Essays In Labour, Monetary, and Experimental Macroeconomics

by

Ibrahim Abuallail

A thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the Doctorate in Philosophy Degree in Economics

Department of Economics
Faculty of Social Sciences
University of Ottawa

Abstract

This thesis brings together three essays that address central questions in applied macroeconomics and empirical labour economics, spanning the identification of job vacancies in the labour market using high-frequency data, the identification of monetary policy shocks in a small open economy, and the determinants of firm price-setting behavior under shocks.

The first chapter examines the measurement of job vacancies, a fundamental input in the analysis of labour market dynamics. Traditional survey-based measures, such as the Job Vacancy and Wage Survey (JVWS), are subject to lags and limitations. By contrast, online job postings offer a timely alternative but raise concerns of accuracy. Using a rich dataset of Canadian job postings from Vicinity Jobs, this work compares online and survey-based vacancy measures and uses algorithms and robust regression with Huber weights to improve the mapping from postings to actual vacancies. Using machine learning forecast models, these methods reduce prediction error by an average of 15% relative to existing approaches, demonstrating the potential of combining online data and econometrics to deliver more timely and reliable vacancy statistics for researchers and policymakers.

The second chapter investigates the transmission of monetary policy shocks in Canada, a small open economy, through the use of factor-augmented models and machine learning. I propose a novel approach that filters economic variables using LASSO-based techniques before factor extraction, which enhances the reliability of the subsequent Factor-Augmented Vector Autoregression (FAVAR) estimates. Imposing sign restrictions and studying the impulse responses, I find that positive monetary policy shocks exert significant contractionary effects on employment, inflation, real GDP, and housing prices, though effects on the exchange rate remain ambiguous. I also highlight important structural breaks in Canada's monetary policy regime and demonstrate the value of combining machine learning with factor models for small open economy analysis.

The third chapter turns to experimental evidence on price-setting behavior, using a laboratory experiment where participants repeatedly set prices under Bertrand competition. The design varies market structure (duopoly versus monopolistic competition) and pricing frictions (flexible prices versus menu costs), and introduces four shocks: a demand shock and three cost shocks differing in size and uncertainty. The experiment shows that large shocks trigger disproportionately larger price adjustments, and that firms' forecasts of market prices is a stronger determinant of pricing compared to their expectations of costs, highlighting the importance of expectations and strategic interaction. These findings contribute to the literature on nominal rigidities, expectations formation, and the transmission of shocks to inflation.

Declaration of Authorship

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. The first chapter was co-authored with Anne-Lore Fraikin, and the third chapter was co-authored with Dr. Isabelle Salle.

I understand that my thesis may be made electronically available to the public.

General Introduction

Understanding how labour markets function, how monetary policy shocks transmit through an economy, and how firms set prices in response to shocks are three interrelated questions at the heart of macroeconomics. Together, they shape employment, inflation, and growth—the very outcomes policymakers seek to stabilize. Yet despite decades of research, persistent challenges remain. Labour market indicators are often measured with delay, limiting their usefulness for real-time policy decisions. Small open economies such as Canada face unique complexities in identifying monetary policy shocks, given their exposure to global forces. And at the firm level, pricing behavior is influenced not only by costs but also by expectations, strategic interactions, and frictions—factors that complicate the transmission of shocks to inflation. This thesis brings these themes together, presenting three essays that contribute to each of these areas by combining novel data sources, machine learning techniques, and controlled laboratory evidence.

The first chapter examines how online job postings can be used to provide more timely and accurate measures of job vacancies. Reliable vacancy data are essential for understanding labour market tightness, diagnosing skills shortages, and designing effective immigration or training policies. In Canada, the Job Vacancy and Wage Survey (JVWS) provides high-quality vacancy measures but with substantial reporting lags. Online job postings, by contrast, are available in real time but contain sectoral biases and challenges that limit their representativeness. Using data from Vicinity Jobs, I develop and compare statistical and machine learning methods—including a robust regression with Huber weights—that significantly reduce prediction errors relative to existing approaches. The results show that online job postings, when corrected for biases and processed with modern econometric tools, can provide a reliable, high-frequency complement to traditional survey data, offering academics and policymakers new insights into labour market conditions.

The second chapter turns to monetary policy in Canada, focusing on how shocks propagate in a small open economy. Unlike the U.S., Canada's economy is more exposed to global commodity prices and exchange rate fluctuations, making the identification of monetary policy shocks particularly challenging. I propose a novel approach that integrates machine learning-based variable selection into the Factor-Augmented Vector Autoregression (FAVAR) framework. By filtering the data through techniques such as LASSO, Random Forests, and Elastic Net before factor extraction, I construct more informative factors that better capture the dimensions of inflation and employment relevant for policy. The resulting sign-restricted FAVAR model yields impulse response functions that are more consistent with macroeconomic theory, showing that contractionary monetary policy shocks in Canada significantly lower employment, real GDP, and inflation, while effects on the exchange rate remain ambiguous. The findings also highlight important structural breaks in Canada's monetary policy regime.

The third chapter investigates how firms adjust prices when confronted with shocks, using a laboratory experiment designed to isolate the roles of expectations, nominal rigidities, and market structure. Participants repeatedly set prices under Bertrand competition, facing four types of shocks: a demand shock and three cost shocks that vary in size and uncertainty.

The design varies whether firms operate in a duopoly or in monopolistic competition, and whether they face flexible prices or menu costs. The results show that larger shocks induce proportionally larger price adjustments, while uncertainty dampens pass-through, consistent with firms adopting a "wait-and-see" strategy. Strikingly, firms' forecasts of the market price consistently exert a stronger influence on pricing decisions than their own expected costs, highlighting the central role of expectations. The results provide clean experimental evidence on the microfoundations of price-setting behavior, with direct implications for understanding inflation dynamics and the transmission of shocks.

The three essays demonstrate the value of combining new data, machine learning methods, and experimental evidence to advance our understanding of core macroeconomic issues. By improving measurement of labour market tightness, refining models of monetary policy transmission in small open economies, and uncovering how expectations and frictions shape pricing decisions, the thesis provides insights that are both academically relevant and of practical importance for policymakers seeking to stabilize employment, inflation, and growth.

The results highlight several actionable insights. High-frequency vacancy indicators derived from online job postings—once corrected for biases—can meaningfully enhance real-time assessments of labour market tightness and better inform decisions around immigration, training, and workforce development. In monetary policy, accounting for structural breaks in Canadian data is crucial for correctly identifying the effects of interest rate changes, and the evidence suggests that the exchange rate channel plays a more limited role than often assumed. Finally, the experimental findings show that firms' expectations—particularly their forecasts of market prices—are central to inflation dynamics, highlighting the need for central banks to monitor and manage firm expectations alongside consumer expectations. Together, these contributions offer policymakers improved tools, clearer empirical guidance, and a broader understanding of how labour markets, monetary policy, and firm behaviour interact to shape employment, inflation, and economic stability.

Acknowledgments

I dedicate this dissertation to my late father, Dr. Mohamed Abdelhalim Abuallail, whose work as a plastic surgeon touched the lives of so many. He will forever be remembered as a hero, a pillar of strength to his family, and the most loving father. Beyond his professional excellence and dedication to his patients, he embodied kindness, generosity, and wisdom in every aspect of his life. His unwavering support and encouragement gave me the courage to pursue my ambitions, and his sharp intellect and charisma inspired all who knew him.

My sincere gratitude goes to my mother, Dr. Ghada Elbishry, whose love, wisdom, and sacrifices have been a constant source of strength and inspiration throughout my life and this journey.

This dissertation would not have been possible without the guidance and encouragement of many individuals. Foremost, my supervisor at the University of Ottawa, Dr. Isabelle L. Salle, who provided unwavering support, continuous guidance, and incredible mentorship throughout the course of my doctoral studies. Her insight, patience, and encouragement not only strengthened this thesis but also profoundly shaped my development as a researcher. I am deeply grateful for the opportunity to learn from her expertise and for her constant belief in my work.

I am sincerely grateful to the members of my committee—Dr. Jonathan Crechet, Dr. Yang Zhang, Dr. Chris Gunn, and Dr. Valentyn Panchenko—for their thoughtful feedback, careful reading, and invaluable suggestions. Their expertise not only strengthened the arguments and clarity of this dissertation but also encouraged me to think more critically and broaden the scope of my research.

I owe my deepest gratitude to my family, whose sacrifices, encouragement, and unwavering support have been the foundation of my journey. To my wife, Laura Abuallail, I am profoundly thankful for your patience, love, and constant encouragement, which sustained me during the most challenging moments. To my daughter, Kenzie Abuallail, you have been a source of joy and motivation, reminding me of what truly matters beyond academic work. I am equally grateful to my sisters, Dr. Hend Abuallail and Dr. Sara Abuallail, whose encouragement, love, and achievements have always inspired me to push further. My accomplishments are as much theirs as they are mine.

Finally, I wish to thank my friends and colleagues at the University of Ottawa, including Motasem Qaddoura, Ghina Abdul Baki, Aaron Tormeti, Paola Santamaria, Sadia Zohaib, Aline Zayat, and Sunny Karim. Their friendship, collaboration, and support have made this journey not only intellectually rewarding but also personally fulfilling. I am grateful for the many conversations, shared challenges, and moments of encouragement along the way, and I look forward to continuing our friendships and collaborations in the years ahead.

In loving memory of my father, Dr. Mohamed Abdelhalim Abuallail

Contents

List of Tables

List of Figures

Chapter 1

Bridging the Gap: Online Job Postings, Survey Data and the Assessment of Job Vacancies*

Abstract

Job vacancy data is essential for economists studying labour market dynamics or business cycles, but the conventional survey approach has limitations, particularly substantial lags. While online job postings offer the potential for a timely assessment of job vacancies, they also pose challenges in terms of accuracy. We critically analyze the measurement of job vacancies in the Canadian labour market, comparing data derived from a large database of online job postings from Vicinity Jobs, with the conventional survey method, the Job Vacancy and Wage Survey (JVWS). We then build and compare algorithms, and use machine-learning models to improve the accuracy of the measurement of actual job vacancies derived from online job postings compared to survey data. We introduce a robust regression with Huber weights, which we find further reduces prediction errors by an average of 15% compared to existing approaches in the literature. Our methods aim to provide academics and policymakers with a higher frequency and yet reliable measurement of job vacancies.

* Special thanks to the Labour Market Information Council (LMIC) for their support

1.1 Introduction

This paper introduces novel machine learning techniques and robust regression methods to address two key challenges in interpreting online job postings data as indicators of job vacancies. First, job vacancy data is often biased towards certain sectors, such as digitally intensive roles, which undermines its reliability for measuring economy-wide vacancies. Second, survey-based vacancy measures are released with considerable delays, limiting their usefulness for real-time labour market analysis. As a result, there is a need to enhance forecasting models that leverage online job postings to produce timely and accurate high-frequency estimates of job vacancies. Reliable job vacancy estimates are crucial for policymakers facing diverse policy questions, ranging from immigration, upskilling initiatives and education investments, to monetary policy decisions considering labor market dynamics. Accurate, timely, and unbiased forecasts would provide significant value to both academics and policymakers

Job vacancies are a signal of a gap in the labour market, where employers seek to be matched with workers with specific skills, qualifications, and experience. An increase in vacancies is indicative of a tightening labour market, which has important implications for wages and inflationary pressures. It can also signify shortages in workers with specific skills and in specific industries, leading to implications for areas such as immigration policy. The identification of these vacancies has long been recognized by economists as crucial to informing policy decisions. For instance, Pissarides (1986) argues that the high rate of unemployment in Britain in the 1980s could have been alleviated by an expansionary fiscal policy, an analysis that entirely depends on the assessment of job vacancies, unemployment flows, and the prevailing matching dynamics. In addition, macro-labour economists who study business cycles recognize the importance of studying vacancy data and worker flows, since these generate fluctuations in unemployment, hours, and in turn, output (Yashiv, 2007).

The conventional way to measure job vacancies in an economy is through employer surveys. For this reason, the Job Vacancy and Wage Survey (JVWS) was launched in Canada in 2014 to assess the level of job vacancies by region and by detailed occupation. The survey includes short job descriptions to allow researchers to match vacancies to National Occupation Classification (NOC) codes. While the JVWS is of high quality, compromises made (including lags in reporting, survey design, and administrative costs) often lead to an incomplete picture of the labour market conditions. Therefore, complementing this data with a measure that uses online job postings can provide a high-frequency real-time access to data to researchers and policymakers, allowing for quicker responses to changes happening in the labour market.

This can be especially important during times of shocks, such as the COVID-19 pandemic. When used alongside official vacancy data, online job posting data can provide valuable additional insights into unfilled positions and can circumvent the need for firms to report NOC codes or additional information such as required worker skills, which can be captured from online job postings. This could, in turn, increase the survey response rate, ease the burden of data collection, and enhance the overall accuracy and comprehensiveness of labour market data.

We explore data from Vicinity Jobs (VJ), which is an online web scraping tool that scrapes data from multiple online job posting websites, and offers data by province, city, and specific location. We find strong evidence that VJ data Granger causes vacancies from JVWS. This motivates the potential use of online job postings as a leading indicator, and hence, timely complement to job vacancies measured by the survey. However, the online job postings data series weighs all job vacancies equally, which does not take into account potential biases explored in Vu et al. (2019), such as over-representation of certain sectors and educational requirements. Online job postings data also contain geographical biases, and estimation problems, such as when employers use one posting to recruit for more than one position, or when postings are available for jobs that the employer is not actively recruiting for, as described in an insights report by the Labour Market Information Council (Labour Market Information Council (LMIC), 2020). To address these biases and improve the reliability of online job postings data as a real-time gauge for job vacancies, we develop and compare several statistical tools, algorithms, and machine learning models based on their ability to reduce errors in predicting the survey vacancies. Only a handful of papers exist in the literature that are specifically designed to address the challenges associated with interpreting online job posting data as job vacancies. Evans et al. (2023) use a winsorization[1] algorithm to improve online job postings forecasts in Australia. Turrell et al. (2022) use a weighting function, computed from the weights of jobs within their sectors in the survey, to improve the representativeness of online job postings using data from the UK.

This paper extends this literature by comparing these two procedures with a novel approach that integrates robust regression and machine learning models to forecast job vacancies using a combined measure. We introduce a weighting function to improve the representativeness of online job postings, implement a winsorization algorithm to address the influence of outliers in the data, and we introduce a robust regression with Huber weights to further improve survey prediction errors. We also discuss a novel approach of combining multiple different methods, such as combining a winsorization algorithm with a weights adjustment

[1]A winsorization technique is a method that deals with outliers by eliminating extreme values in a dataset, and replacing them with certain minimum or maximum values.

function, and we test and apply several machine learning models that use online job postings to predict vacancies. Finally, we combine this improved forecast accuracy with a weighting function to ensure the data is not only more accurate but also more representative of the broader labor market. Our results show that the combination of winsorization and robust regression with Huber weights significantly outperforms other methods, reducing prediction errors by an average of 15% compared to the winsorization algorithm alone.

The rest of this paper is organized as follows: In Section 2, we discuss the data, the conventional approach of measuring job vacancy data through surveys and our online job postings dataset. We also provide a comparison between both data sources. In Section 3, we detail the various algorithms we develop to address the challenges of online job postings data, including the machine learning forecast models, and highlight the key results. Section 4 concludes.

1.2 Data

In this section, we outline the traditional measure of job vacancies and its inherent limitations. We then introduce our proposed complementary dataset of online job postings, acknowledging some of its own constraints, which we aim to address in this paper. Finally, we provide a thorough comparative analysis of the two datasets.

1.2.1 The Canadian Job Vacancy and Wages Survey (JVWS)

Surveys are regarded as the most accurate way to measure vacancies in an economy, and in the Canadian context their reliability is reinforced by their nationally representative scope. The JVWS was launched to enable researchers to identify vacancies by region and detailed occupation. The survey targets businesses with at least one employee. These businesses are identified from the Business Register (BR), which is a Statistics Canada central repository that contains basic information about businesses in Canada. The total number of businesses in the BR as of December 2022 is approximately 1.22 million. The stratified random sample consists of 100,000 business locations, randomly selected on a quarterly basis, and is updated every three months to reflect changes to the BR. This includes adding new business locations and removing ones that no longer exist. Labour Market Information Council (LMIC) (2020) discusses the sample size aspect of the JVWS, arguing that it limits the localness of available observations and excludes several large employers. The quarterly sample is then allocated to the three survey months, allowing for the release of preliminary monthly estimates. Statistics Canada releases more detailed information by occupation and economic region from the

JVWS with a lag of a few months after the reference period.

Selected businesses are asked to complete the survey electronically. Although participation is mandatory, Statistics Canada estimates that the response rate varies between 75% and 85%. While adjustments are made to account for non-response, there remains some concern about the presence of non-sampling errors.

The survey includes brief short job descriptions to allow researchers to match vacancies with National Occupation Classification (NOC) codes, which businesses are not required to report directly due to high administrative costs of the survey.

As mentioned, the JVWS provides researchers with valuable insights into the labour market, including trends and changes in job vacancies. However, survey data have some well-documented limitations, which we tackle with our complementary online job postings dataset. First and foremost, there is a large gap in time reporting. The full detailed survey results are released with a lag of a few months. This makes it impractical to inform real-time policy decisions, reducing the survey's effectiveness as a tool for timely economic and labour market interventions. While Statistics Canada provides preliminary monthly estimates, they are less granular, lack information on specific occupations, are subject to change, and are still published with a few weeks lag.

Second, the sample size is less than 10% of all employers in Canada. This makes it impractical to study vacancies at a detailed local or occupational level because of data suppression for data confidentiality reasons. As such, the full data of JVWS is not publicly available, and researchers and policymakers must go through the Research Data Center (RDC) in order to gain access to it. Third, even though they account for roughly 5% of all employment in Canada, federal, provincial and territorial administration jobs are excluded from the survey. This also includes a considerable number of management occupations that are not captured in the data.

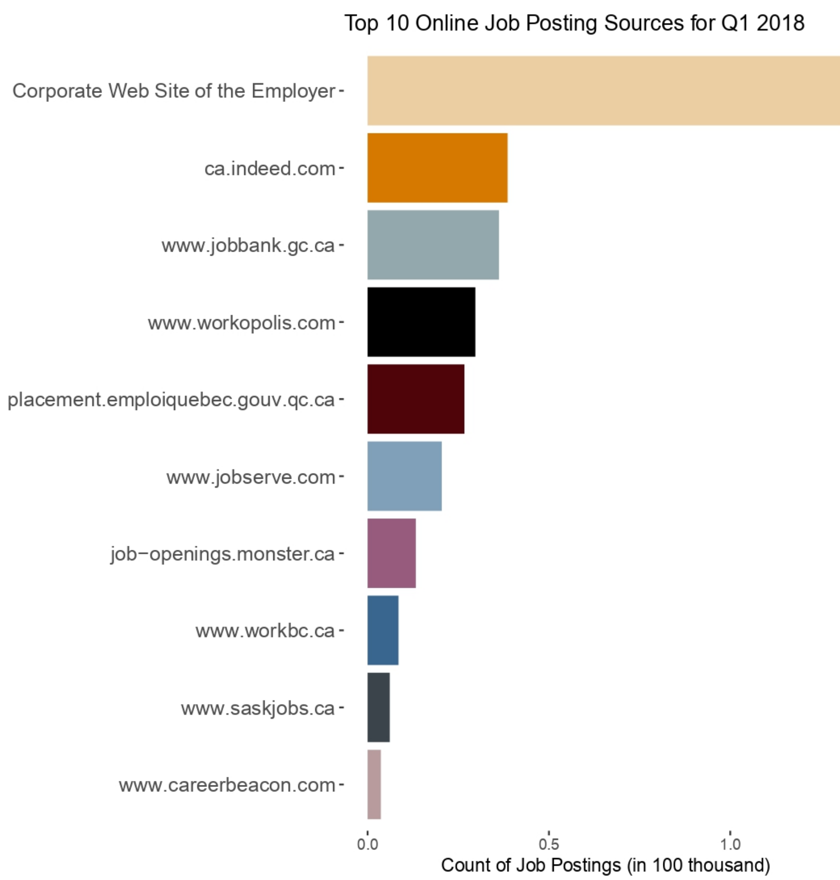
1.2.2 Online Job Postings

To address the aforementioned limitations of survey data in measuring job vacancies, there is a strong impetus to complement them with data from online job postings, which have emerged as a valuable alternative. In this section, we will highlight our source of online job postings.

Vicinity Jobs Data Description

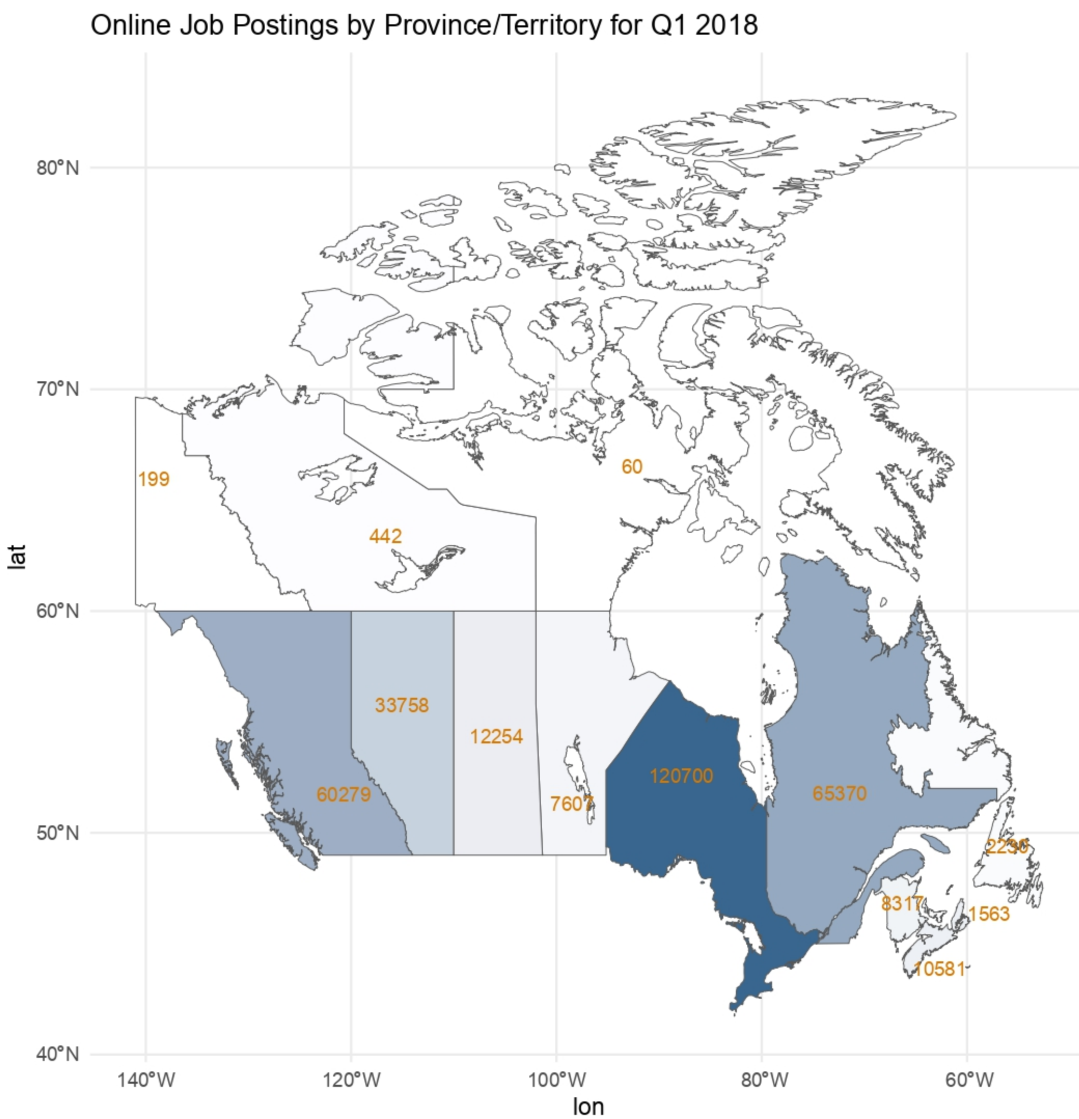
In our research, we explore data from Vicinity Jobs. The dataset contains 16,781,882 job postings for the period between 2018 and the end of 2023, and is collected using Vicinity Job's web-scraping tool that collects online job postings from multiple website sources. Approximately half of the data sources are from actual employer websites. Other third-party non-employer job posting websites may also be included. However, in order to be included, they must go through a manual assessment and have a verification process, such as verifying the employer and ensuring that the job postings they contain are authentic. Some examples of these sources include Indeed.ca and Monster.com. Most of the third-party data sources included are cost-based, meaning that employers must pay to post, or must incur high administrative costs. For instance, the federal job bank, Job Service Canada, includes multiple forms and fields that must be filled in, and requires the inclusion of a social insurance number in order to post there. All of these measures are designed to ensure the authenticity of the job postings. Figure 1-1 shows the top 10 sources of online job posting data in Canada.

Figure 1-1: Top 10 website sources for VJ data. Employer websites and government job boards comprise the majority of the data sources



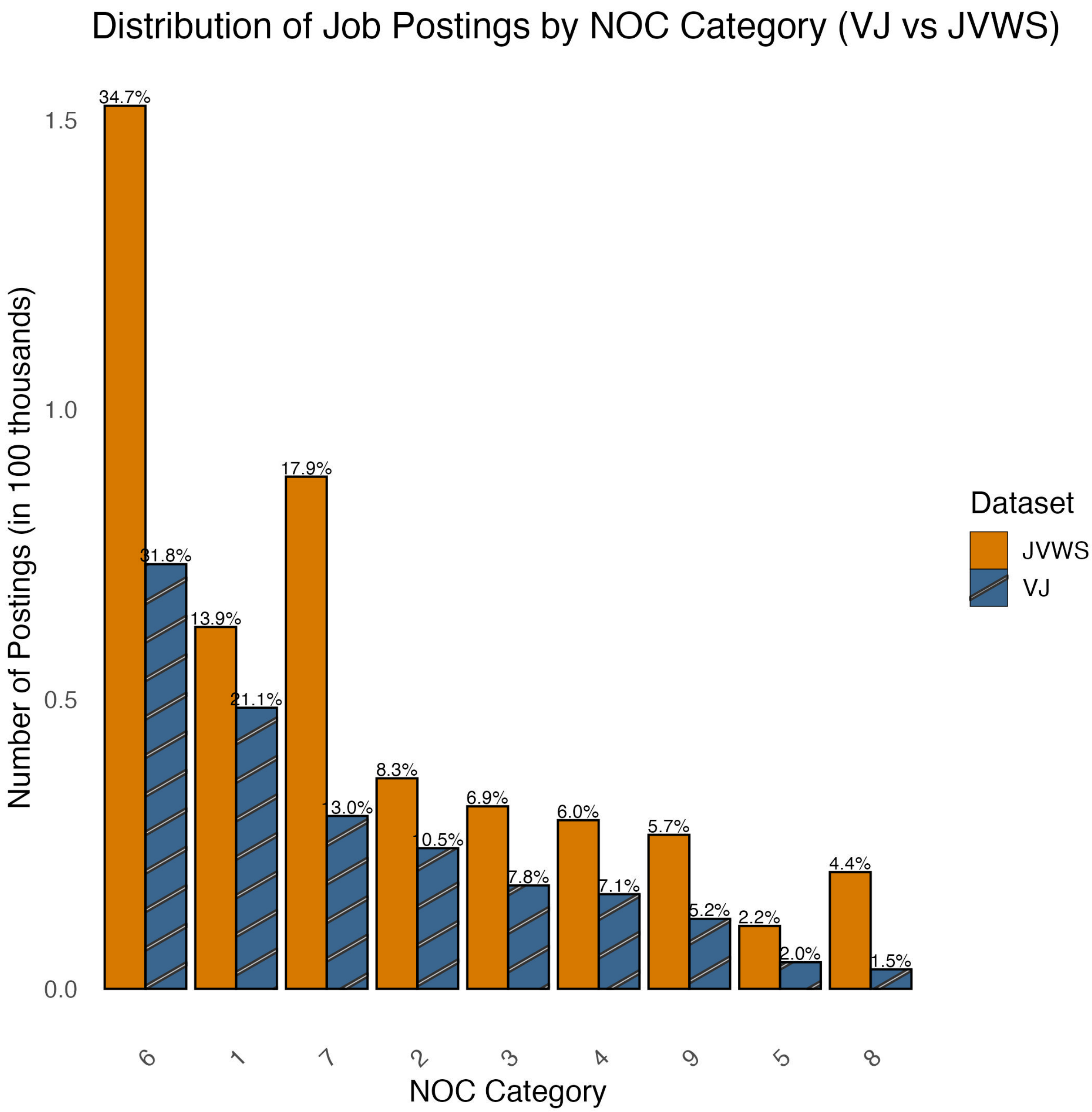
We examine the distribution of online job postings across Canada by province and territory in Figure 1-2. There is a clear regional disparity in the number of job postings, with certain provinces showing a higher volume of postings, reflecting a more robust online job market or the overall size of their economies and their higher demand for labour. The distribution of job postings is generally aligned with the weights of each province relative to the total Canadian GDP, as can be seen in Table 1-A.2, with few exceptions. Unsurprisingly, Ontario, British Columbia, Quebec, and Alberta have the highest number of job postings, with Quebec and BC exhibiting a particularly strong job postings activity that exceeds their GDP weights.

Figure 1-2: Distribution of online job postings by province and territory



We then look at the breakdown of the distribution of online job postings by broad occupational categories. Figure 1-3 highlights a large variation in the frequency of occupational categories in the online data.

Figure 1-3: Distribution of Broad Occupational Categories in Online Job Postings Data and Survey. (Percentages indicate weight of each category) See table 1-A.3 in the appendix for breakdown of NOC categories.



The first four NOC categories (NOC 1 to NOC 4) together account for 46.5% of the total job postings, while all the other categories together account for a combined 53.5% of the total share. By comparison, in the survey, the first four NOC categories have a combined share of only 35.1%, while all the other categories have a combined share of 64.9%. The first four NOC categories include occupations in business, finance, natural and applied sciences, education, and law. These groups include occupations that were found to be among the most digitally intensive (see Abuallail and Vu (2022)). Conversely, the remaining groups include trades, transport and equipment operators and related occupations, manufacturing and utilities occupations, and occupations related to natural resources, agriculture, and production.

We also examine the distribution of employment by broad occupational category in Appendix Table 1-A.4. We find large disparities between the distribution of online job postings and the corresponding share of these jobs in total employment. For example, Business, Finance, and Administration occupations, represent 21.1% of online job postings, but only 15.34% of employment. On the other hand, natural resources, agriculture, and related production occupations, account for 1.5% of job postings, but 2.06% of total employment. These observations motivate our focus on potential biases and limitations in the online job postings data.

Limitations and potential biases

We highlight several of the limitations of online job postings data, which we aim to address in our research. One of the most commonly recognized issues with online job posting datasets is the potential for duplication. Employers frequently post the same job opening on different websites in an attempt to reach a wider audience and attract more applicants. While web scrapers can detect identical text, variation in website formats often compel employers to make changes to their postings. Jijkoun (2016) found that comparing a sample of identical postings from the same employer across different websites using textual shingles[2] results in a duplication detection level as low as 37%. The algorithm was unable to detect the remaining duplicates. This finding highlights the challenges and complexities associated with identifying duplicate postings.

Another common issue documented in the literature on online job postings is the potential for bias in data representation across different occupations. Online job postings are often seen by employers as an effective way to attract candidates for high-skilled white-collar positions. However, this is not necessarily the case for other occupations, where more traditional forms of job advertising, such as physically posted "help wanted" signs or word-of-mouth, remain prevalent. These industries include hospitality and construction. Carnevale et al. (2014) estimate that 80% to 90% of job postings requiring a Bachelor's Degree or above are posted online. In contrast, only 30% to 40% of postings for those requiring a college degree are

[2]Shingles are a sequence of consecutive words as they appear in an online job posting. For instance "We are looking for critical thinkers" is a 6-word shingle, and "for critical thinkers from a diverse" is another shingle.

posted online, and only 40% to 60% of those requiring a high school degree are posted online, as highlighted in Appendix Figure 2-A.1. Furthermore, Labour Market Information Council (LMIC) (2020) found that, when comparing JVWS to Vicinity Jobs data, online job postings tend to over-represent positions requiring university degrees, while under-representing jobs that emphasize on-the-job training. Additionally, they observed that online postings disproportionately under-represent Canada's largest cities —Toronto, Vancouver, and Montreal—compared to other regions across the country.

In addition to the challenges mentioned above, online job postings also face a distinct problem of authenticity. Some employers may post jobs not with the intention of hiring, but to gauge interest or assess talent in a community for future hiring or expansion decisions. Conversely, employers might seek to recruit multiple candidates using a single job posting to minimize costs, as online job posting portals typically charge for each posting (Labour Market Information Council (LMIC), 2020). The effect of pricing on job postings has been studied as early as in the work of Abraham and Wachter (1987). More recently, Cajner et al. (2016) study the effects of changes in pricing on the Conference Board Help Wanted Online series (HWOL), and find that these changes distort the data, proposing adjustments to vacancy series constructed from HWOL to account for this distortion.

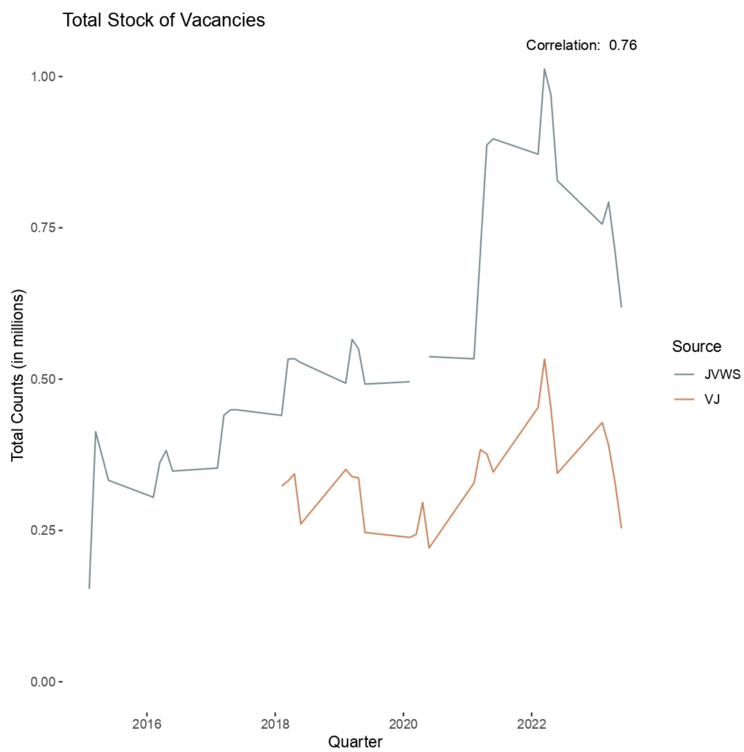
Some job postings may also be posted by scammers, and job scams have been on the rise in recent years. Keerthana et al. (2021) train a machine learning neural network on a dataset containing hundreds of fake jobs and thousands of real jobs. Despite attempting several machine learning models and techniques, the best result was around a 70% success rate in detecting fake jobs.

Finally, online job postings also suffer from a lack of historical data, with the Vicinity Jobs dataset only starting in 2018. This makes it difficult to analyze labour market trends using online job postings, but it is something that will only improve over time as more data is collected.

1.2.3 Comparing Online Job Postings with Survey Data

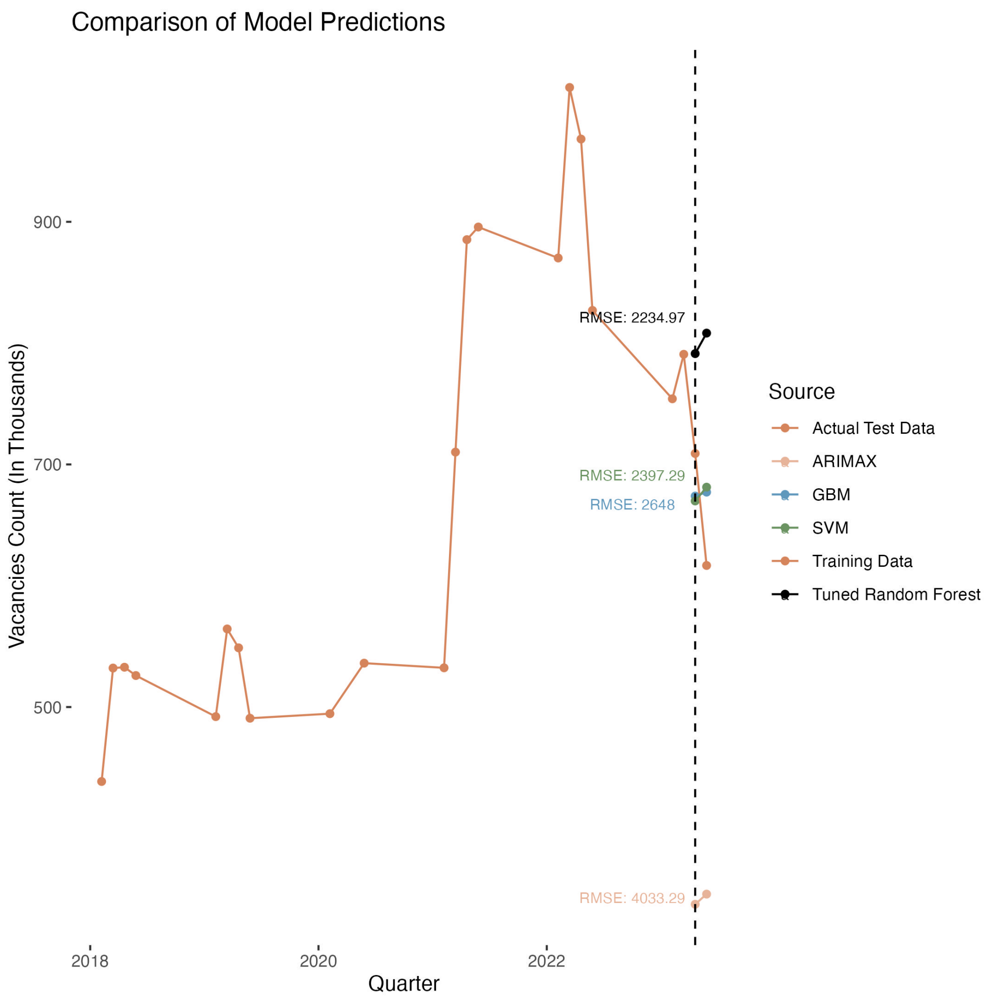
We first examine the total stock of vacancies as calculated from Vicinity Jobs data, and that from the JVWS vacancy data. As can be seen in Figure 2-A.2, there is a lower number of vacancies in online job postings, which could be due to the several issues mentioned earlier, including the fact that many job vacancies do not get posted online. However, we find a positive correlation of 0.76 between the two data sources across the entire time length of the dataset. We also test for and establish stationarity in both datasets using an augmented Dickey-Fuller test. We find a low p-value of 0.01 in both datasets, which indicates rejection of the null hypothesis of unit roots.

Figure 1-4: Total Stock of Vacancies Comparison of VJ and JVWS. Note: JVWS includes two missing values during the second and third quarters of 2020, when the survey was not conducted during the COVID-19 pandemic.



We then aim to assess the initial difference between both raw datasets using machine learning algorithms, which we can then use as a benchmark to compare the performance of our improved algorithms. For this, we fit different machine learning models, evaluate them using root mean squared errors (RMSE), and use the testing part of the sample to evaluate the performance of these forecasts. We use a tuned random forest model, which builds multiple decision trees and averages their predictions to improve accuracy and reduce overfitting, making it robust for handling large datasets. We also use a gradient boosting model (GBM), and a support vector machine (SVM). GBM improves predictions by combining multiple simple models sequentially, correcting errors from previous models, and is effective at capturing important and subtle patterns in the data. SVM works by finding the optimal hyperplane that best separates the data into different classes, which makes it effective in maximizing the margin between data points of different classes, thus improving overall prediction accuracy. We choose these models because of their effectiveness in handling diverse datasets and capturing complex relationships, as well as the balance they provide in accuracy, robustness, and computational efficiency for improving job vacancy forecasts using both JVWS and online job postings data. For a comprehensive comparison of different machine learning models, see Hassan et al. (2018) and Osisanwo et al. (2017). Figure 2-A.3 shows the different forecasts made by all models, and their RMSE values, compared to the testing sample.

Figure 1-5: Comparison of model predictions, showing the actual job vacancies divided into training and test samples divided by the dashed vertical line, and the two-period forecasts of different models compared to the test sample. Forecast errors are computed using Root Mean Squared Errors (RMSE) for all individual data points, aggregated over the quarter. The SVM, GBM, and tuned random forest models produce superior forecasts compared to a standard ARIMAX model, with a reduced RMSE of as much as around 50%.



