor to amplify the private investment crowding-in effect. This appendix takes the argument a step further by demonstrating that this difference in time profiles is fundamentally driven by the composition of procurement. In short, the type of goods the government procures determines the time horizon of the spending plan, which in turn triggers heterogeneous investment responses across industries.

The most significant difference between the Reagan era and Vietnam war period lies in the composition of goods procured from the private sector. As shown in Figure 1.E.5, the majority of the spending increase during the Reagan era was for defense equipment. By its nature, procuring such durable equipment entails multi-year contracts and long production lead times. This incentivizes firms to undertake preemptive capital investment well before delivery in anticipation of certain long-term demand. In contrast, the Vietnam War period was dominated by the procurement of short-term nondurable goods such as petroleum products, and ammunition. These goods are intended for immediate demand fulfillment, thereby exhibiting the characteristics of the short-term time profile providing a relatively weak incentive for sustained investment.

This mechanism is clearly confirmed by the industry-level private investment responses in Figure 1.E.6. During the Reagan era, investment in the defense industries, which were directly linked to these procurement contracts, surged distinctly and led the overall private investment crowding-in. During the Vietnam era, however, a similar, modest investment increase was observed in both defense and non-defense industries, resulting in a weak aggregate-level crowding-in effect.

Overall, the macroeconomic factor of the time profile presented in the main text does not exist in isolation; it is supported by the micro-level mechanism of the detailed composition of procurement. The strong crowding-in during the Reagan era was the result of a procurement composition focused on defense durables that created a long-term time profile, which in turn intensively induced investment in related industries. This finding strongly reinforces the central argument of this paper.

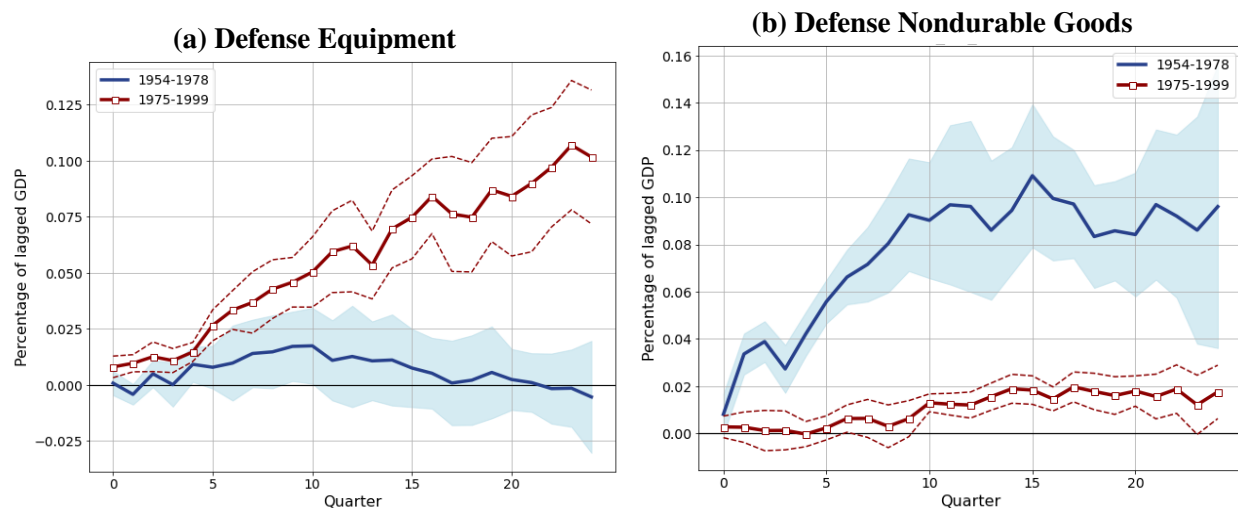


Figure 1.E.5: Responses of Defense Spending by Type of Goods

Notes: The figures show the responses of defense spending by type of goods to DNS for the 1954–1978 and 1975–1999 samples. Defense equipment is the national defense equipment investment in BEA Table 3.9.5 (line 21). Nondurable defense consumption goods is national defense consumption expenditures for intermediate nondurable goods in BEA Table 3.10.5 (line 30). The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

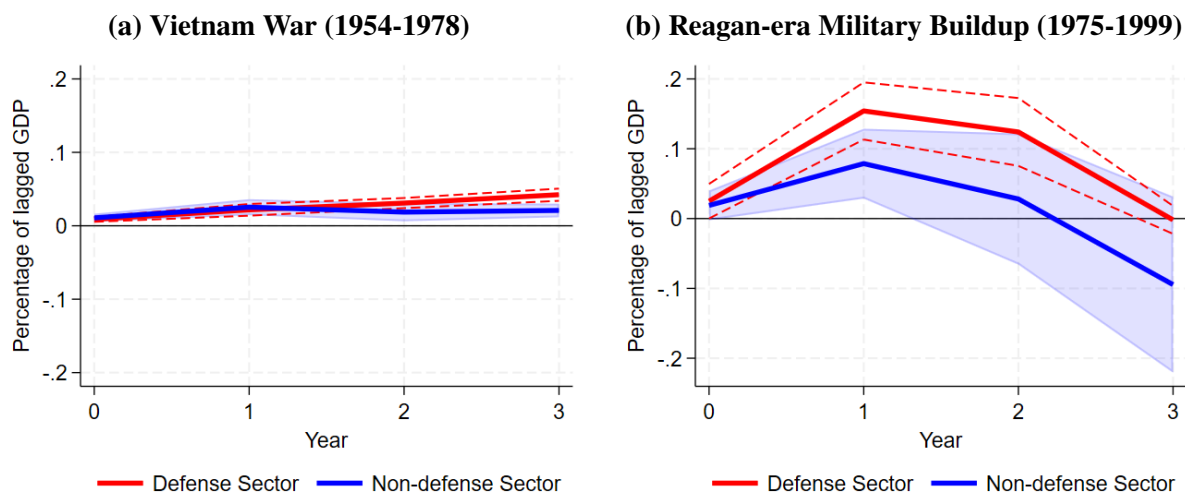
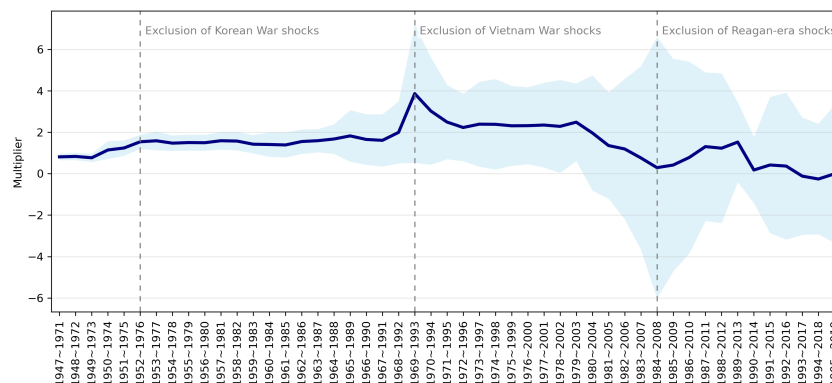


Figure 1.E.6: Responses of Private Fixed Investment by Sector

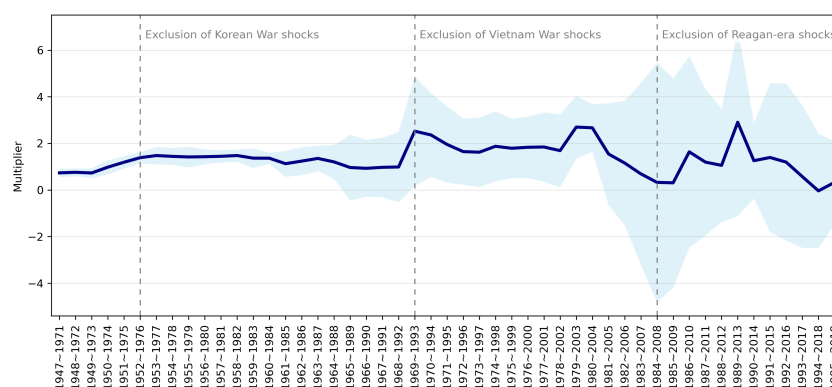
Notes: The figures show the responses of private fixed investment to DNS for the 1954–1978 and 1975–1999 samples. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

1.E.3 Rolling Window Multipliers

(a) 3-year Horizon Cumulative GDP Multiplier



(b) 4-year Horizon Cumulative GDP Multiplier



(c) 5-year Horizon Cumulative GDP Multiplier

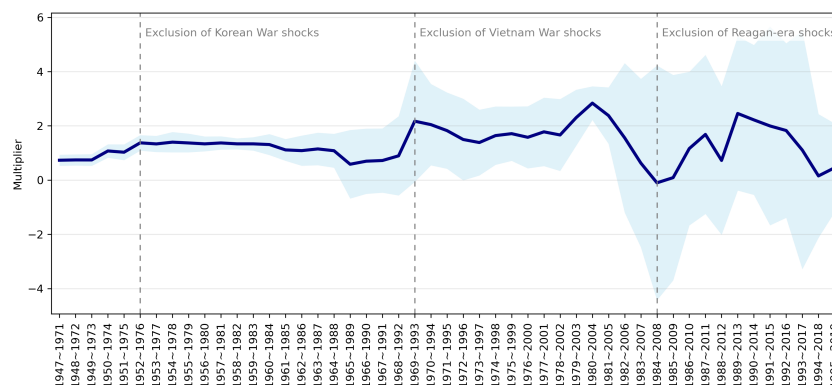


Figure 1.E.7: Rolling window cumulative GDP Multipliers

Notes: The figures show cumulative GDP multipliers at the 3-, 4-, and 5-year horizons, estimated with a 25-year rolling window (1-year step). Shaded area represent 90% confidence intervals.

1.E.4 Weak Effect of Government Spending in the Korean War Sample

When an economy is near its potential output, a positive demand shift tends to have a greater impact on prices rather than real GDP. This happens because labor and capital are already near full capacity, making it difficult to further increase production in the short run.⁶ According to this reasoning, the effects of government spending news also vary depending on the state of the economy. During a recession, government spending news can more easily translate into higher production.⁷ Conversely, when the economy approaches its potential output, the same level of spending tends to fuel inflation rather than real GDP.

Given a convex supply curve, it is necessary to examine the economic conditions in each sample in order to predict the effects of government spending news. Table 1.E.1 shows that economic conditions were closer to full employment during the 1947–1964 sample period, with an average unemployment rate of 4.86%, the lowest among all sample periods. Notably, during the Korean War (1950Q2–1953Q3), when large-scale defense expenditures were undertaken, the unemployment rate averaged a historically low 3.40%. From 1954 to 1978, the average unemployment rate rose slightly to 5.37%. During the Vietnam War period (1965Q1–1973Q4), the unemployment rate averaged 4.5%, higher than during the Korean War. In the early 1960s (1960Q1–1963Q4), a period marked by heightened Cold War tensions due to events such as the Space Race and the Cuban Missile Crisis, the average unemployment rate was 5.9%. The average unemployment rate for the 1975–1999 sample period was the highest at 6.62%. Notably, during the Reagan administration (1981Q1–1989Q1), when defense spending increased significantly, the average unemployment rate was exceptionally high at 7.50%. These results suggest that the effects of government spending are weaker in the 1947–1964 sample, which had the strongest economic conditions.

The magnitude of government spending increases can also play a pivotal role. Under a convex supply curve, a sizable demand surge shifts the equilibrium closer to potential output, pushing

⁶Boehm and Pandalai-Nayar (2022) provides empirical evidence for this by showing that industry supply curves are convex, which implies a significant drop in supply elasticity near potential output.

⁷Auerbach, Gorodnichenko, et al. (2011) and Ramey and Zubairy (2018) show that government spending during economic downturns can have substantial effects on output.

the economy toward its capacity limits. Consequently, the output effects of government spending diminish, while price pressures intensify. Table 1.E.1 shows that the maximum value of real defense spending growth is largest during the 1947–1964 period. Due to the exceptionally large shocks during the Korean War, the maximum value reached 24.86% in 1951Q1. In the 1954–1978 sample, it was a more moderate 6.02% in 1966Q3 due to the influence of the Vietnam War, whereas during the 1975–1999 period, it was the smallest at 4.09% in 1984Q4. Thus, the exceptionally large scale of maximum defense spending growth from 1947 to 1964 suggests a weaker effect of government spending on the real economy.

In summary, the high unemployment rate and the exceptionally large maximum growth in defense spending observed in the 1947–1964 sample suggest a lower multiplier, implying that government spending has a greater impact on prices than on real output. Figure 1.E.8 supports this claim: in the 1947–1964 sample, the price level rises steadily for up to 12 quarters after the shock, with inflation remaining persistently positive. Real variables, by contrast, exhibit relatively weaker responses during this period. The top right panel of Figure 1.E.9 shows that the increase in real GDP over the first 12 quarters is comparatively smaller in 1947–1964. The middle two panels show that the responses of real consumption and the unemployment rate are also noticeably weaker. The bottom two panels show that total hours worked increase only modestly in this sample, whereas nominal wages, like prices, increase more significantly.

Table 1.E.1: Economic Conditions and Defense Spending Growth

	Unemployment Rates			Defense Spending Growth		
	Average	Max	Min	Average	Max	Min
1947-1964	4.86	7.37	2.57	1.21	24.86	-7.03
1954-1978	5.39	8.87	3.40	-0.24	6.02	-6.24
1975-1999	6.62	10.67	4.07	0.29	4.09	-5.20
1995-2019	5.69	9.93	3.60	0.31	4.68	-5.20

Notes: The left part of the table shows the average, maximum, and minimum unemployment rates for each sample period, while the right part shows the average, maximum, and minimum real defense spending growth for each sample period.

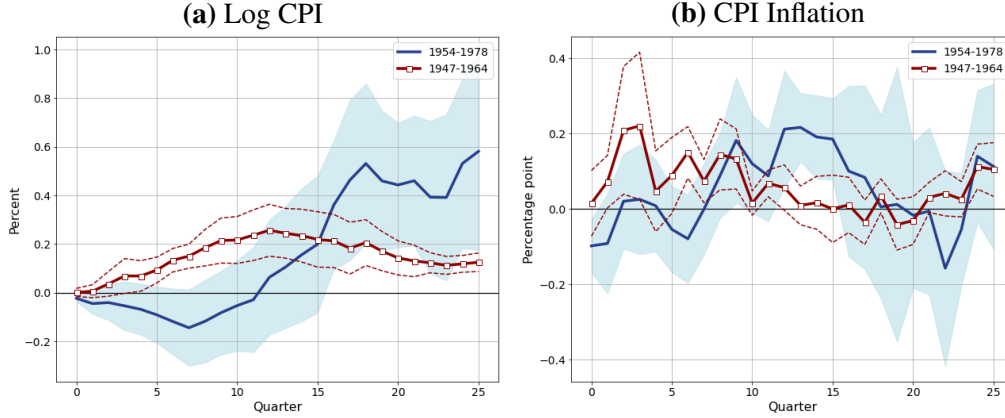


Figure 1.E.8: Responses of CPI

Notes: The figures show the responses of CPI variables to DNS for the 1947–1964 and 1954–1978 samples. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

1.E.5 State Dependent Multipliers

A nonlinear local projection (LP) model can be used to estimate state-dependent government spending multipliers. In this analysis, I examine whether the multiplier is higher during periods of economic slack. The slack state is defined as periods when the unemployment rate exceeds 6.5%.

The model is specified as follows:

$$\sum_{j=0}^h \frac{y_{t+j} - y_{t-1}}{y_{t-1}} = I_{t-1} \left[\gamma_{A,h} + \phi_{A,h}(L) z_{t-1} + m_{A,h} \sum_{j=0}^h \frac{g_{t+j} - g_{t-1}}{y_{t-1}} \right] + (1 - I_{t-1}) \left[\gamma_{B,h} + \phi_{B,h}(L) z_{t-1} + m_{B,h} \sum_{j=0}^h \frac{g_{t+j} - g_{t-1}}{y_{t-1}} \right] + \omega_{t+h}, \text{ for } h = 0, 1, \dots, H \quad (1.E.1)$$

where $I_{t-1} \times DNS_t$ and $(1 - I_{t-1}) \times DNS_t$ are used as an instrument variable for the respective interaction of cumulative government spending with the two state indicators. y is real GDP and g is real government spending. γ_A and γ_B are constant terms for each state. For the baseline model, z

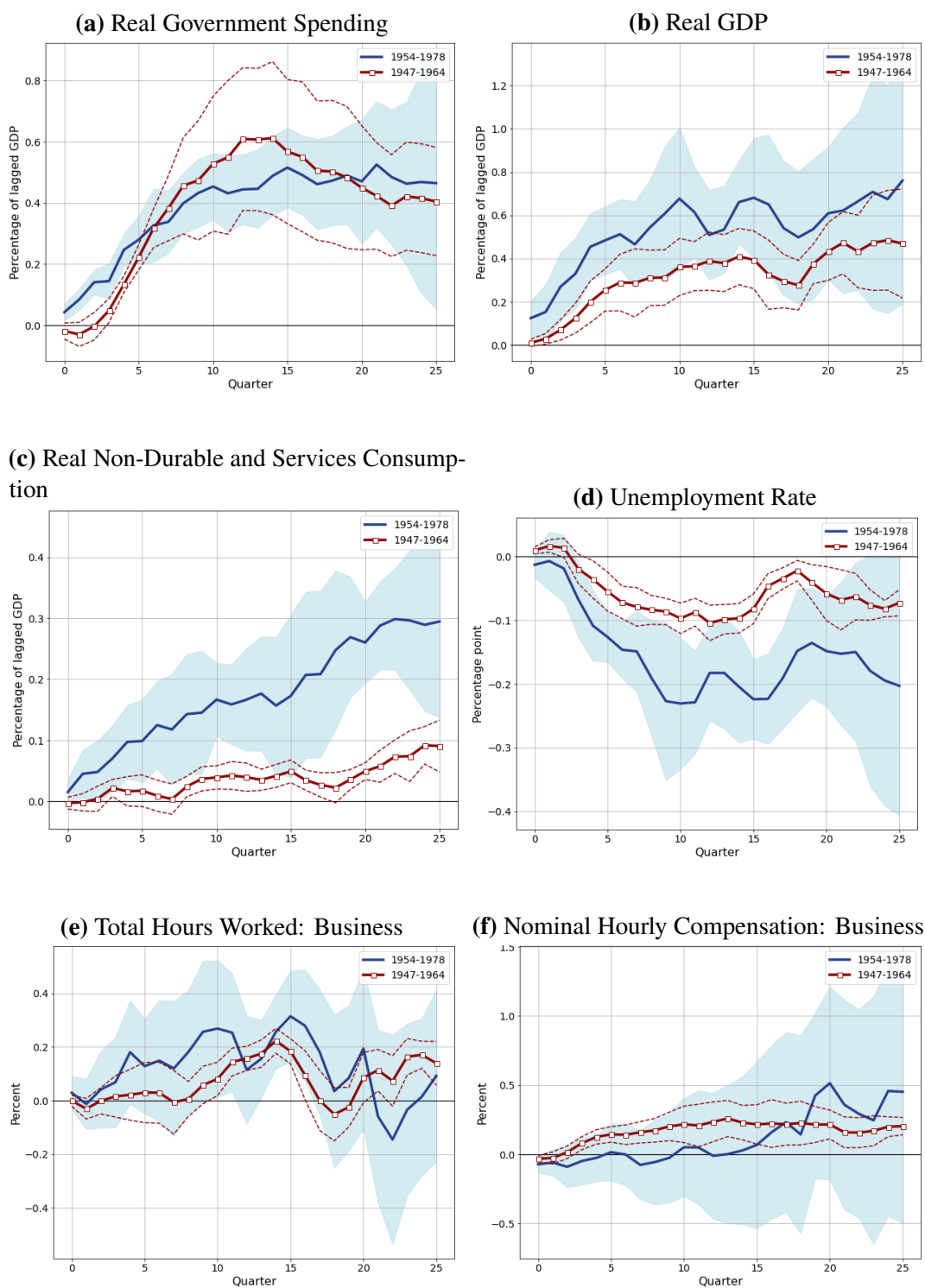


Figure 1.E.9: Responses of Macro Variables

Notes: The figures show the responses of NIPA variables to DNS for the 1954–1978 and 1975–1999 samples. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10% of the real GDP in $t-1$. The shaded areas and dashed lines represent the 90% confidence intervals.

consists of four lags of $y_t - y_{t-1}$ divided by y_{t-1} , $g_t - g_{t-1}$ divided by y_{t-1} , difference in 3-month tbill rate, and DNS. $\phi(L)$ is a fourth-order polynomial in the lag operator.

In the equation, $m_{A,h}$ denotes the cumulative multiplier at horizon h for the slack state, while $m_{B,h}$ denotes the cumulative multiplier at horizon h for the normal state. Table 1.E.2 shows that the cumulative multiplier in the slack state exceeds that in the normal state across horizons from 3 to 6 years. The result is robust in both the full sample (1947–2019) and the post–Korean War sample (1954–2019).

Table 1.E.2: Government Spending Multipliers by State of Slack

	Slack: 1947-2019	Normal: 1947-2019	Slack: 1954-2019	Normal: 1954-2019
3-year horizon	1.83*** (0.41)	0.89*** (0.14)	2.09*** (0.61)	1.45*** (0.46)
4-year horizon	1.67*** (0.36)	0.83*** (0.17)	2.00** (0.47)	1.24*** (0.39)
5-year horizon	1.61*** (0.33)	0.85*** (0.13)	1.93*** (0.41)	1.12*** (0.34)
6-year horizon	1.57*** (0.32)	0.91*** (0.12)	1.81*** (0.37)	1.13*** (0.31)

Notes: The table shows the 3-year, 4-year, 5-year, and 6-year cumulative government spending multipliers by state of slack. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Standard errors are reported in parentheses.

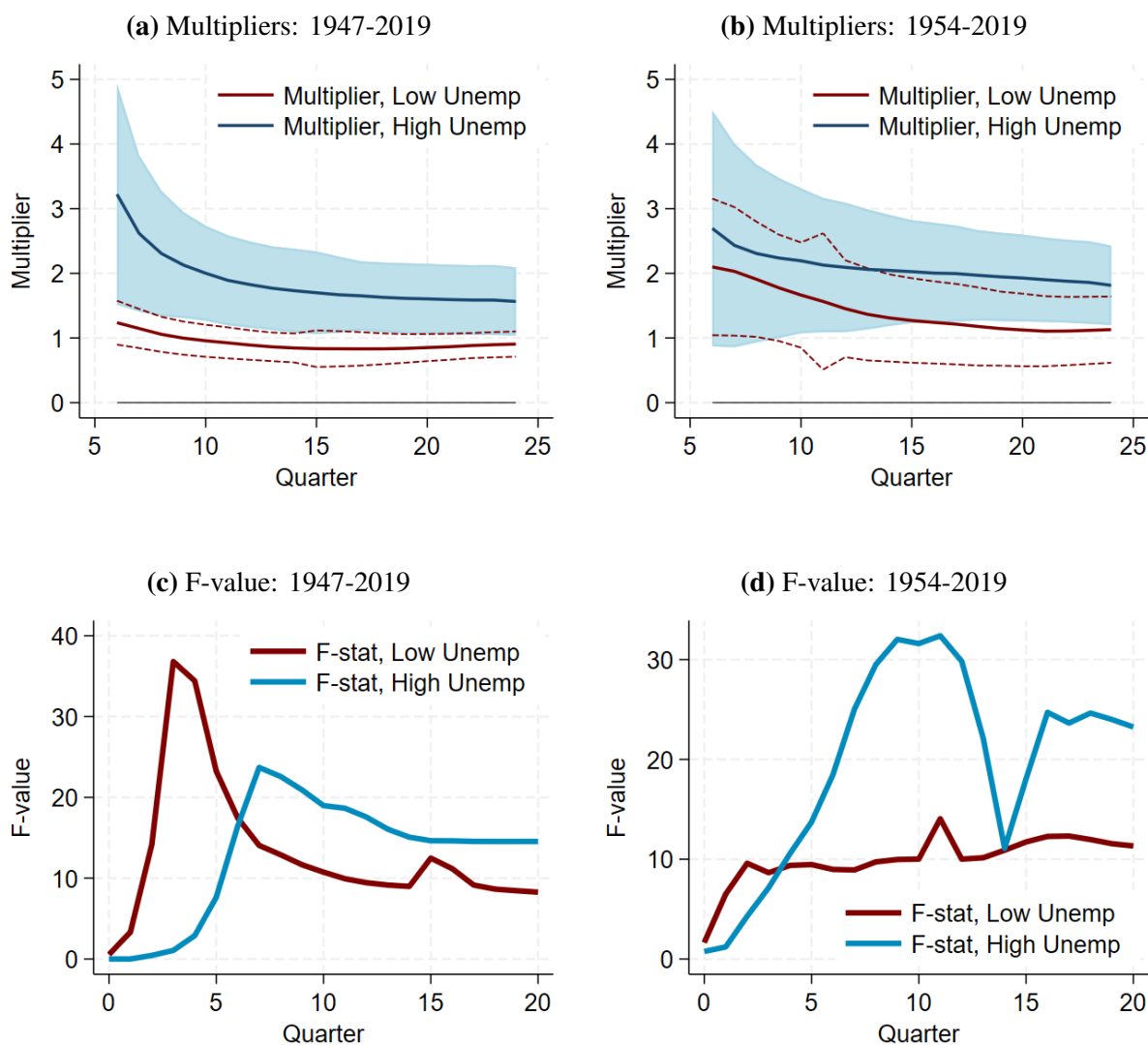


Figure 1.E.10: Government Spending Multipliers by State of Slack

Notes: The figures show the government spending multipliers and first-stage F-values for the slack state and normal state. The left and right panels show the estimation results for the 1947–2019 sample and the 1954–2019 sample, respectively.

1.F Theory: New Keynesian Model with Government Spending News

In this section, I employ a standard New Keynesian (NK) model with government spending news shocks to theoretically rationalize the empirical findings presented in the previous section. In particular, the model examines whether it can account for the stronger crowding-in of private investment observed during the Reagan-era sample (1975–1999), which includes the Volcker disinflation, relative to the Vietnam War period (1954–1978), which spans the Great Inflation episode. The analysis proceeds in two steps. First, I show that when government spending news signals a persistent increase in spending and a back-loaded implementation path (time profile), the crowding-in effect on private investment becomes stronger. Second, I demonstrate that incorporating these characteristics of government spending is essential for jointly explaining the large private investment response and the coexistence of a more active monetary policy regime during the Reagan-era sample.

1.F.1 Model Description

This study’s theoretical analysis is based on the New Keynesian dynamic stochastic general equilibrium (DSGE) model of Leeper et al. (2017), which offers a flexible framework for examining the dynamic effects of fiscal policy across different monetary–fiscal regimes. The main assumptions are: (i) prices and wages follow Calvo-type nominal rigidities, (ii) government spending directly enters household utility, (iii) consumption displays habit formation, and (iv) rule-of-thumb (non-saving) households are excluded.

Another advantage of this model is that Leeper et al. (2017) provide a set of posterior mean parameters, estimated via Bayesian methods. Their estimation is based on U.S. quarterly data from 1955:Q1 to 2007:Q4, which largely encompasses the Vietnam War period (1954–1978) and the Reagan-era sample (1975–1999). Accordingly, I fix all parameters at their posterior means, except

for a subset of key parameters that are varied in the model experiments.