We further break the model predictions down by NOC groups to compare predictions by RMSE across NOC groups that are most and least represented in the online job postings in Figures 1-A.8 and 1-A.9 in the Appendix. We find that forecasts are better, with lower RMSE values, for NOC groups that contain more digital occupations and are more highly represented in the online job postings data.

1.3 Algorithms and Results

We incorporate elements of both the Turrell et al. (2022) and the Evans et al. (2023) algorithms, with appropriate parameter modifications, and we enrich these approaches by adding a robust regression with Huber weights, and by combining multiple approaches simultaneously.

Given that online job postings from Vicinity Jobs are a flow variable, and JVWS vacancy data are a stock variable, we must first convert this flow variable into a stock variable. To do that, we need to calibrate the average time a posted job offer is filled. Using US data, The Josh Bersin Company (2023) finds that the average job posting is filled within 44 days. Given the absence of similar data for Canada, we assume a similar average as in the US, and aggregate over the 44 days before the end of each quarter. This is in-line with the six-week aggregation period that is widely used in the literature. The aggregation function is:

$$VJ_q = \sum_{d \in q} (VJ_d) \tag{1.1}$$

where VJ_d denotes the number of postings observed on day d, and the index $d \in q$ refers to all days that fall within the 44-day window associated with quarter q.

1.3.1 Weighting and Scaling Function

Given that online job postings tend to be more biased towards certain jobs than others, we introduce a weighting function on the job vacancies found in the job postings. The goal of this weighting function is to improve the representativeness of online job postings by adjusting the distribution of vacancies to reflect that observed in survey data. To achieve this, we consider the weight of each job vacancy from the JVWS in the total stock of vacancies for the quarter. Our weighting function, then, becomes:

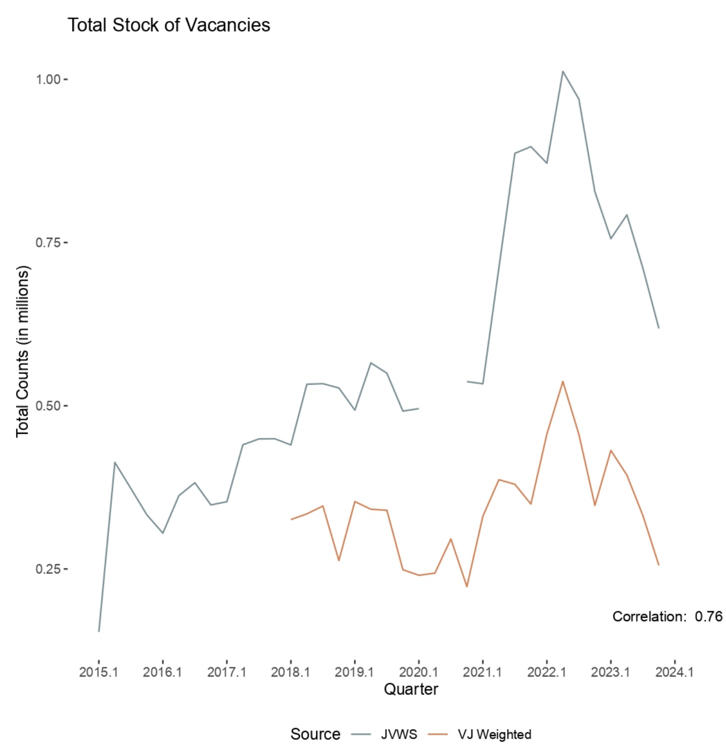
$$w_{i,q} = \frac{(V_{i,q})^{jvws}}{V_q^{jvws}} \tag{1.2}$$

where i stands for a NOC code, q stands for the quarter, $V_{i,q}^{jvws}$ stands for the vacancies from the JVWS survey for a given NOC code in a given quarter, and V_q^{jvws} is the sum of vacancies for all NOC codes within the same quarter. To incorporate the weights, we create a set of matched NOCs for all quarters between VJ and JVWS, and we re-scale the distribution of online job postings in proportion of their relative weights, that is, the scaling factor equals

$(1+w_{i,q})$. This scaling ensures that the distribution of online job postings aligns more closely with survey data, while avoiding potential issues of overfitting if the weights were directly applied to all online job postings. The adjustment also accounts for the under-representation of certain jobs, and the overall gap between vacancies measured from JVWS and online job postings, namely that several alternative recruitment methods such as "Help Wanted" signs also exist.

Appendix Table 1-A.5 shows the effect of the matched and weighted approach on improving the distribution of jobs compared to the survey, using the same example NOC groups mentioned in the previous discussion. We also find that this has little effect on the aggregate number of vacancies measured from online job postings. To see this, we use this scaled VJ data to calculate the error in absolute value, determined by the absolute difference in mean percentage changes between consecutive quarters in Vicinity Jobs data and the same metric for JVWS. The error in absolute value is then compared against the error that would occur without applying the weight and scaling function. As illustrated in Appendix Figure 2-A.4, the reduction in average error is too small to be visually discernible in the graph. On average, the errors fall by only around 0.07%. We also plot the stocks of job vacancies again, using the weighted VJ data instead in Figure 1-6, which shows that the adjustment in online job postings does not substantially change the total number of vacancies measured.

Figure 1-6: Total Stock of Vacancies Comparison of VJ and JVWS



We proceed to predict the aggregated job vacancies using the adjusted VJ data and our selected machine learning models to assess potential improvements in predictions. As expected, the models do not demonstrate increased accuracy in predicting the subsequent two quarters. Figure 1-A.7 in the Appendix shows the forecasted points compared to the test dataset points, and the RMSE values. It is important to note that all results are aggregated for the economy, but are calculated individually by NOC, allowing us to determine model predictions and RMSE for any specific job or broad occupational category in the Canadian economy.

1.3.2 Winsorization Algorithm

To improve the predictions, we develop and test a winsorization algorithm inspired by Evans et al. (2023). The main benefit of this is to address the issue that online job posting platforms with a large number of postings have an out-sized effect on the vacancy results. Figure 1-A.2 shows that there is little difference between sources with small or large number of postings in predicting job vacancies. In other words, the size of an online platform is only marginally linked to its ability to predict vacancies. By directly using quarter-to-quarter changes across all website sources to predict vacancies, we would be implicitly weighing each source's signal according to the change in the source's count of postings. Sources with larger numbers of postings will have a much greater impact on the aggregated data, and thus the predictions or trends derived from that data. There are two main problems with this. First, when a single large source has an increase or decrease in its number of postings, the overall vacancy measurement can significantly shift, indicating a market change. However, this shift may not truly represent market trends if it is driven by the behavior of the source (e.g., changes in scraping policies or job posting strategies) rather than genuine changes in the number of job vacancies. Second, larger sources are likely to have a higher incidence of duplicate postings, which can distort the actual number of unique job vacancies.

To implement the winsorization algorithm, we compute quarter-to-quarter percentage changes in postings for each online job board. Let $j = 1, \ldots, J$ index job posting sources. For each source j and quarter q, the variable $p_{j,q}$ denotes the aggregated number of online job postings for that source in quarter q, constructed using the stock measure defined above, whereas the JVWS would be the released vacancy level per NOC job at that specific quarter. The percentage change between quarters $q-1$ and q is then given by:

$$\Delta p_{jq} = \frac{p_{jq} - p_{j,q-1}}{p_{j,q-1}} \tag{1.3}$$

Then we apply the three-step winsorization described in Evans et al. (2023), by testing different values of the cut-off $k\%$. Finally, we obtain the mean of the winsorized values $\Delta \bar{p_{jq}}^{(w)}$.

To select the cutoff parameter (k), we apply a simulation that tests different possible values and finds the value that minimizes the mean absolute prediction error over the entire dataset. We find that the optimal cut off is 20%. This is higher than the optimal cut off in Evans et al. (2023) of around 10%, and is due to the fact that this parameter is highly data dependent. Figure 1-A.3 in the appendix shows a comparison of the absolute value errors across all quarters when using a 5% winsorization algorithm. Figure 1-A.4 in the appendix shows that this improves for a 20% winsorization cut-off level, with a reduction in the aggregate average error across all quarters of around 27% compared to the 5% cut-off level.

1.3.3 Robust Regression with Huber Weights

Given the clear potential for outliers to significantly impact our understanding of changes in the labour market when using online job postings data, we explore applying a robust regression with Huber weights to the winsorized data using the M-estimation method (Huber et al., 1967). Robust regression methods have been widely used to handle outliers in data, for instance for households, where outliers in consumption and earnings could have an influential impact (Gorodnichenko and Peter, 2007). They are extensively used to control for outliers and influential observations in surveys, such as for firm expectations on macroeconomic variables (Coibion et al., 2018).

Compared to ordinary least squares (OLS), which minimizes the sum of squared residuals, M-estimation minimizes a more general objective function. Consider the regression model

$$y_i = x_i'\beta + \varepsilon_i,$$

where y_i denotes the dependent variable, $x_i = (x_{1i}, \ldots, x_{Ki})'$ is the vector of regressors for observation i, $\beta = (\beta_1, \ldots, \beta_K)'$ is the vector of parameters to be estimated, and $\varepsilon_i = y_i - x_i'\beta$ represents the regression error term. An M-estimator solves

$$S(\beta) = \sum_{i=1}^{n} H(\varepsilon_i),$$

where $H(\cdot)$ denotes a general loss function applied to the residuals; for OLS, this corresponds to $H(\varepsilon_i) = \varepsilon_i^2$. Taking the derivative of $S(\beta)$ with respect to a generic parameter β_k and setting it equal to zero yields

$$\sum_{i=1}^{n} \frac{\partial H(\varepsilon_i)}{\partial \varepsilon_i} x_{ki} = 0. \tag{1.4}$$

It is convenient to define the weight applied to each observation as

$$w_i = \frac{1}{\varepsilon_i} \frac{\partial H(\varepsilon_i)}{\partial \varepsilon_i}.$$

Substituting this definition into equation (1.4) leads to the iteratively reweighted normal equation

$$\sum_{i=1}^{n} w_i \, \varepsilon_i \, x_{ki} = 0, \tag{1.5}$$

which shows that M-estimation can be carried out using an iteratively reweighted least squares (IRLS) procedure. Starting from an initial guess for the weights, one fits a weighted regression, updates the residuals and weights, and repeats these steps until convergence is achieved.

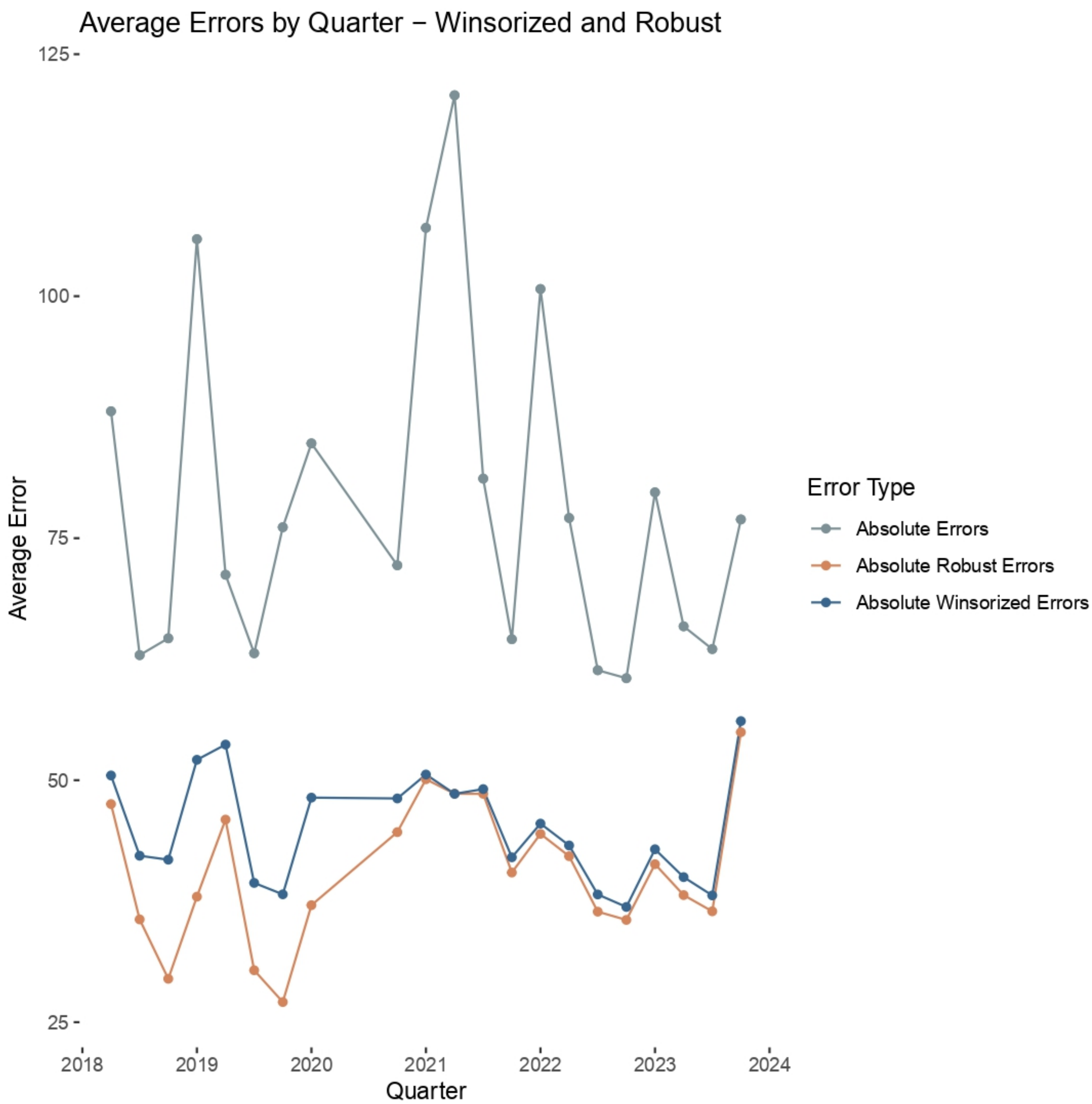
To implement the Huber M-estimator specifically, the loss function is defined as

$$H_\psi(\varepsilon_i) = \begin{cases} \varepsilon_i^2/2, & \text{if } |\varepsilon_i| \leq \psi, \\ \psi|\varepsilon_i| - \psi^2/2, & \text{if } |\varepsilon_i| > \psi, \end{cases} \tag{1.6}$$

where $H_\psi(\cdot)$ denotes the Huber loss function and $\psi > 0$ is the tuning parameter that determines the threshold at which the function transitions from quadratic to linear.

In the empirical application, a robust regression is estimated using only a constant term, where the dependent variable is the quarter-to-quarter percentage change in job postings from the VJ dataset after winsorization. This allows us to obtain a robust measure of central tendency to the percentage movements of jobs across all sources, from one quarter to another. Using this adjustment to all outliers identified through the Huber function, we obtain Figure 1-7, which shows the improvement in errors compared to the winsorization-only algorithm. As in the previous case, we retain all calculations and errors at the NOC level, which allows for the analysis of certain jobs or broad occupational categories. By implementing the robust regression on the winsorized changes, the average absolute value error of prediction across all quarters is reduced by an additional 6%, compared to the average errors calculated using the winsorization-only approach.

Figure 1-7: Errors when using Robust Regression with Huber weights, compared to Winsorization and original errors. The y-axis contains the aggregated average error for each quarter, calculated using the absolute value distance between the measure of vacancies from online job postings and that from the survey.

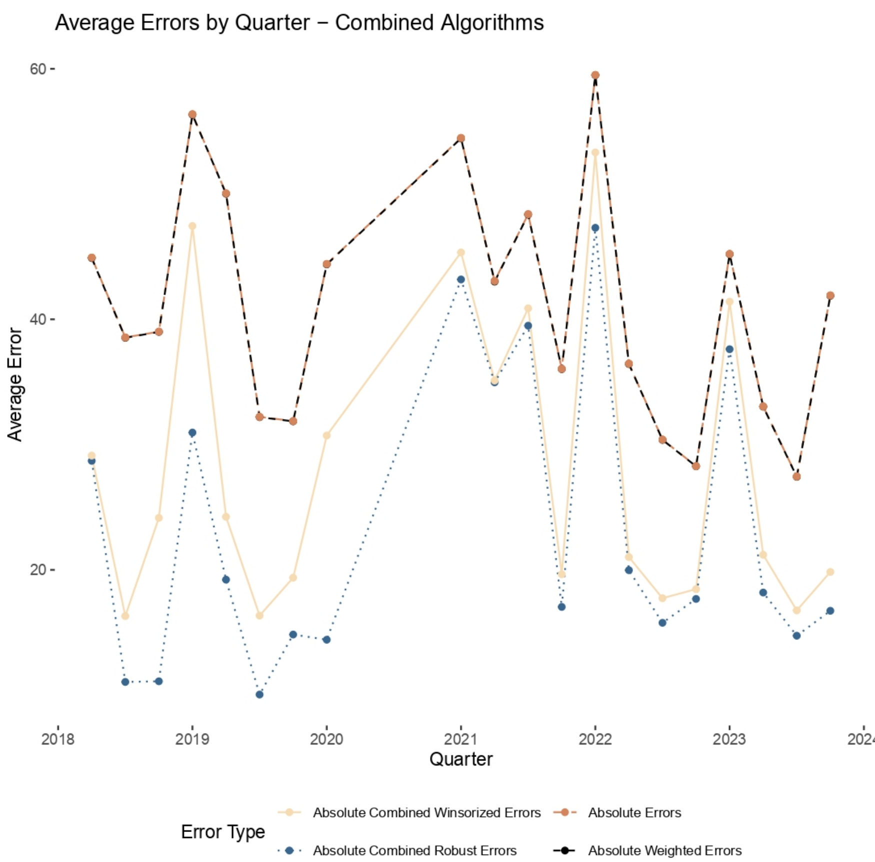


1.3.4 Combined Approaches

Another important contribution of our work is the ability to combine different approaches in an attempt to solve the representation issue, while also improving the prediction of job vacancies using online job postings. We start by combining our weighting function introduced earlier with the winsorization approach. The goal is to ensure that, while we tackle the issues of potential duplication and outsized effects of larger job posting sources, we also account for the fact that, due to the nature of the dataset, job postings are biased towards certain jobs more than others. To create this combined approach, and after winsorizing extreme values by source, we aggregate this data to find the number of job postings corresponding to these winsorized percentage changes. We then apply the weighting function and obtain the absolute value errors of the winsorized-weighted algorithm. We can clearly see in Appendix Figure 1-A.10 that this combined approach further leads to large improvements in the prediction errors, a reduction of around 32% compared to the raw data.

We also apply the robust regression method to the winsorized percentage changes for a winsorized-weighted-robust algorithm. This results in a further 15% reduction in absolute value errors over all quarters. Figure 1-8 shows a comparison of the two combined algorithms with the weighting function approach, and the original errors. We use a Diebold-Mariano test, and we find that this improvement in errors in the combined algorithms approach compared to raw data is statistically significant with a p-value close to 0 (see Appendix Table 1-A.6).

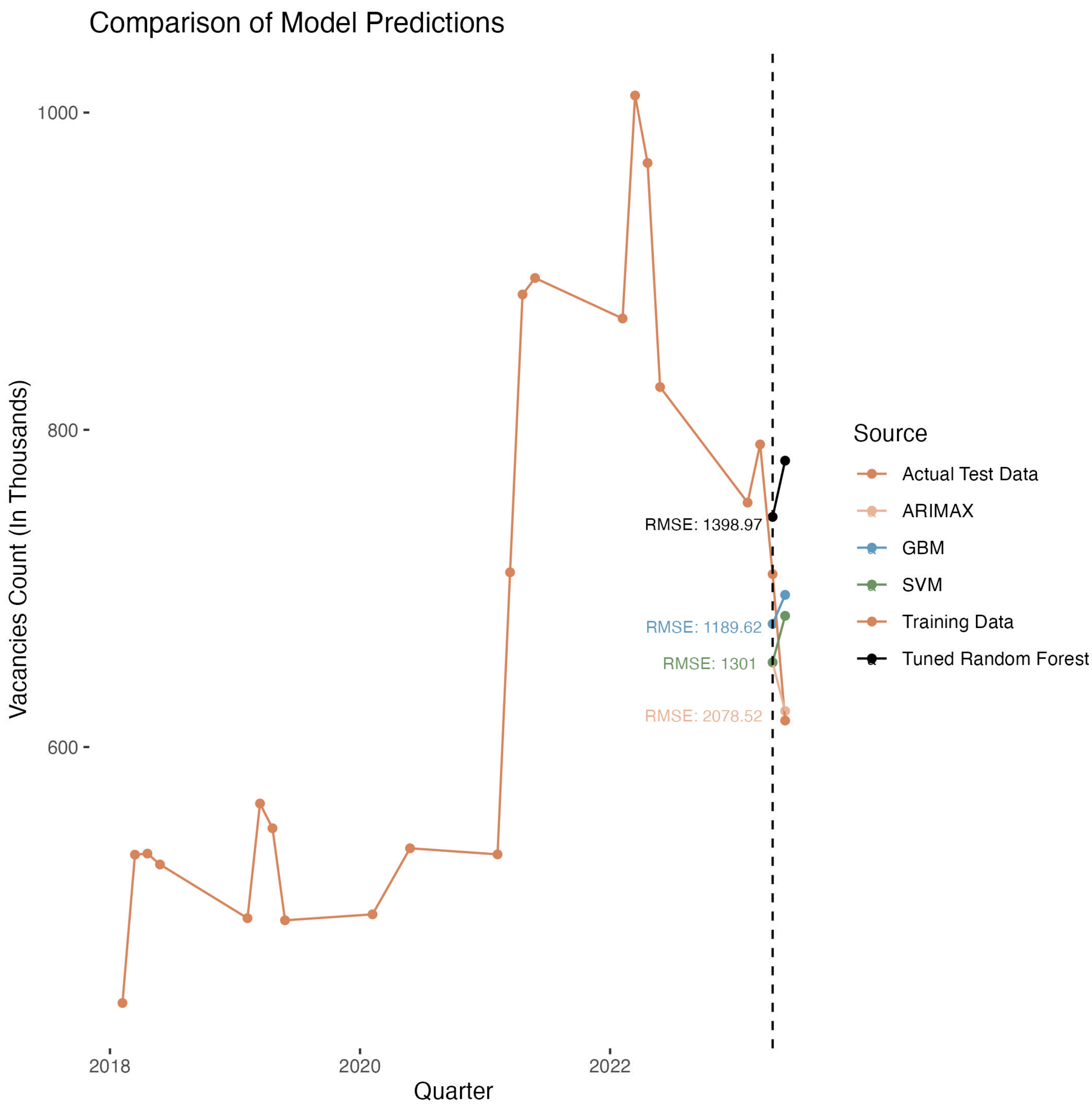
Figure 1-8: Comparison of combined algorithm errors with weighting function errors, and original errors.



1.3.5 Machine Learning Forecasts

Given this improvement, we employ our suite of machine learning models once again to evaluate the potential improvement in predictive accuracy with the adjusted online job postings dataset. Figure 1-9 shows an improvement in machine learning predictive capabilities for JVWS data points within the test dataset, as measured by the RMSE. Among all models tested, the GBM and SVM models perform better, achieving forecasts that more closely approximate the actual data, with a fall of RMSE of around 50%.

Figure 1-9: Performance of Machine Learning Models on combined algorithm data.



1.4 Conclusion

Job vacancy data is crucial for understanding labour market dynamics and business cycles, informing policy decisions, and responding to economic fluctuations. Conventionally, the main approach to measuring vacancies has relied entirely on survey data. We discuss their inherent limitations, including issues of lags in data release, and data granularity. In response to these challenges and the need to complement these datasets, we explore the potential of online job postings as a complementary source of labour market data. Our study of data from Vicinity Jobs highlights the real-time, granular, and expansive nature of online job postings data. However, we also identify and address the inherent limitations and biases in this approach, such as the potential prevalence of duplications, authenticity issues, and the over- or under-representation of certain job categories and geographic areas.

To address these limitations, we develop an innovative and comprehensive approach that employs a suite of machine learning models, a weighting function, a winsorization algorithm to align online job postings more closely with official JVWS data, and a robust regression with Huber weights methodology. We show that a combination of winsorization and robust regression, along with a weighting function can simultaneously improve prediction errors and representativeness, with a decrease of 15% in prediction errors in the combined dataset, compared to using the existing methods in the literature. Addressing these challenges allows for real-time labor market analysis, which is particularly valuable in periods of economic uncertainty or rapid change.

While conventional surveys like the JVWS remain invaluable for their high-quality, depth, and longer data time series, the inclusion of online job postings offers a complementary perspective that is more agile and expansive. This dual approach provides researchers and policymakers with a robust and responsive toolkit for understanding and responding to labor market dynamics. Our research contributes significantly to the ongoing dialogue on labour market data analysis, offering practical solutions to long-standing challenges and setting the stage for more timely, and effective labor market policies and interventions in Canada.

There are several potential extensions to this work. With granular analysis and geographically segmented data, we can explore changes across broad occupational categories in different parts of Canada. We can also examine the differences arising from substituting labour vacancy data for online job postings data, before and after combining algorithmic approaches. Additionally, studying the relationship between job vacancy metrics derived from online job postings and business cycle fluctuations presents another promising direction. This research could investigate the predictive power of these vacancy measures on the output gap or the unemployment rate, improving macroeconomic forecasting capabilities.

1-A Appendix

1-A.1 Who Uses Online Job Postings?

In order to get a sense of the importance of online job postings data, we set out to study which organizations in Canada, whether governments (provincial or federal), think tanks, or others, are actively monitoring or, in some way, using online job postings. While most provincial government websites include job boards and online job postings on their platforms, whether for the purpose of hiring for government jobs or for displaying online postings available in the province, not all governments use this type of data to analyze job market information, which is what we aim to study. While the list in table 1-A.1 is by no means exhaustive, it provides an idea about the prevalence and importance of online job postings data in research, and in labour market insights.

Table 1-A.1: Online Job Postings Data used in Canadian organization work and research.

Name	**Type of Organization**	**Use of OJP**	**Data Sources**
The Conference Board of Canada	Think Tank	Canadian Hiring Index	Vicinity Jobs
Bank of Canada	Government	Research Report	Indeed
OECD	International Organization	Research Report	Lightcast
Government of Canada	Government	Labour Market Insights in Budget 2014	Wanted Analytics
Eco Canada	Non-profit Organization	Environmental Job Market Trends	Gartner TalentNeuron
Government of British Colombia	Government	Estimating and Forecasting Environmental Workers	Environmental Careers Canada (ECO)

1-A.2 Supplemental tables and graphs

Table 1-A.2: Distribution of GDP by province as of end of 2017, using chained 2017 dollars, and Q1 2018 weight of total job postings by province. Source: Statistics Canada and author calculations.

Province	End of 2017 GDP weight	Q1 2018 Job Posting weight
Newfoundland and Labrador	1.56%	0.69%
Prince Edward Island	0.31%	0.48%
Nova Scotia	1.98%	3.27%
New Brunswick	1.63%	2.57%
Quebec	19.41%	20.22%
Ontario	38.40%	37.33%
Manitoba	3.32%	2.35%
Saskatchewan	3.78%	3.79%
Alberta	16.03%	10.44%
British Columbia	13.07%	18.64%
Yukon	0.14%	0.06%
Northwest Territories	0.22%	0.14%
Nunavut	0.15%	0.02%

Table 1-A.3: NOC 2021 Broad Occupational Categories (BOC). Data source: NOC, Statistics Canada

NOC 2021 BOC	Description of Category
0	Legislative and senior management occupations
1	Business, finance and administration occupations
2	Natural and applied sciences and related occupations
3	Health occupations
4	Occupations in education, law and social, community and government services
5	Occupations in art, culture, recreation and sport
6	Sales and service occupations
7	Trades, transport and equipment operators and related occupations
8	Natural resources, agriculture and related production occupations
9	Occupations in manufacturing and utilities

Table 1-A.4: Employment weights in Canada in March 2018 distributed by NOC 2021 Broad Occupational Categories (BOC). Data source: Statistics Canada

NOC 2021 BOC	Weight of Employment in March 2018
0	9.24%
1	15.34%
2	7.68%
3	7.51%
4	10.56%
5	2.85%
6	24.53%
7	15.22%
8	2.06%
9	4.99%

Table 1-A.5: Comparison between matched and weighted VJ, raw VJ, and JVWS data distributions.

	VJ Raw	**VJ Weighted**	**JVWS**
NOCs 1 - 4	46.5%	33.04%	35.1%
All other NOCs	53.5%	66.96%	64.9%

Table 1-A.6: Results of the Diebold-Mariano test comparing the errors from raw data with those from the combined robust algorithm

Diebold-Mariano Test Results	
Test Between	Absolute Errors vs Absolute Combined Robust Errors
DM Statistic	15.999
Forecast Horizon	1
Loss Function Power	2
P-value	$< 2.2 \times 10^{-16}$
Alternative Hypothesis	Two-sided

Figure 1-A.1: Percentage of Job Openings Posted Online by Education Level. Data source: Carnevale et al. (2014)

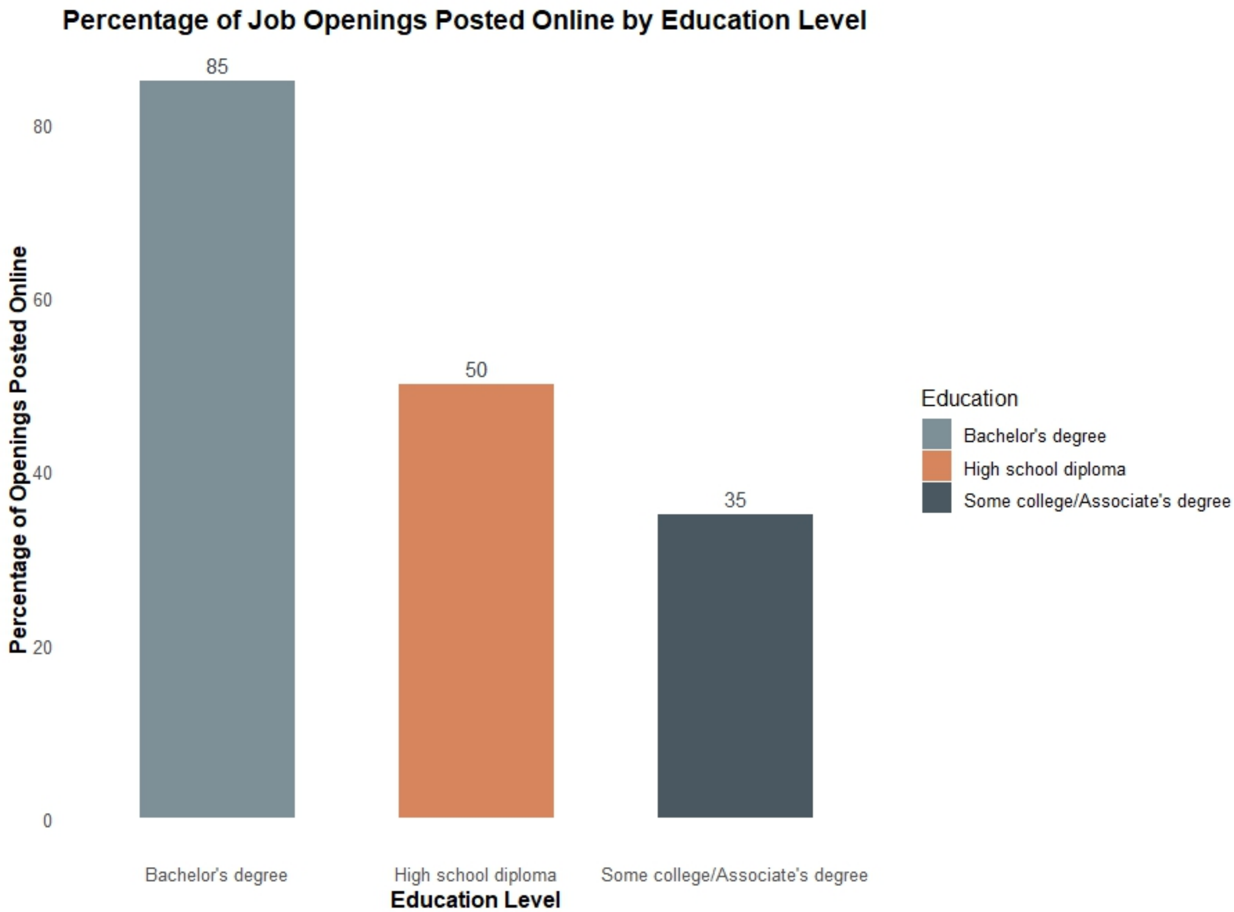


Figure 1-A.2: Comparison of model predictions (weighted VJ data). Note: the slope of the best fit line is -0.0336684

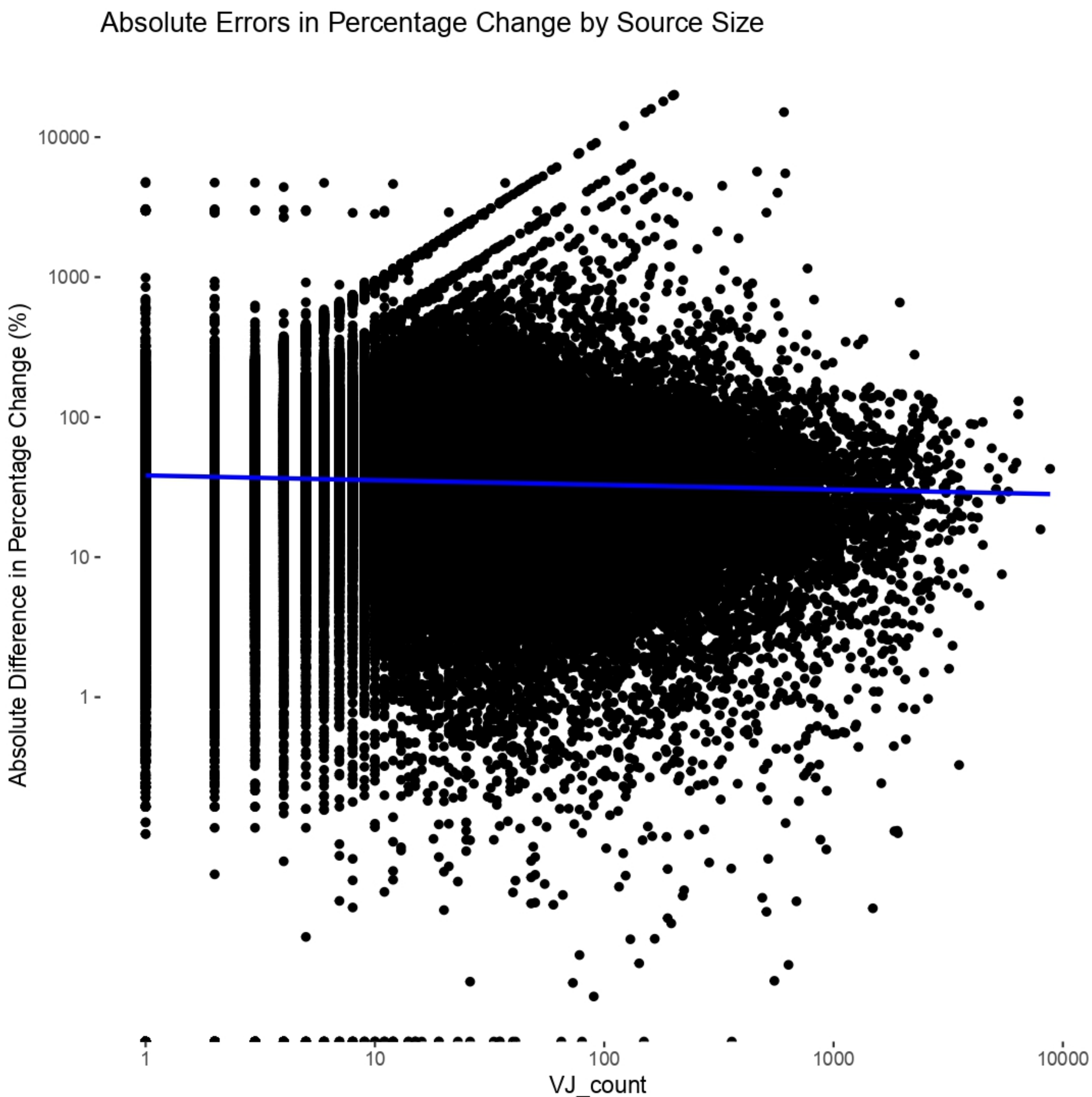


Figure 1-A.3: Absolute Value Errors when using a 5% winsorization algorithm

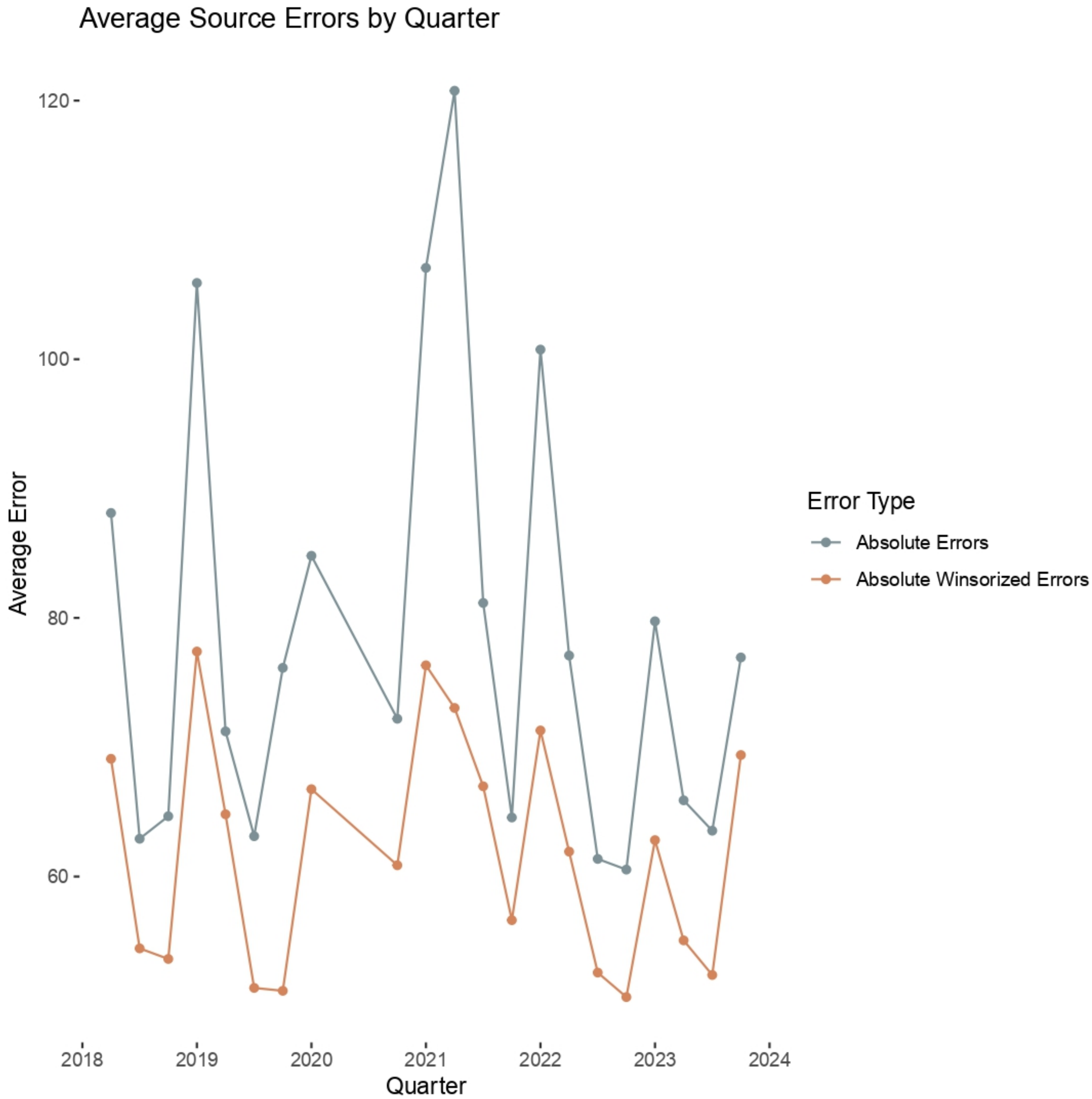


Figure 1-A.4: Absolute Value Errors when using a 20% winsorization algorithm

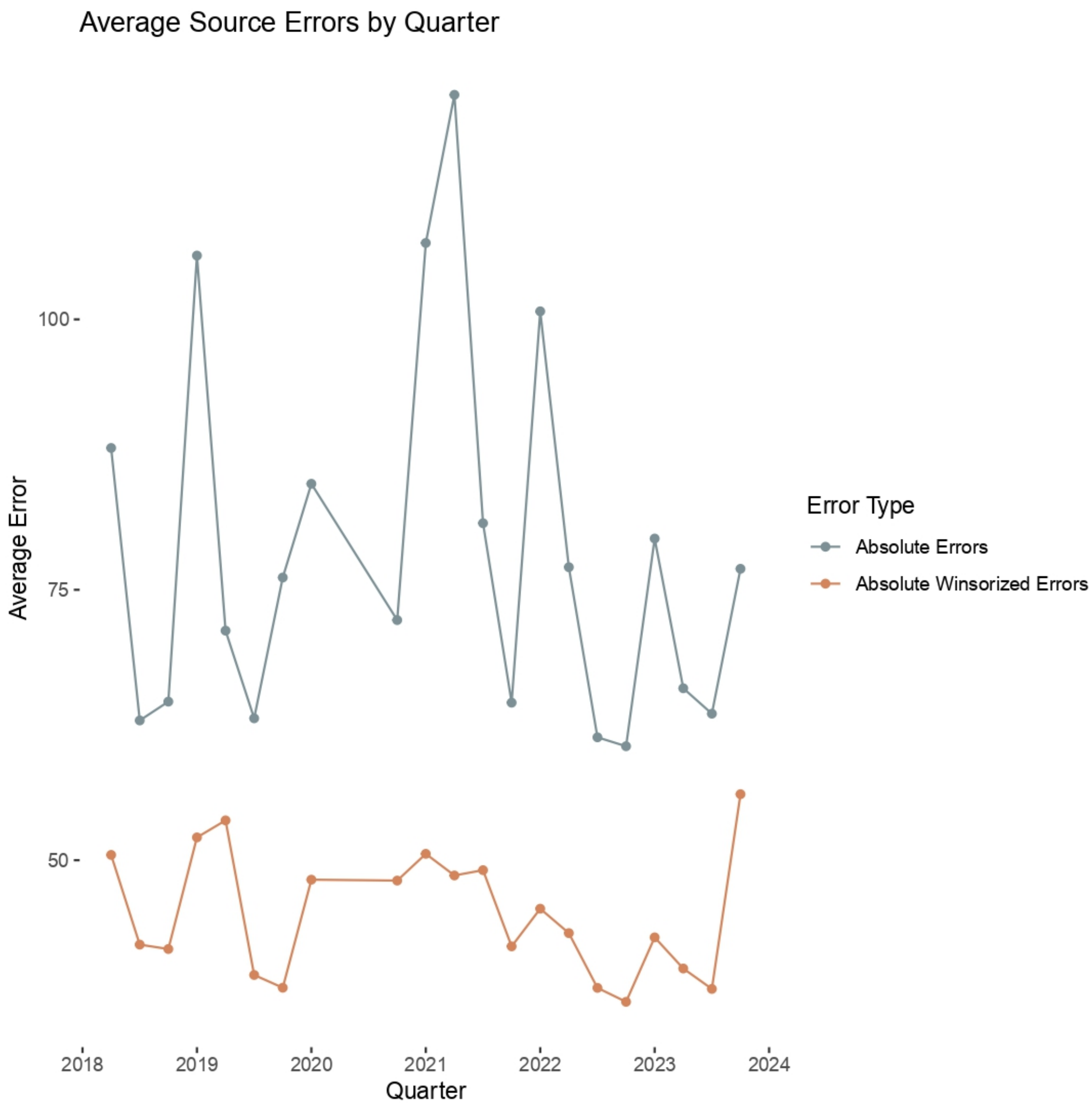


Figure 1-A.5: Aggregated stock of vacancies using Robust Regression, compared to survey vacancies

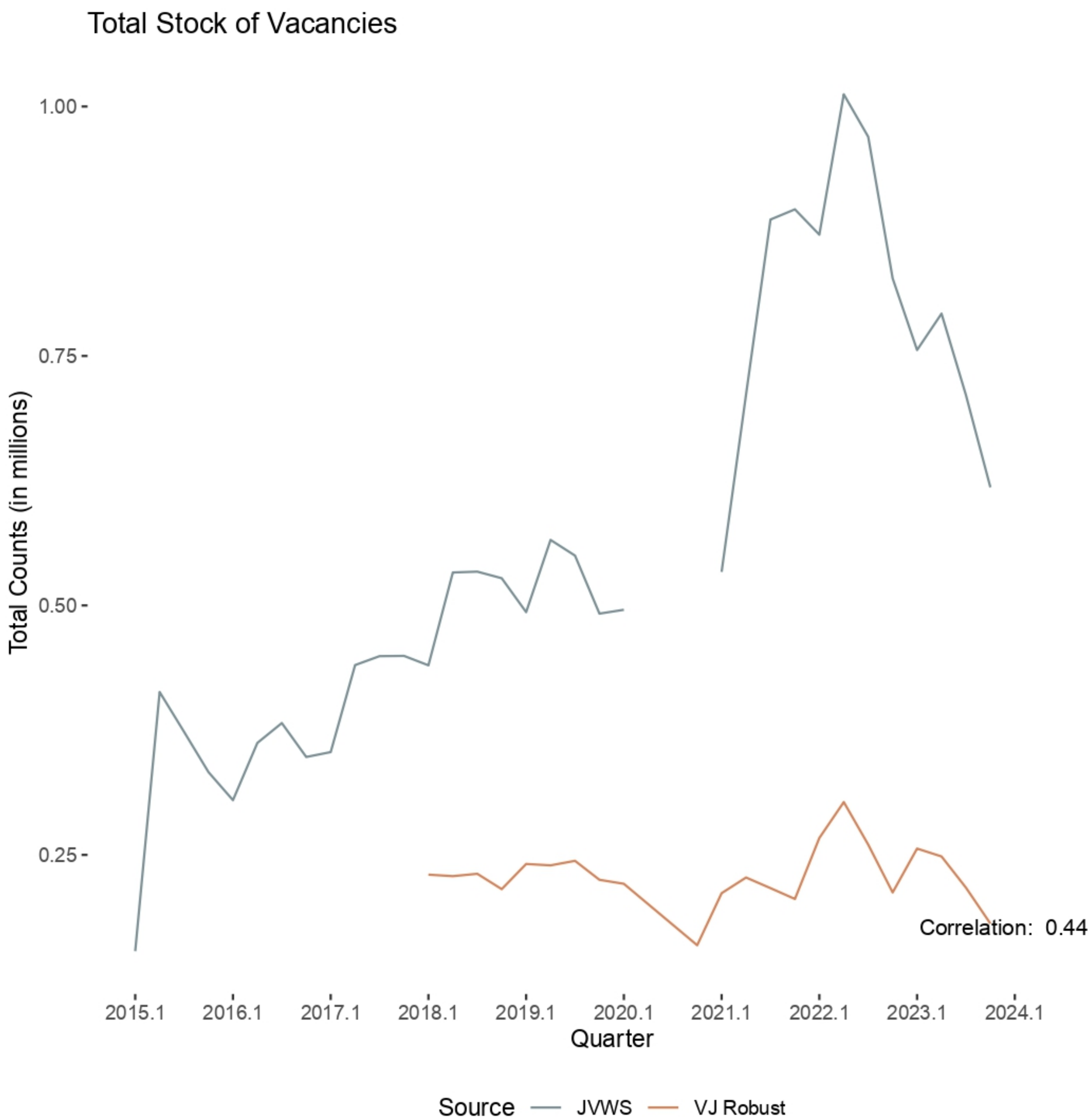


Figure 1-A.6: Absolute Value Error for Weighted VJ compared to VJ. Both lines completely overlapping, indicating no change in errors.

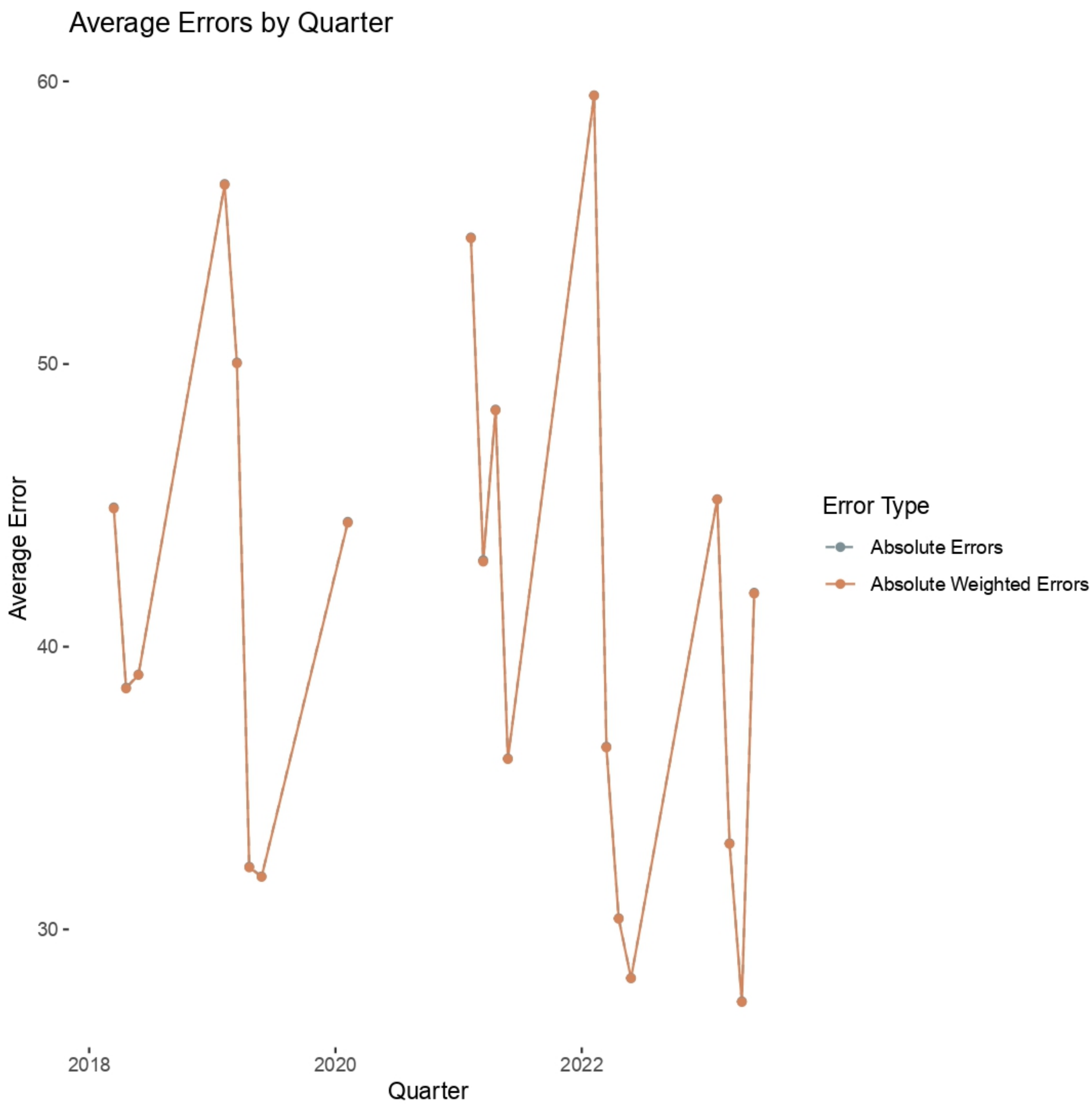


Figure 1-A.7: Comparison of model predictions (weighted VJ data)

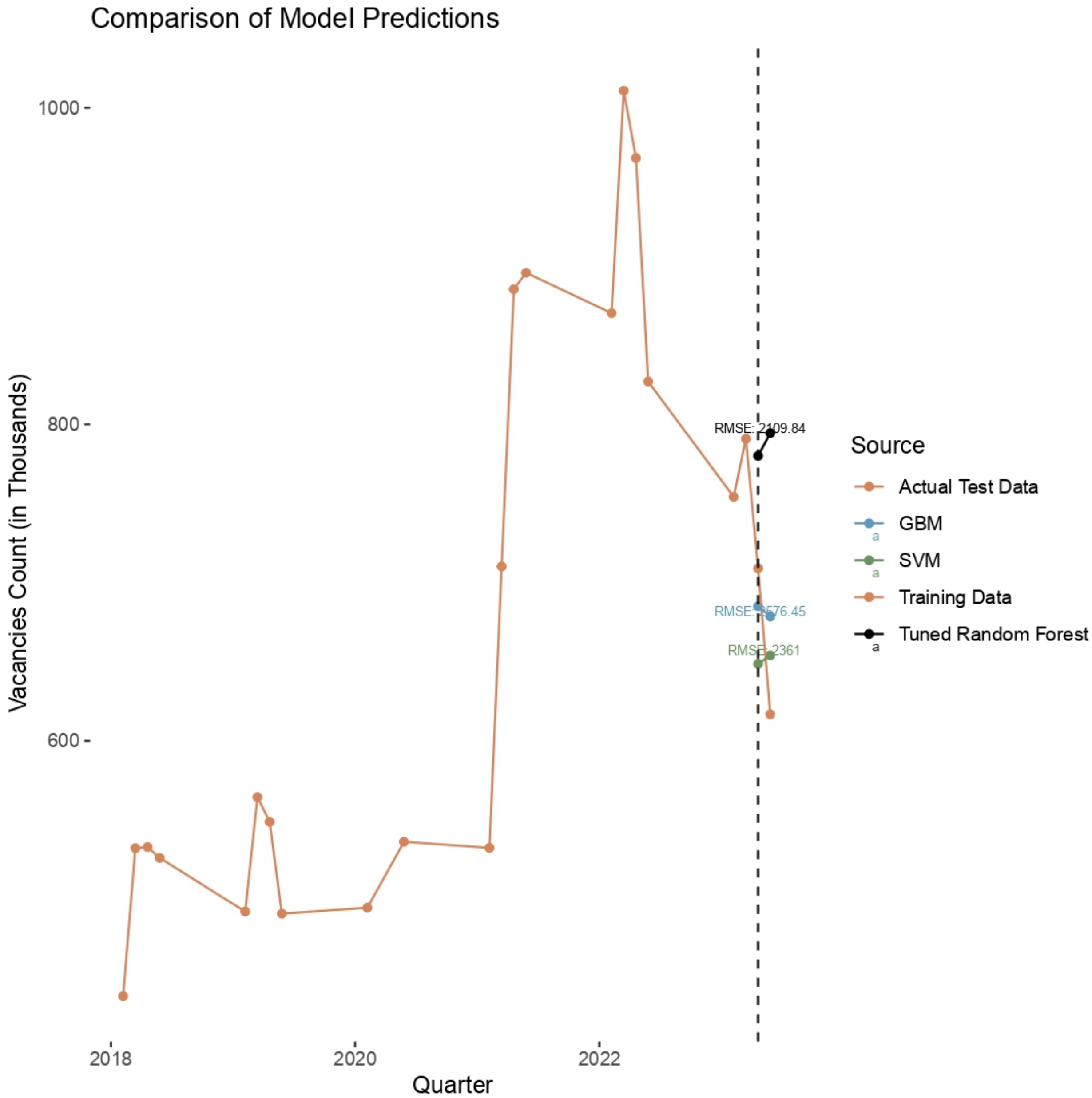


Figure 1-A.8: Comparison of model predictions (using raw VJ data) specifically focused on most represented NOC groups 1 - 4

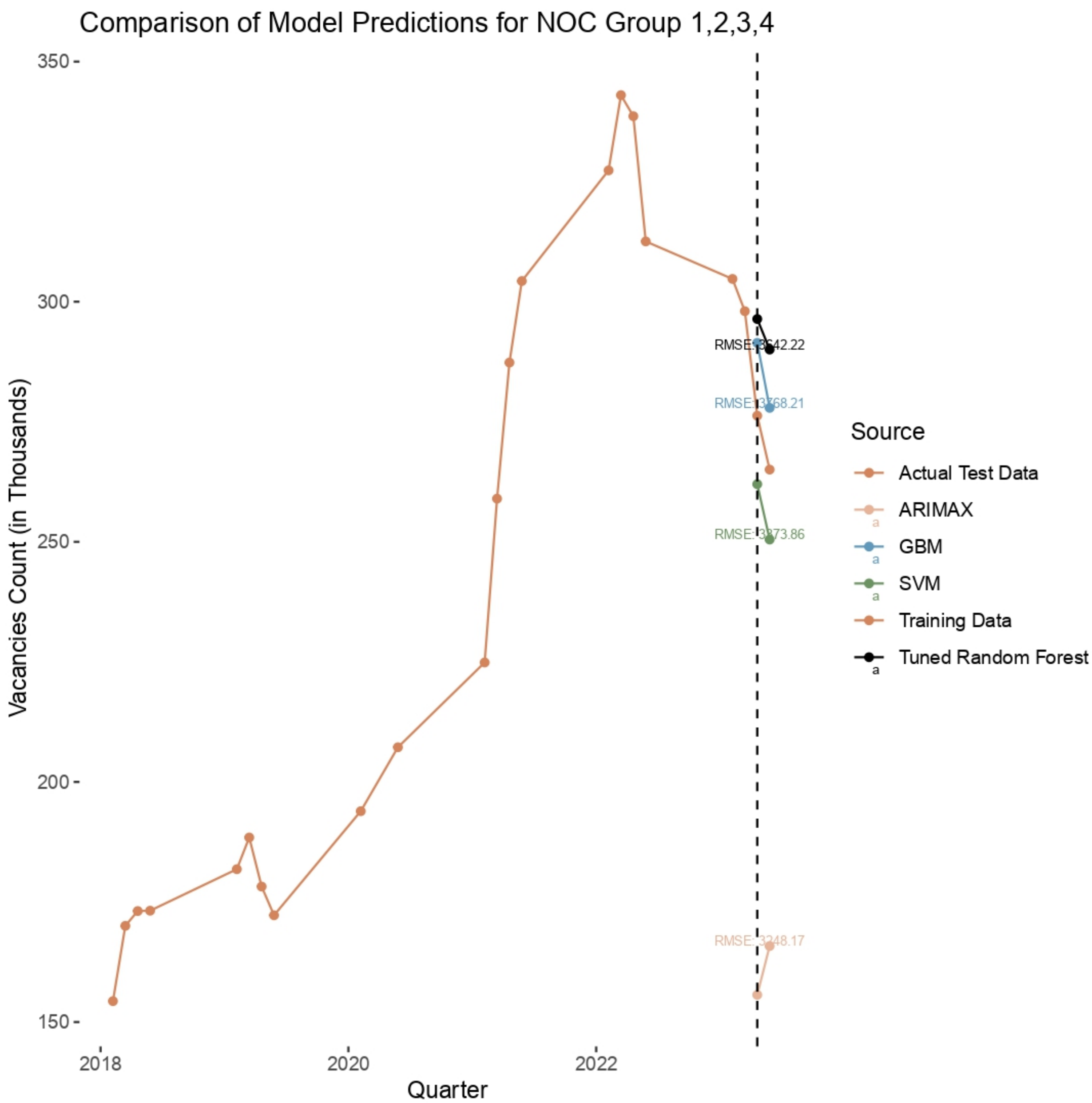


Figure 1-A.9: Comparison of model predictions (using raw VJ data) specifically focused on least represented NOC groups 5 - 9

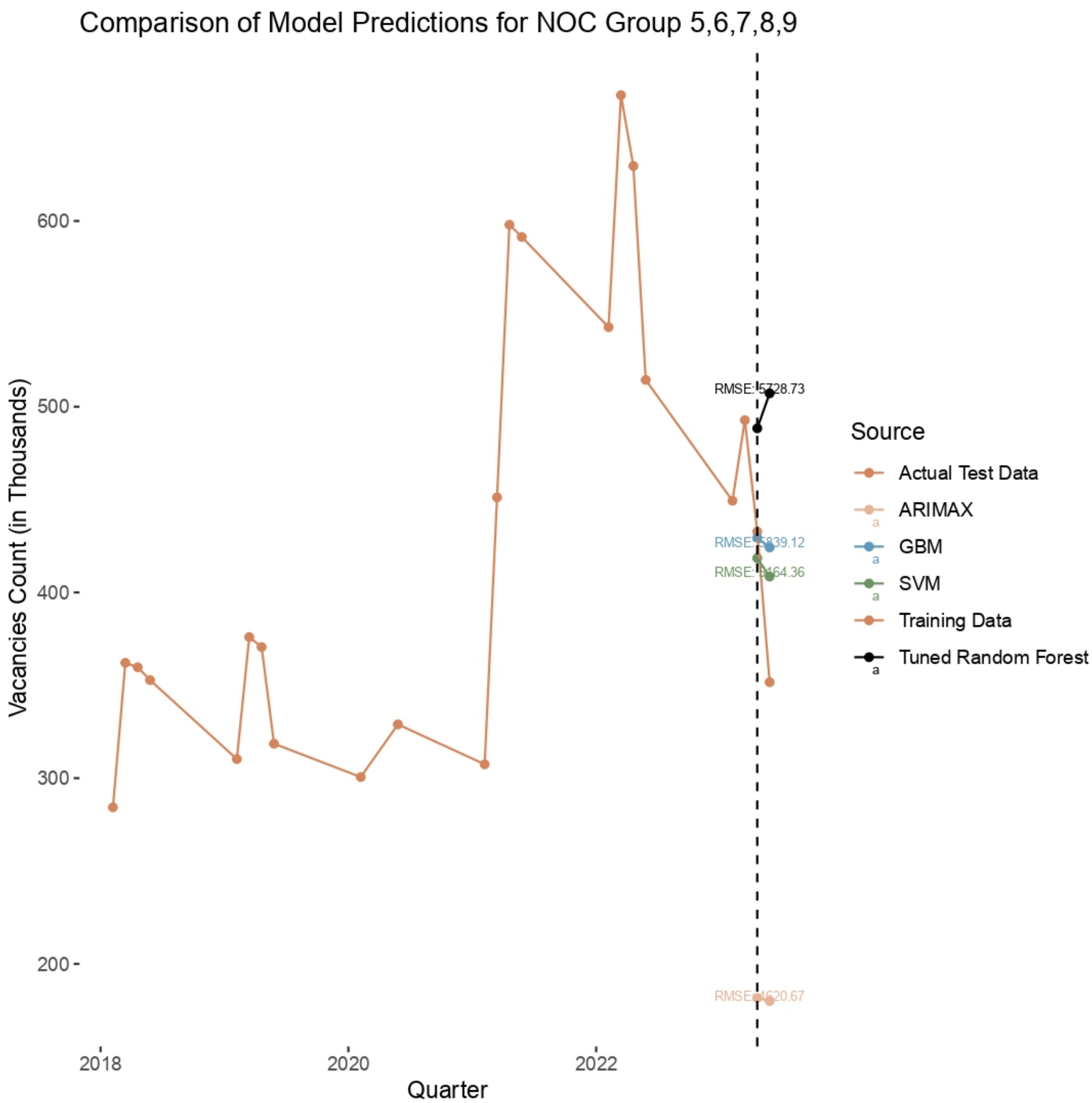


Figure 1-A.10: Winsorized-Weighted combined algorithm's absolute value errors, compared to unchanged VJ data errors

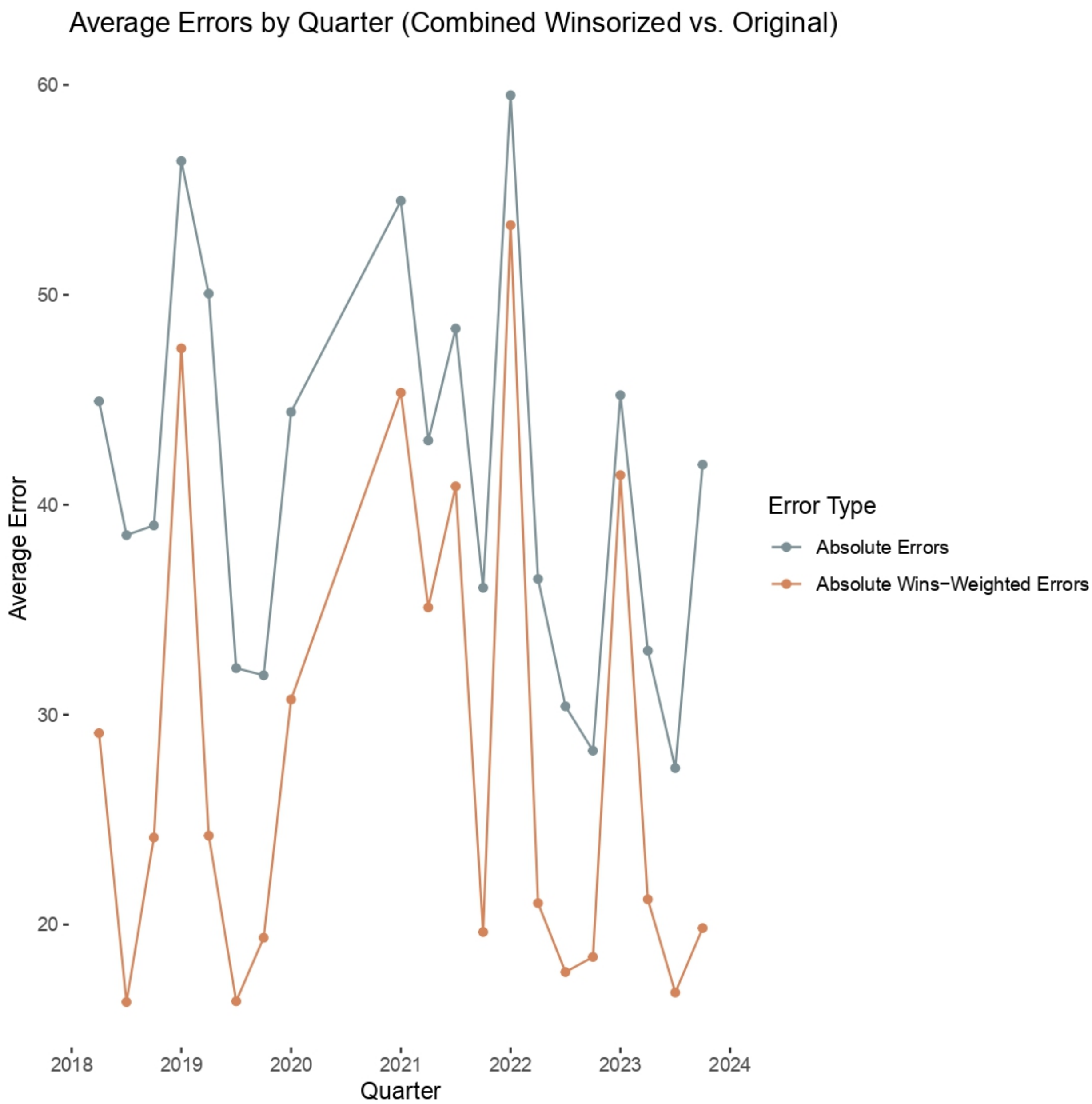
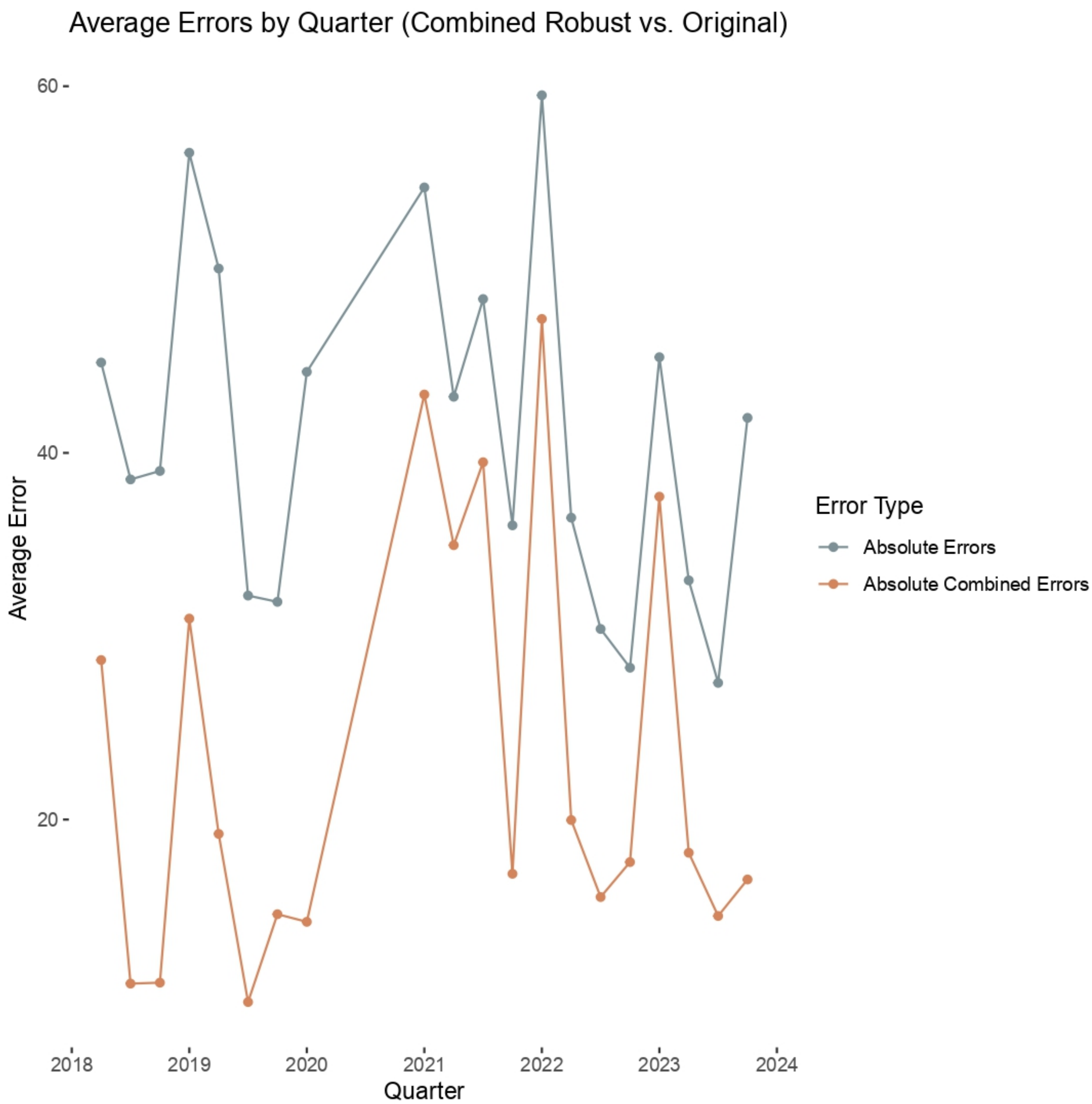


Figure 1-A.11: Winsorized-Weighted-robust combined algorithm's absolute value errors, compared to unchanged VJ data errors



Chapter 2

Using Machine Learning and Factor Analysis to study Monetary Policy Shocks in a Small Open Economy

Abstract

Monetary policy shocks are widely acknowledged to have significant effects on an economy and its macroeconomic variables. This paper uses a novel approach to study the effects of these shocks using machine learning and factor analysis in a sign-restricted Factor-Augmented Vector Autoregression (FAVAR) model, in a small open economy. The approach introduces filtered variables using a combination of Least Absolute Shrinkage and Selection Operator (LASSO) and machine learning algorithms, before creating the factor model, and using the factors in the sign-restricted FAVAR model. The findings highlight that this filtration process is effective in producing a more reliable model, compared to constructing factors without filtration, and that there are important structural breaks to the conduct of monetary policy in Canada. Using this model to produce impulse response functions, a positive shock to monetary policy in Canada leads to statistically significant downward effects on the employment and inflation factors, and on real GDP and housing prices. On the other hand, unlike in Champagne and Sekkel (2018), the effect on the exchange rate is ambiguous.

2.1 Introduction

A central challenge in empirical macroeconomics is how to extract the policy-relevant signal from large information sets. Factor-augmented VARs (FAVARs) provide a practical solution by summarizing multivariate data into a small number of latent factors. In practice, however, the standard approach of constructing factors from the entire data panel treats all series as equally informative for the policy question at hand. This indiscriminate aggregation can dilute signals that matter for monetary policy, introduce noise, and in turn produce factors and impulse responses that are difficult to interpret or that conflict with theoretical priors

To address this challenge, this paper proposes a novel methodology for selecting variables within the FAVAR framework. This is then combined with the application of sign restrictions to analyze the effects of monetary policy shocks in a small open economy. To identify the most informative variables before constructing the factor model, I employ a suite of statistical and machine learning techniques, including LASSO, Random Forests, Elastic Net, Boruta, stepwise regression, and Recursive Feature Elimination (RFE). These methods are applied systematically to select variables that are highly predictive of key targets, such as inflation and unemployment. Results from these methods are combined through a voting mechanism, which aggregates the selections to prioritize variables consistently identified as significant across techniques. I show that this results in an improved factor model, compared to a model that constructs factors from the raw data. I use the comprehensive dataset from Fortin-Gagnon et al. (2022), which provides stationary and standardized monthly macroeconomic data for Canada spanning several decades, and focus on the period from 1996 onward. This is motivated by structural breaks detected in monetary policy conduct in 1992 and 1995, which I find using a Bai-Perron multiple breakpoint estimation procedure. I show that it is crucial to account for these breaks.

After variable selection, I perform factor analysis using Principal Axis Factoring (PAF) with varimax rotation to construct latent factors that capture the underlying economic dynamics of the data. These factors, representing core dimensions such as inflation and employment activity, are then integrated into the FAVAR model to analyze the transmission mechanisms of monetary policy shocks. While it is true that central banks utilize extensive datasets when making policy decisions, and the FAVAR approach allows for including this larger information set in the analysis, I demonstrate that carefully filtering these datasets before constructing factors can enhance the performance of the FAVAR model, compared to constructing factors from the raw data. The filtration process improves Tucker Lewis scores, Root Mean Square Error, and Bayesian Information Criteria for the filtered model, compared to a factor model constructed from the full data set. This improved accuracy of factor representation leads to impulse response functions that are more aligned with expectations from macroeconomic theory, compared to the impulse responses of a standard FAVAR with factors constructed from the full data set.

FAVAR models are particularly well suited for studying small open economies, since they allow for the inclusion of a broad set of variables that influence policy transmission. In small open economies, the effects of monetary policy shocks can be particularly complex due to exposure to global shocks, commodity price fluctuations, and exchange rate dynamics. While the literature has extensively studied the transmission of monetary policy shocks in the U.S., focusing on the Federal Reserve's actions, relatively little research has been conducted on Canada's central bank. Although the Canadian economy shares some similarities with the U.S., key differences, such as the importance and effects of commodity prices, can influence how monetary policy shocks propagate. As Beine et al. (2012) show, commodity prices respond differently to monetary policy shocks in Canada than in the U.S., leading to distinct effects on their respective currencies and broader economies. Earlier studies, such as Barth and Bennett (1974), have further emphasized the importance of studying Canada's unique transmission mechanisms, revealing significant causal relationships between monetary aggregates, and changes in industrial production and GNP within the Canadian context.

A key challenge in standard FAVAR models lies in how the informational content of the data is summarized. Typically, factors are extracted from a large panel of variables using principal components, which identify directions of maximum variance. While this approach effectively summarizes broad co-movements in the data, it does not distinguish between variables based on their direct relevance to monetary policy objectives such as inflation, and employment. In this paper, I enhance this framework by introducing a targeted variable selection step prior to factor extraction. Rather than constructing factors from the full data panel, I first apply a suite of statistical and machine learning techniques to identify variables that are most closely tied to the core objectives of monetary policy, which are inflation control and maximum sustainable employment. Principal components are then computed from this refined set of variables. This filtering step focuses the factor model on policy relevant variation and I show that it improves the interpretability and performance of the resulting FAVAR, particularly in generating impulse responses that align more closely with theoretical expectations.

The results of the model highlight the important role of monetary policy in shaping real economic outcomes. A positive monetary policy shock results in a significant decline in employment, real GDP, and inflation factors. However, unlike Champagne and Sekkel (2018) who find that a positive monetary policy shock in Canada leads to a weakened Canadian dollar, I find that the effects on the value of the Canadian dollar are ambiguous, reflecting the multifaceted dynamics of currency exchange rates in an open economy. This result persists even when restricting the sample to a comparable time period. The comparison with Champagne and Sekkel (2018) is particularly relevant due to their use of a narrative identification strategy—widely seen as a robust alternative to VAR-based methods—their prominence within Canadian empirical monetary policy research, and their shared focus on identifying the macroeconomic effects of monetary policy shocks in Canada.

This paper is divided as follows: Section 2 provides a background and a review of the literature, with a focus on efforts made to identify monetary policy shocks within the Canadian context. Section 3 discusses the data and methodology. Section 4 highlights the key results and tests. Finally, Section 5 concludes.

2.2 Background and Literature Review

It is important to define monetary policy shocks in the context of this research. Christiano et al. (1999) explain that monetary policy shocks can be viewed as the variation in policy makers' responses that are unaccounted for as a reaction to the state of the economy. These shocks can then take the role of the disturbance term in an equation of the form:

$$S_t = f(\Omega_t) + \sigma_s \epsilon_t^s \tag{2.1}$$

where S_t is the policy instrument, f is a function that relates the policy instrument to the information set available (Ω_t), and $\sigma_s \epsilon_t^s$ is the monetary policy shock, with ϵ_t^s normalized to have unit variance, and σ_s being the standard deviation of the shock.

Many researchers have attempted to provide different theoretical explanations for monetary policy shocks as deviations from expectations in monetary policy. Christiano et al. (1996) explain the dichotomy in monetary policy decisions as responses to monetary and nonmonetary developments in an economy. They explain that the monetary responding aspect of the decision can be found using a feedback rule, with the main challenge of researchers being to correctly identify the assumptions on the rule. With a correctly specified rule, monetary policy shocks can then be simply obtained by taking the difference between the actions implied by the feedback rule and the actual actions taken by the central bank in a time series dataset. On the other hand, some of the most commonly discussed explanations are that these deviations represent changes in the preferences of the monetary authority. Kim et al. (2017) discuss the major differences in monetary policy between the Paul Volcker era, and the post-Volcker time, highlighting the major changes in policy that can be attributed to the Fed chairman of 1979-1987. Other potential explanations are that these shocks represent measurement errors in the data available to the monetary policy decision makers, who often rely on preliminary data at the time they make these decisions. On the other hand, Miranda-Agrippino and Ricco (2021) provide evidence of information frictions between monetary policymakers (who are assumed to have superior data) and agents in the economy. They discuss the idea that informationally constrained individuals might see an interest rate hike as either a deviation of the central bank from its monetary policy rule, or a sign that there are stronger than expected fundamentals and that the central bank is simply responding endogenously with its informationally superior data.

2.2.1 VAR Approaches

VARs with contemporaneous relationships between variables (as is typically modeled in economics) present a challenge in their estimation, particularly due to the need to develop some restrictions in order to guarantee a unique solution. This "structural" approach to VARs with different identification strategies using different assumptions is the focus of the pioneering work of Sims (1972) and Bernanke and Blinder (1992). With more variables than restrictions, the SVAR approach requires economic intuition, or assumptions from economic theory, to create more restrictions on the parameters of the model and enable the estimation of the values of the parameters in the structural model. Sims (1972) assumes that the policy variable does not respond contemporaneously to output shocks, arguing that the central bank views past data when deciding on the policy rate. On the other hand, Bernanke and Blinder (1992) assume that output does not respond contemporaneously to monetary policy shocks, and instead only responds with a delay which is represented by lags in the VAR model. Given these timing restrictions, a "recursive" approach was developed, whereby some variables included in the model are assumed to not respond contemporaneously to monetary policy shocks, and some variables are assumed to not enter the information set of the central bank when making its policy rate decision. To see this more clearly, assume the VAR variable Y_t can be partitioned into three different blocks:

$$Y_t = \begin{bmatrix} X_{1t} \\ S_t \\ X_{2t} \end{bmatrix} \tag{2.2}$$

Where X_{1t} represents variables that do not respond contemporaneously to monetary policy shocks, S_t is the policy instrument, and X_{2t} are variables that do not enter the information set of a policy maker when deciding on the policy rate. This is further developed and explained in Christiano et al. (1999) and the results of a 100bps shock can be found in Appendix Figure 2-A.1.

Another approach that yielded different results is the instrumental work of Romer and Romer (2004) (R&R), where they take a narrative approach to calculating the shocks, and find that they play a much more significant role in the economy than previously thought. They also revisit this question in Romer and Romer (2023), and continue to find significant effects for monetary policy on the economy using the latest narrative evidence. The narrative approach relies on using Federal Reserve staff forecasts (commonly referred to as the "Greenbook"), and regressing changes in the intended funds rate around forecast dates on these forecasts. The residuals of this regression, are what R&R consider to be changes that are not made directly in response to information about future economic developments, and can hence be considered monetary policy shocks.

Coibion (2012) examines and compares standard VAR shocks with those of R&R, and further develops an integrated approach whereby a standard VAR model is built with a pre-specified shock series (the R&R shocks). He compares the effects of a monetary policy shock using the standard VAR approach with that of R&R as well as with the integrated approach. The effects of monetary policy shocks turn out to be medium (neither as large as in R&R baseline models, nor as low as in standard VAR models). This can be seen in Appendix Figure 2-A.2.