Before turning to the model analysis, it is necessary to clarify the composition of government spending the model represents. Within the framework of Leeper et al. (2017), government spending is modeled as a composite of private intermediate goods, which closely corresponds to procurement spending in the empirical analysis. Therefore, the model is well suited for analyzing the Vietnam War and the Reagan-era samples, as these periods were characterized by a high share of procurement spending and exhibited crowding-in effects.⁸

Households. Savers optimize utility from composite consumption, $C_t^{*,S}(j)$, consisting of private, $C_t^S(j)$, and public, G_t , consumption goods. α_G is the degree of substitutability of the consumption goods: when $\alpha_G < 0$, private and public consumption are complements; when $\alpha_G > 0$, the goods are substitutes.

$$E_t \sum_{t=0}^{\infty} \beta^t u_t^b \left(\ln(C_t^{*,S}(j) - \theta \tilde{C}_{t-1}^{*,S}) - \frac{(L_t^S(j))^{1+\xi}}{1+\xi} \right) \quad (1.F.1)$$

$$C_t^{*,S}(j) \equiv C_t^S(j) + \alpha_G G_t$$

$$\begin{aligned} & P_t(1 + \tau_t^C)C_t^S(j) + P_t I_t^S(j) + P_t^B B_t(j) + R_t^{-1} B_{s,t}(j) \\ &= (1 + \rho P_t^B)B_{t-1}(j) + B_{s,t-1}(j) + (1 - \tau_t^L) \int_0^1 W_t(l) L_t^S(j, l) dl \\ &+ (1 - \tau_t^K) R_t^K v_t(j) \bar{K}_{t-1}^S(j) - \psi(v_t) \bar{K}_{t-1}^S + P_t Z_t^S(j) + D_t(j) \end{aligned} \quad (1.F.2)$$

Non-savers have the same preferences as savers. Non-savers are rule-of-thumb agents who consume their entire disposable income each period, which consists of after-tax labor income and lump-sum transfers from the government, Z^N .

⁸To explain the War on Terror period, when crowding-out dominated, we would need to incorporate non-procurement spending, which involves government production technologies. I leave this extension for future research.

$$(1 + \tau_t^C)P_t C_t^N(j) = (1 - \tau_t^L) \int_0^1 W_t(l) L_t^N(j, l) dl + P_t Z_t^N(j) \quad (1.F.3)$$

Firms. A perfectly competitive final goods producer uses a continuum of intermediate goods, $Y_t(i)$, to produce the final good, Y_t , with the constant-return-to-scale technology.

$$\left(\int_0^1 Y_t(i)^{\frac{1}{1+\eta_t^p}} di \right)^{1+\eta_t^p} \geq Y_t \quad (1.F.4)$$

Dixit-Stiglitz aggregation yields the demand is:

$$Y_t(i) = Y_t \left(\frac{\bar{P}_t(i)}{\bar{P}_t} \right)^{-\frac{1+\eta_t^p}{\eta_t^p}}. \quad (1.F.5)$$

Intermediate goods producers are monopolistic competitors in their product market. Firm i has access to the technology.

$$Y_t(i) = K_t(i)^\alpha (A_t L_t(i))^{1-\alpha} - A_t \Omega, \quad (1.F.6)$$

where $\alpha \in [0, 1]$ and $\Omega > 0$ represents fixed costs to production that grow at the rate of technological progress.

Prices follow a Calvo (1983) mechanism, under which an intermediate firm has a probability $(1 - \omega_p)$ each period of being able to reoptimize its price. Firms that reoptimize their price in period t maximize expected discounted nominal profits subject to the demand for $Y_t(i)$. Given the production function, average and marginal costs coincide, which allows expected discounted nominal profits to be written as:

$$E_t \sum_{s=0}^{\infty} (\beta \omega_p)^s \frac{\lambda_{t+s}}{\lambda_t} \left[\left(\prod_{k=1}^s \pi_{t+k-1}^{\chi_p} \pi^{1-\chi_p} \right) P_t(i) Y_{t+s}(i) - M C_{t+s} Y_{t+s}(i) \right] \quad (1.F.7)$$

Aggregation. Aggregate consumption is:

$$C_t = \int_0^1 C_t(j) dj = (1 - \mu)C_t^S + \mu C_t^N \quad (1.F.8)$$

Goods market clearing is:

$$Y_t = C_t + I_t + G_t + \psi(v_t)\bar{K}_{t-1}. \quad (1.F.9)$$

Policy rules. Fiscal and monetary policy rules play a central role in the model experiments, as selected coefficients in these rules are allowed to vary to match the empirical patterns observed in each subsample. Thus, I present only the policy rule equations here. The full set of model equations appears in Appendix 1.F.1. First, government spending follows the fiscal rule given in Equation 1.F.10.

$$\hat{g}_t = \rho_G \hat{g}_{t-1} - (1 - \rho_G)\gamma_G \hat{s}_{t-1}^b + u_t^G \quad (1.F.10)$$

$$s_{t-1}^b \equiv \frac{P_{t-1}^B B_{t-1}}{P_{t-1} Y_{t-1}}$$

where $\hat{g}_t$ is log-deviation of government spending from its steady-state value, ρ_G is the persistence of government spending, $\hat{s}_{t-1}^b$ is the log-deviation of debt-to-GDP ratio from its steady-state value, γ_G is the intensity of fiscal feedback to the debt-to GDP ratio, and u_t^G is a government spending shock. s_{t-1}^b is defined as the ratio of the nominal market value of government bonds outstanding ($P_{t-1}^B B_{t-1}$) to nominal GDP ($P_{t-1} Y_{t-1}$). Next, monetary policy follows a standard monetary policy rule in Equation 1.F.11.

$$\hat{R}_t = \rho_r \hat{R}_{t-1} + (1 - \rho_r)[\phi_\pi \hat{\pi}_t + \phi_y \hat{y}_t] + u_t^m \quad (1.F.11)$$

$$u_t^m = \rho_m u_{t-1}^m + \epsilon_t^m$$

where $\hat{R}_t$ is the nominal interest rate, $\hat{\pi}_t$ is quarterly inflation, and $\hat{y}_t$ is output gap, and u_t^m is a monetary policy shock that follows an AR(1) process. Parameter ϕ_π is the interest rate responses to

inflation, ϕ_y the interest rate responses to output gap, and ρ_r is the interest-rate smoothing.

Policy regimes. Monetary–fiscal regimes are defined by how strongly monetary policy reacts to economic conditions and how government spending responds to government debt. In the monetary-led regime (regime M), the central bank takes an active role in stabilizing prices and output by responding strongly to inflation and the output gap, high values of the ϕ_π and ϕ_y . Concurrently, fiscal policy plays a passive role, ensuring debt sustainability through strong feedback to government debt, high value of the γ_G . Conversely, in a fiscal-led regime (regime F), government spending rules show little to no feedback to debt, allowing fiscal policy to be more active in influencing the economy. In this scenario, monetary policy tends to be passive, exhibiting a weak response to inflation and the output gap.

Government spending news. To consider the *time profile* of government spending in the model, I introduce a government spending news shock. This specification is designed to replicate the empirical impulse response path of government spending. To this end, I incorporate two key elements into the government spending process. First, in contrast to the standard model where a contemporaneous shock leads to an immediate jump and gradual decay, a positive news shock here generates a *hump-shaped* response. This response is characterized by a gradual rise followed by a peak, reflecting situations in which the government announces plans to expand spending over several years. Second, an *implementation lag* arises, representing a delay of k periods between the announcement of the spending plan (the news) and the actual onset of the spending increase.

To incorporate these two features, I follow Ascari et al. (2023), who adopt the specification proposed by Ramey (2009). Specifically, the stochastic process for government spending, u_t^G , is assumed to follow an AR(3) process. This process is driven by a 'news' shock, ϵ_t^G , which only begins to affect spending after a k -period delay:

$$u_t^G = \rho_1^\eta u_{t-1}^G + \rho_2^\eta u_{t-2}^G + \rho_3^\eta u_{t-3}^G + \epsilon_{t-k}^G \quad (1.F.12)$$

$$\rho_1^\eta > 1, \rho_2^\eta < 0, \rho_3^\eta < 0$$

In this specification, ϵ_t^G represents the 'news' arriving at time t about future spending. The parameter k explicitly captures the *implementation lag*: news announced at time t only begins to impact the level of government spending u^G at time $t + k$. The *hump-shaped* dynamics after the spending begins are generated by the autoregressive coefficients. The coefficient ρ_1^η is set to be greater than 1 ($\rho_1^\eta > 1$) to produce the initial, gradual increase in spending that forms the upward-sloping part of the hump. To ensure the overall process remains stationary, the parameters ρ_2^η and ρ_3^η are assigned negative values. These coefficients counteract the explosive root from ρ_1^η , causing the spending path to peak and then revert to its steady state, thereby completing the hump-shaped response.

1.F.2 Regime Selection

Leeper et al. (2017) provide *posterior estimates* for two policy regimes using U.S. data from 1955:I to 2007:IV. The regimes differ in the priors on the policy rule coefficients: under Regime M, the prior means for the interest rate response to inflation ϕ_π and for the responses of government consumption and transfers to debt γ_G , γ_Z are set higher, whereas under Regime F they are set lower. Table 1.F.1 reports the posterior means and 90% credible sets for both regimes. As expected, the most pronounced cross-regime differences in the posteriors arise for ϕ_π , γ_G , γ_Z ; however, the prior differences also spill over to other parameters. For example, interest rate inertia ρ_r is substantially higher under Regime M, and the interest rate's response to output ϕ_y is modestly larger.

Table 1.F.1: Estimated Parameters in Leeper et al. (2017)

Parameter	Regime M		Regime F	
	Mean	90% CS	Mean	90% CS
<i>Preference and HHs</i>				
ξ , inverse Frisch labor elasticity	1.54	[0.92, 2.14]	2.32	[1.49, 3.20]
θ , habit formation	0.99	[0.98, 1.00]	0.99	[0.98, 1.00]
α_G , G in utility	-0.19	[-0.36, -0.02]	-0.16	[-0.34, 0.02]
<i>Frictions and production</i>				
100γ , ss tech growth	0.24	[0.18, 0.31]	0.27	[0.20, 0.33]
ψ , capital utilization	0.13	[0.08, 0.17]	0.13	[0.06, 0.19]
s , inv. adj. cost	5.21	[3.68, 6.71]	3.97	[2.47, 5.36]
ω_p , price stickiness	0.89	[0.86, 0.91]	0.95	[0.94, 0.96]
ω_w , wage stickiness	0.87	[0.83, 0.92]	0.85	[0.81, 0.89]
χ_p , price indexation	0.06	[0.01, 0.11]	0.06	[0.01, 0.11]
χ_w , wage indexation	0.09	[0.03, 0.15]	0.09	[0.03, 0.15]
<i>Monetary policy</i>				
ϕ_π , interest rate response to inflation	1.14	[0.98, 1.31]	0.15	[0.08, 0.23]
ϕ_y , interest rate response to output	0.18	[0.13, 0.22]	0.17	[0.14, 0.20]
ρ_r , lagged interest rate response	0.76	[0.71, 0.81]	0.15	[0.06, 0.23]
<i>Fiscal policy</i>				
γ_G , govt. cons. resp. to debt	0.21	[0.13, 0.30]	0.0000	[-0.0017, 0.0016]
γ_Z , transfer resp. to debt	-0.03	[-0.13, 0.08]	0.0000	[-0.0017, 0.0016]
ρ_G , lagged govt. cons. resp.	0.98	[0.98, 0.99]	0.98	[0.98, 0.99]
ρ_Z , lagged transfer resp.	NE	—	0.98	[0.97, 0.99]

Notes: The table shows the posterior means and 90% credible sets of the estimated parameters for Regime M and Regime F reported in Leeper et al. (2017). The estimates are based on data from 1955:I to 2007:IV.

The differences in the estimated parameter values across the two regimes affect the range of γ_G and ϕ_π consistent with a unique equilibrium. Figure 1.F.1 shows the determinacy regions (white areas) obtained by fixing all parameters other than γ_G and ϕ_π at their posterior estimates for each regime. In the left panel, using the Regime F parameters, as γ_G approaches zero, the determinacy condition transitions to one with an upper bound on ϕ_π (passive monetary policy), with this bound forming at around $\phi_\pi \approx 0.1$. In contrast, in the right panel, using the Regime M parameters, as γ_G exceeds 0.2, the determinacy condition shifts to one with a lower bound on ϕ_π (active monetary policy), with the upper bound forming around $\phi_\pi \approx 0.3$.

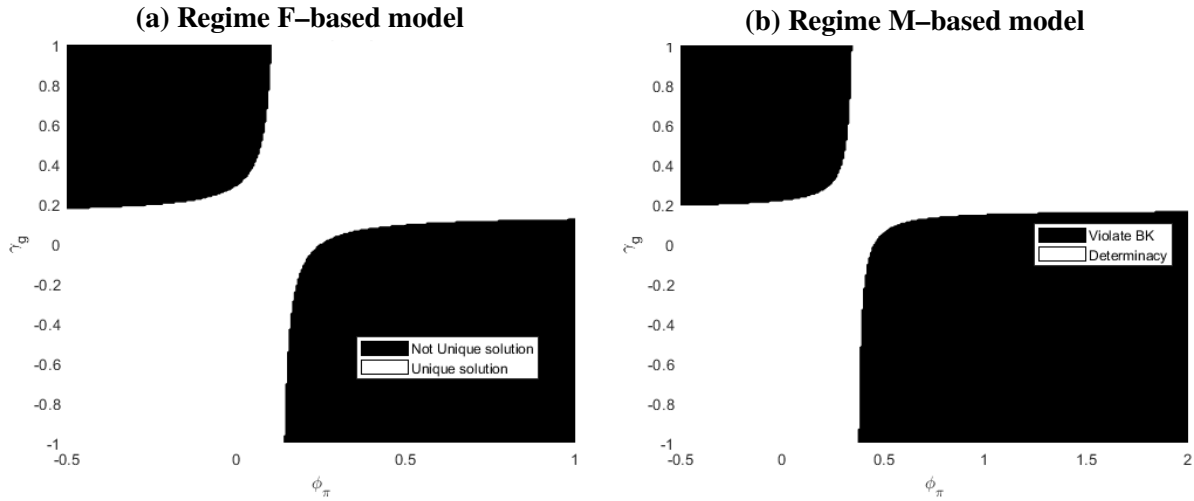


Figure 1.F.1: Determinacy Region of Leeper et al. (2017)

Notes: This figure shows the determinacy region for γ_g and ϕ_π in the model of Leeper et al. (2017). All other parameters are set at the posterior means estimated from the 1955–2007 sample for the two regimes. The left panel shows the determinacy region using Regime F parameters, while the right panel shows the determinacy region using Regime M parameters.

Next, to evaluate which regime better fits the Vietnam War sample and the Reagan-era sample, I perform impulse response function (IRF) matching for each sample under both regimes. I allow ϕ_π to vary freely and select the value that best matches the empirical IRFs. One complication is that, in both regimes, when γ_G is near the interval between 0 and 0.2, a wide range of ϕ_π values satisfies the determinacy condition, making the regime boundary ambiguous. To impose a sharp boundary, I exclude this interval and set $\gamma_G = 0.00$ for Regime F and $\gamma_G = 0.25$ for Regime M. I then select ϕ_π

by IRF matching within the determinacy region for each regime.

The figure shows that the Vietnam War sample (1954–1978) is better accounted for by Regime F, whereas the Reagan-era sample (1975–1999) is better accounted for by Regime M. In the top-left panel, even at the lower bound of the determinacy-admissible range for Regime M, $\phi_\pi = 0.16$, the model's positive response of private investment is far smaller than the empirical IRF, implying that matching would require an even lower ϕ_π , which lies outside the admissible range for Regime M. By contrast, the top-right panel shows that under Regime F with $\phi_\pi = 0$, the model and data align closely for the investment IRFs. In the bottom-left panel (Reagan-era sample), the model under Regime M matches the empirical IRFs well at $\phi_\pi = 0.60$. Conversely, the bottom-right panel shows that under Regime F, even at its upper bound $\phi_\pi = 0.25$, the model's investment IRF rises well above the empirical counterpart, implying that ϕ_π would need to be higher than the admissible range under Regime F. Taken together, these comparisons favor Regime F for the Vietnam War sample and Regime M for the Reagan-era sample.

Table 1.F.2: Parameter Values from the IRF-matching

Parameter		Regime F		Regime M	
		1954–1978	1975–1999	1954–1978	1975–1999
<i>News Shock Process</i>					
ρ_1^η	coefficient on u_{t-1}^G	1.382	1.425	1.384	1.426
ρ_2^η	coefficient on u_{t-2}^G	-0.18	-0.18	-0.18	-0.18
ρ_3^η	coefficient on u_{t-3}^G	-0.25	-0.25	-0.25	-0.25
k	news horizon	0	0	0	0
<i>Policy Rule</i>					
ϕ_π	interest rate resp. to inflation	0.00	0.25	0.16	0.60
ϕ_y	interest rate resp. to output gap	0.17	0.17	0.18	0.18
ρ_r	lagged interest rate response	0.15	0.15	0.76	0.76
γ_G	gov. cons. resp. to debt	0.00	0.00	0.25	0.25

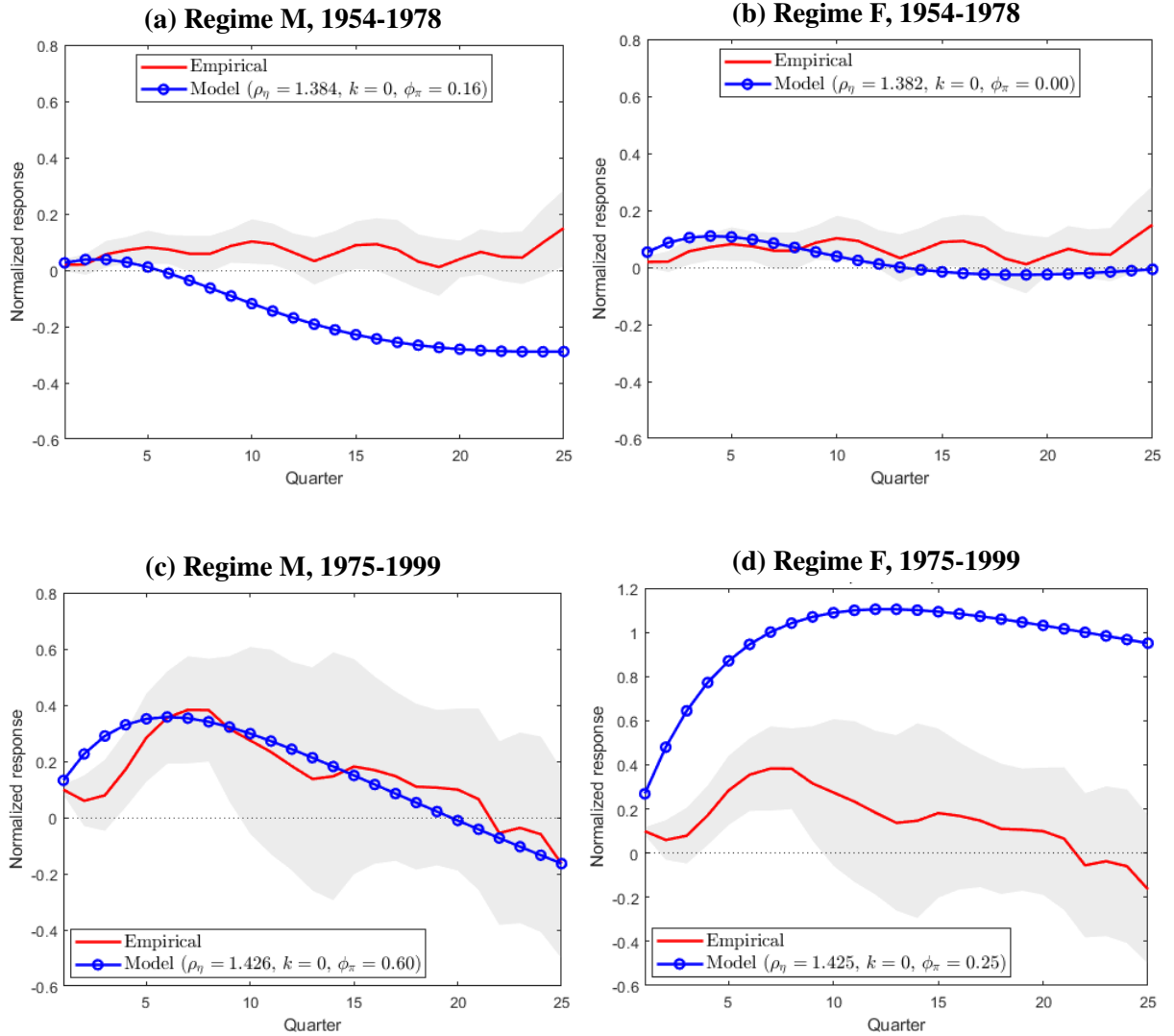


Figure 1.F.2: Replication of Empirical IRFs– Private Investment

Notes: The figures show the response of private investment in NK DSGE model. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10 units of goods.

1.F.3 Role of the Time Profile of Government Spending

To analyze how the time profile of government spending affects the crowding-in effect on private investment, I conduct model simulations using alternative parameter values in Equation 1.F.12. For each parameter combination, I plot the responses of government spending and private investment to a news shock. As in the empirical analysis, the shock is normalized so that the six-year cumulative increase in government spending equals ten units of output. The specific parameter values are summarized in Table 1.F.3.

First, I compare the results across three values of $\rho_1^\eta \in \{1.41, 1.42, 1.43\}$. Following Ramey (2009), ρ_2^η and ρ_3^η are fixed at -0.18 and -0.25 , respectively. A higher ρ_1^η implies a more persistent and back-loaded increase in government spending, with a larger share of spending occurring further in the future relative to the initial news. Next, I vary the implementation-lag parameter $k \in \{0, 4, 8\}$. As k rises, the spending path becomes increasingly back-loaded while maintaining its persistence. For the policy-rule parameters, the interest-rate response to inflation is set to 1.0.⁹ All remaining parameters follow the posterior estimates for the monetary-dominant regime (Regime M) reported in Leeper et al. (2017).¹⁰

The top panels of Figure 1.F.3 show that when the news shock implies a more persistent and back-loaded government spending path, private investment exhibits a stronger initial crowding-in effect. When $\rho_1^\eta = 1.41$, the response of government spending begins to flatten around quarter 15, and no significant crowding-in is observed. In contrast, when $\rho_1^\eta = 1.43$, government spending continues to rise through quarter 24, accompanied by a pronounced initial increase in private investment. These results suggest that firms expand their initial investment more aggressively when they expect a sustained rise in government demand well into the future.

The bottom panels of Figure 1.F.3 show that a longer implementation lag, which implies a more back-loaded spending path, leads to a stronger initial crowding-in effect on private investment. In

⁹The positive relationship between the time profile of spending and the degree of crowding-in holds regardless of this parameter choice.

¹⁰Results are qualitatively similar under the fiscal-dominant regime (Regime F).

contrast to the previous exercise, this simulation isolates the effect of back-loaded implementation while holding persistence fixed. When the implementation lag is zero ($k = 0$), government spending rises immediately, and private investment declines, exhibiting no crowding-in. When the lag is positive, however, the increase in government spending is delayed by k quarters, and private investment rises significantly during the initial period before government spending begins to materialize. These results suggest that firms' forward-looking behavior drives the initial rise in private investment prior to the eventual crowding-out effects that occur once government spending materializes.

Table 1.F.3: Parameter Values for Simulation of News Effects

Parameter		Baseline	Experiment Range
<i>News Shock Process</i>			
ρ_1^η	coefficient on u_{t-1}^G	1.42	{1.41, 1.42, 1.43}
ρ_2^η	coefficient on u_{t-2}^G	-0.18	—
ρ_3^η	coefficient on u_{t-3}^G	-0.25	—
k	implementation lag	0	{0, 4, 8}
<i>Policy Rule</i>			
ϕ_π	interest rate resp. to inflation	1.00	—
ϕ_y	interest rate resp. to output gap	0.18	—
ρ_r	lagged interest rate response	0.76	—
γ_G	gov. spend. resp. to debt	0.21	—

1.F.4 IRF Matching: Private Investment

This section uses the model to explain the empirical IRFs observed during the Vietnam War and the Reagan era. Given the historical contexts of these two periods, the empirical results point to a puzzle that contradicts the predictions of the prevailing monetary–fiscal policy regimes. Specifically, the Vietnam War period corresponds to the pre-Volcker era, when monetary policy responded only weakly to inflation and President Johnson financed the war without raising taxes. Hence, this period can be characterized as a "Regime F". In contrast, the Reagan era followed the Volcker disinflation, during which a strong anti-inflationary monetary policy became firmly established. Accordingly, it

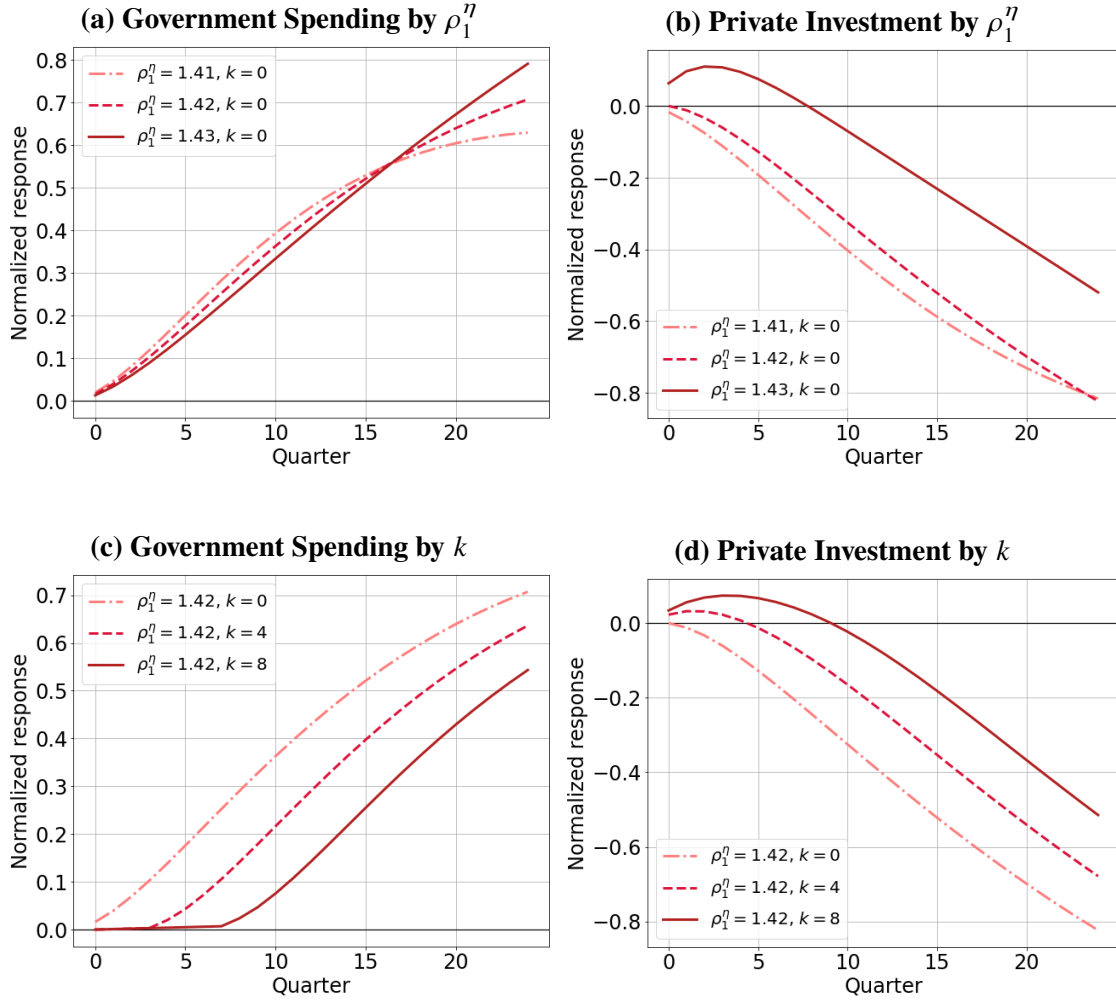


Figure 1.F.3: Simulation by time profile of spending

Notes: The figures show the response of government spending and private investment in NK DSGE model. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10 units of goods.

can be classified as a "Regime M".¹¹ Standard theory would predict a stronger crowding-in effect under Regime F, where the central bank accommodates fiscal expansion. However, in the empirical results, the opposite holds true: the Reagan-era sample exhibits a much stronger crowding-in of private investment than the Vietnam War sample.