On the other hand, Kuttner (2001) uses evidence from the Fed funds futures market to measure these shocks in the form of surprises, or the differences between actual policy rate changes and what was expected and already accounted for in the market. The results could also be useful in assessing shocks even when unconventional monetary policy tools are used, such as forward guidance. For example, market reactions to surprises in forward guidance notes, even when there are no changes made to the policy rate, could indicate a shock at that time. This analysis became crucial during periods of the zero low bound, where unconventional monetary policy tools become necessary when decision makers face a liquidity trap.

Many studies have used different forms of shocks and built upon the model introduced in Bernanke et al. (2005), where they develop a Factor Augmented VAR (FAVAR). This approach allows the inclusion of many variables in the model (not a limited number, as in the Structural VAR approach). The ability to include many more macroeconomic variables compared to the standard VAR models reduces the risk of omitted variables, potentially improving the results. However, Ramey (2016) highlights a shortcoming of this approach, which is that all the variables must be transformed to be stationary.

Meanwhile a sizeable line of literature has been focused on the estimation of impulse responses using a different way. Jordà (2005) developed a "local projections" method that involves dividing the input data into smaller segments, and using these segments to estimate the impulse responses in a localized manner. The method is based on the idea that the impulse response changes slowly over time, and therefore the impulse response estimated within a window is a good approximation of the impulse response over that time period. By using multiple windows, the method provides a more comprehensive estimate of the impulse response, compared to VAR models, over the entire data set (see Appendix Figure 2-A.3).

Finally, another recent influential study by Miranda-Agrippino and Ricco (2021) focuses on calculating monetary policy shocks through an instrument that accounts for non-nested information sets between the Feds and private agents. To do this, the authors first build a series of monthly market surprises using the sum of the daily series in Gürkaynak et al. (2005). They then regress these surprises on their lags (as a method to control for autocorrelation), and on the Greenbook estimates (to control for the central bank's private information). Specifically, the regression is:

$$mps_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i mps_{t-i} + \sum_{j=-1}^{3} \theta_j F_t^{cb} x_{q+j} + \sum_{j=-1}^{2} \zeta_j [F_t^{cb} x_{q+j} - F_{t-1}^{cb} x_{q+j}] + z_t \tag{2.3}$$

Where mps_t denotes the market-based monetary surprises, $F_t^{cb} x_{q+j}$ is the Greenbook forecast for quarter $q + j$ made at time t. The difference term denotes the revised forecast between two consecutive meetings. x_q includes macroeconomic variables (output, inflation, and unemployment in this case).

Using this measure of shocks (z_t from the regression), the authors then introduce a Baysian Local Projections (BLP) approach as a robust method that combines features of VARs and LPs, a choice the authors see as one that considers a trade-off between potential bias and variance. Appendix Figure 2-A.4 shows that they find that contractionary monetary policy is significantly recessionary, and find no evidence of a price puzzle, a term used to describe impulse responses that show prices increasing following monetary policy tightening. This puzzle appears in Coibion (2012) with a standard VAR.

2.2.2 The Case for Canada

There are a number of ways in which differences between the Canadian and US economies might lead to differences in monetary policy and its transmission and effect on the economy. First and foremost, the size of these two economies (as measured by GDP) is one major point of difference. 2021 World Bank Data indicates that the US GDP is around $23 trillion, whereas the Canadian GDP is around $2 trillion. Parry (1998) explains that this difference in size between economies has major implications for monetary policy. For instance, Canadian policymakers must treat foreign or global factors affecting interest rates or prices as exogenous, whereas the US retains some influence to affect global interest rates and prices due to its sheer size.

Another important difference between both economies, is the trade profiles of both countries. While the US has many global trade partners, Canada relies almost exclusively on the US for its trade (around 73% of all exports according to World Bank data). For this reason, the exchange rate has been an important indicator for Bank of Canada over the years, even after abandoning systematic intervention in the market in the 1990s. In fact, Zettelmeyer (2004) highlights instances in which Bank of Canada reacted with interest rate changes in response to exchange rate movements even after adopting inflation targeting. Ragan (2005) explains that exchange rate movements could have implications on aggregate demand in Canada. Accordingly, the Bank would determine its response potentially based on the reasons behind these fluctuations in exchange rates. Changes in global commodity prices also play a stronger role in these foreign exchange fluctuations for Canada, given that it relies more heavily on exporting fuel, minerals, and other natural resources, which reflects in potentially more action to adjust the policy rate in response to commodity price fluctuations from Bank of Canada, compared to the Federal Reserve.

Finally, the differences in the composition of the banking sector between both countries has potential implications for the transmission of monetary policy, especially through the credit channel as discussed in Bernanke and Blinder (1988). More specifically, Juurikkala et al. (2011) provide an explanation for the bank lending channel, and use an empirical model to highlight the fact that differences in bank sizes and numbers can lead to different potential amplification effects in the transmission of monetary policy.

Recognizing all these important differences, Champagne and Sekkel (2018) use similar narrative evidence approach (to R&R) to construct a measure of monetary policy shocks for Canada, by comparing the forecasts from Bank of Canada staff projections for the period between 1974-2015 with the actual policy rate data. Alstadheim et al. (2013) discuss the importance of the regime switch that the Bank of Canada implemented in the 1990s, where the Bank started inflation targeting and abandoned systematic intervention in the foreign exchange market. Champagne and Sekkel (2018) take this into account (see Appendix Figure 2-A.5) as a structural break in the data, and focus their analysis on the period following inflation targeting.

On the other hand, Claus and Dungey (2016) examine yield curve movements (rotations and shifts) in Canada to test the validity of different hypotheses on the reaction of the term structure to monetary policy surprises. They use a theoretical macroeconomic model to test the effects of monetary policy actions on different yields. They find that despite efforts by central banks to improve transparency and to communicate clearly to the pubirlic, abrupt changes in yield curves following monetary policy announcements seem to indicate that many surprises still exist in the data. With a focus on different potential identification strategies for these policy surprises (using the Rigobon and Sack (2004) framework for heteroskedastic identification, or the Thornton (2014) homoskedastic correction for news bias approach), the authors estimate the impact of monetary policy surprises on the term structure. While they do not use this measure as a monetary policy shock in a model to calculate the impulse responses of other macroeconomic variables, their measure can nonetheless potentially be used for this purpose.

Finally, Koeppl et al. (2024) provide a novel approach to identifying monetary policy shocks in Canada by utilizing changes in the Nelson-Siegel yield curve factors on policy announcement dates, as well as during speeches and other communications from the Bank of Canada. Rather than focusing solely on short-term policy rates, they examine the entire yield curve. Their findings demonstrate that monetary policy shocks, while effective in influencing inflation and GDP, often lead to twists in the yield curve, with varying impacts across maturities. Importantly, they also highlight the role of communication, including speeches by senior Bank officials, in shaping expectations and amplifying the effects of monetary policy. Their work provides further evidence that the transmission mechanisms of monetary policy in Canada are distinct, reflecting the Canada's unique economic structure and reliance on external factors.

2.3 Data and Methodology

This section discusses the dataset, including some of its key features and descriptive statistics, and an explanation of the methodology used. This includes the FAVAR model with sign restrictions, and an explanation of the filtration process for variables developed for the factor model.

2.3.1 Data

The analysis in this paper uses the comprehensive Canadian macroeconomic database developed by Fortin-Gagnon et al. (2022). This dataset includes hundreds of stationary and standardized monthly economic indicators, covering a wide range of economic activities such as production, labour, housing, trade, financial flows, prices, and stock markets. The dataset spans several decades and provides a balanced panel suitable for analyzing macroeconomic fluctuations. Importantly, it accounts for structural adjustments, seasonal variations, and the integration of older and newer series. Seasonality is addressed using the SEATS model-based decomposition method developed by the US Census Bureau, and checked using the Kruskal-Wallis test (Kruskal and Wallis, 1952). Missing data was filled using an expectation-maximization algorithm as in Stock and Watson (2002) and McCracken and Ng (2016).

Figure 3-2 shows some of the main variables in this dataset. The dataset is stationary by design, and stationarity is confirmed in Appendix Table 2-A.1 using Augmented Dickey-Fuller (ADF) and Phillips Peron (PP) tests.

Figure 2-1: Combined time series of some of the main variables from the dataset from 1992 to 2024

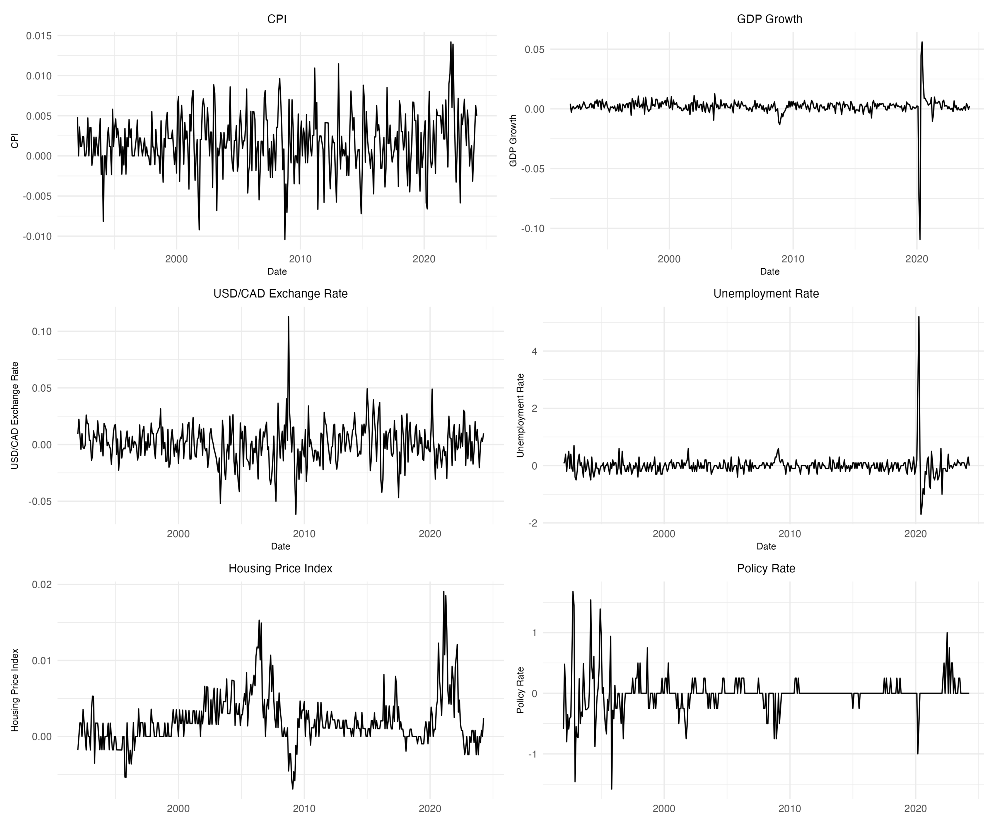


Table 2-1 provides a summary statistics for some of the factors computed from this data set. These are initial latent factors that are calculated using PAF from the entire data set, and are based on the descriptions in Fortin-Gagnon et al. (2022).

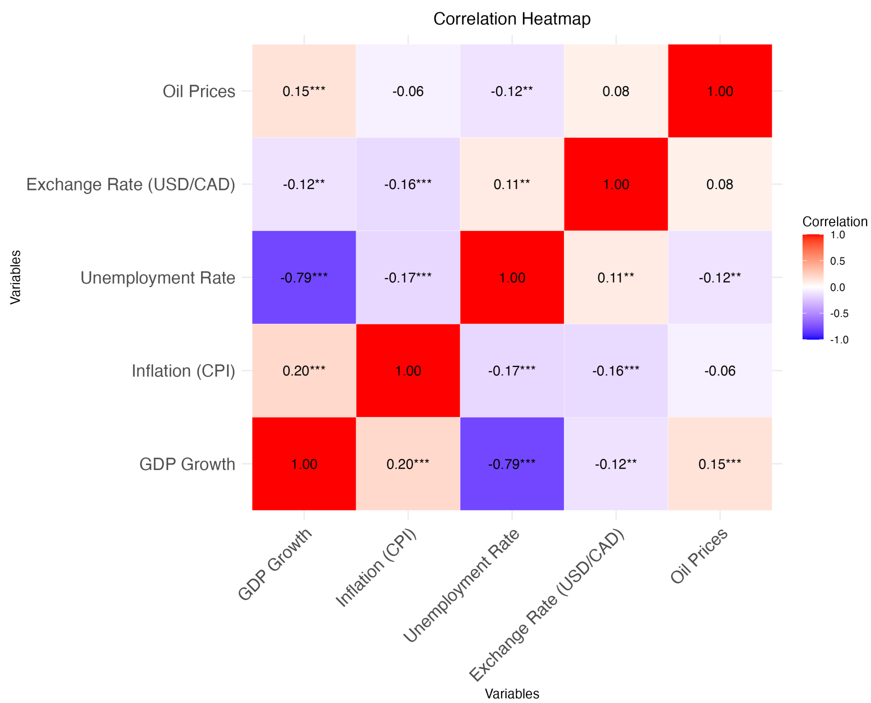
Table 2-1: Summary Statistics for factors produced from the data.

Factor	Mean	SD	Median
Production	0.303	0.079	0.002
Interest Rates	0.077	0.381	-0.002
Housing	0.016	0.112	0.003
Commodities	0.018	0.092	0.003
Logistics	0.002	0.148	0.007
Inflation	0.141	0.159	0.003

Figure 2-2 presents a heatmap of pairwise correlations among selected key macroeconomic indicators (e.g., GDP growth, inflation, and unemployment), offering a snapshot of the interrelationships between various dimensions of the economy. GDP growth exhibits a strong negative correlation with the unemployment rate, consistent with standard macroeconomic theory linking strong output performance to lower unemployment. This includes Okun's Law, which posits that higher economic output is typically associated with lower levels of unemployment. Inflation shows a moderate positive correlation with GDP growth, aligning with the notion that periods of robust economic activity can place upward pressure on prices and reflecting the dynamics associated with the Phillips curve, which links inflationary pressures to economic activity and labor market tightness.

In contrast, variables such as the exchange rate and oil prices display more modest relationships with domestic indicators, suggesting that these external or commodity-related shocks may influence the economy, but not as strongly as core macroeconomic fundamentals. This correlation structure reinforces the expectation that while these economic indicators are interrelated, they capture distinct facets of economic behavior and can thus serve as valuable inputs in factor-based modeling approaches to disentangle the key drivers of economic dynamics and understand the propagation of shocks across the economy.

Figure 2-2: Heatmap of correlations among select macroeconomic variables. Correlation test results in brackets. *** indicates significance at the 99% confidence level, ** at the 95% level, and * at the 90% level.



Over the past three decades, both inflation and GDP growth in Canada have remained relatively stable, hovering around modest positive and negative values. This steady pattern persisted until the onset of the COVID-19 pandemic, when GDP growth experienced a sharp plunge followed by a rapid recovery (see Figure 2-3).

Similarly, rolling 12-month standard deviations reveal that volatility has remained subdued throughout the pre-pandemic era. Only during the COVID-19 surge did economic turbulence intensify, as supply chain disruptions and government-mandated lockdowns, and subsequent easing measures caused a sharp but temporary spike in GDP growth volatility (Figure 2-4).

Figure 2-3: Time series showing GDP growth and CPI over the past 3 decades. Grey areas show potentially large shock events, such as the 2008 financial crisis and the COVID-19 pandemic.

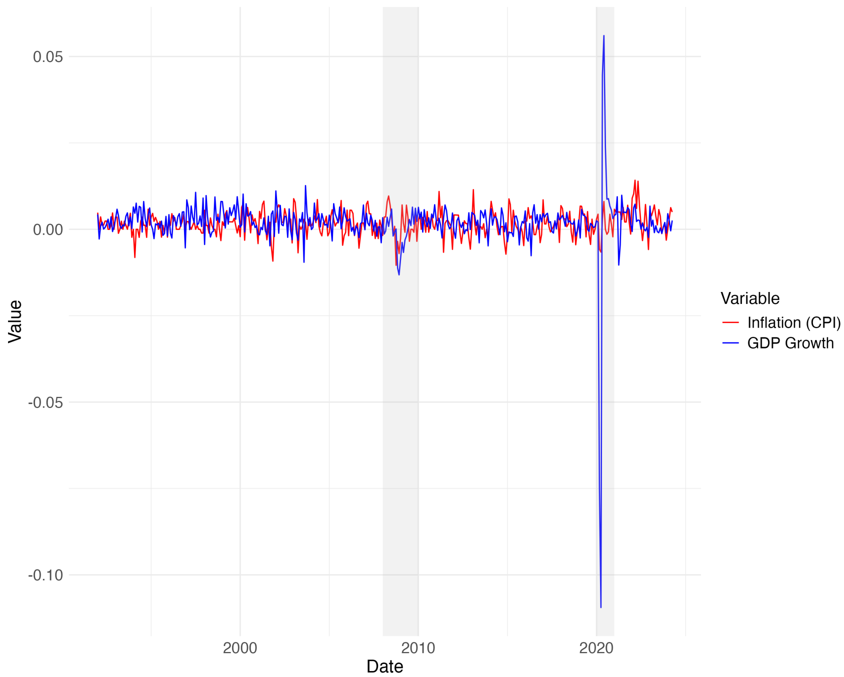
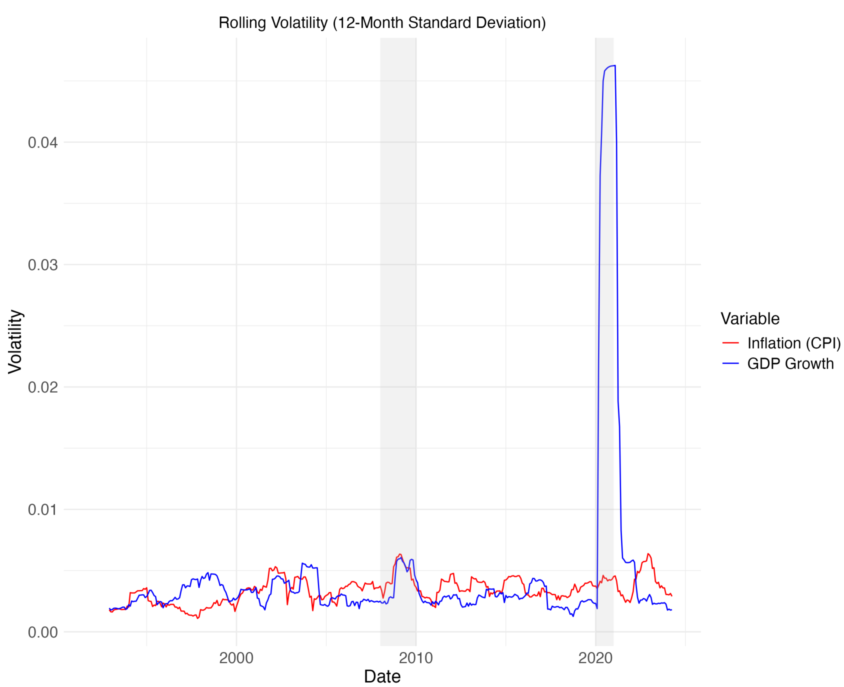


Figure 2-4: 12-month rolling standard deviation for GDP growth and CPI. Grey areas show potentially large shock events, such as the 2008 financial crisis and the COVID-19 pandemic.



However, despite the COVID-19 pandemic being a considerable shock to GDP growth and the overall economy, there seems to be little evidence of a structural break in the conduct of monetary policy at the time. Instead, there are two structural breaks that can be detected in the time series, one in 1992 and the other in 1995. These two breaks were detected using the Bai-Perron multiple breakpoint estimation procedure, which partitions the time series into distinct regimes. In the baseline scenario, consider a time series $\{y_t\}_{t=1}^{T}$ and a simple linear model

$$y_t = \beta_0 + \beta_1 t + \epsilon_t, \tag{2.4}$$

where t is time, β_0 and β_1 are parameters, and ϵ_t is a stochastic error term.

The multiple breakpoint method finds breakpoints $T_1, T_2, \ldots, T_m$ at which the parameters of the model may change. After introducing these breakpoints, the model can be expressed as piecewise linear segments:

$$y_t = \begin{cases} \beta_{0,1} + \beta_{1,1} t + \epsilon_t & \text{for } t = 1, \ldots, T_1 \\ \beta_{0,2} + \beta_{1,2} t + \epsilon_t & \text{for } t = T_1 + 1, \ldots, T_2 \\ \vdots & \\ \beta_{0,m+1} + \beta_{1,m+1} t + \epsilon_t & \text{for } t = T_m + 1, \ldots, T. \end{cases} \tag{2.5}$$

Here, each segment defined by (T_j, T_{j+1}) has its own set of parameters $(\beta_{0,j+1}, \beta_{1,j+1})$. The estimation procedure selects both the number and positions of these breakpoints to minimize the overall residual variation while ensuring each regime contains a sufficient number of observations. In doing so, the fitted model segments and corresponding breakpoints, such as those given by (2.5), highlight structural changes in the underlying data-generating process that are not evident in a single-regime model like (2.4).

The first break appears to coincide with the initial years of formal inflation targeting and the emergence of the economy from the 1990-1992 recession. This is similar to Champagne and Sekkel (2018), who use a Chow test to find a significant structural break during inflation targeting. Although the Bank of Canada introduced inflation targeting in 1991, it was not until 1992 that inflation targets were firmly established, and the Bank began refining its focus on core inflation measures. This may have been reflected in operational changes, such as adjustments in interest rate policy and communication strategies, responding to a slowly improving domestic economy and a gradually recovering global environment. By the mid-

1990s, the second structural break may be linked to Canada's significant fiscal consolidation in 1995, when drastic measures aimed at reducing the country's large debt and deficits alleviated pressures on monetary policy. As the government brought public finances under control, the Bank of Canada could shift its attention more squarely to sustaining low inflation, no longer needing to compensate for fiscal imbalances with higher interest rates. Moreover, global financial turbulence in the mid-1990s (for example, the fallout from the Mexican crisis) could have led to monetary policy adjustments aimed at stabilizing the exchange rate. Taken together, these structural breaks in 1992 and 1995 likely reflect both the lagged effects of implementing a new policy regime and the evolving interplay of domestic fiscal conditions, global economic conditions, and maturing inflation-targeting practices. I show that ignoring these breaks produces a weaker factor model, and the results supporting this claim are presented in Section 4.

2.3.2 Methodology

A key contribution of this paper is the systematic process used to select variables for the FAVAR model. This involves integrating multiple methods and then applying a voting mechanism to identify variables that are consistently selected across these methods. By aggregating the strengths of different approaches, this process reduces the risk of selecting variables based on any single method's inherent biases.

I begin with a dataset of N macroeconomic indicators, denoted as

$$\mathbb{X}_t \equiv \{X_{1,t}, X_{2,t}, \ldots, X_{N,t}\} \quad \text{for } t = 1, \ldots, T \tag{2.6}$$

and a set of target variables Y_t, which are inflation and unemployment. The goal is to select a subset of $\mathbb{X}_t$ that provides the most predictive power for Y_t. To achieve this, I implement several distinct selection methods. These include regularization-based regressions such as the LASSO,

$$\min_{\{\beta_i\}} \sum_{t=1}^{T} \left(Y_t - \beta_0 - \sum_{i=1}^{N} \beta_i X_{i,t} \right)^2 + \lambda \sum_{i=1}^{N} |\beta_i| \tag{2.7}$$

where β_0 is the intercept term, β_i represents the regression coefficients for each predictor $X_{i,t}$, and $\lambda > 0$ is the regularization parameter that controls the degree of sparsity by penalizing the absolute values of the coefficients, $\sum_{i=1}^{N} |\beta_i|$. This penalty has the effect of shrinking some coefficients β_i toward zero, thereby reducing their influence or effectively excluding the corresponding predictors from the model.

Elastic Net combines both LASSO and Ridge penalties to balance between sparsity and coefficient shrinkage. The optimization problem is given by:

$$\min_{\{\beta_i\}} \sum_{t=1}^{T}(Y_t - \beta_0 - \sum_{i=1}^{N} \beta_i X_{i,t})^2 + \alpha\lambda \sum_{i=1}^{N} |\beta_i| + (1-\alpha)\lambda \sum_{i=1}^{N} \beta_i^2 \tag{2.8}$$

where α is the mixing parameter that determines the weight between the LASSO penalty ($\alpha\lambda$) and the Ridge penalty ($(1-\alpha)\lambda$). By adjusting α, the Elastic Net can promote sparsity while also handling multicollinearity among predictors. These methods impose penalties on the regression coefficients to promote sparsity and retain only the most relevant predictors. Relevance is defined as the degree to which a predictor $X_{i,t}$ contributes to explaining the variance in the target variable Y_t. This is evaluated based on several criteria: the magnitude of the regression coefficients, predictive power, and statistical significance.

I also incorporate a Random Forest model, an ensemble method that captures nonlinearities and interactions among predictors. It operates by constructing multiple decision trees during training, each built on a bootstrap sample of the data. At each split in a decision tree, Random Forest randomly selects a subset of predictors to consider. The final prediction is obtained by averaging the outcomes of all the individual trees. I train a Random Forest model for each target variable, specifying the target as a function of all available predictors in the dataset. I then rank the predictors by their importance scores. Predictor importance is quantified based on the reduction in node impurity achieved by splits involving each variable across the ensemble of trees. Node impurity is measured by the reduction in variance when a split is made on a given variable. Specifically, each split on a predictor results in a reduction in variance at that node, and the total reduction is aggregated across all trees for that predictor. The resulting value represents the importance score, where a higher score indicates a greater contribution to improving the model's predictive accuracy. To determine the appropriate threshold for selecting predictors, I consider the top X variables based on their importance scores, where X is determined iteratively. For each value of X, I construct a factor model using the selected predictors and evaluate the resulting model using information criteria, such as the Bayesian Information Criterion (BIC). The threshold X that results in the best factor model, as indicated by the lowest information criteria values, is selected as the cutoff.

The Boruta algorithm refines this process further by comparing the importance of each predictor to that of deliberately permuted "shadow" variables, which have no real association with the target variable. Each original predictor's importance score is compared to the highest importance score among the shadow variables. Only predictors with importance scores significantly higher than those of the shadow variables are retained, as measured by Z-test scores. Variables with similar or lower scores are excluded.

To complement these approaches, I use Stepwise Regression, which iteratively refines the predictor set based on BIC scores. Beginning with a full model that includes all available predictors for a given target variable, the algorithm evaluates the contribution of each variable by considering its removal. A variable is removed from the model if its exclusion reduces the BIC, indicating that its presence does not significantly enhance the model fit. Conversely, the algorithm also evaluates variables excluded at earlier steps and considers adding them back if their inclusion reduces the BIC. By iterating through these steps, Stepwise Regression systematically eliminates redundant predictors while retaining those that substantively contribute to explaining the target variable, as measured by the information criteria.

In addition to Stepwise Regression, I employ Recursive Feature Elimination (RFE). This process begins with the full set of predictors and iteratively reduces the feature set. At each step, the model is retrained on the remaining predictors, and importance scores are recalculated based on their contribution to the model's predictive accuracy. The elimination process continues until the feature set is reduced to an optimal subset, determined based on model performance. To evaluate model performance during RFE, I use ten-fold cross-validation. In this approach, the dataset is divided into ten approximately equal-sized subsets, or "folds." The model is trained on nine folds and tested on the remaining fold, repeating this process ten times so that each fold is used exactly once as the test set. The performance metrics from these iterations are averaged to provide a reliable overall assessment of the model. This ensures the model is evaluated on diverse subsets of the data, reducing the influence of any single unrepresentative train-test split.

From these various methods, multiple subsets of selected variables emerge. From these various methods, multiple subsets of selected variables emerge. Let $S_1, S_2, S_3, S_4, S_5, S_6$ denote the sets of variables identified by LASSO, Elastic Net, Random Forest, Boruta, Stepwise, and Recursive Feature Elimination (RFE), respectively. For every candidate variable X_i, I compute a vote count:

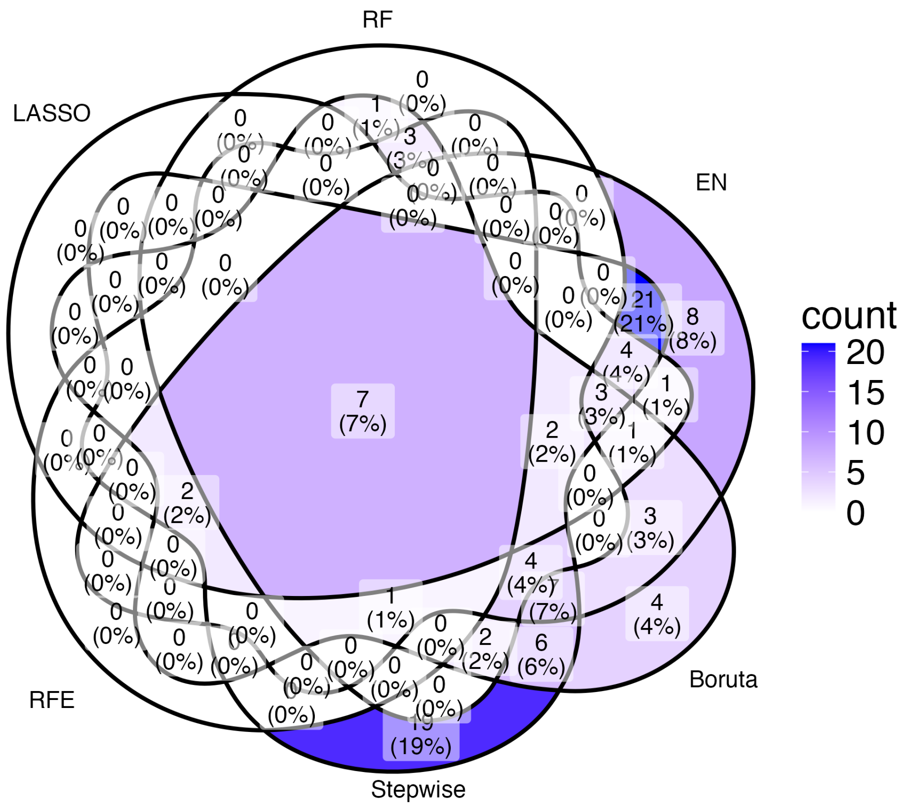
$$V_i = \sum_{j=1}^{6} I(X_i \in S_j) \tag{2.9}$$

where $I(\cdot)$ is an indicator function that equals one if X_i appears in the j-th subset.

Variables with higher vote counts are chosen more frequently across methods, reflecting their relevance. By ranking variables based on V_i and selecting the top M (where M denotes the number of variables chosen), I obtain a subset of predictors with broad support across multiple selection criteria. The choice of M is not fixed a priori. Instead, I search over various values of M and different numbers of factors to be extracted later. For each potential M, I select the top M variables and then perform factor extraction using Principal Axis Factoring followed by a Varimax rotation, which is an orthogonal rotation that ensures extracted factors are orthogonal and interpretable (see Osborne (2015) for a more detailed explanation). Figure 2-5 shows a Venn diagram of the different methods and the number of variables selected by each method. An alternative way to see this is also available in Appendix Figure 2-A.6 From this method, $M = 20$ emerges as the optimal choice. There is a total of 7 variables that were selected by all methods, whereas some variables were selected by one or more methods combined.

I then evaluate the resulting factor model using a range of criteria, such as the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA), and the Bayesian Information Criterion (BIC). I also check for pathological estimation issues, such as Heywood cases which occur when the model results imply that more than 100% of the variance of a variable is explained by the factors, which is theoretically impossible and can occur when the choice of the number of factors is inappropriate. These cases are then excluded when exploring multiple M values and factor dimensions.

Figure 2-5: Venn Diagram for variables selected by different methods. Subsets around the edges represent sets of variables that are only selected by a single method. Subsets closer to the center represent choices by several methods, and the center represents variables chosen by all methods.



Once the final set of M variables is determined, I estimate a factor model

$$W_t = \Lambda F_t + u_t \tag{2.10}$$

where W_t is the M-dimensional vector of selected variables at time t, Λ is an $M \times r$ loading matrix, $\mathbf{F}_t$ is an r-dimensional vector of latent factors, and $\mathbf{u}_t$ represents idiosyncratic noise. Tables 2-2 and 2-3 show the factors extracted using this method, and the main contributing variables with their factor loadings. Employment, energy and oil markets, inflation, housing and real estate related variables are all selected as crucial variables for the Canadian economy. These factors then serve as inputs to the sign-restircted FAVAR model.

Table 2-2: Factor Analysis Results: Part 1 - 8 Factors and Contributing Variables. This shows the factors that were selected from the factor model constructed from the data after applying the selection function and filtering. The table also shows the main variables contributing to each factor as well as their loadings.

Factor	**Main Contributing Variables (Factor Loadings)**
Employment and Economic Activity (PA1)	Total Employment (0.96), Service Employment (0.95), Part-time Employment (0.81), Sales Index (0.86), Transportation Imports (0.80)
Energy Prices and Oil Markets (PA5)	Oil Price Index (0.61), Global Oil Benchmark (0.89)
Durable Goods Consumption (PA6)	Clothing Prices (0.93), Durable Goods Prices (0.87)
CPI Services (PA4)	Service Sector Prices (0.96)

Table 2-3: Factor Analysis Results: Part 2 - 8 Factors and Contributing Variables. This shows the factors that were selected from the factor model constructed from the data after applying the selection function and filtering. The table also shows the main variables contributing to each factor as well as their loadings.

Factor	**Main Contributing Variables (Factor Loadings)**
Housing and Real Estate (PA8)	Shelter Costs (0.51), Money Supply (-0.14), Long-Term Unemployment (0.59)
Inflation Dynamics (PA2)	Core Inflation Excluding Food (0.70), Core Inflation Excluding Energy (0.55), General Goods Prices (0.67)
Unemployment Duration (PA7)	Short-Term Unemployment (-0.78), Medium-Term Unemployment (-0.61)
Banking and Mortgage Rates (PA3)	Policy Interest Rate (0.17), Mortgage Credit (0.39)

FAVAR Model After obtaining the set of factors, I include the employment, inflation, housing, and oil factors, along with key observed macroeconomic variables (real GDP growth, USD/CAD exchange rate, and the central bank policy rate), into a Factor-Augmented Vector Autoregressive (FAVAR) model. Let $\mathbf{Y}_t$ denote the set of m observed macroeconomic indicators. Stacking the r factors $\mathbf{F}_t$ and the m observed variables $\mathbf{Y}_t$ produces a combined $(r+m)$-dimensional vector:

$$\mathbf{Z}_t = \begin{pmatrix} \mathbf{F}_t \\ \mathbf{Y}_t \end{pmatrix}. \tag{2.11}$$

The first step is to then estimate a reduced-form VAR(p) model:

$$\mathbf{Z}_t = \mathbf{A}_0 + \mathbf{A}_1\mathbf{Z}_{t-1} + \cdots + \mathbf{A}_p\mathbf{Z}_{t-p} + \mathbf{e}_t \tag{2.12}$$

where $\mathbf{A}_0$ is a vector of intercepts, $\mathbf{A}_i$ are $(r+m) \times (r+m)$ coefficient matrices for $i = 1, \dots, p$, and $\mathbf{e}_t$ is a $(r+m)$-dimensional vector of reduced-form errors. Data transformations ensure stationarity, and information criteria guide the choice of p, as will be shown in the Tests Section 4.1. Estimating this reduced-form VAR by Ordinary Least Squares (OLS) yields the residual covariance matrix

$$\hat{\Sigma} = \frac{1}{T-p}\sum_t \hat{\mathbf{e}}_t\hat{\mathbf{e}}_t' \tag{2.13}$$

where $\hat{\mathbf{e}}_t$ are the estimated residuals. The covariance matrix $\hat{\Sigma}$ captures the relationships between the reduced-form errors. To link these errors to structural shocks, this matrix is factorized using a Cholesky decomposition as:

$$\hat{\Sigma} = \hat{C}\hat{C}' \tag{2.14}$$

where $\hat{C}$ is a lower triangular matrix that provides an initial decomposition of $\hat{\Sigma}$. Then orthonormal rotation matrices, denoted as $\mathbf{Q}$, are applied to $\hat{C}$. An orthonormal matrix $\mathbf{Q}$ satisfies the property:

$$\mathbf{Q}\mathbf{Q}' = \mathbf{I} \tag{2.15}$$

where $\mathbf{I}$ is the identity matrix. The orthonormal property ensures that $\mathbf{Q}$ does not distort the variance-covariance structure of the residuals, while allowing for alternative representations of the relationships between variables.

By applying these rotations, a candidate impact matrix is generated as:

$$\mathbf{P} = \hat{C}\mathbf{Q}. \tag{2.16}$$

The matrix $\mathbf{P}$ describes how structural shocks, such as the monetary policy shock, propagate through the system, where each column of $\mathbf{P}$ corresponds to the impact of a specific structural shock on all variables in the model. Among the randomly generated rotations, only those that satisfy the specified sign patterns on impact are retained, yielding a set of admissible structural models. The sign restrictions applied to a monetary policy shock can be summarized in Table 2-4, and are in line with Uhlig (2005), with the employment factor also specified to react negatively. This follows from the Phillips curve, where contractionary monetary policy leads to lower labour demand. This is also supported by empirical evidence, see for example Champagne and Sekkel (2018), and Koeppl et al. (2024).

Table 2-4: Sign Restrictions for Monetary Policy Shock. Interest rate reacts positively to monetary policy shock, while inflation and employment react negatively

Variable	**Reaction to Shock**
Interest Rates	$+$
Inflation	$-$
Employment	$-$

For each admissible model, I compute impulse response functions (IRFs) that trace out the dynamic responses of both the observed variables and the latent factors to the monetary policy shock.

2.4 Tests and Results

This section highlights the main tests and results, including a discussion on optimal fit and the impulse response functions generated from the sign restricted FAVAR model.

2.4.1 Tests

There are two main sets of tests conducted. The first is to assess the appropriate number of lags to use in the VAR model. The second is to ensure and verify that the key variables in the dataset are stationary, which is a critical assumption in VAR analyses.

Number of Lags in the Model

I start by comparing several models with different number of lags and choosing the optimal model based on different information criteria. First I look at the Akaike Information Criterion (AIC) test. The AIC is one of the most commonly used statistical measures for model selection in the context of parametric models. It is designed to find a balance between the goodness of fit of a model and its complexity, penalizing overfitted models that are complex without offering more relevant information, and instead favoring simplicity in explaining the data.

AIC takes into account both likelihood (how well a model fits the observed data), and the number of parameters used in the model. The formula for AIC within the context of the VAR model is:

$$AIC(n) = \ln\det \sum_u (n) + \frac{2}{T} n K^2 \tag{2.17}$$

where $\sum_u (n)$ represents the covariance matrix of the residuals, T is the sample size, K is the number of variables in the model, and n is the lag length

Models are ranked and compared according to their value of AIC, with a lower AIC value indicating a relatively (compared to other models) more favorable trade-off between model fit and complexity.

Next, I consider the Hannan-Quinn Information Criterion (HQ), introduced by Hannan and Quinn (1979). HQ modifies the AIC by applying a stronger penalty for model complexity, particularly for small sample sizes. Therefore, in general HQ information criterion tends to favor simpler models compared to AIC. The formula for HQ is:

$$HQ(n) = \ln\det \sum_u (n) + \frac{2\ln\ln T}{T} n K^2 \tag{2.18}$$

The third test is the Schwarz Criterion (SC), also known as the Bayesian Information Criterion (BIC). Proposed by Zellner (1962), SC is particularly useful for macroeconomic data, as noted by Stock and Watson (2016). SC imposes an even stronger penalty on model complexity than HQ, favoring simpler models. Its formula is:

$$SC(n) = \ln\det \sum_u (n) + \frac{\ln T}{T} n K^2.$$

.

Finally, the Final Prediction Error Criterion (FPE), developed by Akaike, evaluates the model's predictive accuracy on simulated data. The FPE formula is:

$$FPE(n) = \left(\frac{T + n^*}{T - n^*}\right)^K \ln \det \sum_u (n) \tag{2.19}$$

where n^* is the total number of parameters in each equation.

The FPE criterion identifies five lags as the optimal choice. However, three out of the four criteria (AIC, HQ and SC) independently identify the six-lag model as optimal for this analysis, which is therefore selected as the preferred specification for this analysis.

Unit Root Tests

Stationarity can be crucial to obtain meaningful and reliable results, especially within the context of VAR models. Several studies on non-stationary time series show that it can lead to spurious results, which is documented in the early work of Granger and Newbold (1974). In order to ensure that the data is stationary, I run two main commonly used time series tests. First, I use the Augmented Dickey Fuller (ADF) test. The null hypothesis of the ADF test is that the time series being tested contains a unit root (non-stationary). The ADF test estimates the following regression model:

$$\Delta y_t = \rho y_{t-1} + \sum(\beta_i \Delta y_{t-1}) + \epsilon_t \tag{2.20}$$

From the test, I obtain the test statistics to be compared with the critical values from the Dickey-Fuller distribution.