This empirical puzzle can be resolved by incorporating the *time profile* of government spending into the model. To verify this mechanism, I conduct a two-step IRF matching exercise. In the first step, I estimate (ρ_1^η) and k to match the empirical government spending IRF. In the second step, I estimate the monetary policy response to inflation (ϕ_π) to match the private investment IRF.¹² As reported in Table 1.F.4, the estimated ρ_1^η for the Reagan era (1.426) is higher than that for the Vietnam period (1.382), while k is estimated to be zero for both samples. Meanwhile, the estimated monetary policy response ϕ_π is 0.00 for the Vietnam period and 0.65 for the Reagan era, consistent with their historical characterizations.¹³

As shown in Figure 1.F.4, the model-implied IRFs closely replicate the empirical responses. Notably, the model captures the stronger rise in private investment during the Reagan era, despite of a more active monetary policy. This outcome reflects the persistent and back-loaded nature of spending amplified investment responses strongly enough to offset the dampening effects of tighter monetary policy. Ultimately, an analysis that focuses solely on the policy regime yields misleading predictions. Accurately identifying the effects requires accounting for the time profile of government spending.

¹¹The fiscal stance during the Reagan era is less clear than monetary stance. However, compared with the Vietnam War, fiscal policy placed greater emphasis on debt sustainability. Although the primary balance remained in deficit until the early 1980s, it turned into a surplus by the late 1980s. As Cochrane (2022) notes, despite the common perception of "Reagan deficits," fiscal policy during this period was not particularly loose, and the magnitude of the primary deficit was not exceptional.

¹² ρ_1^η is simulated over the range from 1.38 to 1.43 in increments of 0.001, and k is simulated from 0 to 4. The parameter ϕ_π is simulated over the range from 0.00 to 1.20 in increments of 0.01. Parameter values of ϕ_π that do not satisfy the determinacy condition are excluded from the analysis.

¹³These results suggest that the Vietnam War sample corresponds to Regime F, while the Reagan-era sample corresponds to Regime M. Since the monetary policy rule includes a positive response to output gap ($\phi_y > 0$), ϕ_π can be less than 1 even under Regime M. For a detailed discussion of the regime selection, see Appendix 1.F.2.

Table 1.F.4: Parameter Values from the IRF-matching

Parameter		LTW(2017)		IRF-Matched	
		Regime F	Regime M	1954–1978	1975–1999
<i>News Shock Process</i>					
ρ_1^η	coefficient on u_{t-1}^G	0.106	0.127	1.382	1.426
ρ_2^η	coefficient on u_{t-2}^G	0	0	-0.18	-0.18
ρ_3^η	coefficient on u_{t-3}^G	0	0	-0.25	-0.25
k	news horizon	0	0	0	0
<i>Policy Rule</i>					
ϕ_π	interest rate resp. to inflation	0.15	1.14	0.00	0.65
ϕ_y	interest rate resp. to output gap	0.17	0.18	0.17	0.18
ρ_r	lagged interest rate response	0.15	0.76	0.15	0.76
γ_G	gov. cons. resp. to debt	0.00	0.21	0.00	0.21

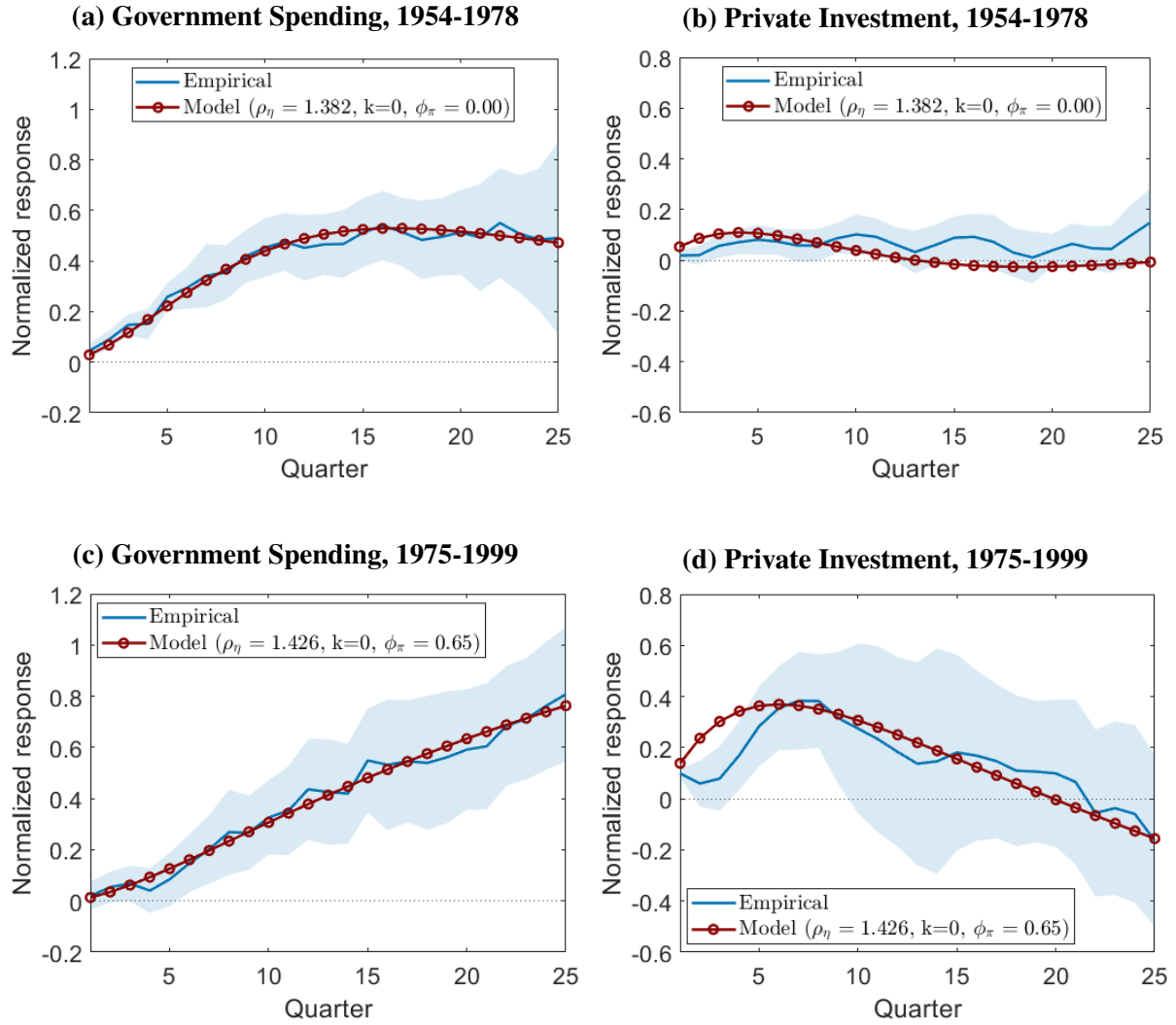


Figure 1.F.4: Replication of Empirical IRFs

Notes: The figures show the response of government spending and private investment in NK DSGE model. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10 units of goods.

1.F.5 IRF Matching: Private Consumption

To account for private consumption IRFs in addition to private investment, I extend the model to include non-saver households (rule-of-thumb consumers). A higher share of non-saver households (μ_{HH}) raises the marginal propensity to consume (MPC), thereby amplifying the effect of government spending on private consumption. I therefore treat μ_{HH} and the interest rate response to inflation (ϕ_π) as free parameters and search for the combination that best matches the empirical IRFs of private investment and private consumption. All other parameter settings are identical to those in the main text.

Figure 1.F.5 shows that the optimal μ_{HH} in the Reagan-era sample (1975–1999) is higher than in the Vietnam War sample (1954–1978). The left panels demonstrate that when $\mu_{HH} = 0.0$ and the monetary policy response to inflation is very weak ($\phi_\pi \approx 0$), the model reproduces the muted consumption and investment responses observed in the Vietnam War sample. In contrast, the right panels show that when $\mu_{HH} = 0.55$ and the monetary policy response to inflation is strong ($\phi_\pi = 0.6$), the model captures the large consumption and investment responses observed in the Reagan-era sample.

Table 1.F.5: Parameter Values from the IRF-matching

Parameter		1954–1978	1975–1999
<i>Households</i>			
μ_{HH}	proportion of non-savers	0.00	0.55
<i>Policy Rule</i>			
ϕ_π	interest rate resp. to inflation	0.00	0.60
ϕ_y	interest rate resp. to output gap	0.17	0.18
ρ_r	lagged interest rate response	0.15	0.76
γ_G	gov. cons. resp. to debt	0.00	0.21

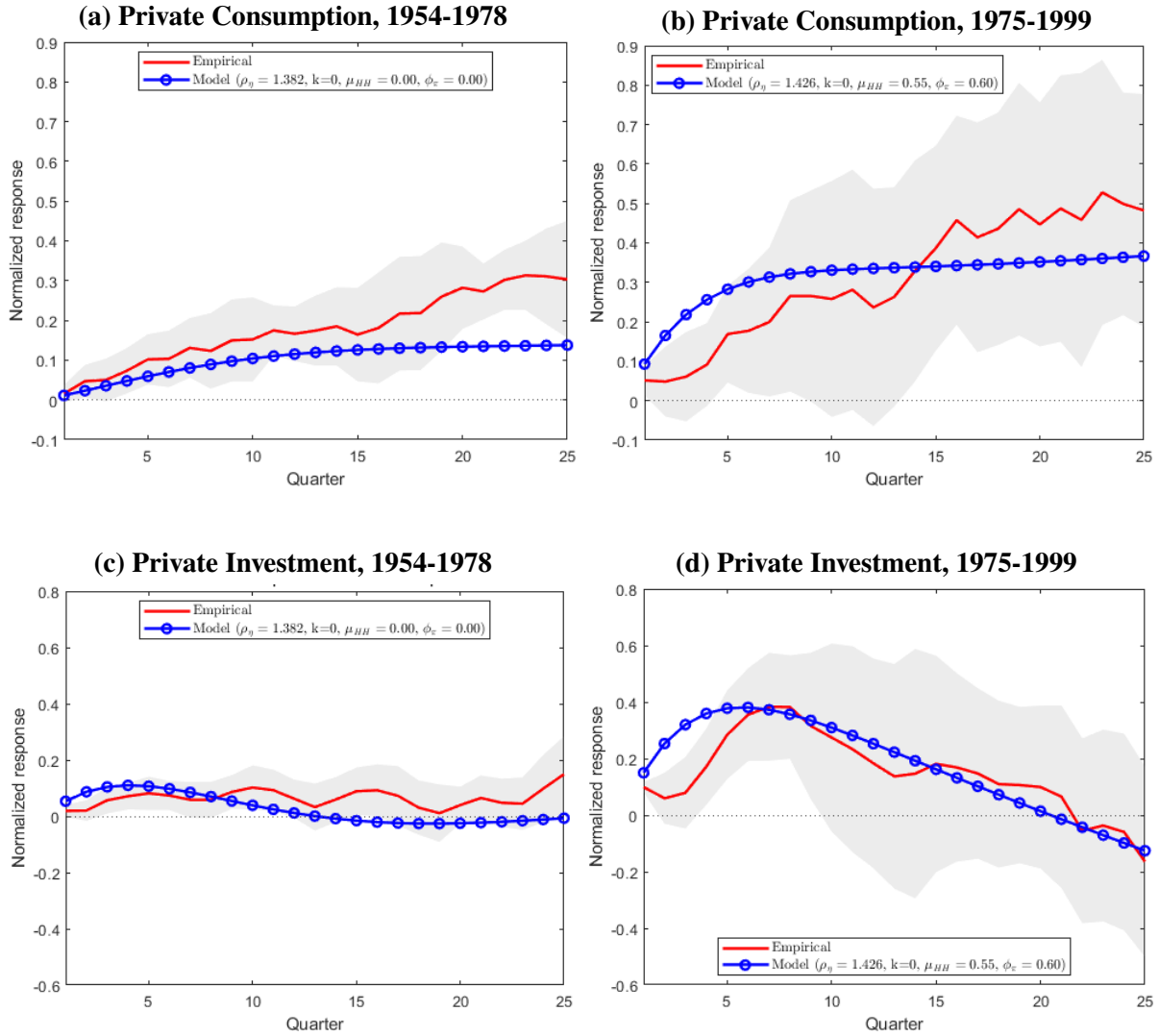


Figure 1.F.5: Replication of Empirical IRFs

Notes: The figures show the response of private consumption and private investment in NK DSGE model. The scale is adjusted such that the 6-year cumulative response of government spending corresponds to 10 units of goods.

Appendices to Chapter 2

2.A Summary Statistics of Procurement Contracts

The matched sample consists of procurement events involving listed firms over the period 2000–2024. As is typical in procurement data, the distribution of contract amounts is highly right-skewed, with a small number of very large awards accounting for a substantial share of total obligated value. The gap between the mean and the median is therefore large, underscoring the importance of scaling by firm size.

The summary statistics are presented in three ways. First, statistics by contract type compare definitive contracts, IDV awards, and DOs under IDVs to show how scale and distributions differ across contractual structures. Second, statistics by procurement object distinguish products, R&D services, and non-R&D services in order to document heterogeneity by procurement object. Third, comparisons of firm characteristics between treatment and control groups show how firms experiencing procurement events differ from non-event firms. The table by contract type reports all major contract forms so that the overall structure of procurement remains visible even if later analyses adopt a narrower baseline sample.

Table 2.A.1: Summary Statistics of Procurement Contracts

Contract Type	Events	Mean	SD	<i>p</i> ₂₅	<i>p</i> ₅₀	<i>p</i> ₇₅	<i>p</i> ₉₉	Max
<i>Panel A: All</i>								
Definitive	14,174	4,325	47,685	161	500	2,000	61,303	5,295,000
DO under multi-award IDV	93,929	776	6,452	11	49	246	12,161	1,055,910
DO under single-award IDV	104,177	1,232	12,085	12	60	389	15,704	914,246
<i>Panel B: By PSC Type</i>								
Definitive								
Product	6,509	4,287	68,128	154	420	1,700	50,681	5,295,000
R&D service	3,993	2,476	12,646	233	549	1,520	34,910	416,485
Non-R&D service	3,844	6,115	18,849	100	537	4,000	91,457	291,089
DO under multi-award IDV								
Product	63,646	479	6,794	8	28	118	6,857	1,055,910
R&D service	2,360	937	3,793	71	207	597	10,132	80,622
Non-R&D service	30,943	1,297	5,404	29	138	654	19,416	246,849
DO under single-award IDV								
Product	71,769	1,178	12,486	9	40	281	14,196	890,682
R&D service	3,356	1,015	6,549	52	161	518	15,151	241,546
Non-R&D service	34,730	1,162	10,532	24	106	453	16,478	914,246
<i>Panel C: By Firm Sector</i>								
Definitive								
Goods sector	10,304	4,864	55,393	193	558	2,285	66,855	5,295,000
Service sector	2,046	3,602	13,870	154	486	2,030	59,030	268,768
DO under multi-award IDV								
Goods sector	51,540	791	7,914	12	47	229	11,661	1,055,910
Service sector	18,218	1,189	4,627	24	120	591	19,004	143,379
DO under single-award IDV								
Goods sector	58,711	1,213	11,582	9	40	242	19,365	914,246
Service sector	17,365	1,227	8,154	45	169	636	15,653	661,475
<i>Panel D: By 4-Digit Sector Match</i>								
Definitive								
Non-match	10,831	3,349	14,929	150	443	1,590	54,586	416,485
Match	3,453	7,280	92,958	204	836	3,644	96,803	5,295,000
DO under multi-award IDV								
Non-match	68,695	806	6,968	11	48	238	12,769	1,055,910
Match	27,564	640	4,534	10	47	244	9,808	466,054
DO under single-award IDV								
Non-match	82,038	1,152	13,059	12	59	339	14,278	914,246
Match	29,250	1,165	6,470	9	60	606	16,380	490,501

Notes: Obligation amounts in thousands of dollars (*federal_action_obligation*). DO = Delivery Order; IDV = Indefinite Delivery Vehicle; PSC = Product or Service Code. Panel B classifies contracts by PSC. Panel C classifies firms into Goods and Service sectors by primary SIC code. Panel D defines Match as the contract's 4-digit sector matching the firm's primary sector.

2.B Balance Tables

Table 2.B.1: Daily Firm Characteristics: Treatment vs. Control (Unconditional)

Variable	Treatment Mean	Control Mean	Difference	SMD	Obs.
<i>Panel A: All contracts</i>					$N_t=106,615; N_c=5,759,318$
Excess return	-0.003	-0.014	0.011	0.005	5,865,933
log(Market cap)	15.722	14.011	1.711	0.799	5,854,879
<i>Panel B: Industry mismatch</i>					$N_t=78,227; N_c=5,759,318$
Excess return	0.003	-0.014	0.018	0.008	5,837,545
log(Market cap)	15.755	14.011	1.744	0.813	5,826,491
<i>Panel C: Industry match</i>					$N_t=30,939; N_c=5,759,318$
Excess return	-0.018	-0.014	-0.004	0.002	5,790,257
log(Market cap)	15.704	14.011	1.693	0.795	5,779,203
<i>Panel D: R&D service</i>					$N_t=6,264; N_c=5,759,318$
Excess return	0.012	-0.014	0.026	0.011	5,765,582
log(Market cap)	16.093	14.011	2.082	0.981	5,754,528
<i>Panel E: Non-R&D service</i>					$N_t=34,482; N_c=5,759,318$
Excess return	0.002	-0.014	0.017	0.007	5,793,800
log(Market cap)	15.841	14.011	1.831	0.853	5,782,746
<i>Panel F: Product</i>					$N_t=69,887; N_c=5,759,318$
Excess return	-0.006	-0.014	0.008	0.004	5,829,205
log(Market cap)	15.675	14.011	1.664	0.780	5,818,151

Notes: Unconditional comparison of treatment and control firm-days without fixed-effect adjustment. Excess return is measured on the day prior to the event. log(Market cap) is the average of log daily market capitalization over the prior 10 trading days. |SMD| is the absolute Standardized Mean Difference (SMD), defined as $|\bar{X}_t - \bar{X}_c| / \sqrt{(s_t^2 + s_c^2)/2}$, where absolute values above 0.25 is a standard threshold for assessing imbalance (Stuart and Rubin, 2008; Imbens and Rubin, 2015).

Table 2.B.2: Quarterly Firm Characteristics: Treatment vs. Control (Unconditional)

Variable	Treatment Mean	Control Mean	Difference	SMD	Obs.
<i>Panel A: All contracts</i>					$N_t=19,640; N_c=83,519$
log(Assets)	8.037	6.912	1.124	0.482	103,159
Market-to-book ratio	1.816	2.149	-0.333	0.210	99,345
Leverage	1.051	1.081	-0.030	0.009	100,607
ROA	0.020	0.006	0.013	0.286	98,038
Cash ratio	0.269	-0.315	0.584	0.229	95,800
<i>Panel B: Industry mismatch</i>					$N_t=15,012; N_c=83,519$
log(Assets)	8.112	6.912	1.200	0.514	98,531
Market-to-book ratio	1.812	2.149	-0.337	0.212	94,795
Leverage	1.058	1.081	-0.023	0.007	96,018
ROA	0.019	0.006	0.013	0.275	93,529
Cash ratio	0.269	-0.315	0.585	0.227	91,324
<i>Panel C: Industry match</i>					$N_t=7,641; N_c=83,519$
log(Assets)	8.292	6.912	1.379	0.594	91,160
Market-to-book ratio	1.774	2.149	-0.375	0.243	87,768
Leverage	1.031	1.081	-0.050	0.016	88,787
ROA	0.021	0.006	0.015	0.336	86,535
Cash ratio	0.337	-0.315	0.652	0.266	84,462
<i>Panel D: R&D service</i>					$N_t=2,703; N_c=83,519$
log(Assets)	8.375	6.912	1.463	0.608	86,222
Market-to-book ratio	1.891	2.149	-0.258	0.163	82,931
Leverage	1.113	1.081	0.032	0.008	83,859
ROA	0.015	0.006	0.009	0.176	81,718
Cash ratio	0.174	-0.315	0.489	0.178	79,649
<i>Panel E: Non-R&D service</i>					$N_t=8,672; N_c=83,519$
log(Assets)	8.382	6.912	1.469	0.636	92,191
Market-to-book ratio	1.707	2.149	-0.442	0.286	88,699
Leverage	1.032	1.081	-0.049	0.016	89,729
ROA	0.020	0.006	0.014	0.306	87,529
Cash ratio	0.362	-0.315	0.678	0.275	85,364
<i>Panel F: Product</i>					$N_t=13,376; N_c=83,519$
log(Assets)	8.207	6.912	1.295	0.555	96,895
Market-to-book ratio	1.832	2.149	-0.317	0.202	93,244
Leverage	1.042	1.081	-0.039	0.012	94,469
ROA	0.021	0.006	0.015	0.335	91,969
Cash ratio	0.291	-0.315	0.606	0.242	89,850

Notes: Unconditional comparison of treatment and control firm-quarters without fixed-effect adjustment. All variables are lagged one quarter. The control group consists of non-event firm-quarters from firms with at least one contract. |SMD| is the absolute Standardized Mean Difference (SMD), defined as $|\bar{X}_t - \bar{X}_c| / \sqrt{(s_t^2 + s_c^2)/2}$, where absolute values above 0.25 is a standard threshold for assessing imbalance (Stuart and Rubin, 2008; Imbens and Rubin, 2015).

Table 2.B.3: Daily Firm Characteristics: Treatment vs. Control (FE-Adjusted)

Variable	Treatment Mean	Control Mean	Difference	SMD	Obs.
<i>Panel A: All contracts</i>					$N_t=100,510; N_c=5,477,994$
Excess return	0.004	0.000	0.004	0.002	5,578,504
log(Market cap)	-0.001	0.000	-0.001	0.003	5,567,389
<i>Panel B: Industry mismatch</i>					$N_t=73,014; N_c=5,477,994$
Excess return	0.004	0.000	0.005	0.002	5,551,008
log(Market cap)	0.003	0.000	0.002	0.004	5,539,893
<i>Panel C: Industry match</i>					$N_t=29,930; N_c=5,477,994$
Excess return	0.003	0.000	0.003	0.002	5,507,924
log(Market cap)	0.001	0.000	0.001	0.002	5,496,813
<i>Panel D: R&D service</i>					$N_t=5,942; N_c=5,477,994$
Excess return	-0.012	0.000	-0.012	0.006	5,483,936
log(Market cap)	0.032	0.000	0.032	0.058	5,472,824
<i>Panel E: Non-R&D service</i>					$N_t=32,559; N_c=5,477,994$
Excess return	0.011	0.000	0.011	0.005	5,510,553
log(Market cap)	0.025	0.000	0.025	0.044	5,499,442
<i>Panel F: Product</i>					$N_t=65,622; N_c=5,477,994$
Excess return	0.002	0.000	0.002	0.001	5,543,616
log(Market cap)	-0.016	0.000	-0.016	0.029	5,532,501

Notes: All variables are residualized using firm fixed effects and industry (4-digit NAICS)×date fixed effects. Excess return is the residual of the one-day lagged excess return; log(Market cap) is the residual of the average log daily market capitalization over the prior 10 trading days. |SMD| is the absolute Standardized Mean Difference (SMD), defined as $|\bar{X}_t - \bar{X}_c| / \sqrt{(s_t^2 + s_c^2)/2}$, where absolute values above 0.25 is a standard threshold for assessing imbalance (Stuart and Rubin, 2008; Imbens and Rubin, 2015). The control group is common across all panels and consists of firm-days with no contract event.

Table 2.B.4: Quarterly Firm Characteristics: Treatment vs. Control (FE-Adjusted)

Variable	Treatment Mean	Control Mean	Difference	SMD	Obs.
<i>Panel A: All contracts</i>					$N_t=18,893; N_c=78,694$
log(Assets)	0.025	-0.006	0.031	0.067	97,587
Market-to-book ratio	-0.003	0.001	-0.003	0.003	93,930
Leverage	0.013	-0.003	0.016	0.005	95,066
ROA	0.000	0.000	0.000	0.011	92,457
Cash ratio	0.006	-0.002	0.008	0.005	90,207
<i>Panel B: Industry mismatch</i>					$N_t=14,363; N_c=78,694$
log(Assets)	0.036	-0.006	0.042	0.089	93,057
Market-to-book ratio	-0.011	0.001	-0.012	0.012	89,477
Leverage	0.014	-0.003	0.017	0.005	90,576
ROA	0.000	0.000	0.000	0.011	88,043
Cash ratio	-0.001	-0.002	0.000	0.000	85,827
<i>Panel C: Industry match</i>					$N_t=7,507; N_c=78,694$
log(Assets)	0.011	-0.006	0.017	0.037	86,201
Market-to-book ratio	0.003	0.001	0.002	0.002	82,953
Leverage	-0.005	-0.003	-0.002	0.001	83,868
ROA	0.000	0.000	0.000	0.011	81,602
Cash ratio	0.031	-0.002	0.033	0.019	79,519
<i>Panel D: R&D service</i>					$N_t=2,662; N_c=78,694$
log(Assets)	0.032	-0.006	0.038	0.079	81,356
Market-to-book ratio	-0.018	0.001	-0.018	0.019	78,206
Leverage	0.040	-0.003	0.043	0.011	79,030
ROA	-0.001	0.000	-0.001	0.021	76,864
Cash ratio	0.002	-0.002	0.003	0.002	74,785
<i>Panel E: Non-R&D service</i>					$N_t=8,445; N_c=78,694$
log(Assets)	0.024	-0.006	0.030	0.066	87,139
Market-to-book ratio	-0.003	0.001	-0.003	0.003	83,786
Leverage	0.002	-0.003	0.005	0.002	84,711
ROA	0.000	0.000	0.000	0.015	82,473
Cash ratio	0.003	-0.002	0.005	0.003	80,301
<i>Panel F: Product</i>					$N_t=12,865; N_c=78,694$
log(Assets)	0.024	-0.006	0.030	0.065	91,559
Market-to-book ratio	-0.004	0.001	-0.005	0.005	88,069
Leverage	0.013	-0.003	0.016	0.005	89,169
ROA	0.000	0.000	0.000	0.007	86,648
Cash ratio	0.008	-0.002	0.010	0.006	84,512

Notes: All variables are residualized using firm fixed effects and industry (4-digit NAICS)×quarter fixed effects and are lagged one quarter. |SMD| is the absolute Standardized Mean Difference (SMD), defined as $|\bar{X}_t - \bar{X}_c| / \sqrt{(s_t^2 + s_c^2)/2}$, where absolute values above 0.25 is a standard threshold for assessing imbalance (Stuart and Rubin, 2008; Imbens and Rubin, 2015). The control group is common across all panels and consists of firm-quarters with no contract event in the quarter.