Then, I use the Phillips-Peron (PP) test. It deals with the autocorrelation in the error terms in a different way, compared to the ADF test. Instead of adding lagged difference terms to the test regression and relying on information criteria to choose the number of lags, the PP test uses a nonparametric statistical method to estimate the test regression and correct for autocorrelation in the error terms. This implies that it does not rely on a specific model for the error terms. This makes it more robust to certain forms of autocorrelation. The main regression of the PP test is:

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \epsilon_t \tag{2.21}$$

Where y_t is the variable being tested, α, β, γ are parameters to be estimated, and ϵ_t is the error term.

The results of both tests are in Appendix Table 2-A.1, and confirm the same conclusion that there is no evidence of unit roots, all variables are stationary.

2.4.2 Results

There are four candidate models to consider based on two key dimensions: whether the factors are constructed from raw data or the filtered subset, and whether the full dataset or post-1995 data is used. The latter consideration accounts for structural breaks identified through the Bai-Perron test. Tables 2-5 and 2-6 summarize the key attributes of these models, including the number of factors, filtered variables, main factors identified, and key variables in each factor.

Table 2-5: Summary of factor models constructed from the full data, and from the filtered set - Using the full time series.

Model Details	**Raw Full Series**	**Filtered Full Series**
Number of Factors	3	7
Number of Filtered Variables	-	20
Main Factors (Explained Variance)	Economic Growth (18%), Labor Dynamics (7%), Price Stability (7%)	Employment Trends (28%), Consumer Prices (9%), Long-term Inflation Expectations (8%)
Key Variables in Factors	GDP growth, Unemployment rate, CPI	Employment rate, CPI goods, CPI excluding food

Table 2-6: Summary of factor models constructed from the full data, and from the filtered set - Using the 1996-onward time series to account for structural breaks

Model Details	**Raw 1996-onward**	**Filtered 1996-onward**
Number of Factors	7	8
Number of Filtered Variables	-	20
Main Factors (Explained Variance)	Economic Growth (18%), Industrial Output (5%), Labor Dynamics (5%)	Employment Trends (27%), Energy Prices (8%), Inflation (7%)
Key Variables in Factors	GDP growth, SPI, Unemployment rate	Employment rate, WTI petroleum price, CPI goods

These four optimal models are then compared among each other using several goodness-of-fit measures and information criteria. Figure 2-6 shows that the model that constructs factors using the filtered set and considers the structural breaks detected in the data demonstrates the highest Tucker-Lewis Index (TLI). The TLI measures model fit by comparing the chi-square statistic of the model to a null model, penalizing for model complexity. Higher TLI values indicate better model fit:

$$\mathrm{TLI} = 1 - \frac{\chi^2/\mathrm{df}_{\mathrm{model}}}{\chi^2/\mathrm{df}_{\mathrm{null}}} \tag{2.22}$$

where χ^2 is the chi-square statistic and df represents the degrees of freedom. The filtered 1996 model achieves a TLI of 0.921, outperforming all other optimal models.

Figure 2-6: Tucker Lewis Comparison of the four optimal factor models selected. This shows that the filtered model that accounts for the detected structural breaks, including data from 1996 onwards, has the highest TLI score.

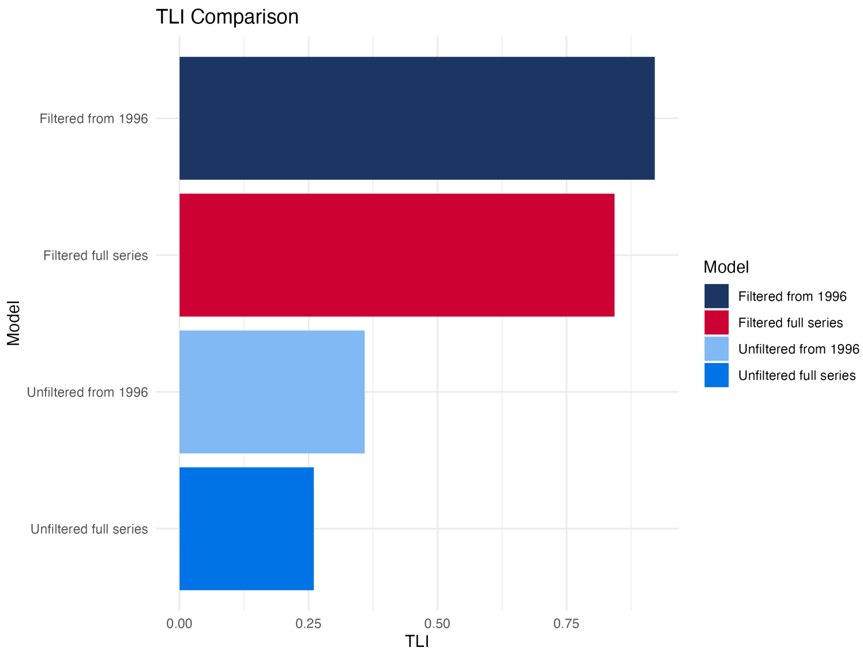
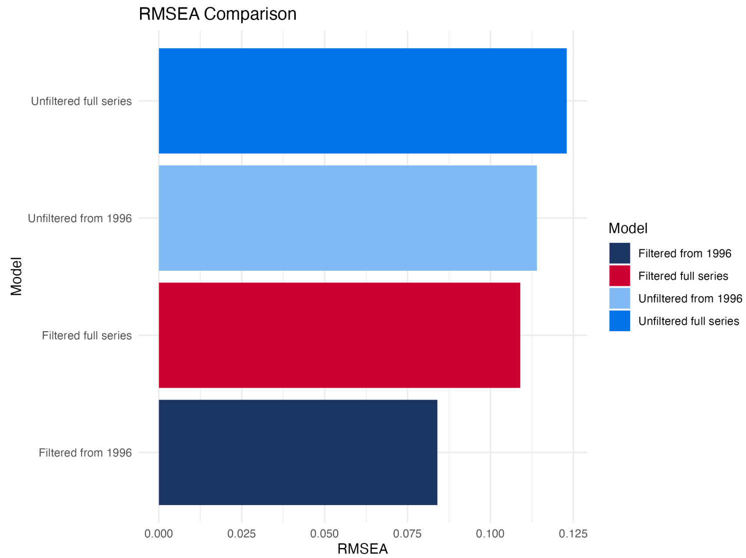


Figure 2-7 highlights that the filtered 1996 model has the lowest Root Mean Squared Error of Approximation (RMSEA), indicating better model fit. The RMSEA is defined as:

$$\text{RMSEA} = \sqrt{\frac{\chi^2 - \text{df}}{\text{df}(n-1)}} \tag{2.23}$$

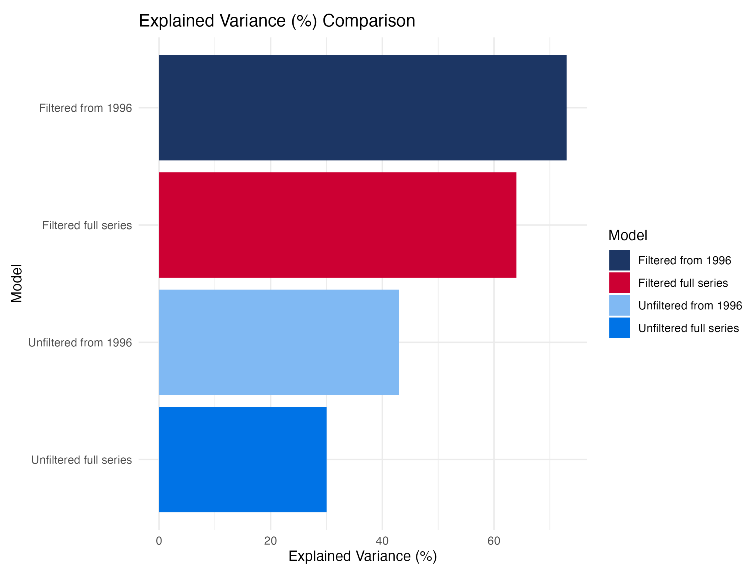
where n is the sample size. A lower RMSEA value implies a closer fit to the data. The filtered 1996 model achieves an RMSEA of 0.084, significantly better than the other models.

Figure 2-7: Root Mean Squared Error Comparison of the four optimal factor models selected.



Considering which model has a higher percentage of explained variation in the target variables, inflation and unemployment, the filtered 1996 clearly produces the highest (see figure 2-8). This model captures 73% of the total variation, significantly outperforming the other models.

Figure 2-8: Total Variation in inflation and unemployment explained by the four optimal models selected

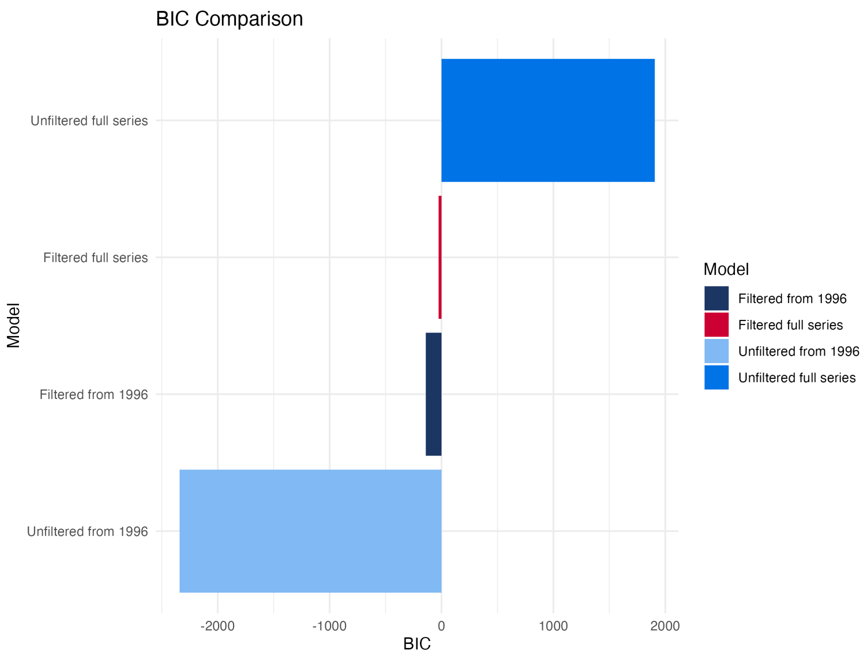


Finally, comparing the BIC values for all four models, it is clear that the 1996 models produce more parsimonious models. The BIC is defined as:

$$\text{BIC} = -2 \cdot \ln(L) + k \cdot \ln(n) \tag{2.24}$$

where L is the likelihood of the model, k is the number of parameters, and n is the sample size. Lower BIC values indicate more parsimonious models. Figure 2-9 compares the Bayesian Information Criterion (BIC) values across the models. While the filtered 1996 model does not achieve the lowest BIC, it produces the second-best value, reinforcing the importance of accounting for structural breaks and variable filtration.

Figure 2-9: Bayesian Information Criteria Comparison of the four optimal models selected



Impulse Response Functions (IRFs) from the Filtered 1996 Model Figure 2-10 presents a panel of impulse response functions (IRFs) generated using the filtered 1996 model within a sign-restricted Factor-Augmented Vector Autoregressive (FAVAR) framework. The panel illustrates the dynamic responses of six key macroeconomic variables to a contractionary monetary policy shock, represented by a one-standard-deviation increase in the policy rate. This shock simulates an interest rate hike by the central bank. Appendix Figure 2-A.8 presents a corresponding panel of impulse response functions estimated using a sample period closely aligned with that of Champagne and Sekkel (2018). The results are broadly consistent, though the responses exhibit slightly lower statistical significance—an expected outcome given the reduced sample size.

The top-left graph in the panel shows the response of the employment factor. The shock leads to a pronounced and immediate decline in employment, with the effects being statistically signficant and persisting for several months before stabilizing. This highlights the sensitivity of labour markets to interest rate changes, as higher borrowing costs reduce business activity and subsequently employment levels. The steep initial decline aligns with theoretical expectations, reflecting the strong connection between monetary policy and labour market dynamics.

In the top-right graph, the response of GDP growth is depicted. A contractionary monetary policy shock causes a modest decline in output, which peaks within the first few months and quickly begins to dissipate. This relatively short-lived impact suggests that while monetary tightening does exert downward pressure on aggregate output, other stabilizing forces, such as fiscal policies or external demand, help the economy recover in a timely manner. The magnitude of the response is smaller compared to the employment factor, suggesting that GDP is less directly affected by interest rate changes in the short term.

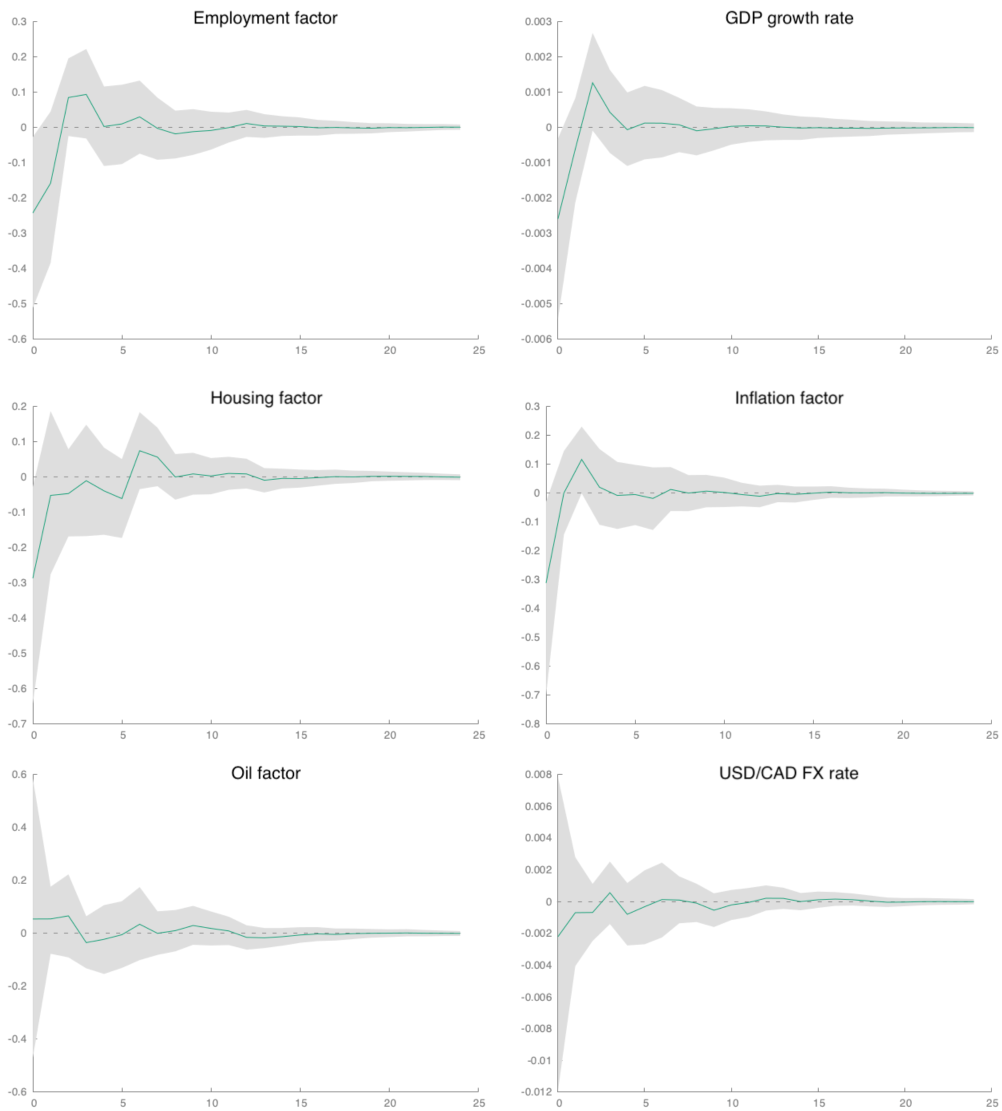
The middle-left graph highlights the reaction of the housing factor. Among all the variables analyzed, housing exhibits a strong response to the monetary policy shock. The increase in interest rates significantly raises borrowing costs, leading to reduced housing demand and a contraction in housing market activity. This emphasizes the critical role of monetary policy in shaping housing market outcomes and its broader implications for household wealth and consumption.

Inflation, displayed in the middle-right graph, reacts sharply and negatively to the monetary policy shock, consistent with the central bank's objective of reducing inflationary pressures. The immediate decline in inflation reflects the effectiveness of monetary tightening in curbing aggregate demand.

The bottom-left graph shows the response of the oil factor. Compared to other variables, the oil factor exhibits a more muted, statistically insignificant, and transient reaction to the monetary policy shock. While there is an initial decline, the response quickly fades, suggesting that oil prices are influenced more by factors beyond monetary policy, such as non-domestic supply and demand conditions. This highlights the limited role of central bank interventions in directly affecting commodity markets.

Lastly, the bottom-right graph presents the response of the exchange rate. While the point estimate suggests a slight appreciation of the Canadian dollar following the shock, the confidence intervals indicate that the effect is not statistically significant. This lack of significance makes it challenging to draw firm conclusions about the direction of the exchange rate response. This suggests that while monetary policy may influence exchange rates, other factors such as global risk sentiment and trade dynamics likely play a more dominant role in determining currency movements.

Figure 2-10: Impulse Response Functions (IRFs) of key macroeconomic variables to a monetary policy shock. This has been estimated using the model with factors constructed from the final optimal FAVAR model selected, which constructs factors from the filtered set, and for 1996-onward data to consider structural breaks.



For comparison, the impulse responses of the FAVAR model with factors constructed from the raw data without filtration are included in Appendix Figure 2-A.7. In this model, responses across variables have larger confidence bands and do not show statistical significance for impulse responses after monetary policy shocks. This holds even when a lower level of significance bands, 68% confidence interval, is considered. In addition, several variables such as the employment factor, GDP growth, and inflation factor exhibit much smaller effects than expected from theory (including Phillips curve) and empirical evidence, for example, (Champagne and Sekkel, 2018) and (Koeppl et al., 2024).

2.5 Conclusion

This paper introduces a novel variable filtration process that is applied before constructing the factor model for the analysis of monetary policy shocks in a FAVAR framework. The approach combines advanced machine learning techniques, such as LASSO and Random Forests, with traditional econometric methods like Recursive Feature Elimination and Stepwise Regression. The results show that this filtration process enhances the reliability and fit of the factor model, as evidenced by improved goodness-of-fit and information criteria. By systematically filtering the dataset to focus on the most informative variables, the methodology ensures that the constructed factors capture the underlying economic structure more effectively, compared to a factor model constructed from the raw data directly.

The paper also investigates the transmission of monetary policy shocks in Canada, a small open economy with distinct macroeconomic characteristics and unique channels for monetary policy transmission. While considerable research has been devoted to understanding the transmission of U.S. monetary policy shocks, the Canadian context has received comparatively less attention. This paper addresses this gap by developing a comprehensive framework that accounts for structural breaks, integrates large-scale macroeconomic data, and employs a novel filtration process for variable selection in a Factor-Augmented Vector Autoregressive (FAVAR) model.

The analysis identifies two structural breaks in 1992 and 1995, corresponding to the Bank of Canada's adoption of inflation targeting and significant fiscal reforms in the mid-1990s. Accounting for these structural breaks is crucial for improving the reliability of factor models. The results demonstrate that models which ignore these regime changes perform significantly worse in terms of fit and robustness, highlighting the importance of capturing shifts in monetary policy conduct when studying the effects of monetary shocks.

Building on these contributions, the paper employs a sign-restricted Factor-Augmented Vector Autoregressive (FAVAR) model to trace the dynamic responses of key macroeconomic variables to monetary policy shocks. By leveraging information from the filtered factor model and applying sign restrictions grounded in economic theory, the FAVAR framework provides interpretable impulse response functions (IRFs). The results indicate that a contractionary monetary policy shock leads to statistically significant declines in employment, GDP, inflation, and housing market activity. These findings are consistent with theoretical expectations and underscore the central bank's ability to influence real economic outcomes through monetary policy.

Interestingly, and unlike recent empirical findings in the literature, the responses of the exchange rate and the oil factor to monetary policy shocks are less pronounced. The exchange rate response is ambiguous, reflecting the complex interplay of global risk sentiment, trade dynamics, and other external factors in determining currency movements. Similarly, the oil factor exhibits a muted and statistically insignificant response, suggesting that global commodity markets are influenced more by external supply and demand conditions than by domestic monetary policy. The findings highlight the potential of combining machine learning, factor modeling, and sign-restricted VAR frameworks to enhance the study of monetary policy shocks.

2-A Appendix

Table 2-A.1: Summary of stationarity test results. The values marked with $*$, $**$, and $***$ denote significance at the 10%, 5%, and 1% levels, respectively. The ADF and PP tests examine the null hypothesis of a unit root (non-stationarity). A rejection of the null indicates that the series is stationary.

Variable	ADF_Stat	PP_Stat
GDP Growth	-17.297***	-14.083***
Industrial Production	-15.528***	-17.297***
Oil Prices (Canada)	-20.207***	-35.081***
Employment	-17.498***	-13.307***
Unemployment Rate	-14.564***	-12.734***
Bank Rate (Long-Term)	-11.369***	-14.659***
CPI (All Items)	-12.770***	-15.526***
Exchange Rate (USD/CAD)	-12.441***	-14.749***
Money Supply (M3)	-11.478***	-16.917***
Export Balance (BP)	-15.455***	-18.315***
Building Permits Index	-21.266***	-41.401***
House Prices	-5.623***	-8.417***
Money Supply (M2)	-7.912***	-10.300***
Total Reserves	-14.119***	-22.738***
5-Year Mortgage Rate	-13.704***	-16.958***
3-Month Treasury Bills	-10.283***	-20.942***

Table 2-A.2: Number of lags selected based on different information criteria

Information Criterion	Number of Lags Selected
Akaike Information Criterion (AIC)	6
Hannan-Quinn Information Criterion (HQ)	6
Schwarz Criteron (SC)	6
Final Prediction Error Criterion (FPE)	6

Figure 2-A.1: IRFs from Recursive Approach from Christiano et al. (1999).

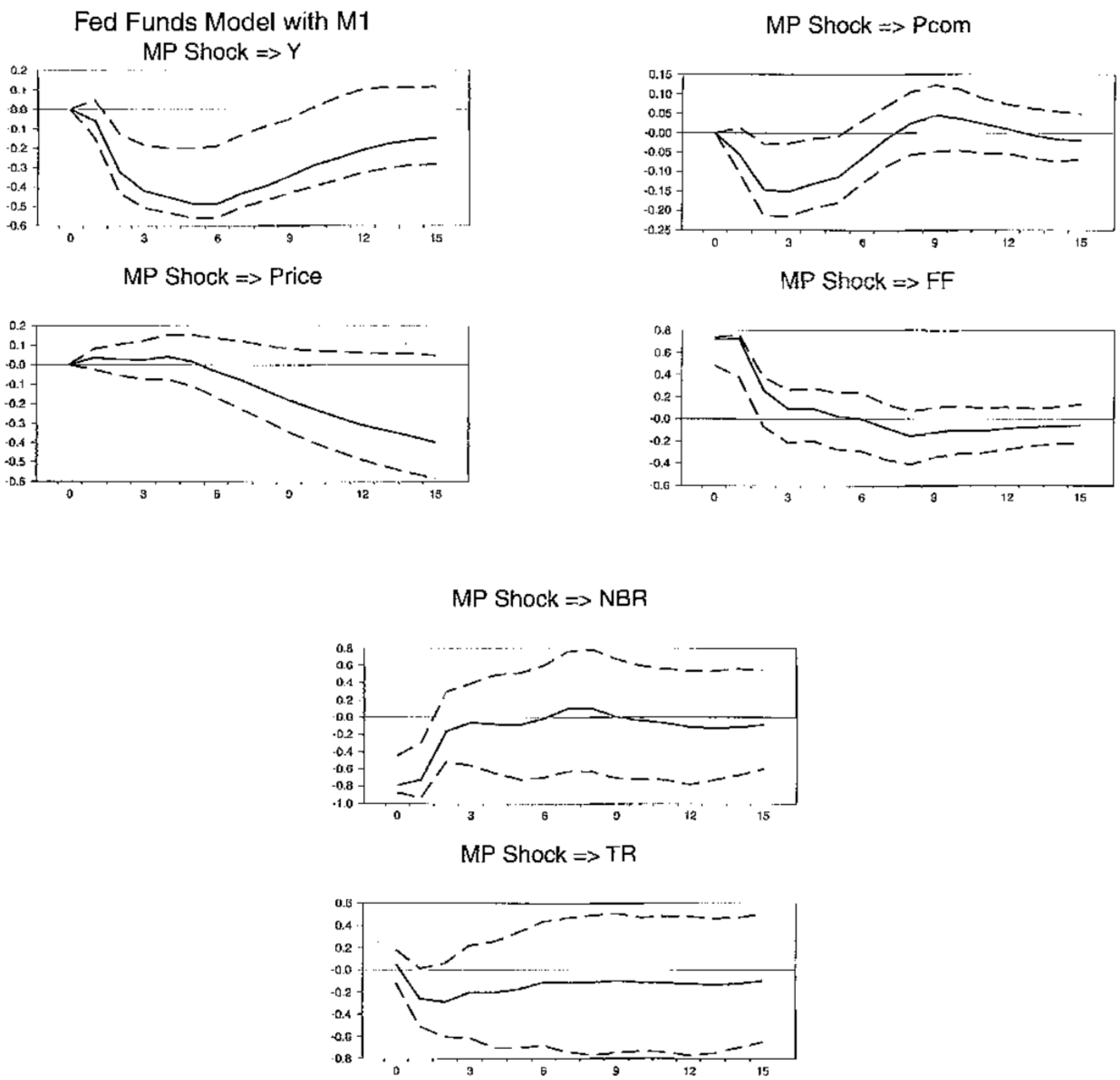


Figure 2-A.2: IRFs from standard VAR, VAR with R&R shocks, and R&R baseline from Coibion (2012)

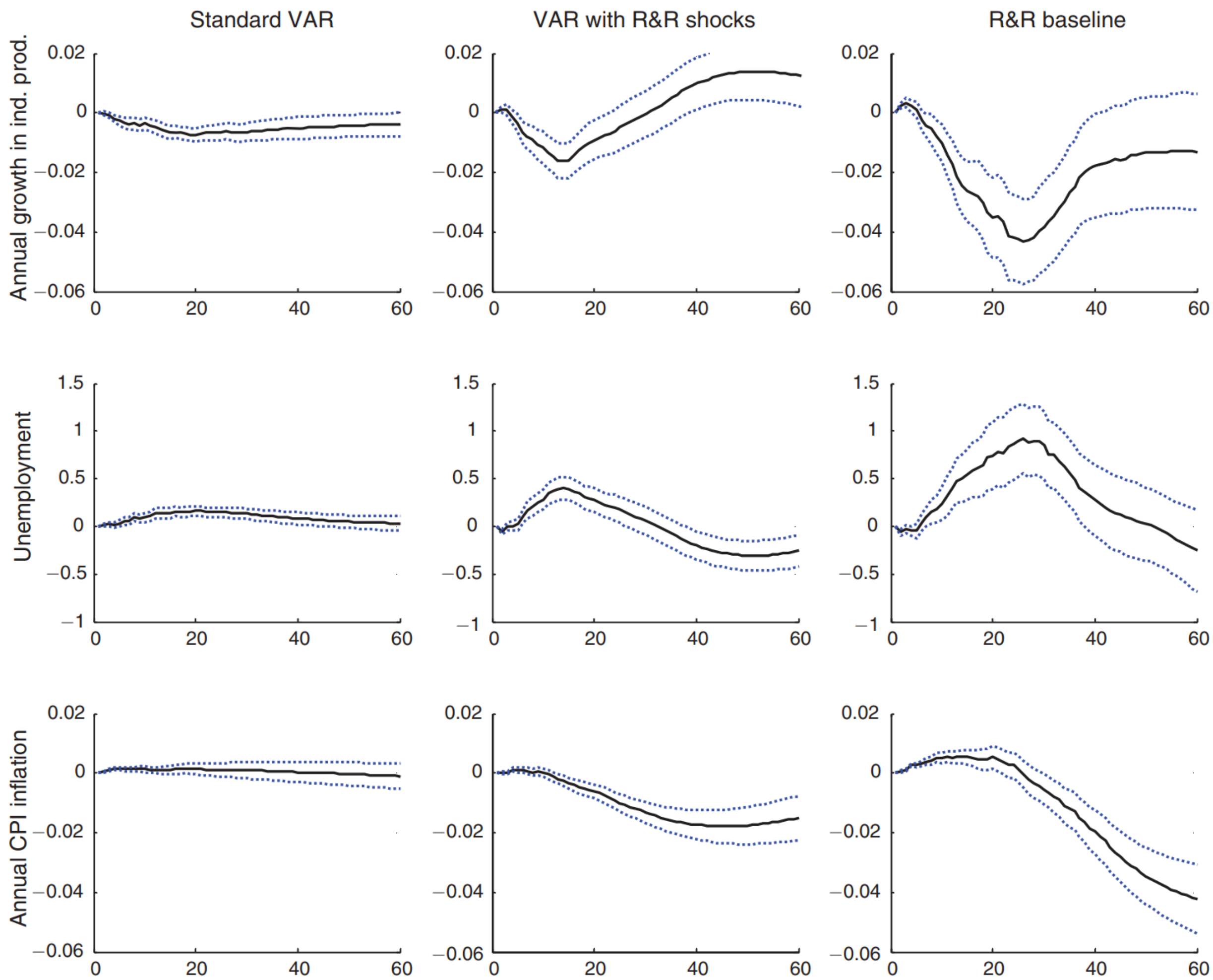


Figure 2-A.3: IRFs using Local Projections Approach from Jordà (2005)

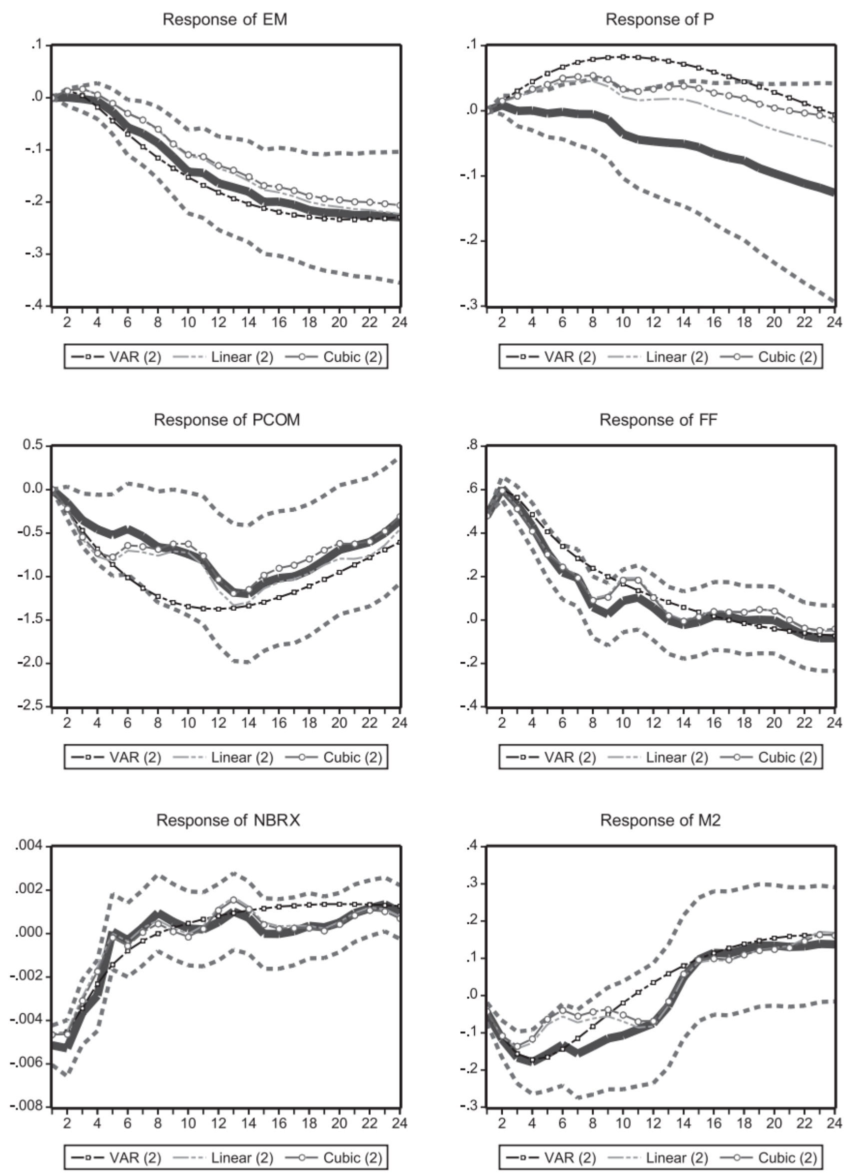


Figure 2-A.4: IRFs for Miranda-Agrippino and Ricco (2021). Top row: VAR (dashed) vs BLP (solid). Bottom row: LP (dotted) vs BLP (solid)

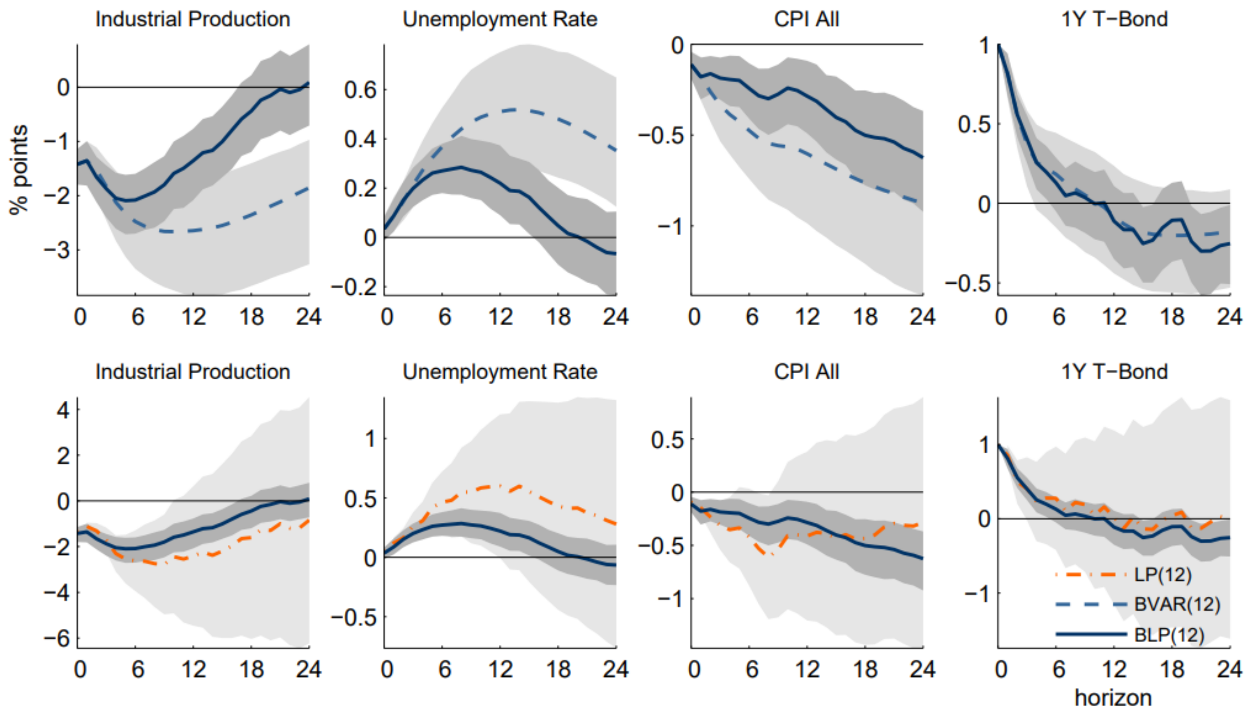


Figure 2-A.5: IRFs for Canadian Narrative Approach from Champagne and Sekkel (2018)

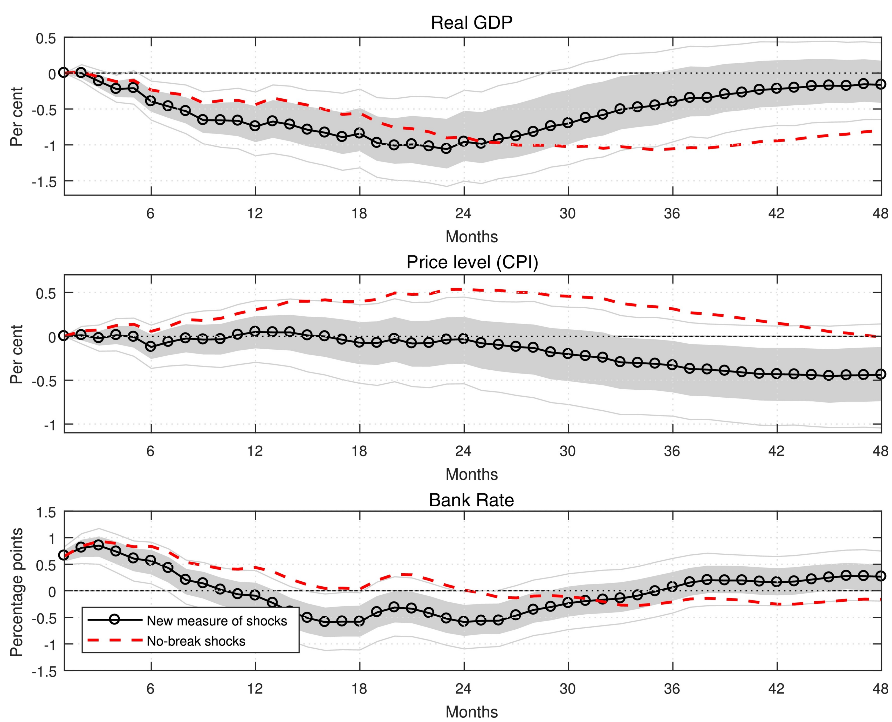


Figure 2-A.6: Bar Graph for variables selected by different methods.

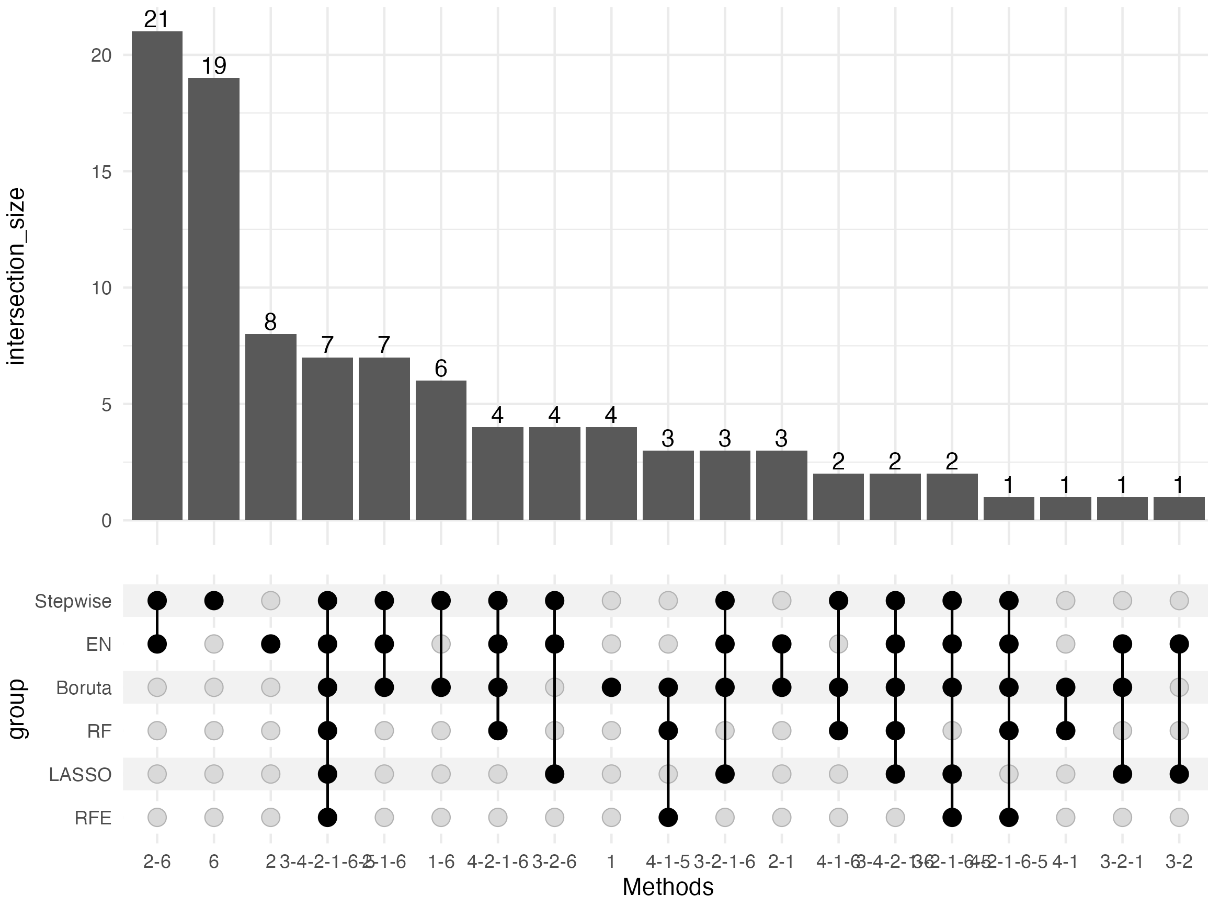


Figure 2-A.7: Impulse Response Functions (IRFs) of Key Macroeconomic Variables to a BANK_RATE_L Shock, Estimated Using the model with factors constructed from the full data set.