2.C Sensitivity analysis

2.C.1 Procurement Shock Denominator

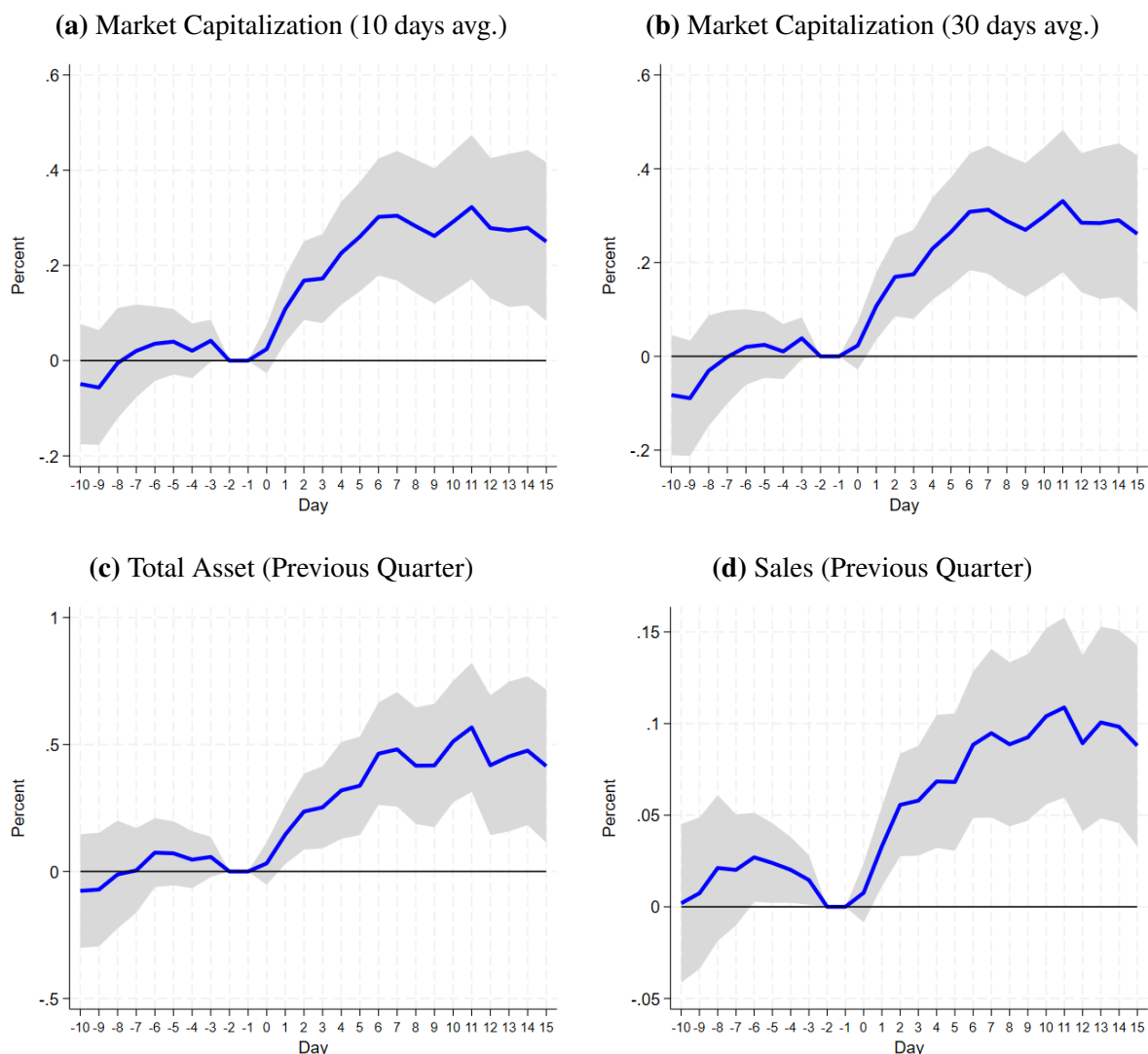


Figure 2.C.1: Daily Stock Price Responses by Shock Denominator

Notes: This figure plots the estimated coefficients from local projections of excess returns on definitive procurement shocks. The x-axis represents trading days relative to the action date ($t = 0$). The procurement shock is defined as the contract obligation amount divided by a firm-level normalizer. Panel (a) uses the firm's average market capitalization over the 10 trading days prior to the action date as the normalizer. Panel (b) uses the firm's average market capitalization over the 30 trading days prior to the action date. Panel (c) uses the firm's total assets in the previous quarter. Panel (d) uses the firm's sales in the previous quarter. Shaded areas denote 90% confidence intervals. Driscoll–Kraay standard errors are computed with bandwidth 12.

2.C.2 Sector Specialization Criterion

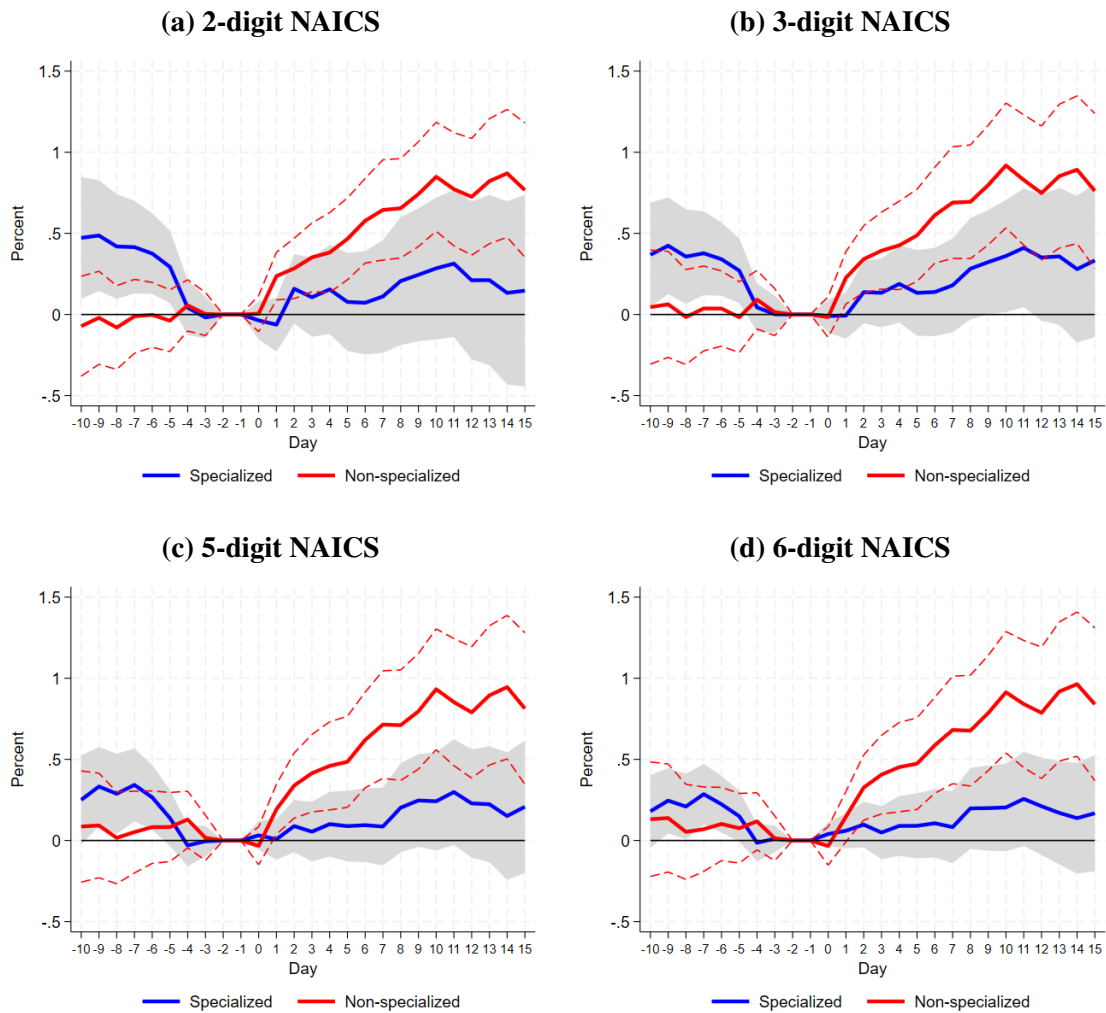


Figure 2.C.2: Daily Stock Price Responses by Sector Specialization

Notes: This figure compares stock price responses for specialized-sector contracts, where the contract's N-digit NAICS code matches the firm's primary industry, with non-specialized-sector contracts. Shaded areas denote 90% confidence intervals. Driscoll–Kraay standard errors with bandwidth 12.

2.D Quarterly Investment Responses by Contract Form

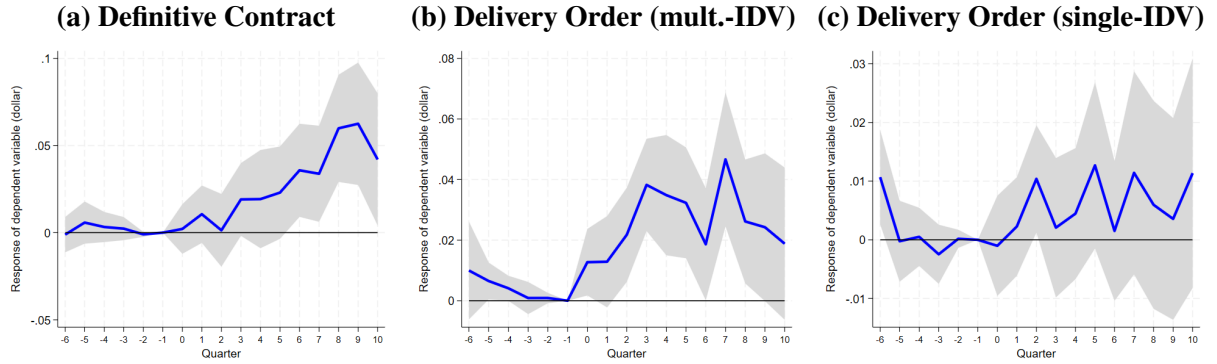


Figure 2.D.1: Quarterly CAPX Responses by Contract Form

Notes: This figure plots the estimated coefficients from local projections of CAPX on relative procurement shocks. Panels (a) and (b) show responses for definitive contracts and delivery orders (DO) under multiple-award IDVs, respectively. Panel (c) displays the response for DOs under single-award IDVs. Shaded areas denote 90% confidence intervals. Driscoll–Kraay standard errors with bandwidth 4.

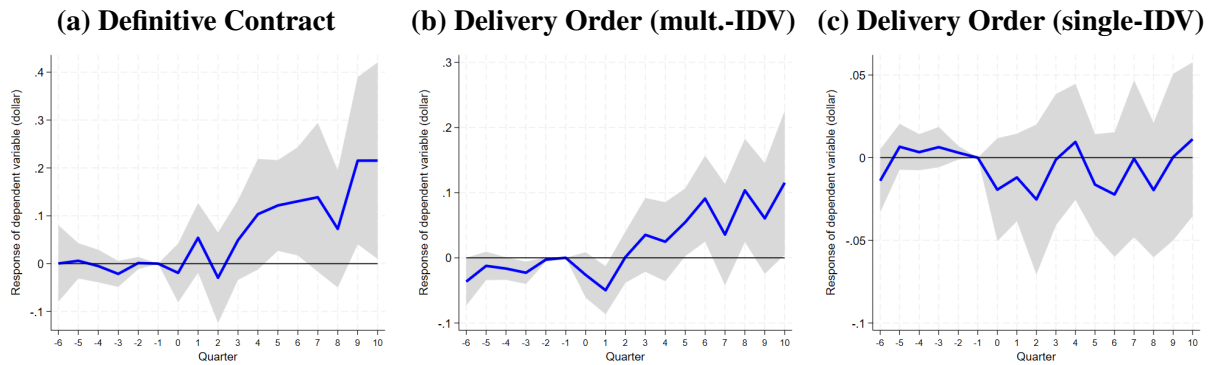


Figure 2.D.2: Quarterly R&D Responses by Contract Form

Notes: This figure plots the estimated coefficients from local projections of R&D on relative procurement shocks. Panels (a) and (b) show responses for definitive contracts and delivery orders (DO) under multiple-award IDVs, respectively. Panel (c) displays the response for DOs under single-award IDVs. Shaded areas denote 90% confidence intervals. Driscoll–Kraay standard errors with bandwidth 4.

2.E Stock Price Responses to IDV Framework Awards

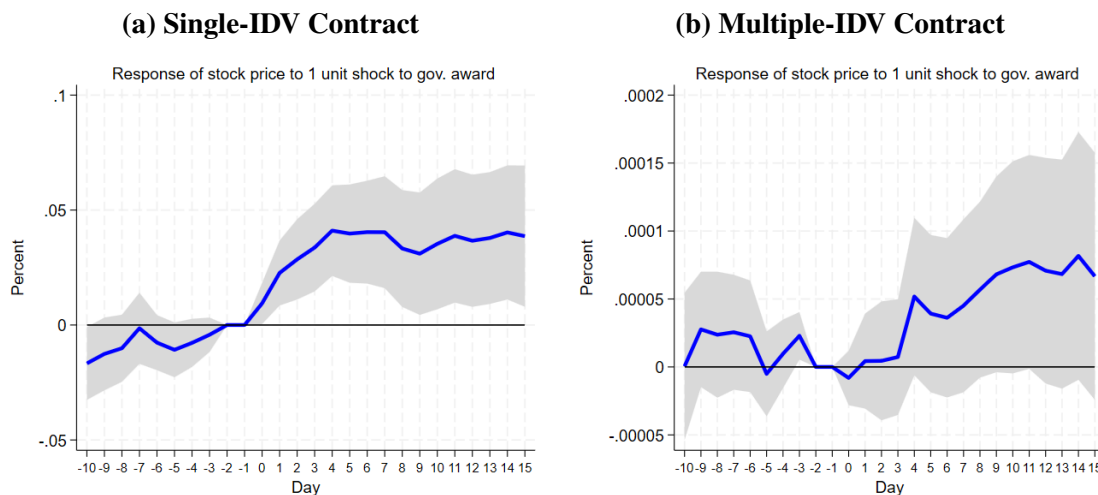


Figure 2.E.1: Daily Stock Price Responses: Indefinite Delivery Vehicle

Notes: This figure plots the estimated coefficients from local projections of excess returns on relative IDV(Indefinite Delivery Vehcecles) contract shocks. The x-axis represents the trading days relative to the action date ($t = 0$). Panel (a) show responses for single-award IDVs and Panel (b) shows responses for multiple-award IDVs. Shaded areas denote 90% confidence intervals. Driscoll–Kraay standard errors with bandwidth 12.