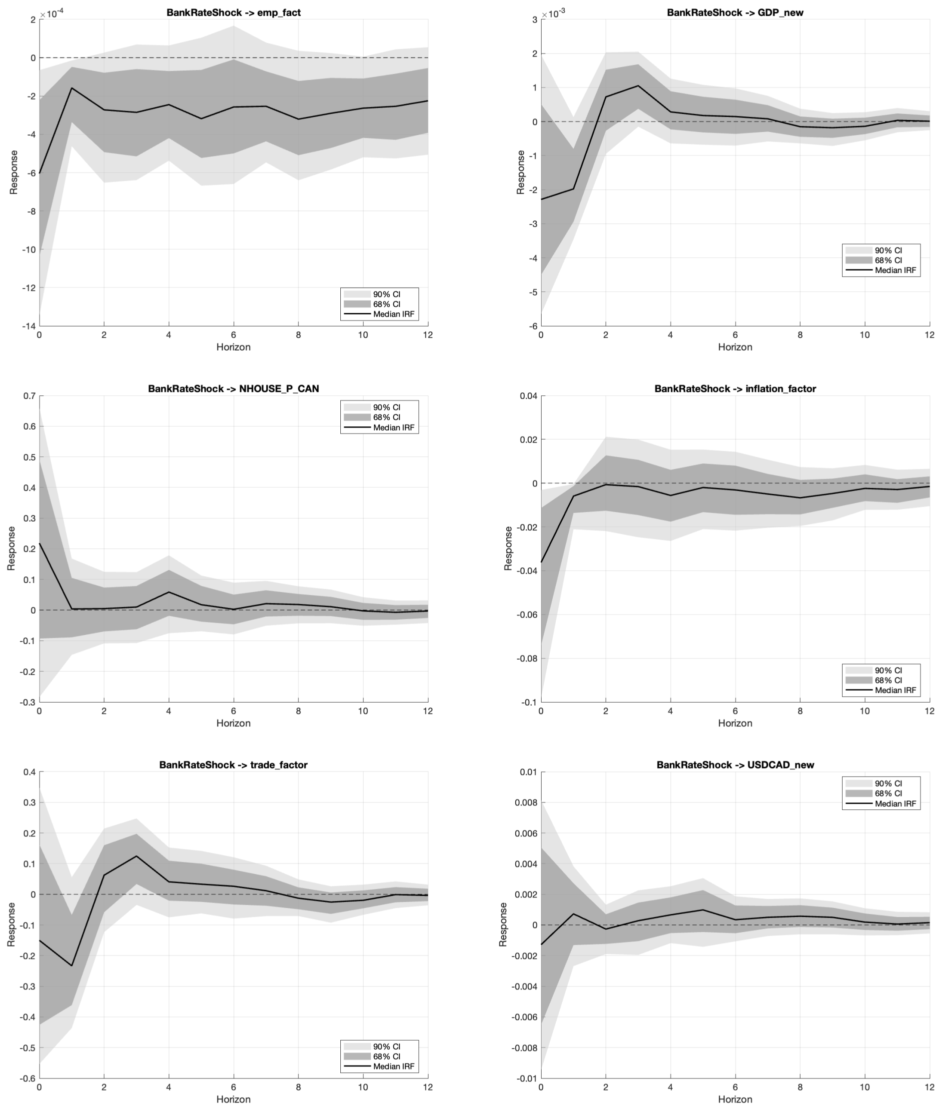
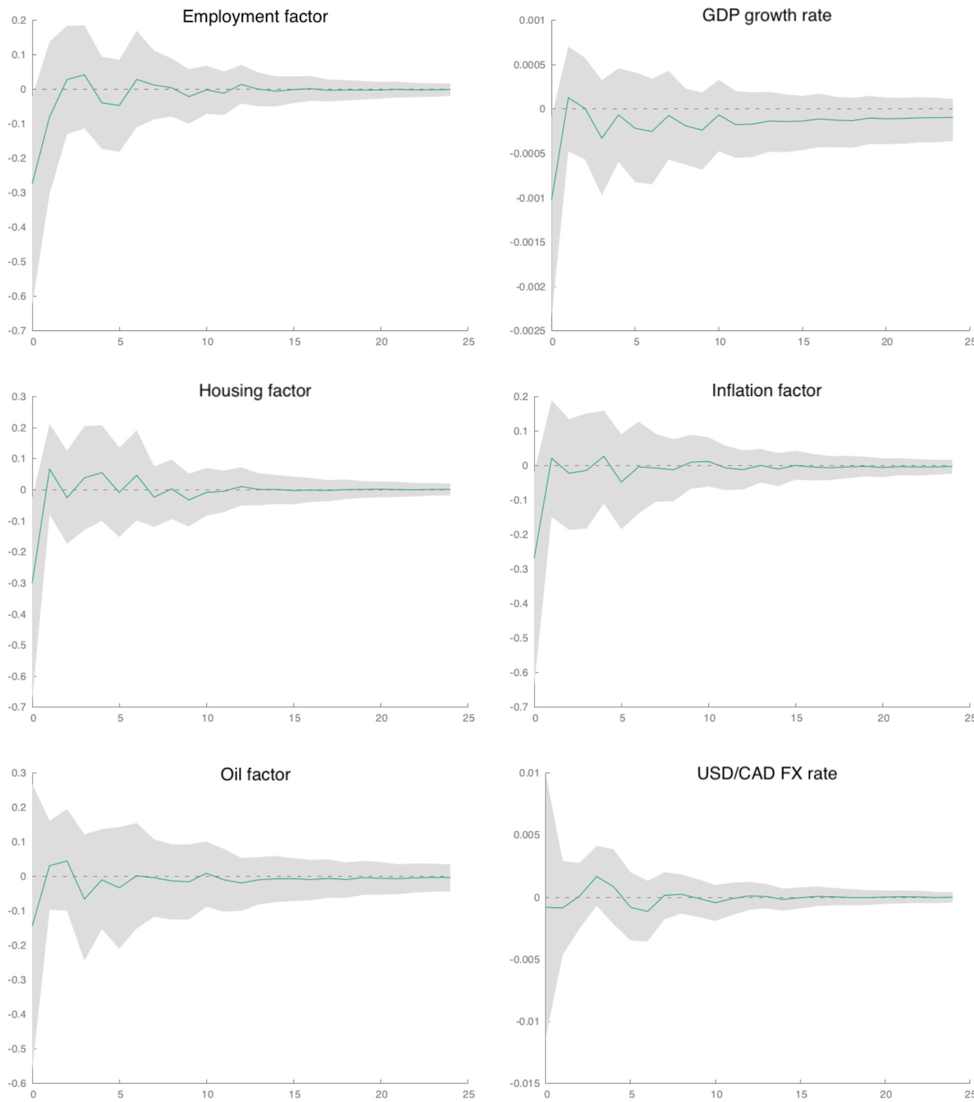


Figure 2-A.8: Impulse Response Functions (IRFs) of Key Macroeconomic Variables to a monetary policy shock. This is estimated using the optimal FAVAR model selected, with factors constructed from the filtered set. The time period is restricted to match Champagne and Sekkel (2018) for comparability. Results hold with less statistical significance, which can be explained by the smaller sample size used.



Chapter 3

Understanding Firm Price Setting: Evidence from a Lab Experiment*

Abstract

This paper examines how firms set prices in response to economic shocks using a controlled lab experiment. Participants engage in repeated Bertrand competition under varying market structures—duopoly versus monopolistic competition—and pricing frictions, with either flexible prices or menu costs. Firms face four types of shocks: a demand shock and three cost shocks differing in size and uncertainty (large, small, and Markov process). By precisely controlling costs and shocks, we provide clean evidence on the roles of expectations, strategic interaction, and frictions in shaping pricing behavior. We find that larger shocks induce proportionally larger price adjustments, compared to smaller shocks. We also find that under uncertainty, such as in a world with dynamically changing tariff policies, firms pass through a smaller fraction of expected cost changes, consistent with a "wait-and-see" strategy. We show that firms' forecasts of the market price consistently have a stronger influence on pricing decisions than their expected costs. These findings highlight the importance of expectations, nominal rigidities, and uncertainty in price-setting, with direct implications for understanding inflation dynamics and the transmission of economic shocks.

* The experiment has been pre-registered on the AEA RCT registry under project number AAEARCTR-0015767 and has been granted ethical approval by the Ethics Committee Economics and Business at the University of Amsterdam under project number EB- 12945. We are grateful to the Faculty of Economics & Business of the University of Amsterdam for financing this project through the Research Priorities Area (RPA) 'Complex Human Systems Lab.'

3.1 Introduction

The dynamics of price-setting are central to understanding inflation behavior and the transmission of monetary policy. Recent episodes of high inflation—triggered by supply chain disruptions, energy shocks, and rapid demand shifts—have highlighted the importance of firm-level pricing decisions in shaping macroeconomic outcomes. A key insight from both empirical and theoretical work is that prices do not necessarily adjust smoothly or uniformly: instead, the frequency and size of price may change in response to underlying economic conditions. These observations challenge models based on time-dependent pricing and call for a closer examination of how firms form expectations, respond to shocks, and navigate frictions such as menu costs or market power. Theoretically, these microeconomic behaviors have powerful implications. The degree of nominal rigidity affects the slope of the Phillips curve and the effectiveness of monetary policy. A steep Phillips curve implies that inflation responds strongly to output gaps, while a flatter curve suggests that inflation is more inertial, possibly requiring more aggressive policy to bring it under control. At the same time, price rigidities shape how shocks propagate across the supply chain and whether firms absorb or pass on changes in costs to consumers.

Yet despite the growing availability of micro-price datasets (including scanner data and surveys), empirical work faces persistent limitations. Data often lack direct measures of marginal cost, expectations, or the timing and nature of shocks. Prices may adjust due to unobserved quality changes, pricing algorithms, or internal firm processes unrelated to external shocks. Moreover, disentangling responses to supply versus demand shocks is often infeasible in the field, and estimating pass-through behavior can conflate firm heterogeneity with strategic responses.

To overcome these challenges, we design a laboratory experiment that places subjects in a Bertrand competition under controlled conditions. The experiment isolates key features of real-world pricing environments—strategic interaction, uncertainty, and frictions—while allowing us to observe firm-level costs, expectations, and decisions with precision. Subjects repeatedly set prices and forecast market prices under Bertrand competition, facing four types of shocks: (1) a persistent two-state Markov cost shock representing tariff uncertainty, (2) a large autoregressive AR(1) cost shock, (3) a small AR(1) cost shock, and (4) a demand shock. These shocks are randomly ordered across games but are common across participants, allowing for within-subject comparisons and identification of behavioral responses.

The experiment employs a 2×2 between-subjects design that varies market structure—duopoly vs. monopolistic competition—and the presence of nominal rigidities: either no adjustment cost (flexible pricing) or the imposition of a monetary menu cost whenever subjects decide to change their price. This design enables us to test how price-setting varies with market power and the cost of changing prices, as well as how different types of uncertainty and shock persistence influence behavior. Participants are incentivized to forecast market prices accurately and to set profit-maximizing prices in each round. The rest of the simulated market follows rational expectations and best-response behavior, allowing the experiment to mimic a partial equilibrium environment with realistic strategic dynamics.

A key strength of our study is the combination of rigorous internal and external validation. Internal validity is supported by the controlled laboratory environment, careful experimental design, incentivized decision-making, and the use of repeated interactions to allow for learning and strategic adjustment. For one treatment, we also recruited active entrepreneurs—individuals with real-world experience in setting prices and managing costs—to participate under identical experimental conditions. Their behavior serves as an external benchmark, allowing us to assess whether the patterns observed among student participants hold in a population with practical business experience. The strong similarity of results across both groups provides confidence that our findings are not merely an artifact of the student sample, but are relevant to decision-makers in real markets.

We address four central research questions. First, what is the magnitude of contemporaneous and cumulative pass-through from expected costs to prices across market structures and pricing frictions? Second, how important are firms' own market price forecasts relative to their expected costs in determining prices? Third, how do menu costs affect price inertia and the distribution of adjustment over time? Fourth, how can uncertainty alter the sensitivity of prices to expected costs, relative to cost-push shock and demand shock scenarios? Our data also allow us to study learning dynamics: whether firms improve forecast accuracy over time, and whether those expectations guide their pricing behavior.

The estimates reveal four clear patterns. First, expected costs have a positive and statistically significant effect on prices across all treatments, with the contemporaneous coefficients ranging from 0.22 to 0.35 across different treatments. Second, firms place more weight on their own price forecasts than on expected costs when setting prices. The forecast coefficient exceeds that on expected cost in every treatment and is always highly significant. Third, past prices strongly influence current pricing decisions, with the effect particularly pronounced in treatments with menu costs, showing that nominal rigidities reinforce persistence in pricing. Fourth, firms adjust prices less in response to expected costs under a Markov shock than under other shocks, consistent with greater uncertainty dampening the pass-through of expected cost changes.

Dynamic specifications show that while the immediate pass-through of cost changes varies across treatments, cumulative pass-through over the following periods is substantial—ranging from about 0.71 in the most inertial settings to nearly 0.88 in the most flexible—suggesting that nominal frictions may delay the overall adjustment of prices to costs when firms expect shocks to resolve over time. Moreover, market price forecast accuracy improves over time, with a small but significant positive effect of experience, consistent with firms learning about their environment as the game progresses. These patterns show the importance of expectations, inertia, and uncertainty in shaping pricing behavior.

The rest of this paper is organized as follows: Section 2 reviews the relevant literature and situates our contribution within the broader context of firm price-setting research. Section 3 describes the experimental design, outlining the underlying model, the structure of the shock simulations, and summary statistics on participants' earnings and completion times. Section 4 presents the empirical analysis, beginning with descriptive evidence, followed by impulse responses from local projections, estimates of the determinants of price setting, cumulative pass-through effects, and an examination of forecast accuracy and learning dynamics. Section 5 concludes.

3.2 Literature Review

Understanding nominal price rigidity and firm pricing behavior under uncertainty has been a central theme in macroeconomics for decades. Classical New Keynesian models, relying heavily on time-dependent pricing such as the Calvo (1983) framework, have faced challenges in matching micro-level price dynamics observed in empirical data. Recognizing these limitations, many studies turned to state-dependent pricing models, which allow firms to adjust prices in response to economic conditions, such as cost shocks or margin pressures. Models with menu costs, for example in Golosov and Lucas Jr (2007), embed fixed costs of changing prices and generate infrequent but economically consequential adjustments. A key puzzle has then become reconciling the coexistence of frequent temporary price fluctuations observed empirically, with persistent underlying price stickiness. Micro price datasets—CPI microdata, scanner data, and online prices—establish a common set of facts: prices seem to change much more often than suggested by many models; both increases and decreases are common; and the frequency and size of changes covary with inflation and sectoral shocks. Foundational contributions include Nakamura and Steinsson (2008), Klenow and Kryvtsov (2008), and subsequent surveys, e.g., Klenow and Malin (2010), which emphasize that movements in inflation are linked to both the incidence and magnitude of adjustments. Online data further show that e-commerce and uniform national pricing alter hazards and pass-through to aggregate shocks (Cavallo, 2018a,b)

During the post-pandemic inflation surge in 2021–2023, micro evidence documents a clear shift toward greater price flexibility and faster pass-through. Using U.S. CPI microdata, Montag and Villar Vallenas (2025) show that the frequency of both increases and decreases rose relative to the 2010s, with faster propagation of cost shocks to retail prices. Euro-area studies similarly report higher frequencies of repricing (especially for energy-intensive items), with the surge explained by changes in both the frequency and size of adjustments, and with notable within-retailer synchronization (e.g., Rudolf and Seiler, 2022; Dedola et al., 2025). Furthermore, comparing pre-pandemic and post-pandemic data in Gagliardone et al. (2025) shows that large aggregate shocks affect the proportional passthrough of costs differently, compared to smaller shocks, highlighting the endogenous link between shock size and price adjustments. These studies imply that there is more evidence of a state-dependent environment in the post-pandemic era, compared to the low-inflation decade.

A growing strand of literature uses scanner data and retailer microdata to study how the size of shocks shapes firm behavior. Two important insights emerge. First, evidence shows that large shocks trigger faster and more extensive pass-through than smaller shocks, consistent with state-dependent models where inaction regions are overcome only once shocks are sufficiently large. Cavallo et al. (2024) demonstrate that during the recent inflation surge, the speed of price adjustment increased disproportionately following large shocks, with firms repricing more frequently and synchronously compared to smaller disturbances. Second, micro evidence often shows that even after shocks occur, prices do not fully adjust immediately, with pass-through materializing gradually over several periods. The delayed response documented in Sangani (2025) illustrates this inertia: a large share of pass-through occurs contemporaneously, but sizable lagged effects remain.

Evidence also suggests that firms in more concentrated markets exhibit different adjustment dynamics than those in competitive environments. For instance, more concentrated firms tend to pass through cost shocks less fully in the short run, but maintain higher and more stable markups over time (e.g., Genakos and Pagliero, 2022; Ritz, 2024). This implies that the competitive environment shapes both the timing and extent of pass-through, with concentrated markets generating more rigidity in the short run but potentially more persistent inflationary effects. These findings indicate that firm behavior under shocks depends not only on their size and persistence, but also on the strategic environment in which firms operate.

Lab markets complement micro data by isolating mechanisms behind inertia. Experiments consistently show state dependence with inaction bands and lumpy adjustments (consistent with Ss-type pricing), learning and expectations formation that shape pass-through over time, and strategic interactions that affect the timing and synchronization of changes. Evidence includes pricing with bounded rationality and strategic complementarities (Fehr and Tyran, 2008), empirical Ss-style thresholds with dispersion from cognitive costs and attention limits (Magnani et al., 2016; LeBlanc et al., 2016), and price-setting under shocks in DSGE-like environments (Noussair et al., 2015). Our study builds on these insights by experimentally exploring firm price-setting across different shock environments, market structures, and nominal rigidity regimes. We contribute novel evidence on the mechanisms driving price adjustment, markup dynamics, and learning in markets facing several different economic shocks.

3.3 Experimental Design

We describe the experimental setup in three parts. We first outline the underlying theoretical model that motivates the design. We then present the treatments and hypotheses, highlighting how market structure and nominal rigidities shape behavior. Finally, we describe the implementation details, including subject pool, procedures, shocks, timing, and payoff rules.

3.3.1 Model

Similar to Assenza et al. (2015), we implement a linear demand structure derived from a representative consumer with quadratic utility over differentiated goods. Each firm's demand is given by:

$$q_i = \alpha - \beta p_i + \theta \bar{p} + v_t, \tag{3.1}$$

where v_t is a **demand shock**, q_i is the quantity sold by firm i, p_i is the firm's price, and $\bar{p}$ is the average price in the market. The term $\theta > 0$ captures the degree of substitutability and spillover across firms. This demand structure ensures that higher relative prices reduce quantity sold, but market-wide price increases mitigate the effect of individual price hikes, consistent with a Dixit–Stiglitz framework.

Firms face a linear cost function:

$$c_{i,t} q_{i,t} = c q_{i,t} + u_t, \tag{3.2}$$

with u_t a **cost-push** shock, and $c > 0$ the marginal unit cost of production.

This structure induces strategic interactions among firms' pricing decisions and supports a Nash equilibrium where price depends on marginal cost, the number of firms, and the structure of demand. In treatments with menu costs, firms must weigh the gain from price adjustment against the fixed cost, allowing us to examine both intensive and extensive margins of price changes. Setting $p_{i,t}$ to maximize profits $\pi_{i,t} = (p_{i,t} - c_{i,t})q_{i,t}$ conditional on I_t gives:

$$p^*_{i,t} = \frac{\alpha + \beta c + \tilde{E}_{i,t}(v_t) + \beta \tilde{E}_{i,t}(u_t) + \theta \tilde{E}_{i,t}(\bar{p}_t)}{2\beta}. \tag{3.3}$$

The market follows rational expectations with the Nash equilibrium:

$$p^*_N = \frac{\alpha + c\left(\beta - \frac{\theta}{n}\right)}{2\beta - \theta\left(1 + \frac{1}{n}\right)}. \tag{3.4}$$

The derivation of the Nash equilibrium price equation can be found in Appendix 0-B. We set $\alpha = 50$, and $c = 20$, both with white noise fluctuations of mean 0 and standard deviation 1, and $\beta = 2$, $\theta = 1.5$. This was done to obtain benchmark Nash equilibrium prices of around 43 ECU in a duopoly setting, and around 36 ECU in monopolistic competition.

3.3.2 Treatments and Hypotheses

The design employs a 2×2 between-subject structure that varies both nominal rigidity and market structure. On the rigidity dimension, we compare flexible pricing, in which price changes are costless, to menu cost pricing, in which changing prices incurs a small fixed cost in experimental currency units (ECU). The fixed cost is set at 100 ECU in duopoly settings and 125 ECU in monopolistic competition to generate a similar effect on pricing. On the competition dimension, we compare duopolies, with two firms, to a case of monopolistic competition, where the number of firms is set at 100. Subjects are randomly assigned to one of the four treatment conditions.

Table 3-1: Participant counts by treatment and mode of participation

Treatment	**Participants**
Duopoly, Flexible Prices (Menu = 0)	66 (44 students in lab, 22 students online), and 63 entrepreneurs online
Monopolistic Competition, Flexible Prices (Menu = 0)	63 (43 students in lab, 20 students online)
Duopoly, Menu Cost (Menu = 100)	63 (43 students in lab, 20 students online)
Monopolistic Competition, Menu Cost (Menu = 125)	66 (43 students in lab, 23 students online)

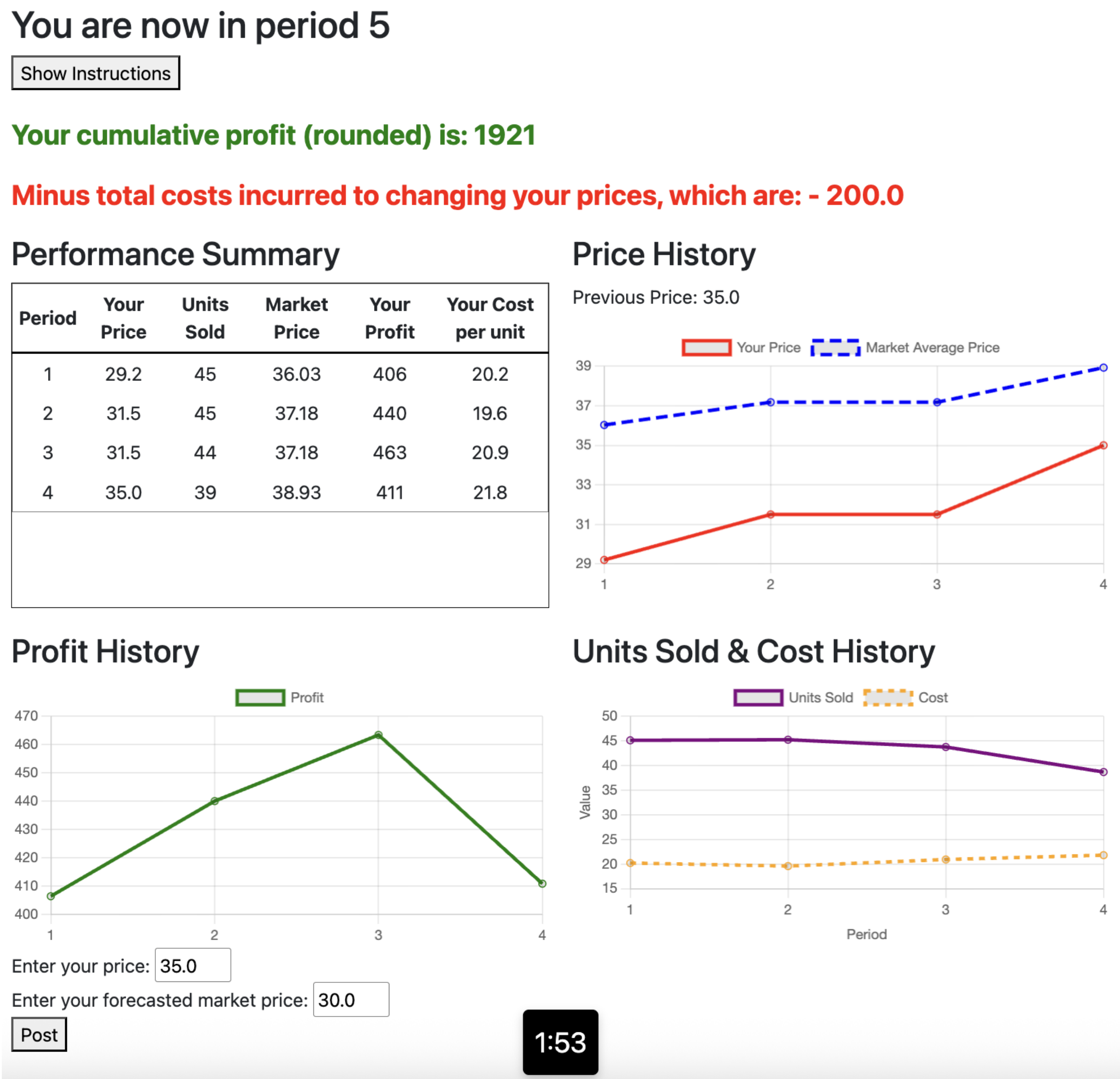
The design is intended to test four central hypotheses. Larger shocks are expected to trigger more frequent and substantial price changes than smaller shocks. The presence of menu costs should reduce the frequency of price adjustments but increase the magnitude of each change. Market structure is expected to shape markups and pass-through, with monopolistic competition generating lower markups than duopoly. Finally, firms' forecasts of the market price are expected to exert a stronger influence on pricing decisions than expectations about their own costs.

3.3.3 Implementation

The experiment with student subjects was carried out at the lab of the Center for Research in Experimental Economics and Political Decision-making (CREED), at the University of Amsterdam, and online, between April 14th, 2025 and April 18th, 2025. Entrepreneur subjects participated online on July 9th, 2025, through Playstudies, an institution in Spain that provides services to researchers in experimental and behavioral studies. The experiment was coded in Python, with the graphical user interface coded using HTML, JavaScript, and CSS. It was run on oTree, a large open-source behavioral research platform. Importantly, this is an individual decision-making experiment: participants act as the sole human decision-maker in their market. All other firms in the market are automated and follow profit-maximizing pricing rules based on rational expectations.

Participants start by reading the instructions in Appendix 0-A, after which they answer a quiz to check their understanding of the key elements of the game. The quiz contains questions about how their margin and profits are calculated in the game, the timing of when they observe each period's costs, and how they can earn the forecasting bonus. They must be able to answer all questions correctly to be allowed to proceed. Then, subjects go through a total of 110 periods. The first 10 periods serve as training rounds and do not affect earnings. The subsequent 100 periods make up the main experimental data, split into four blocks of 25 periods each. Each block corresponds to a different macroeconomic shock: a demand shock, a large AR(1) cost shock, a small AR(1) cost shock, and a two-state Markov cost shock. The order of shocks is randomized across subjects, and subjects are informed at the start of each block when a new sequence begins.

Figure 3-1: Main experiment screen for participants in menu cost treatments.



Period	Your Price	Units Sold	Market Price	Your Profit	Your Cost per unit
1	29.2	45	36.03	406	20.2
2	31.5	45	37.18	440	19.6
3	31.5	44	37.18	463	20.9
4	35.0	39	38.93	411	21.8

In each period, participants observe their past unit production costs, prices and profits, and set both their own price and a forecast of the market price. For each period, participants have a maximum time of 2 minutes to enter their prices and market price forecasts. If they fail to input a price within the time, their past period's price is automatically used and they receive no payout for any profit associated with that period. If this happens in the first period of any block, a random price is selected for the participant. The experiment instructions are available to review at any time throughout the experiment. Their profit is determined by the difference between their price and cost, multiplied by the quantity sold. Forecast accuracy is incentivized: participants earn a bonus if their forecast is within 1 ECU of the realized market price. Two of the four blocks are randomly selected for the profit-based payment, and one block is used for the forecasting bonus.

The shocks were implemented exactly as described in the design. In the two AR(1) cost shock treatments, the cost level for all firms shifts upward at the shock period, and then gradually reverts toward baseline, simulating persistent supply shocks of different magnitudes. In the demand shock treatment, the intercept of the demand function (α) temporarily declines following the same AR(1) process, simulating a contraction in aggregate demand. The AR(1) process is given by

$$x_t = a + bx_{t-1} + \varepsilon_t \tag{3.5}$$

where $a = 4$ is a constant intercept term, $b = 0.8$ represents the persistence of the process, and ε_t is a white noise error term with zero mean and constant variance. In the Markov cost shock treatment, the marginal cost switches stochastically between high and low states with a fixed transition probability of 0.3, capturing an environment of cost uncertainty. Participants receive on-screen announcements the period after the shock occurs, and they have access to historical data from all prior rounds.

Figure 3-2: Simulation of the cost (c_{it}) for the cost-push shocks and demand (α) variable trajectory under the defined AR(1) and Markov processes

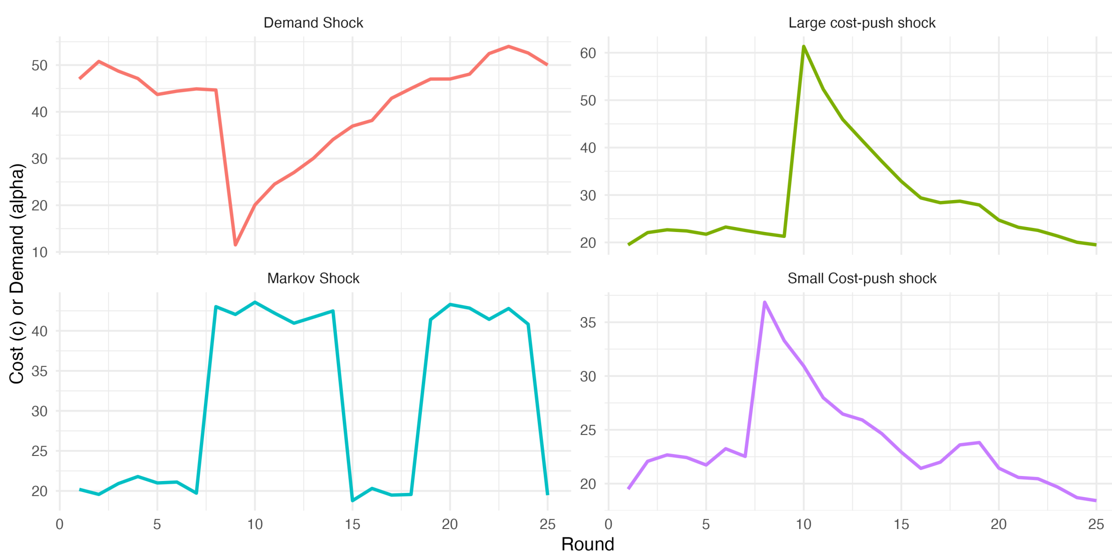
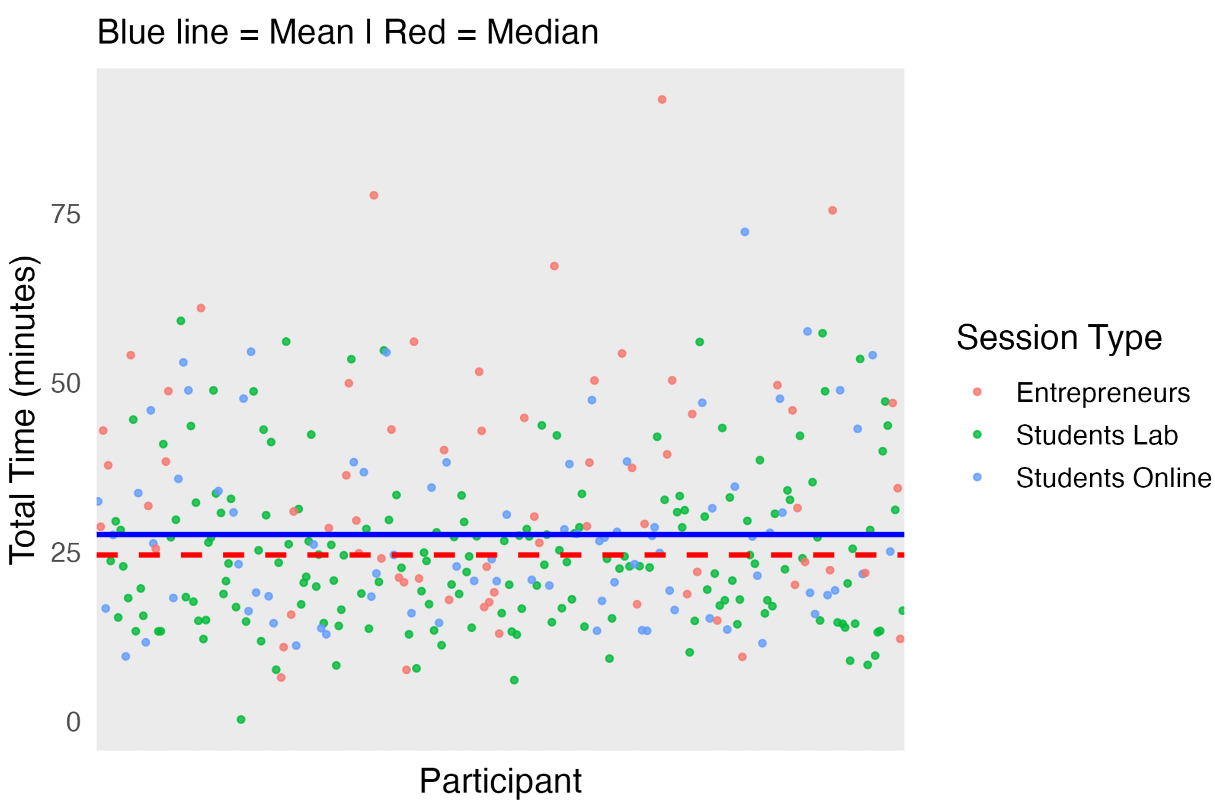


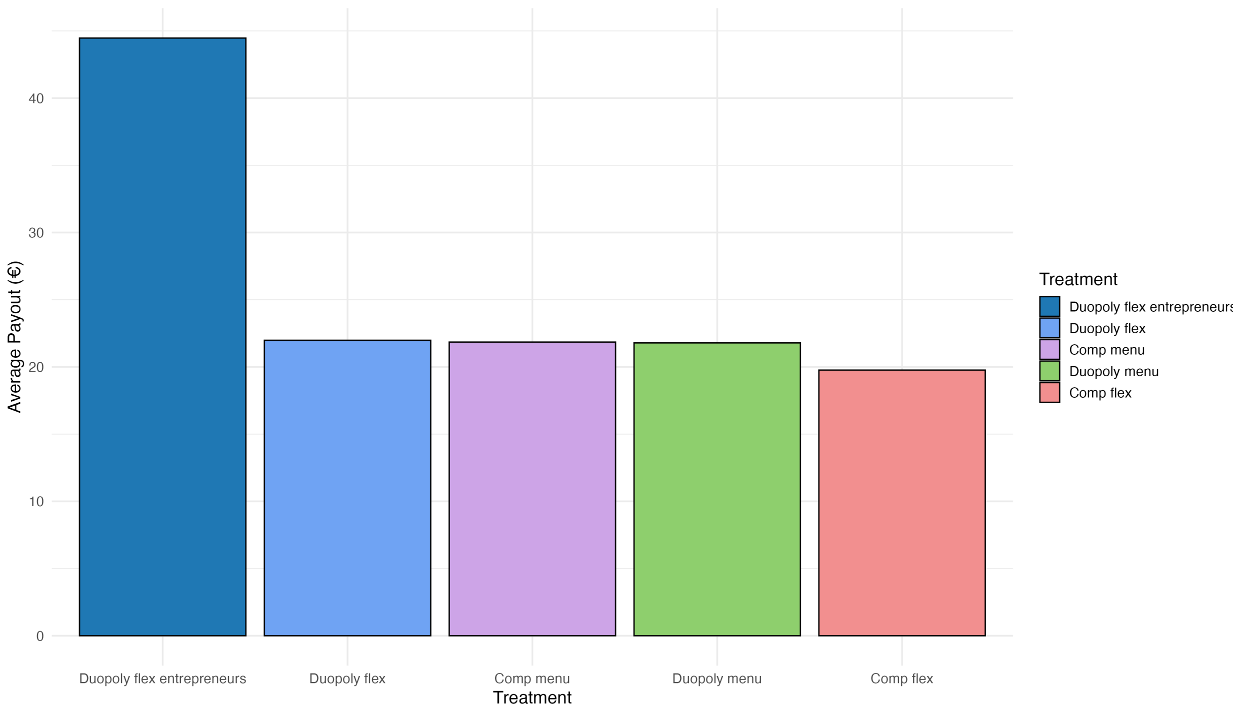
Figure 3-3 shows the time it took to complete the full game for all participants, including students who played in the physical lab and those who participated online. There was a clear difference between the average time it took students to complete the full game, and that of entrepreneurs. For students, the average was around 26 minutes, with a median time of around 24 minutes. For entrepreneurs, the average time was around 55 minutes, with a median time of around 31 minutes. The longer average completion time among entrepreneurs likely reflects differences in decision-making style and familiarity with experimental settings. Whereas students, accustomed to digital tasks and academic experiments, tended to proceed more quickly through the rounds, entrepreneurs may have taken more time to deliberate, reflect on strategy, and ensure they understood the rules.

Figure 3-3: Time spent by all participants, average time, and median time. 4 clear outliers were removed from the data



Subjects were paid for the profits they made in two randomly selected sessions out of the four sessions that they played, and one selected session for the forecast bonus. Average payout students received is around 20 EUR, with slight variations across treatments as in Figure 3-4. This was based on a conversion rate of 1 EUR per 2,000 ECU, 0.25 EUR bonus per correct market forecast, and a 7 EUR show-up fee. For entrepreneurs, a different conversion rate of 1 EUR per 600 ECU was used, and this resulted in a higher average payout of around 44 EUR and median 46 EUR.

Figure 3-4: Average payout (in Euros) for all participants



3.4 Data and Results

To investigate how firms adjust their pricing in the face of different economic disturbances, we analyze their responses to the set of cost and demand shocks using a combination of descriptive and econometric methods. First, we examine the experimental data descriptively, highlighting core patterns in pricing behavior across treatments and shock types—focusing on metrics such as the frequency and magnitude of price adjustments, the average distance from Nash equilibrium pricing, and the evolution of markups. Second, we construct impulse response functions (IRFs) using local projections to trace how prices evolve dynamically following the onset of a shock. Third, we employ fixed-effects panel regressions to study the determinants of price setting, including the role of nominal rigidities, and to estimate the degree of cost pass-through, including a focus on cumulative level pass-through. Finally, we explore the presence of learning dynamics over time. Some of the key findings are:

Finding 1: Firms exhibit a disproportionately stronger and more persistent adjustment in response to large cost-push shocks compared to smaller shocks. This asymmetry suggests that firms perceive large shocks as requiring larger responses to preserve margins and highlights the non-linear nature of firm responses, consistent with models of state-dependent pricing and other findings in the literature.

Finding 2: Expected costs have a positive and significant effect on prices across all treatments. While this contemporaneous pass-through is significant, it is incomplete, and there is evidence of inertia in price setting. Firms continue to adjust their prices for a few periods following a shock.

Finding 3: When changing their prices, firms place a higher weight on their expectations of market prices than on their expectations of costs. This highlights the importance of strategic considerations, including competition, in pricing decisions.

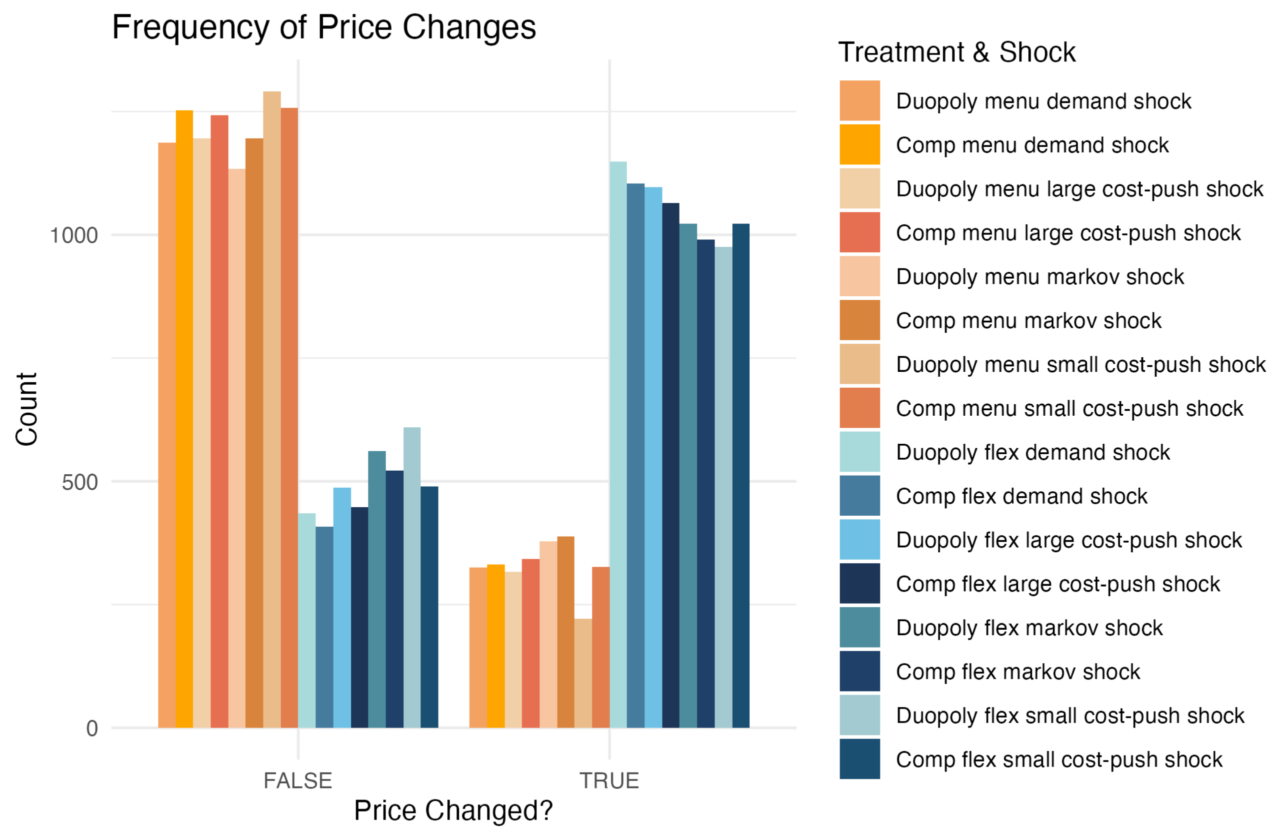
Finding 4: Past prices strongly influence current pricing decisions. This effect is particularly pronounced in menu cost treatments, which shows that nominal rigidities have a strong influence in reinforcing persistence in pricing.

Finding 5: When faced with uncertainty in costs, such as in a dynamically changing tariffs world, firms choose to adopt a careful "wait-and-see" strategy, changing their prices when cost changes materialize rather than immediately in response to uncertainty.

3.4.1 Descriptive Analysis of Experimental Data

Price adjustment behavior differs largely across different institutional settings, varying in market structure and price rigidity. The frequency of price changes, shown in Figure 3-5, is roughly twice as high in flexible pricing treatments compared to those with menu costs. This shows that the small menu costs introduced had a large impact on firms' price decision-making. Importantly, this can also be seen across all types of shock sessions, suggesting that the presence of adjustment costs systematically dampens the propensity to reoptimize prices across different scenarios.

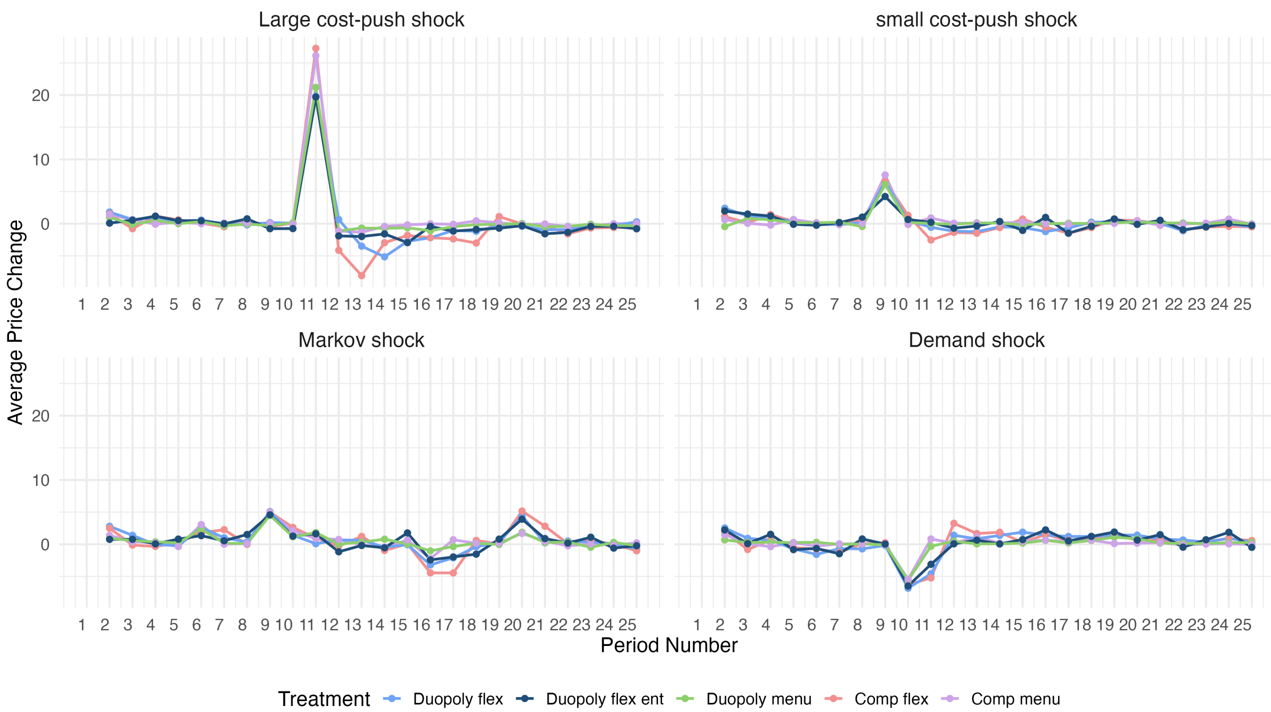
Figure 3-5: Frequency of price changes across treatments for all different sessions.



We also observe large differences in the magnitude of price adjustments across treatments and shock types. As shown in Figure 3-6, average price changes spike immediately after shock announcements, exactly one round after the actual shock occurs. This shows that there is a prompt response once firms are informed that a shock occurred. Interestingly, the average increase in the large cost shock treatment is proportionally greater than in the small cost treatment, suggesting that firms respond more aggressively when profitability is more clearly threatened. This pattern is consistent with Cavallo et al. (2024), who show that large energy shocks in 2022 generated a much faster and stronger pass-through of costs into prices compared to smaller shocks. Large cost shocks push many firms' markups far below desired levels, inducing both a greater incidence and a larger size of price increases.

Under the Markov (uncertainty) shocks, price responses appear more dispersed and muted, consistent with firms' uncertainty about the persistence or direction of the cost disturbance. Unlike the large and small cost-push shocks, which provide firms with clear signals about the size and direction of the cost change, the stochastic nature of the Markov shock complicates decision-making. Firms face the risk of adjusting prices prematurely, only to see costs revert, which would leave them worse off due to unnecessary menu costs or lost competitiveness. This is aligned with theories of state-dependent pricing under uncertainty, where greater volatility in the underlying cost process reduces the perceived benefits of immediate adjustment relative to the risks of overshooting.

Figure 3-6: Average price change by treatment and session



Patterns in firms' markups over time reveal further insights into their adjustment behavior. As shown in Figure 3-7, average markups fall sharply in the period when the shock occurs. This initial drop reflects the fact that firms do not anticipate the exact timing of the shock, and thus fail to adjust their prices immediately, resulting in a temporary erosion of profit margins. In the round that follows, firms revise their prices upward, restoring markups toward pre-shock levels.

Differences across treatments are especially pronounced during cost-push shocks. Firms in menu cost environments tend to maintain higher average markups than those in flexible price settings. This suggests that when price adjustment is costly, firms may adopt a precautionary strategy by setting higher prices to buffer against future shocks or to compensate for infrequent adjustments. The result is a pricing approach that sustains elevated margins despite temporary shocks.

Interestingly, under the Markov (uncertainty) shock scenario, markup dynamics suggest that firms may be more careful about changing their prices in the face of uncertainty, choosing to actively monitor and selectively respond as the shock materializes into actual cost changes, rather than in response to the uncertainty itself. This holds regardless of whether or not firms face menu costs, and can be linked to the overall reduced frequency of price changes observed in Markov shock sessions.

Figure 3-7: Average markup by treatment and session

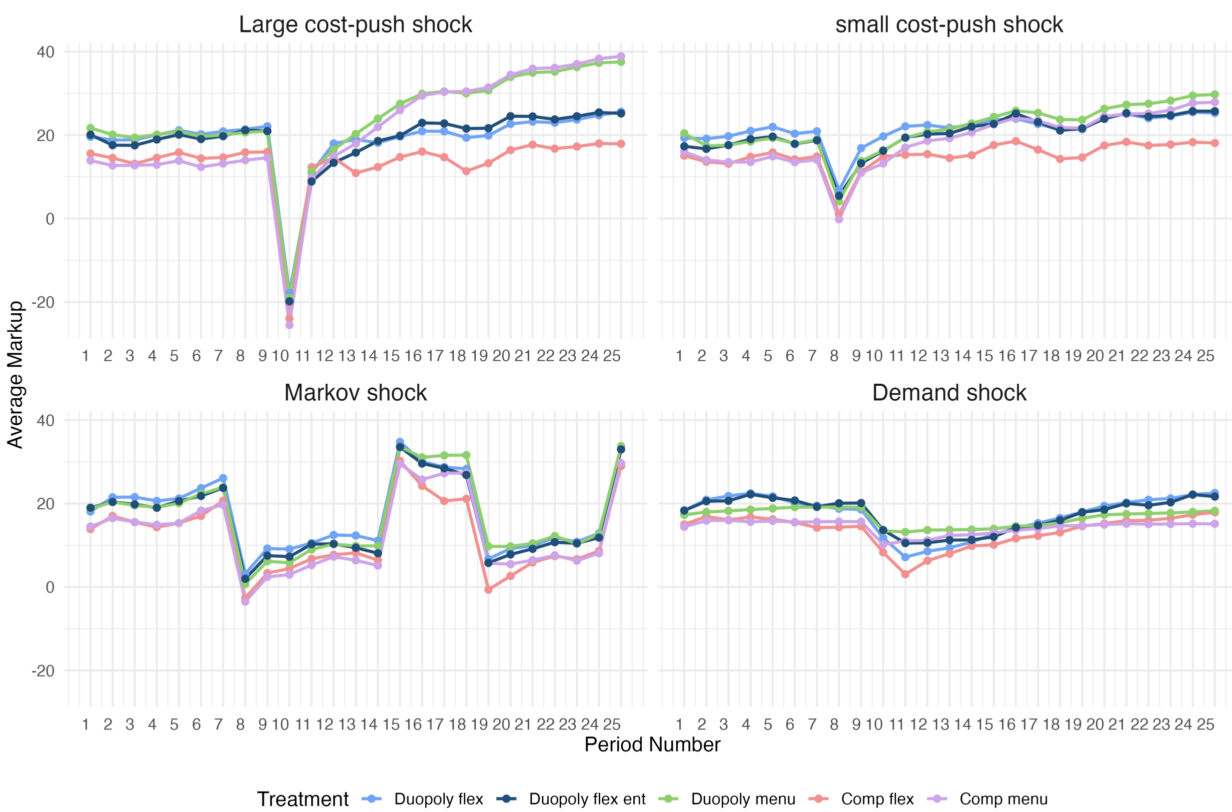
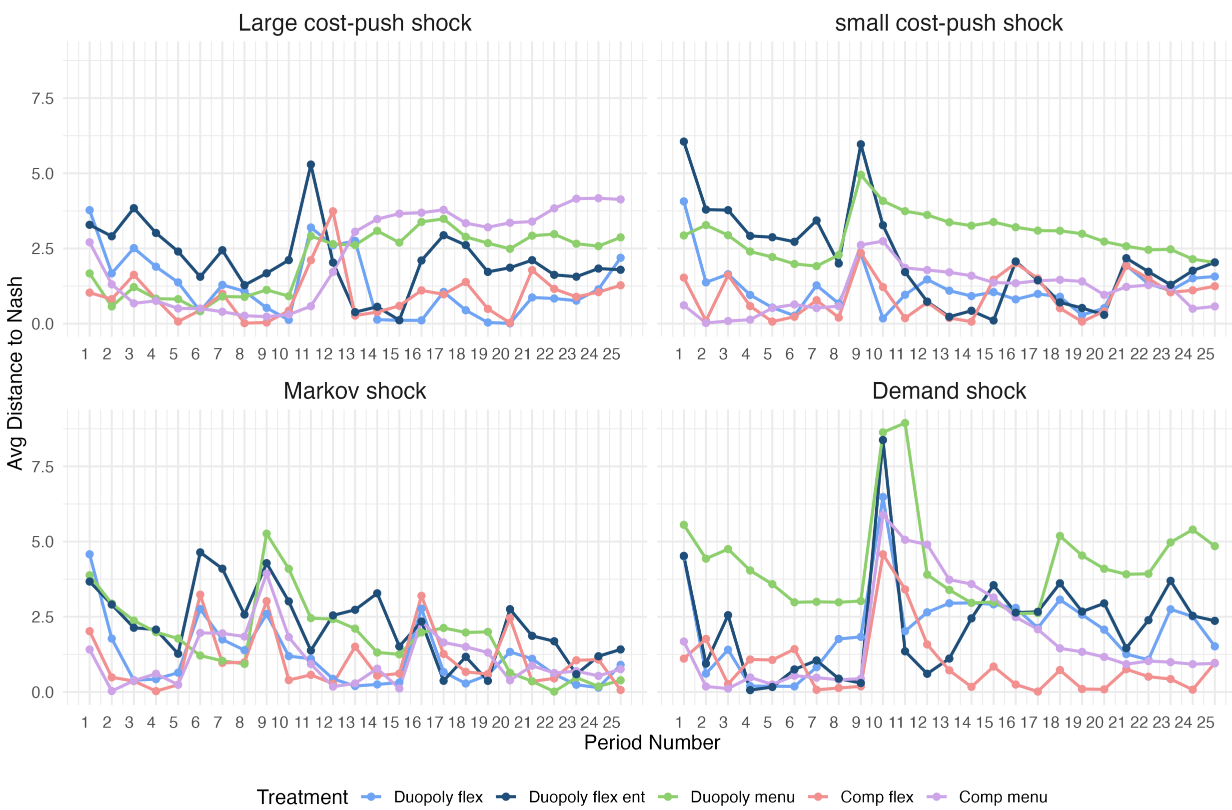


Figure 3-8 shows that the average distance between firms' actual prices and the Nash equilibrium price varies across treatments. Across all panels, shocks are followed by clear departures from equilibrium, reflecting the fact that firms are initially unaware of the shock period. However, the speed and extent of reversion differ sharply by both shock type and institutional environment. The large cost-push shock generates more persistent departures from equilibrium, compared to the small cost-push shock. Nominal rigidities introduced by menu costs seem to amplify deviations from Nash. Treatments with menu costs—particularly in the duopoly case—show both higher peaks and slower convergence back toward equilibrium.

This suggests that competition plays a role in aligning behavior with equilibrium. In competitive treatments—where more firms interact and market power is diluted—firms appear to adopt pricing strategies that are more consistent with Nash equilibrium predictions, possibly due to stronger incentives to respond to market signals. By contrast, in duopoly treatments, strategic complexity or coordination dynamics may lead to more persistent deviations from equilibrium pricing.

Figure 3-8: Average distance to nash price for different treatments and sessions.



3.4.2 Impulse Response Estimation via Local Projections

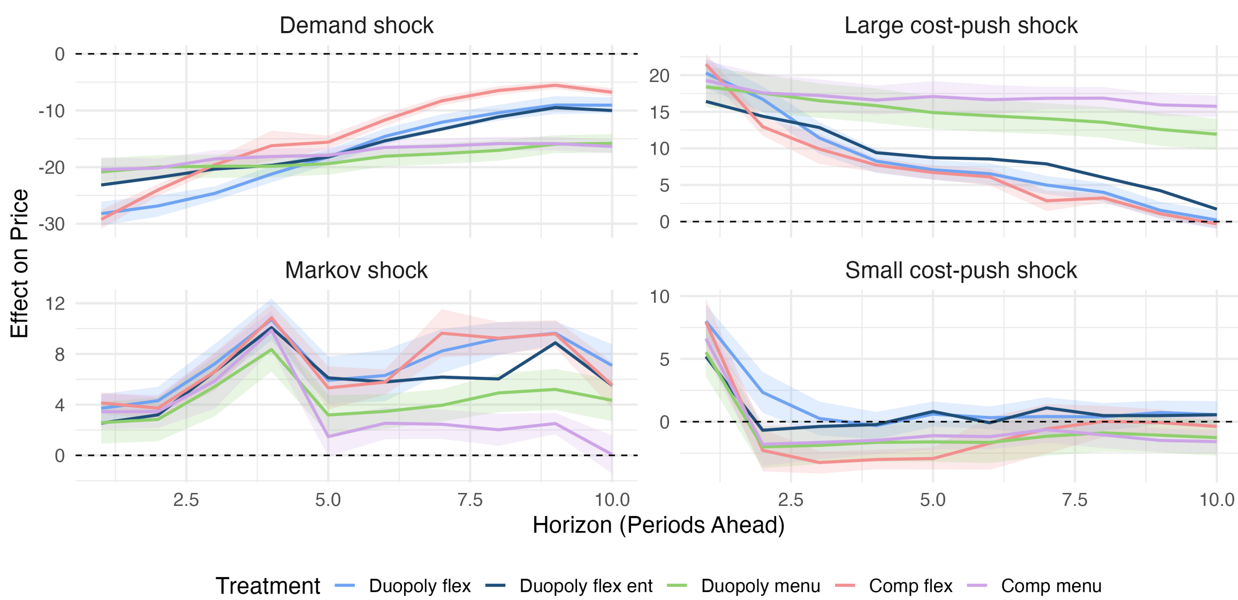
To study the dynamic causal effect of shocks on firm pricing decisions, we estimate impulse response functions (IRFs) using the local projection method introduced by Jordà (2005). Let $p_{i,t}$ denote the price set by firm i in round t, and let $D^s_{i,t}$ be an indicator equal to one if firm i receives a shock of type $s \in \mathcal{S}$ at round t, and zero otherwise. For each horizon $h = 1, 2, \ldots, H$, we estimate the following equation:

$$p_{i,t+h} = \alpha^s_h D^s_{i,t} + \lambda_t + \mu_i + \varepsilon^{(h)}_{i,t}, \tag{3.6}$$

where α^s_h captures the effect of the shock on prices h periods later. λ_t represents the round-in-game fixed effects, and μ_i firm fixed effect which accounts for time-invariant heterogeneity in pricing behavior across firms. The residual term $\varepsilon_{i,t}^{(h)}$ captures all remaining variation.

This specification is estimated separately for each shock type s and each treatment condition defined by market structure (duopoly versus competition) and the presence or absence of menu costs. For each regression, the coefficient α_h^s traces out the impulse response of prices at horizon h to a shock of type s. Because pricing decisions may be serially correlated within firms, we cluster standard errors at the firm level to obtain valid inference. Figure 3-9 shows the impulse responses for the 4 different shocks across all 4 different treatments.

Figure 3-9: Impulse responses from local projections for different treatments and shocks.



A few key observations emerge. First, In the case of uncertainty, firms seem to dynamically adjust their prices as information becomes available. This ”wait-and-see” approach during times of uncertainty is well documented in the literature, including in Aruoba et al. (2024), who show that firms frequently delay full adjustment until additional information resolves.

Second, the sign, size and persistence of the impulse responses vary with the nature of the shock, and these shock-specific patterns interact with pricing frictions. Large cost-push shock produces a much larger effect as a proportion of the shock on prices, compared to the small cost-push shock, which are not only smaller but also a lot less persistent. This results aligns with Cavallo et al. (2024), who find that large shocks have a larger impact on firm pricing than small shocks.

Third, flexible price treatments display the largest impact responses: without adjustment costs, firms can move sharply toward the new optimal price, knowing they can revise again at no cost as conditions change. By contrast, when menu costs are present, initial adjustments are more restrained. Because firms know shocks will gradually resolve, a larger price change would risk overshooting and force further costly revisions. Smaller, more measured changes therefore minimize the likelihood of repeated adjustments.

3.4.3 Determinants of price setting

To understand the determinants of firms' pricing decisions under different experimental conditions, we estimate a fixed-effects model that incorporates both cost expectations and behavioral variables. The regression equation is specified as follows:

$$\begin{aligned} \mathrm{Y}_{it} = {} & \beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,i,t-1} + \beta_4 X_{4,i,t-1} + \beta_5 X_{5,t-1} + \sum_{s \neq \text{baseline}} \gamma_s D_{s,it} \\ & + \sum_{s \neq \text{baseline}} \delta_s (X_{1,it} \times D_{s,it}) + \mu_i + \lambda_t + \varepsilon_{it} \end{aligned} \tag{3.7}$$

The dependent variable is the price set by firm i in round t. $X_{1,it}$ denotes the firm's expected unit cost under rational expectations, while $X_{2,it}$ corresponds to the firm's forecast of the market price. To capture dynamic pricing behavior, the model includes $X_{3,i,t-1}$, the lagged price set by the firm; $X_{4,i,t-1}$, the lagged profit earned; and $X_{5,t-1}$, the lagged market price observed. The terms $D_{s,it}$ are dummy variables indicating the presence of specific shock types—small cost shock, Markov cost shock, or demand shock—with the large cost shock omitted as the baseline category. The interaction terms $X_{1,it} \times D_{s,it}$ allow the pass-through of expected cost to vary depending on the shock type. Firm fixed effects μ_i account for unobserved heterogeneity across firms, and round fixed effects λ_t control for time-specific decisions common to all firms. Standard errors are clustered at the firm level to correct for serial correlation in pricing decisions and ensure that inference remain valid in the presence of within-firm dependence across periods.

We draw four core results from the regression evidence. First, expected costs have a clear and statistically significant effect on prices in every treatment, consistent with the basic pass-through mechanism found in scanner price data in Hong and Li (2017). The magnitude of the response is less than one-to-one, which mirrors the incomplete pass-through widely observed in field data, where competitive pressures, strategic considerations, and frictions prevent full adjustment of prices to underlying cost changes.

Second, the elasticity of prices with respect to firms' own price expectations exceeds the elasticity with respect to expected costs across all treatments. This provides direct evidence of strategic complementarities and markup management: when firms anticipate where prices are heading (their own and their rivals'), that signal weighs more heavily in the short run than their expectations of contemporaneous costs. This pattern aligns with theoretical models of oligopolistic competition, where firms are reluctant to deviate far from expected market prices and instead adjust in tandem with anticipated movements. The survey evidence after the experiment reinforces this interpretation. When asked to describe their main pricing strategy, subjects overwhelmingly referred to the "market price" as the central benchmark, far more frequently than cost-related considerations or idiosyncratic heuristics. Appendix Figure 0-C.4 illustrates this striking dominance: across all four treatments, "market price" is by far the most common collocation in participants' open-ended responses, while terms related to costs, such as "cost production," "cost per unit", or minor adjustments, "slightly higher," "stay close", appear less frequently.

Third, lagged prices are significant in all treatments but larger under treatments with menu costs, indicating greater persistence and thus stronger nominal rigidities when adjustment is costly. This is exactly the propagation expected from menu-cost models: past prices anchor current decisions more tightly when firms internalize the fixed cost of changing prices, producing stickier dynamics. The result highlights the importance of frictions in shaping not just the timing of adjustments but also their magnitude and persistence. In flexible-price treatments, firms re-optimize more freely in each round, so lagged prices matter less. Under menu costs, however, the influence of past prices is magnified, as firms weigh the expense of deviating from established price paths. This evidence is consistent with the classic (S,s) models of price-setting, where menu costs generate inaction bands and greater dependence on historical pricing.

Finally, we find evidence of reduced pass-through to prices when costs follow a Markov (uncertainty) process. Relative to deterministic cost variation, firms translate a smaller share of expected cost changes into prices under uncertainty, aligning with "wait-and-see" behavior documented in the literature. When volatility about the path of fundamentals rises, the option value of inaction grows and price responses are muted and more gradual.

Table 3-2: Regression results for model with different treatments, including external validation results from entrepreneurs session. Estimates shown with standard errors in parentheses.

Variable	**Duopoly Flex**	**Duopoly Flex Ent**	**Duopoly Menu**	**Comp Flex**	**Comp Menu**
Expected Cost	0.350***	0.287***	0.282***	0.217***	0.282***
	(0.084)	(0.056)	(0.032)	(0.071)	(0.036)
Small Cost Shock	0.406	0.498	0.938	0.073	1.013
	(0.680)	(1.043)	(0.736)	(0.837)	(0.811)
Markov Shock	1.318*	1.019	2.426***	-0.063	4.109***
	(0.732)	(0.931)	(0.604)	(0.681)	(0.687)
Demand Shock	-0.266	-0.290	-1.018***	-0.104	-1.333***
	(0.196)	(0.260)	(0.225)	(0.158)	(0.224)
Forecast	0.499***	0.394***	0.389***	0.675***	0.397***
	(0.122)	(0.058)	(0.067)	(0.085)	(0.050)
Lagged Price	0.473***	0.620***	0.715***	0.292***	0.528***
	(0.057)	(0.065)	(0.085)	(0.060)	(0.034)
Lagged Profit	0.001	-0.000	0.000	-0.001	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Lagged Market Price	-0.239***	-0.229**	-0.244***	-0.174***	-0.057
	(0.077)	(0.096)	(0.059)	(0.043)	(0.035)
ExpCost × Small Cost	0.013	-0.008	-0.040	0.023	-0.060*
	(0.032)	(0.041)	(0.032)	(0.038)	(0.035)
ExpCost × Markov Shock	-0.052**	-0.049*	-0.128***	0.008	-0.203***
	(0.024)	(0.029)	(0.026)	(0.023)	(0.029)
Observations	6,336	6,048	6,048	6,048	6,336
Adj. R^2	0.857	0.820	0.925	0.828	0.892
Within R^2	0.827	0.760	0.891	0.803	0.856

Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Standard errors clustered at firm level.

3.4.4 Cumulative Level Pass Through

Inspired by the recent work of Sangani (2025), we examine whether cost pass-through is instantaneous or exhibits inertia. Table 3-3 shows that while much of the pass-through occurs contemporaneously, significant lagged effects remain. The statistically significant coefficients on lagged expected cost changes indicate a delayed adjustment process: firms do not fully incorporate new cost information immediately but instead adjust prices gradually over several periods, with diminishing intensity.

These dynamics also highlight price rigidity differences. In menu cost settings, higher-order lags of expected costs weigh more heavily on pricing than in flexible-price environments. Firms avoid large, immediate adjustments that would quickly misalign prices as costs revert, since repeated changes would trigger further adjustment costs. Instead, they spread responses across periods, tolerating short-run deviations from optimal pricing while incorporating the gradual cost decline into staged adjustments.

Table 3-3: Cumulative Pass-Through Regression Results. Estimates shown with standard errors in parentheses.

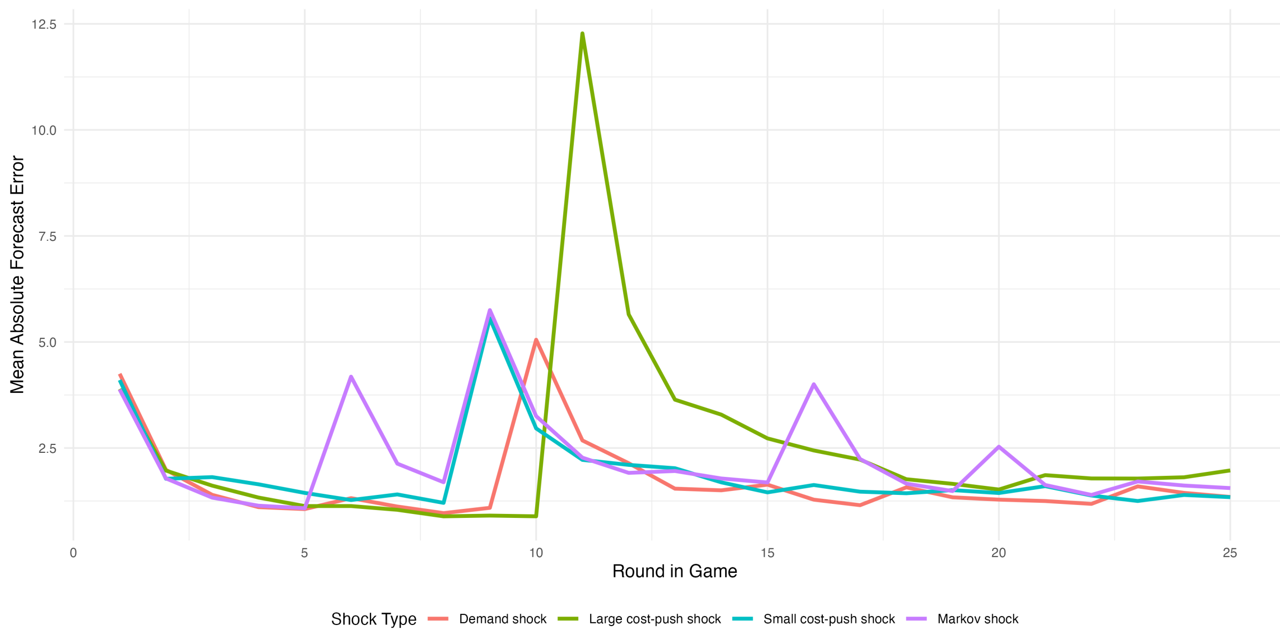
Variable	**Duopoly Flex**	**Duopoly Ent**	**Duopoly Menu**	**Comp Flex**	**Comp Menu**
	0.585^{***}	0.547^{***}	0.551^{***}	0.779^{***}	0.667^{***}
*d_exp_cost_lag*0	(0.051)	(0.050)	(0.051)	(0.051)	(0.048)
	0.152^{***}	0.087^{**}	0.059^{**}	0.137^{**}	0.060
*d_exp_cost_lag*1	(0.027)	(0.029)	(0.019)	(0.046)	(0.037)
	0.028^{*}	0.057^{**}	0.051^{***}	-0.073^{***}	0.035^{*}
*d_exp_cost_lag*2	(0.016)	(0.016)	(0.010)	(0.017)	(0.019)
	-0.041^{*}	0.018	$0.025^{\cdot}$	-0.005	0.029^{**}
*d_exp_cost_lag*3	(0.022)	(0.011)	(0.014)	(0.019)	(0.010)
	-0.007	-0.010	0.046^{***}	0.030^{*}	0.056^{***}
*d_exp_cost_lag*4	(0.024)	(0.026)	(0.008)	(0.016)	(0.013)
	-0.003	$0.041^{\cdot}$	0.021^{*}	-0.003	0.034^{***}
*d_exp_cost_lag*5	(0.009)	(0.025)	(0.009)	(0.012)	(0.008)
Cumulative Pass-Through	0.714 (0.039)	0.740 (0.046)	0.754 (0.053)	0.865 (0.020)	0.881 (0.041)
Observations	5,016	4,788	4,788	4,788	5,016
RMSE	4.544	5.593	3.683	4.901	4.280
Adj. R^2	0.248	0.151	0.300	0.336	0.317
Within R^2	0.257	0.161	0.308	0.345	0.326

Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$, $^{\cdot}p < 0.15$; standard errors clustered at firm level.

3.4.5 Forecast Accuracy and Learning Dynamics

Given the central role of forecasts in determining firms' pricing decisions, we examine whether participants improve the accuracy of their forecasts over time. Figure 3-10 displays the average absolute forecast error by round, separately for each shock type. The figure suggests a general improvement in forecast accuracy over the course of the game, as reflected by the declining trend in forecast errors after the initial rounds. Sharp spikes correspond to the onset of unexpected shocks, after which participants gradually adjust their expectations. This pattern is consistent with a learning process in which firms update their forecasts in response to new information revealed by surprise shocks.

Figure 3-10: Mean absolute forecast error among different sessions.



We define forecast accuracy as:

$$\text{accuracy}_{i,t} = -\left|f_{i,t} - p_t^{\text{market}}\right|, \tag{3.8}$$

where $f_{i,t}$ is firm i's forecast of the market price in round t, and p_t^{market} is the realized market price.

We estimate the following fixed-effects model:

$$\mathrm{Y}_{it} = \beta_1 \cdot \mathrm{X}_{1,it} + \beta_2 \cdot \mathrm{D}_{it} + \alpha_i + \lambda_s + \varepsilon_{it} \tag{3.9}$$

where Yit denotes forecast accuracy. The variable $X1, it$ measures the round number within each game, allowing us to test for learning over time; a positive coefficient indicates improvement in forecast accuracy with experience. The dummy D_{it} equals one in the first round in which a shock occurs, capturing the mechanical increase in forecast errors due to unanticipated shocks. Participant fixed effects α_i account for persistent differences in forecasting ability across individuals, while treatment fixed effects λ_s control for differences in predictability across shock types. Standard errors are clustered at the participant level to account for serial correlation within participants over rounds.

Table 3-4 presents the regression results. The coefficient on Round in Game is positive and statistically significant, indicating that forecast accuracy improves over time. Each additional round is associated with a 0.018 unit increase in accuracy, consistent with a learning effect as participants accumulate experience. The coefficient on Shock Onset is large, negative, and highly significant, reflecting the immediate decline in forecast accuracy when participants are newly exposed to a shock. This stems from the fact that participants are unaware of the shock period or the shock itself until after it has occurred.

Table 3-4: Forecast Accuracy Regression. Evidence of Learning Over Time.

Variable	**Estimate**	**Std. Error**
Forecast Accuracy and Learning		
Round in Game	0.018^{**}	0.008
Shock Onset	-4.825^{***}	0.208

Observations: 25,800; Fixed Effects: Participant (258), Treatment (4).
Adj. R^2: 0.402; Within R^2: 0.072; RMSE: 3.475.
Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.
Standard errors clustered at the participant level.

3.5 Conclusion

This paper uses a controlled laboratory experiment to investigate how firms set prices in response to economic shocks, providing novel insights into the micro-foundations of nominal rigidities and inflation dynamics. By placing human participants in a simulated Bertrand competition and exposing them to different market structures (duopoly and monopolistic competition) and pricing frictions (flexible vs. menu costs), we were able to precisely observe how firms' expectations, costs, and strategic interactions shape their pricing decisions. Our findings provide clean empirical evidence on several key questions that are difficult to address with field data alone.

Our results highlight several key insights. First, firms' expected costs positively influence prices, but the pass-through is incomplete and varies with shock type and institutional setting. Large shocks trigger more frequent and substantial price adjustments, while small shocks elicit more moderate and less persistent responses. Second, firms place greater weight on their own forecasts of market prices than on cost expectations, indicating that strategic considerations and anticipated competitor behavior are central in determining price changes. Third, menu costs introduce inertia: past prices play a stronger role in pricing decisions under adjustment costs, producing persistent deviations from equilibrium and slower convergence following shocks. Fourth, uncertainty—modeled via Markov cost shocks—dampens pass-through, as firms adopt a cautious, "wait-and-see" strategy to avoid premature or costly price changes.

These results offer important implications for macroeconomic theory and policy. The observed behavioral regularities—including Ss-type pricing, the interaction between menu costs and market power, and the central role of expectations—provide support to state-dependent models in capturing real-world pricing dynamics, and highlight that the effectiveness of monetary policy may depend on the prevailing market structure and the nature of the shocks hitting the economy.

0-A Experimental Instructions

Overview

This is an experiment about price decision-making. If you follow the instructions carefully and make good decisions, you can earn a considerable amount of money, which will be paid to you via electronic transfer after the experiment ends. You can access these instructions at any time during the experiment by clicking the *Instructions* button.

Your Role in the Experiment

You will operate a company in a market that produces and sells a product for profit. This is an **individual decision-making** experiment: all other firms and customers are automated. **You are the only human subject in your market**.

To produce your product, you incur a **unit cost of production**, which varies over time:

- Each producer (including you) experiences **small, unpredictable** changes in costs due to **firm-specific** factors, such as a broken machine or a delivery delay.
- Occasionally, **persistent cost changes** affect all producers equally. They may be caused by supply-chain or trade disruptions, international crises, or other global reasons. These are announced via on-screen messages.

The Market You Operate In

[colback=blue!5,colframe=blue!40!black,title=If player is in treatment with 2 firms:] There are 2 firms operating in the market, including yourself. The **market price** is the **average** of all firms' prices. Since the number of firms is small in the market, your price has a **relatively large influence** on the market price.

[colback=blue!5,colframe=blue!40!black,title=If player is in treatment with 100 firms:] There are 100 firms operating in the market, including yourself. The **market price** is the **average** of all firms' prices. Since there are many firms in the market, your price has a **relatively small influence** on the market price.

Your customers are companies that buy products from all firms in the market as inputs to produce their own product. For your company, **higher prices relative to the market price mean fewer sales but higher profit per unit.** Still, each company in your market produces a slightly unique version of the product, so customers prefer to buy a mix that includes yours.

The demand for your products also varies over time due to two elements:

- Your customers experience **small, unpredictable** changes in their budget due to **firm-specific** factors, such as short-term cash flow problems or unexpected windfall.
- Your customers may occasionally face a **downturn**, due for instance to financial constraints or declining sales, in which case they are forced to **reduce** their spending with your company. On-screen messages will notify you when such a situation occurs.

Your Price-Setting Task and Profit-Based Payoff

At the **beginning** of each period, you **set your unit price**, while observing your unit production costs, your quantities sold, your prices, the market prices, your resulting profits, and any messages about recent large cost shocks **up until the last period** (these market data will be displayed in a table and graphs).

The higher your **margin on each unit sold**, that is the difference between your unit production costs and your unit price, the higher your profit. Your total profit equals your margin multiplied by the quantities you sell in any period:

$$\text{Your profit} = (\text{Your unit price} - \text{Your unit cost}) \times \text{Number of sold units}$$

From period 2 onward, your previous price will be **pre-filled**. You can change it by editing the input and clicking *Post*.

All amounts are in **ECU (Experimental Currency Units)**. Use a **decimal point** (e.g., 12.5), not a comma, when entering numbers.

[colback=blue!5,colframe=blue!40!black,title=If player is in treatment with menu cost 100 or 125:] **Be mindful!!** Changing your price is **costly**: Each time you change your price — **regardless of how much** — a **cost of (100 or 125) ECU** is subtracted from your profit. If you don't change your price, you don't pay this cost.

Your Price-Forecasting Task and Forecasting Bonus

At the **beginning** of each period, you will also **forecast the market price** for that same period. Each time your forecast turns out **within 1 ECU** of the realized market price, you earn a 0.25 Euro bonus.

Example:

At the beginning of period 5, you observe your past prices, unit costs, quantities sold, corresponding profits, market prices, and potential messages about cost changes up until period 4. You then post your unit price and forecast of the market price for period 5. At the beginning of the **next period**, that is in period 6, you will observe what your unit costs turned out to be in period 5, the realized market price which, together with your price, determines your quantities sold and your resulting profit in period 5. You will also observe whether your forecast was accurate enough for a bonus.

Sequence of Markets

You will play 110 periods:

- The first 10 periods are **training** periods: they will **not** count towards your earnings, and there is no time limit.
- The next 100 periods are divided into **four independent sequences of 25 periods** and **do count** towards your earnings. It will be clear to you when these paid sequences start and when a sequence ends and a new one starts. At the beginning of each sequence, all values are reset. Initial costs are always close to 20 ECU.
  You will have **2 minutes** per period. If you do not post your price in time, your previous price will be used and you will not earn any profit or bonus.

Your Final Earnings

Not all sequences count towards your earnings. **Two** sequences have been randomly drawn for your profit-based payoff and **one** for your forecasting payoff. Your cumulative profits of the two selected sequences will be converted into Euros at a rate of **2,000 ECU for 1 Euro**. Forecasting bonuses from the selected sequence are added. You also receive a 7 Euro show-up fee.

Once you finish reading these instructions, you will complete a short quiz. After answering correctly, you will begin the training periods.