Appendices to Chapter 3

3.A Additional Descriptive Evidence

Figure 3.A.1 reports the unweighted version of the U.S. county-level bin-scatter plot. The relationship is weaker than in the population-weighted plot reported in the main text, but it remains positive.

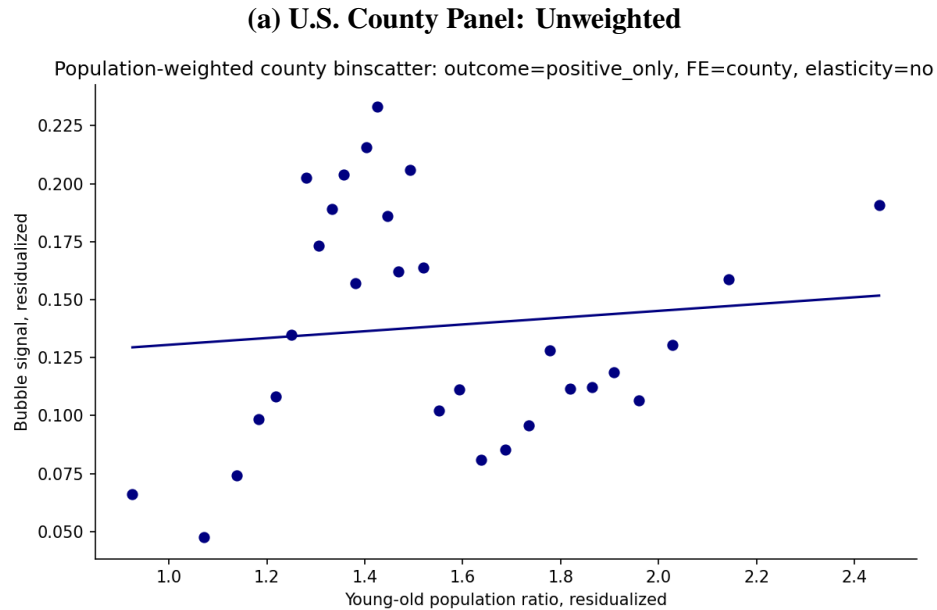


Figure 3.A.1: Unweighted Bin Scatter Plot of Large Housing Booms and Young-to-Old Population Ratio: U.S. County Panel

Notes: This figure presents the unweighted U.S. county-level bin-scatter plot of the large housing boom probability against the young-to-old population ratio. Observations are grouped into 30 equal-sized bins based on the young-to-old ratio, and the plot displays the mean large housing boom probability within each bin along with a linear fit.

3.B Supplementary Logit Specifications

Table 3.B.1: Logit Model: Country Panel

	(1)	(2)	(3)	(4)	(5)
YO ratio	0.35**	1.64***	2.27***	1.90***	2.05**
(Marginal Effect)	0.04**	0.19***	0.26***	0.22***	0.24**
GDP per Cap.			0.80**		3.60**
(Marginal Effect)			0.09**		0.43**
Constant	-2.40***	-4.51***	-13.81***	-3.26***	-42.36**
Observations	1,262	1,262	1,225	1,262	1,225
Country FE	No	Yes	Yes	Yes	Yes
Time FE	No	No	No	Yes	Yes
Pseudo R^2	0.0053	0.0352	0.0408	0.1743	0.1884

Notes: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Standard errors are clustered at the country level.

Table 3.B.2: Logit Model: U.S. County Panel

	(1)	(2)	(3)	(4)
YO ratio	0.21***	0.29***	0.15***	0.27***
	(0.03)	(0.04)	(0.05)	(0.07)
(Marginal Effect)	0.02***	0.03***	0.01***	0.02***
Observations	101,567	37,205	101,567	37,205
Supply Elasticity	No	Yes	No	Yes
County FE	No	No	No	No
Time FE	No	No	Yes	Yes
Pseudo R^2	0.0020	0.0034	0.3223	0.3557

Notes: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. All regressions are weighted by county population. Standard errors are clustered at the county level. County fixed-effect logit specifications were omitted because they are high-dimensional nonlinear FE models prone to perfect prediction.

3.C HMDA Additional Evidence

3.C.1 Pretrend

Because the HMDA analysis is interpreted as a triple-difference design, it is useful to examine whether the relative application intensity of young applicants was already moving before a county enters a boom episode. We therefore estimate a short event-study specification around the *first* boom onset in each county. Let $\tau_{c,t}$ denote event time relative to the first year in which the county-level boom indicator switches from zero to one. We estimate

$$y_{c,a,t} = \sum_{k \in \{-2,0,1\}} \beta_k (\mathbf{1}\{\tau_{c,t} = k\} \times Young_a) + \alpha_{c,a} + \lambda_{t,a} + \mu_{c,t} + \varepsilon_{c,a,t}, \quad (3.C.1)$$

where the omitted category is $\tau_{c,t} = -1$. Thus each coefficient measures the young-versus-old application gap at event time k relative to the year immediately preceding boom onset.

This exercise matters because a clean event-study pattern would typically display little movement in the pre-period. If the pre-period coefficients differ materially from zero, the onset-year estimate may partly reflect dynamics that begin before the boom indicator turns on.

The onset and post-onset coefficients are positive and statistically significant, which is consistent with stronger mortgage application intensity among young applicants during boom county-years. At the same time, the coefficient at event time -2 is negative. This means that the young-versus-old application gap was smaller two years before onset than in the year immediately before onset. The event-study evidence therefore suggests some pre-boom dynamics in relative young application intensity rather than perfectly flat pretrends.

There are two reasons to interpret this pattern cautiously. First, the HMDA sample with publicly available applicant age is short (2018–2024), so the number of clean pre-period observations is limited. Second, timing is measured at the annual frequency and the boom indicator is constructed from annual deviations of county house prices from trend. If boom conditions begin to form near the

Table 3.C.1: HMDA Event Study Around First Boom Onset

	(1)	(2)
Young group	25–34	25–44
Comparison group	45–74	45–74
Reference period	$\tau_{c,t} = -1$	$\tau_{c,t} = -1$
$\mathbf{1}\{\tau_{c,t} = -2\} \times Young_a$	-1.231*** (0.296)	-1.114*** (0.258)
$\mathbf{1}\{\tau_{c,t} = 0\} \times Young_a$	0.715*** (0.197)	1.042*** (0.223)
$\mathbf{1}\{\tau_{c,t} = 1\} \times Young_a$	1.023*** (0.240)	1.038*** (0.233)
Observations	87,464	109,796
Counties	3,230	3,230
County-year cells	22,426	22,463
R^2	0.8913	0.8773

Notes: The outcome is mortgage applications per 1,000 age-specific population. Each column estimates equation (3.C.1). The omitted event time is the year immediately preceding the first boom onset in a county ($\tau_{c,t} = -1$). Regressions are weighted by age-specific population and absorb county-age, year-age, and county-year fixed effects. Standard errors, clustered at the county level, are reported in parentheses. *** indicates significance at the 1% level.

end of year $t - 1$, then the immediate pre-period is not a clean untreated benchmark. We therefore interpret the HMDA results as mechanism-consistent evidence on age-specific participation during boom periods, rather than as a clean event-study design with flat pretrends.

3.C.2 Alternative Boom Definitions and Sample Restrictions

We next assess whether the HMDA results are sensitive to how boom county-years are defined and to simple sample restrictions. The first set of checks tightens the boom criterion by requiring larger deviations of county house prices from trend. The second set trims the estimation sample by dropping county-age cells with very small age-specific populations.

Table 3.C.2: HMDA Robustness: Alternative Boom Thresholds

	(1)	(2)	(3)	(4)
Young group	25–34	25–44	25–34	25–44
Comparison group	45–74	45–74	45–74	45–74
Boom definition	> 1.5 SD	> 1.5 SD	> 2.0 SD	> 2.0 SD
Boom _{c,t} × Young _a	0.950*** (0.217)	1.223*** (0.262)	1.650*** (0.435)	2.404*** (0.546)
Observations	76,328	95,587	76,328	95,587
Counties	2,779	2,779	2,779	2,779
County-year cells	19,099	19,099	19,099	19,099
County-age cells	11,113	13,892	11,113	13,892
R^2	0.8895	0.8754	0.8895	0.8754

Notes: The outcome is mortgage applications per 1,000 age-specific population. Each column estimates equation (3.5) using a stricter boom indicator based on the county-specific standard deviation of detrended annual house prices. Regressions are weighted by age-specific population and absorb county-age, year-age, and county-year fixed effects. Standard errors, clustered at the county level, are reported in parentheses. *** indicates significance at the 1% level.

These additional checks show that the HMDA result is not driven by a particular boom cutoff or by thin county-age cells. The coefficients remain positive and economically meaningful under stricter boom definitions and when the sample is restricted to larger age groups. Overall, the application-margin result appears robust to alternative boom definitions and basic sample restrictions.

Table 3.C.3: HMDA Robustness: Sample Restrictions

	(1)	(2)
Young group	25–34	25–44
Comparison group	45–74	45–74
Restriction	Pop. age > 1,000	Pop. age > 1,000
Boom _{c,t} × Young _a	1.014*** (0.181)	1.365*** (0.231)
Observations	65,814	82,942
Counties	2,697	2,697
County-year cells	18,489	18,494
County-age cells	9,672	12,183
R ²	0.8903	0.8759

Notes: The first two columns require each county-age cell to contain at least 1,000 people. Regressions are weighted by age-specific population and absorb county-age, year-age, and county-year fixed effects. Standard errors, clustered at the county level, are reported in parentheses. * and *** indicate significance at the 10% and 1% levels, respectively.

3.C.3 Risk-Taking Outcomes Among Young Borrowers

The preceding HMDA results show that young applicants increase mortgage-market participation more intensively during boom county-years. We next ask whether this stronger participation is accompanied by riskier mortgage choices among young borrowers. To examine this intensive-margin channel, we re-estimate the HMDA triple-difference specification using age-group averages of loan-to-value (LTV) and debt-to-income (DTI) ratios as outcomes. Specifically, we estimate equation (3.5) with $y_{c,a,t}$ equal to the county-age-year mean LTV ratio or the county-age-year mean DTI ratio among mortgage applications. As in the baseline HMDA specification, regressions absorb county-age, year-age, and county-year fixed effects, are weighted by age-specific population, and cluster standard errors at the county level.

Because mean LTV and mean DTI could be sensitive to a small number of extreme observations, we focus on a trimmed robustness exercise. For each continuous outcome, we remove county-age-year cells in the top and bottom 1% of the full-sample outcome distribution and then re-estimate the same triple-difference specification. This check is designed to ask whether the risk-taking pattern is

driven by outliers or instead reflects a more systematic shift in the composition of young borrowers during booms.

Table 3.C.4: HMDA Risk-Taking During Boom County-Years: Trimmed Mean Outcomes

	(1)	(2)	(3)	(4)
<i>Panel A: Young group = 25–34</i>				
Comparison group	55–74	55–74	45–74	45–74
Outcome	Mean LTV	Mean DTI	Mean LTV	Mean DTI
Trim rule	1% tails removed	1% tails removed	1% tails removed	1% tails removed
Boom _{c,t} × Young _a	0.461*** (0.095)	0.115** (0.052)	0.408*** (0.078)	0.098** (0.041)
Observations	56,186	56,590	75,172	75,649
Counties	2,779	2,779	2,779	2,779
County-year cells	19,097	19,098	19,099	19,098
County-age cells	8,333	8,332	11,110	11,110
R ²	0.9374	0.8913	0.9256	0.8727
<i>Panel B: Young group = 25–44</i>				
Comparison group	55–74	55–74	45–74	45–74
Outcome	Mean LTV	Mean DTI	Mean LTV	Mean DTI
Trim rule	1% tails removed	1% tails removed	1% tails removed	1% tails removed
Boom _{c,t} × Young _a	0.431*** (0.100)	0.090* (0.047)	0.376*** (0.082)	0.072** (0.035)
Observations	75,222	75,794	94,208	94,853
Counties	2,779	2,779	2,779	2,779
County-year cells	19,099	19,099	19,099	19,099
County-age cells	11,112	11,110	13,889	13,888
R ²	0.9203	0.8814	0.9114	0.8685

Notes: The outcomes are county-age-year mean loan-to-value ratios and county-age-year mean debt-to-income ratios among mortgage applications. Each column estimates equation (3.5) after removing the top and bottom 1% of the corresponding outcome distribution in the full sample. Regressions are weighted by age-specific population and absorb county-age, year-age, and county-year fixed effects. Standard errors, clustered at the county level, are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 3.C.4 shows that the qualitative pattern is stable after trimming extreme LTV and DTI observations. For mean LTV, the coefficient ranges from 0.376 to 0.461, implying that in boom county-years the average LTV of young applicants is about 0.38–0.46 percentage points higher than that of the comparison group, even after removing the top and bottom 1% of county-age-year mean outcomes. Relative to the comparison groups' weighted mean LTV levels in the estimation sample,

this corresponds to an increase of roughly 0.5%. For mean DTI, the coefficients range from 0.072 to 0.115 percentage points, which is economically modest but generally positive and statistically significant in three of the four specifications. Relative to the comparison groups' weighted mean DTI levels, these effects are on the order of about 0.2%–0.3%.

Taken together, these results suggest that the main HMDA finding is not solely an application-margin phenomenon. Young applicants are more active in boom county-years, and conditional on applying, they also appear to tilt modestly toward riskier mortgage positions, especially along the LTV margin.