0-B Nash Price Equation Derivation

We start with the demand function for firm i:

$$q_i = \alpha - \beta p_i + \theta \bar{p}, \tag{0-B.1}$$

and the profit function:

$$\pi_i = (p_i - c)q_i = (p_i - c)\left[\alpha - \beta p_i + \theta \bar{p}\right], \tag{0-B.2}$$

where $\alpha > 0$ and $\beta > \theta/n > 0$ (with n firms in the market).

Nash Equilibrium Price

Since the market average is given by

$$\bar{p} = \frac{p_i + (n-1)p}{n}, \tag{0-B.3}$$

a change in firm i's price affects the average price by

$$\frac{d\bar{p}}{dp_i} = \frac{1}{n}. \tag{0-B.4}$$

Thus the profit function (with $\bar{p}$ depending on p_i) is

$$\pi_i(p_i) = (p_i - c)\left[\alpha - \beta p_i + \theta\,\bar{p}(p_i)\right]. \tag{0-B.5}$$

Taking the derivative with respect to p_i we have

$$\frac{d\pi_i}{dp_i} = \frac{d}{dp_i}(p_i - c)\left[\alpha - \beta p_i + \theta \bar{p}\right] + (p_i - c)\frac{d}{dp_i}\left[\alpha - \beta p_i + \theta \bar{p}\right]. \tag{0-B.6}$$

Since

$$\frac{d}{dp_i}(p_i - c) = 1, \tag{0-B.7}$$

and

$$\frac{d}{dp_i}\left[\alpha - \beta p_i + \theta \bar{p}\right] = -\beta + \theta\,\frac{d\bar{p}}{dp_i} = -\beta + \frac{\theta}{n}, \tag{0-B.8}$$

the first-order condition becomes

$$[\alpha - \beta p_i + \theta \bar{p}] + (p_i - c)\left(-\beta + \frac{\theta}{n}\right) = 0. \tag{0-B.9}$$

In a symmetric equilibrium, where $p_i = p$ and $\bar{p} = p$, the condition is

$$\alpha - \beta p + \theta p + (p - c)\left(-\beta + \frac{\theta}{n}\right) = 0. \tag{0-B.10}$$

This can be written as

$$\alpha + p(\theta - \beta) - (p - c)\left(\beta - \frac{\theta}{n}\right) = 0. \tag{0-B.11}$$

Expanding the term gives:

$$\alpha + p(\theta - \beta) - p\left(\beta - \frac{\theta}{n}\right) + c\left(\beta - \frac{\theta}{n}\right) = 0. \tag{0-B.12}$$

Grouping the p terms we obtain:

$$\alpha + p\left[(\theta - \beta) - \left(\beta - \frac{\theta}{n}\right)\right] + c\left(\beta - \frac{\theta}{n}\right) = 0. \tag{0-B.13}$$

Focusing on the bracket:

$$(\theta - \beta) - \left(\beta - \frac{\theta}{n}\right) = \theta - \beta - \beta + \frac{\theta}{n} = \theta\left(1 + \frac{1}{n}\right) - 2\beta. \tag{0-B.14}$$

Thus, the first-order condition becomes

$$\alpha + p\left[\theta\left(1 + \frac{1}{n}\right) - 2\beta\right] + c\left(\beta - \frac{\theta}{n}\right) = 0. \tag{0-B.15}$$

Solving for p yields the Nash equilibrium price:

$$p_N^* = \frac{\alpha + c\left(\beta - \frac{\theta}{n}\right)}{2\beta - \theta\left(1 + \frac{1}{n}\right)}. \tag{0-B.16}$$

0-C Supplemental Tables and Graphs

Table 0-C.1: Extended Pricing Model — Using delta profit instead of lagged profit. Estimates shown with standard errors in parentheses.

Variable	**Duopoly Flex**	**Duopoly Menu**	**Comp Flex**	**Comp Menu**
Expected Cost	0.259***	0.235***	0.219***	0.294***
	(0.062)	(0.026)	(0.057)	(0.025)
Small Cost Shock	-0.068	0.557	0.358	0.530
	(0.710)	(0.671)	(0.795)	(0.877)
Markov Shock	1.729**	2.988***	-0.322	4.350***
	(0.755)	(0.619)	(0.655)	(0.732)
Demand Shock	-0.256	-0.864***	0.074	-0.974***
	(0.174)	(0.207)	(0.210)	(0.257)
Forecast	0.496***	0.386***	0.664***	0.363***
	(0.123)	(0.067)	(0.085)	(0.046)
Lagged Price	0.420***	0.717***	0.312***	0.550***
	(0.048)	(0.076)	(0.058)	(0.035)
Delta Profit	-0.002***	-0.003***	-0.001***	-0.004***
	(0.001)	(0.001)	(0.000)	(0.001)
Lagged Market Price	-0.110	-0.209***	-0.164***	-0.035
	(0.072)	(0.060)	(0.040)	(0.034)
ExpCost × Small Cost	0.032	-0.019	0.010	-0.037
	(0.032)	(0.029)	(0.035)	(0.036)
ExpCost × Markov Shock	-0.062**	-0.138***	0.016	-0.204***
	(0.026)	(0.026)	(0.023)	(0.030)
Observations	6,072	5,796	5,796	6,072
Adj. R^2	0.855	0.930	0.831	0.896
Within R^2	0.828	0.901	0.807	0.865

Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Standard errors clustered at firm level.
The interaction ExpCost × Demand Shock is omitted (collinearity).

Table 0-C.2: Extended Pricing Model — Using Expected Cost from adaptive learning. Estimates shown with standard errors in parentheses.

Variable	**Duopoly Flex**	**Duopoly Menu**	**Comp Flex**	**Comp Menu**
Expected Cost (learning)	0.339***	0.278***	0.194***	0.269***
	(0.081)	(0.033)	(0.068)	(0.038)
Small Cost Shock	0.596	1.003	0.220	1.148
	(0.687)	(0.758)	(0.854)	(0.830)
Markov Shock	5.323***	4.719***	2.889***	5.500***
	(1.331)	(0.587)	(1.075)	(0.681)
Demand Shock	-0.107	-0.930***	-0.002	-1.276***
	(0.208)	(0.228)	(0.169)	(0.219)
Forecast	0.504***	0.390***	0.685***	0.398***
	(0.122)	(0.068)	(0.083)	(0.051)
Lagged Price	0.454***	0.712***	0.290***	0.529***
	(0.054)	(0.085)	(0.060)	(0.034)
Lagged Profit	0.001	0.000	-0.001	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Lagged Market Price	-0.207**	-0.235***	-0.155***	-0.050
	(0.089)	(0.058)	(0.043)	(0.035)
ExpCost × Small Cost	0.005	-0.041	0.017	-0.063*
	(0.032)	(0.033)	(0.039)	(0.035)
ExpCost × Markov Shock	-0.192***	-0.206***	-0.094**	-0.247***
	(0.046)	(0.027)	(0.037)	(0.030)
Observations	6,336	6,048	6,048	6,336
RMSE	3.313	3.070	3.168	3.382
Adj. R^2	0.856	0.925	0.828	0.892
Within R^2	0.826	0.891	0.802	0.855

Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Standard errors clustered at firm level.

The interaction expected_cost × demand shock is omitted (collinearity).

Table 0-C.3: Extended Pricing Model — Using Naive Expectations for Expected Cost. Estimates shown with standard errors in parentheses.

Variable	**Duopoly Flex**	**Duopoly Menu**	**Comp Flex**	**Comp Menu**
Expected Cost (naive)	0.270***	0.222***	0.153***	0.217***
	(0.064)	(0.026)	(0.054)	(0.030)
Small Cost Shock	0.528	0.797	0.222	0.781
	(0.559)	(0.619)	(0.690)	(0.679)
Markov Shock	3.787***	3.489***	2.000**	4.317***
	(1.016)	(0.478)	(0.818)	(0.561)
Demand Shock	-0.184	-1.005***	-0.047	-1.363***
	(0.203)	(0.228)	(0.163)	(0.220)
Forecast	0.504***	0.390***	0.685***	0.397***
	(0.122)	(0.068)	(0.083)	(0.050)
Lagged Price	0.454***	0.714***	0.290***	0.529***
	(0.054)	(0.085)	(0.060)	(0.034)
Lagged Profit	0.001	0.000	-0.001	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Lagged Market Price	-0.208**	-0.240***	-0.155***	-0.053
	(0.089)	(0.059)	(0.043)	(0.035)
ExpCost × Small Cost	0.007	-0.033	0.016	-0.048*
	(0.025)	(0.026)	(0.031)	(0.028)
ExpCost × Markov Shock	-0.119***	-0.148***	-0.052**	-0.192***
	(0.031)	(0.021)	(0.024)	(0.023)
Observations	6,336	6,048	6,048	6,336
RMSE	3.313	3.069	3.169	3.380
Adj. R^2	0.856	0.925	0.828	0.892
Within R^2	0.826	0.891	0.802	0.856

Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Standard errors clustered at firm level.
The interaction expected_cost_naive × demand shock is omitted (collinearity).

Figure 0-C.1: Demand shock announcement to subjects

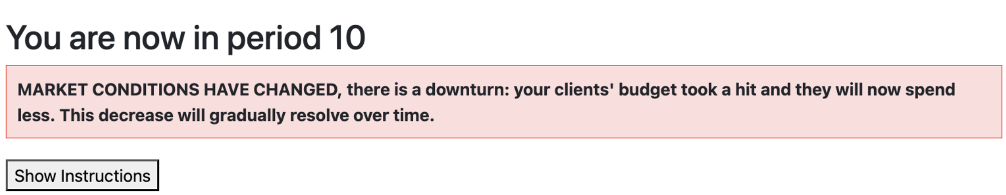


Figure 0-C.2: Markov shock announcement to subjects

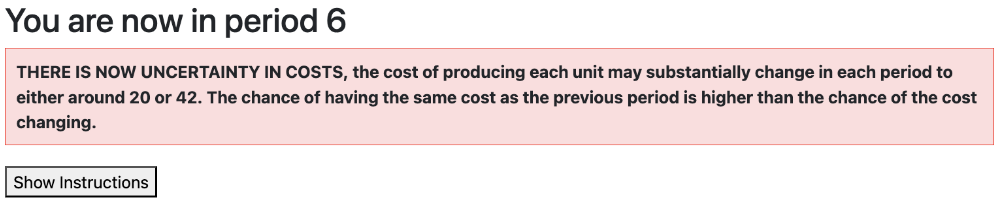


Figure 0-C.3: Cost-push shock announcement to subjects

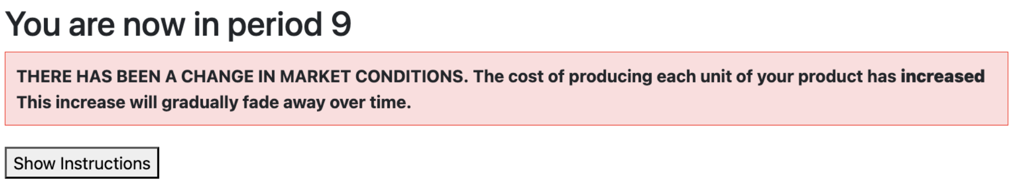


Figure 0-C.4: Collocations of responses among subjects to a question about their main strategy after completing the experiment.

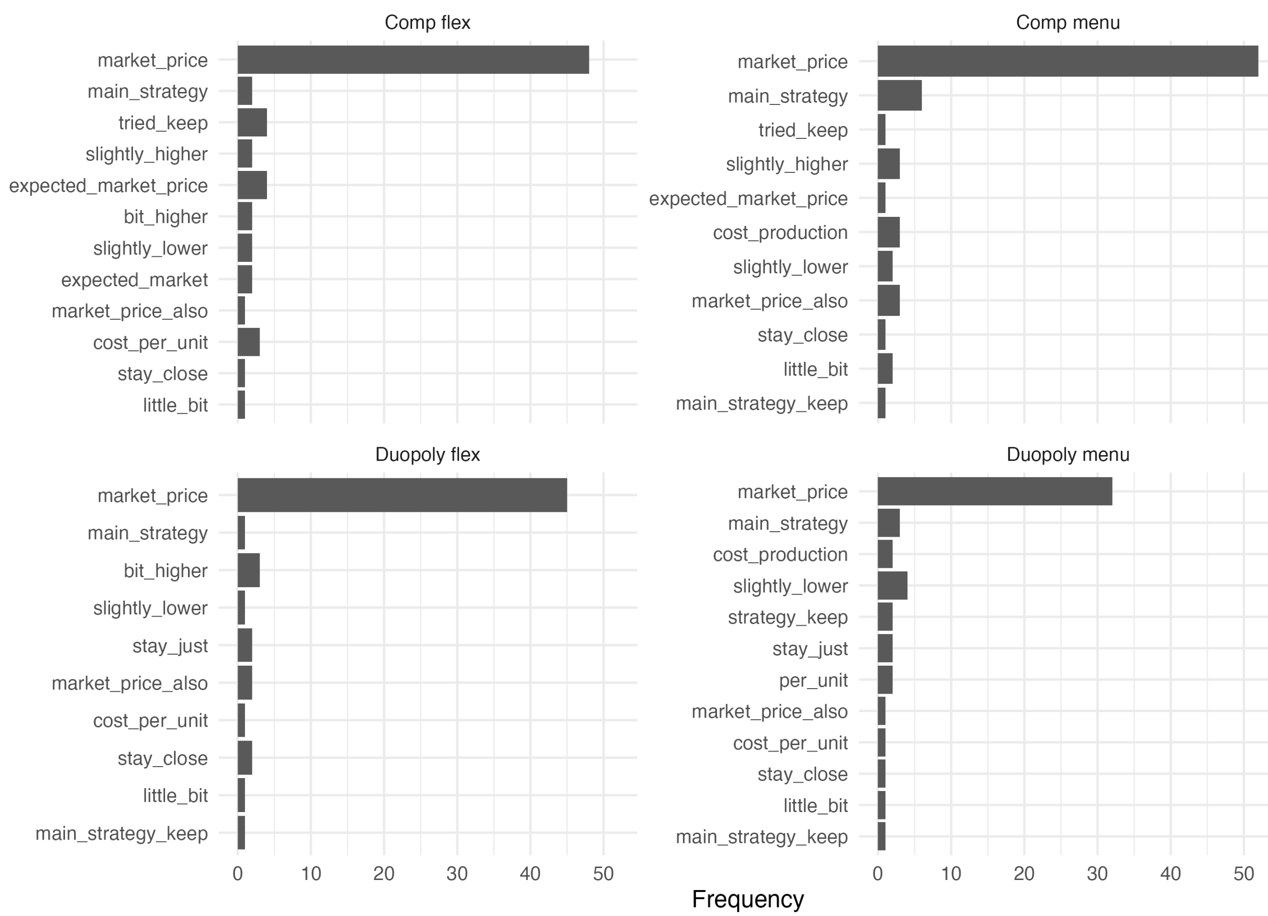


Figure 0-C.5: Individual price changes and average price change across treatments and sessions

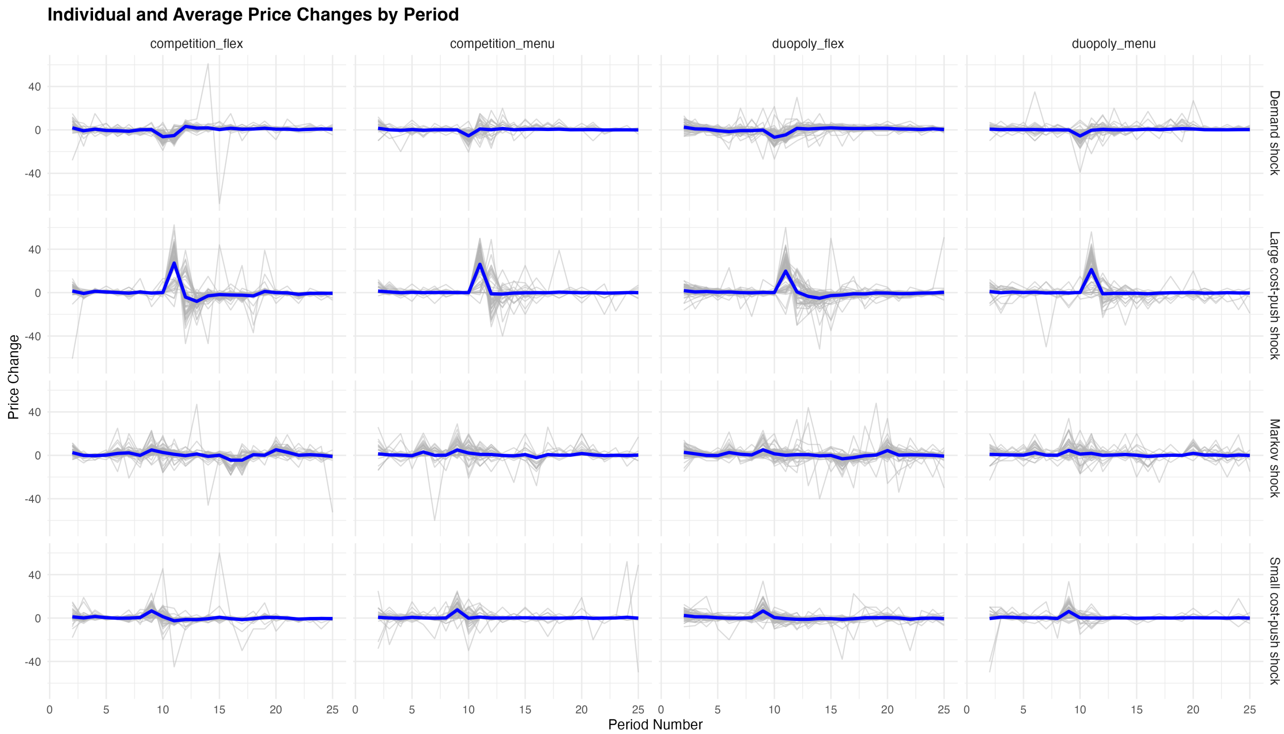
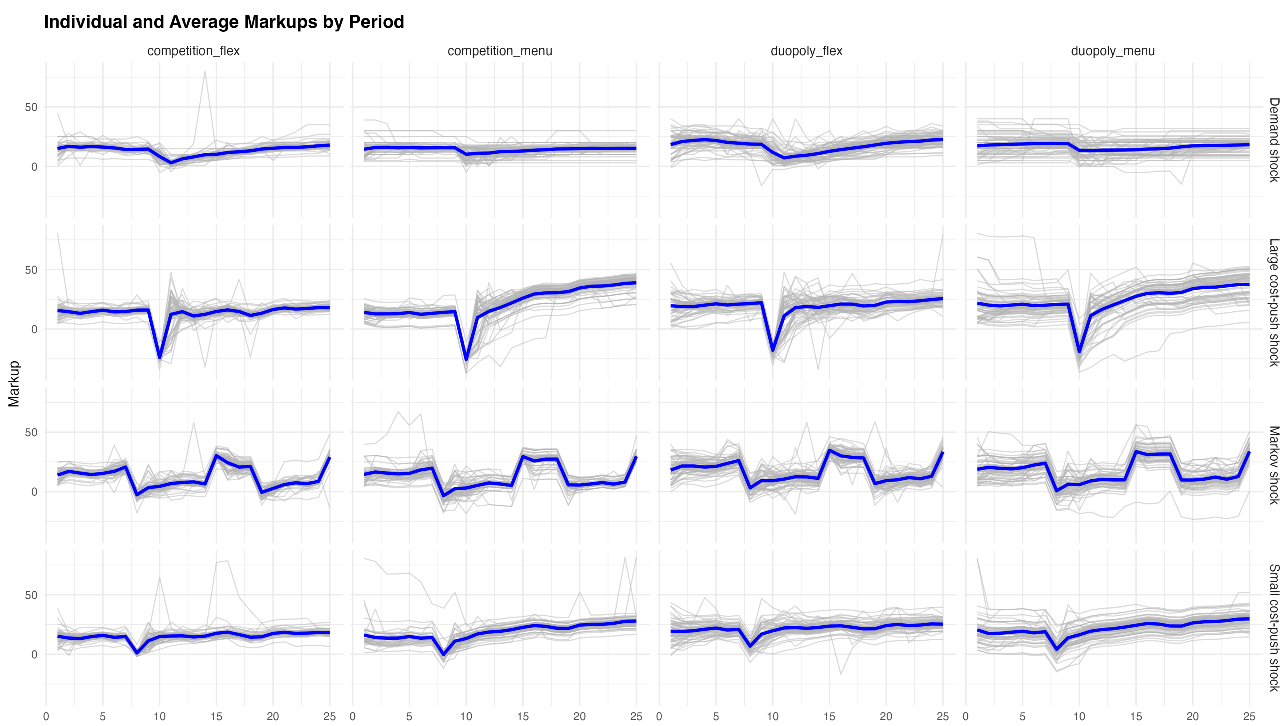


Figure 0-C.6: Individual markups and average markup across treatments and sessions



0-D Supplemental Tables and Graphs for Entrepreneurs

Figure 0-D.7: Time spent by entrepreneurs, average time, and median time. 1 clear outlier removed from data

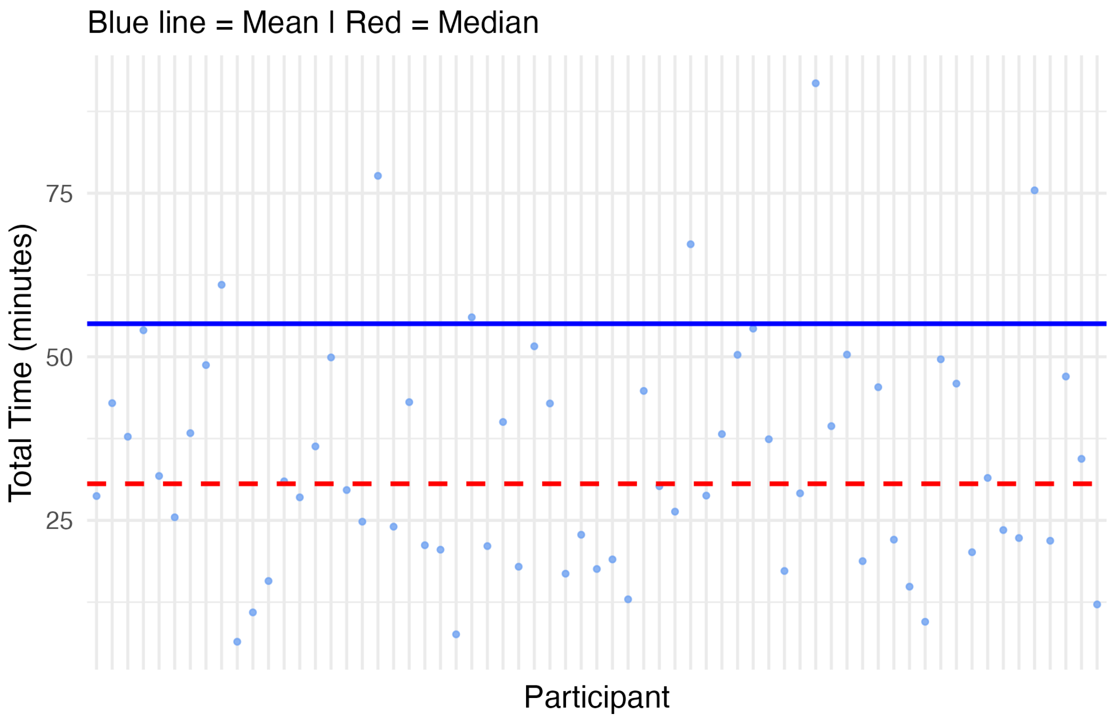


Figure 0-D.8: Payout by entrepreneurs, average payout, and median payout.

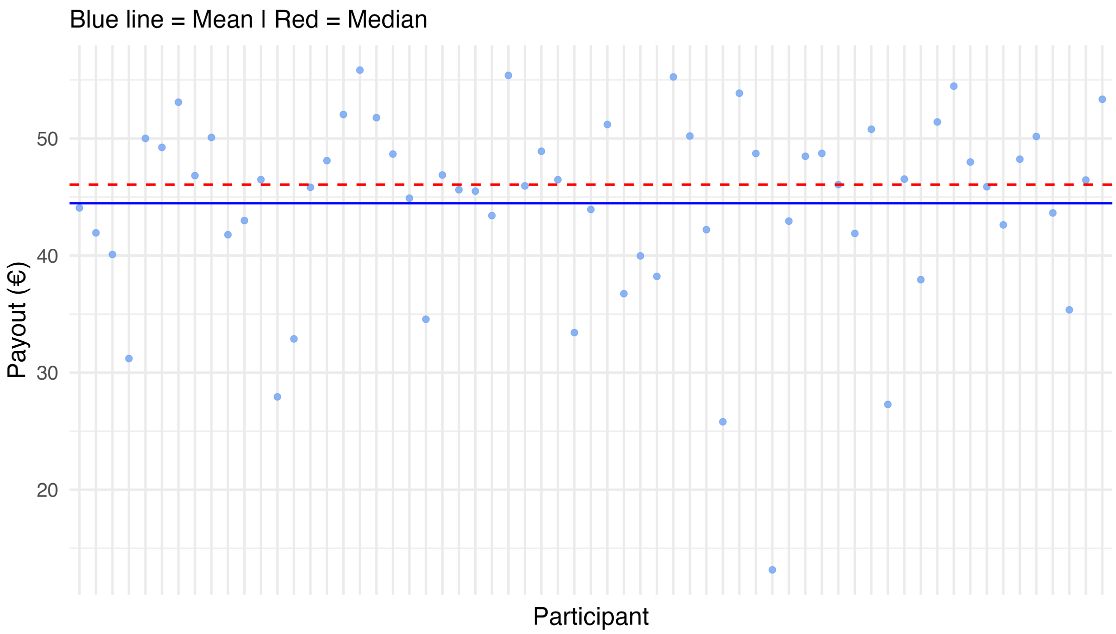


Figure 0-D.9: Time spent by students, average time, and median time. Clear outliers were removed from the data

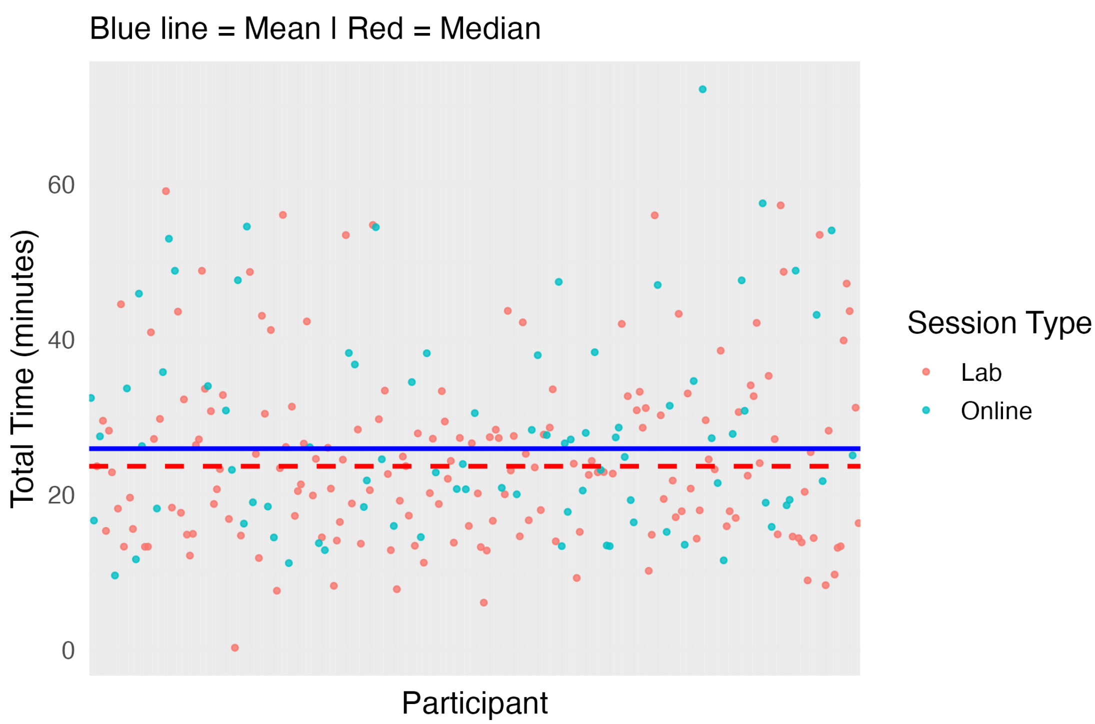


Figure 0-D.10: Frequency of price changes among entrepreneur subjects.

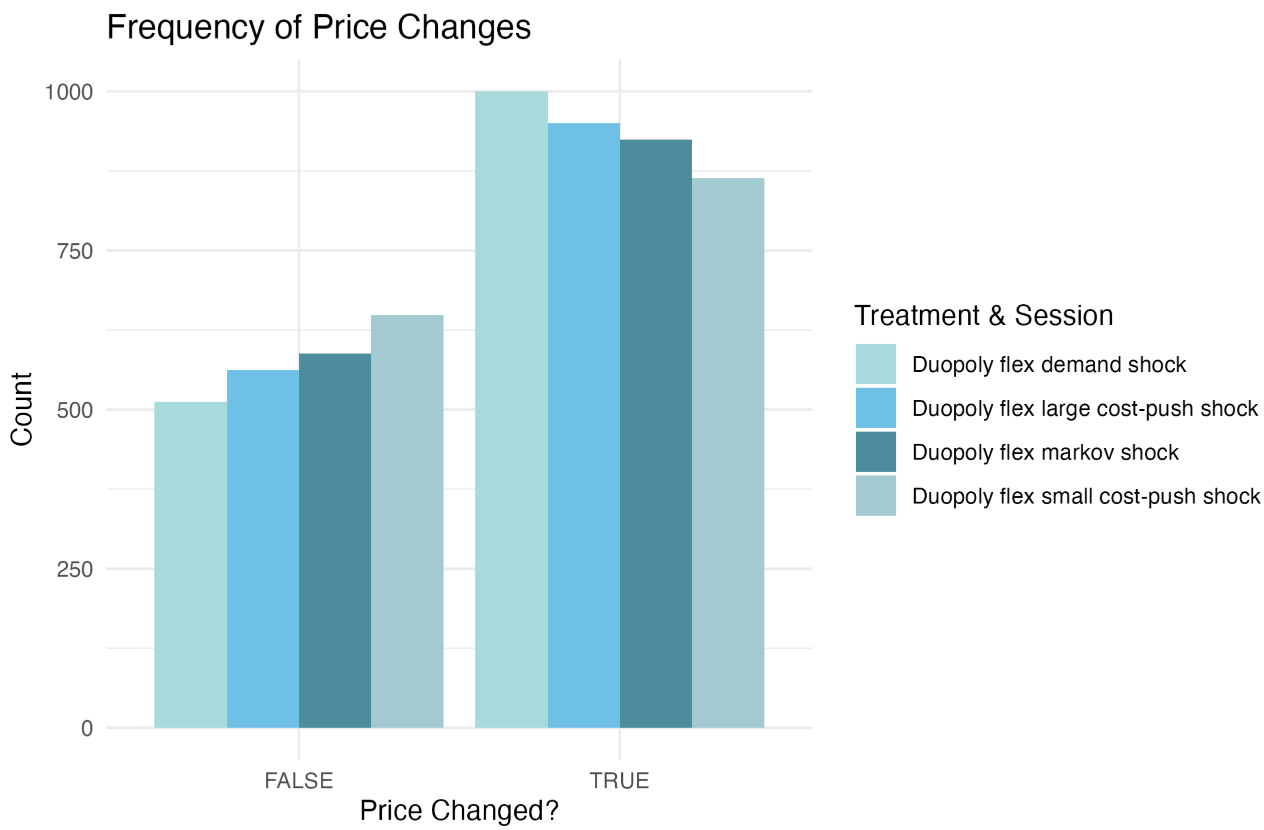


Figure 0-D.11: Average price changes among entrepreneur subjects

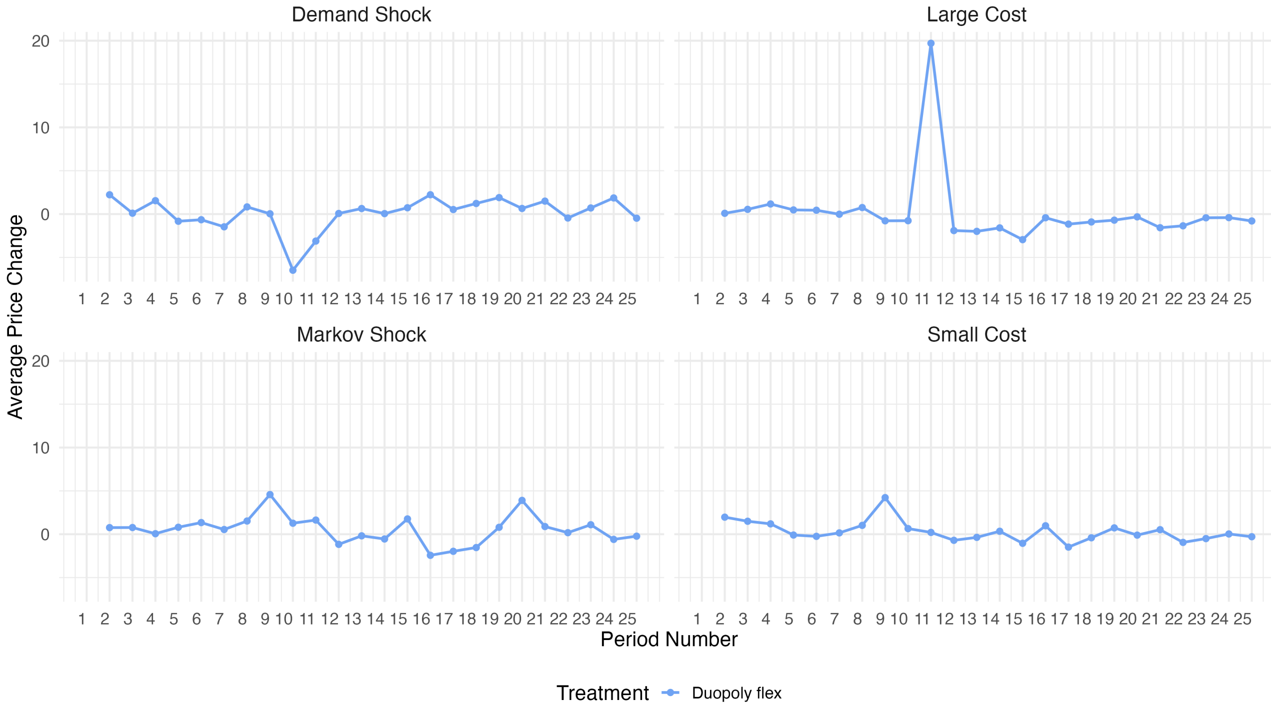


Figure 0-D.12: Average markup among entrepreneur subjects

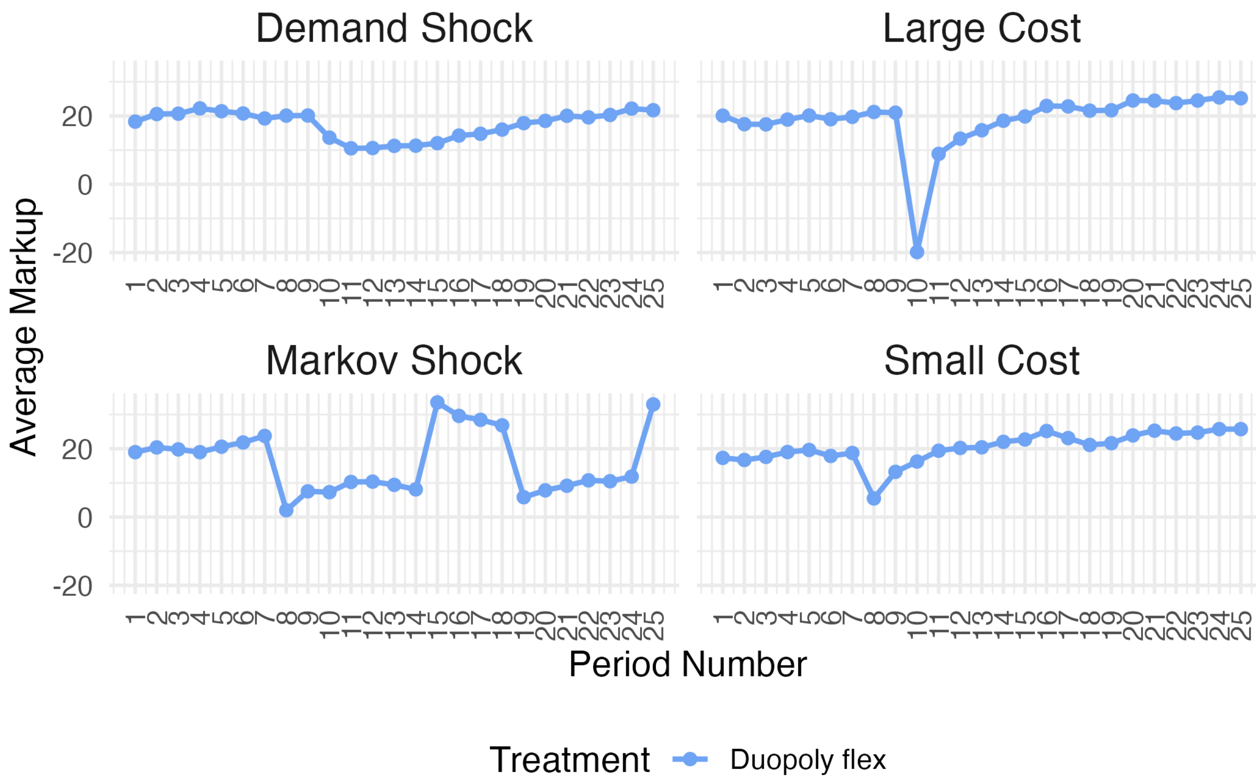


Figure 0-D.13: Average distance to nash among entrepreneur subjects.

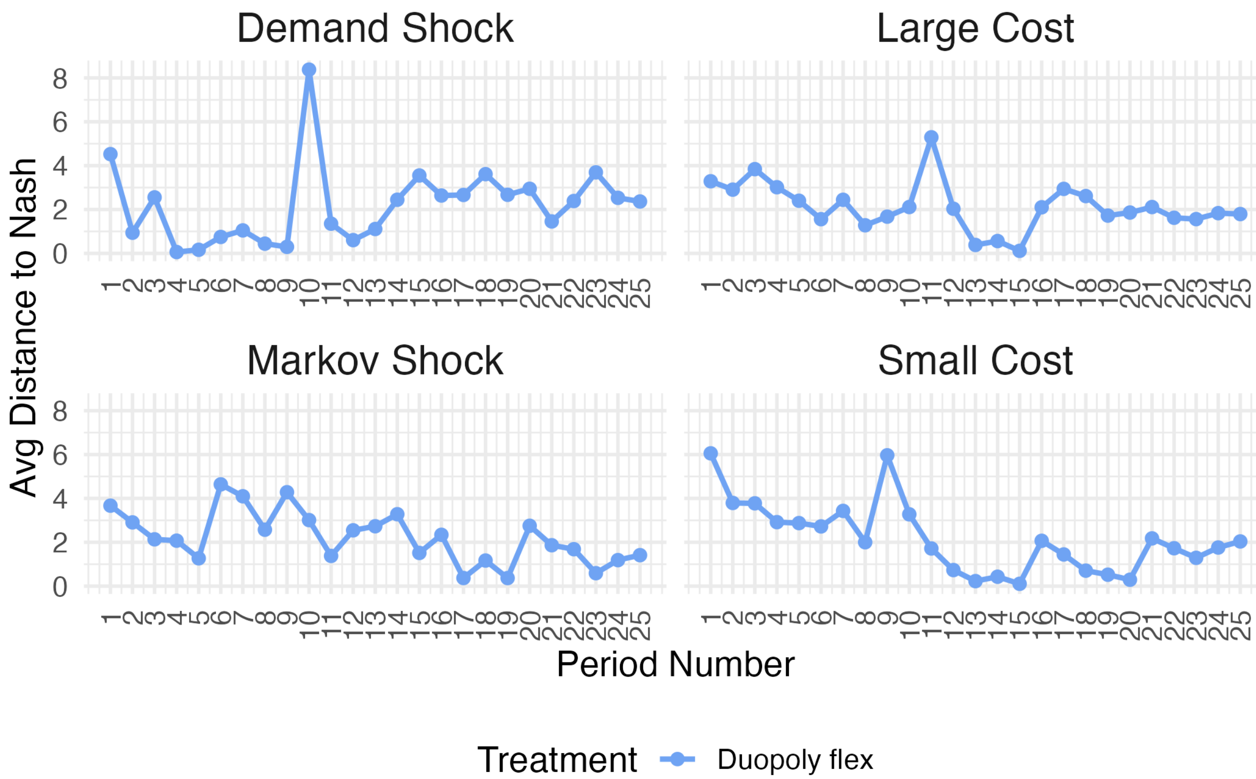


Figure 0-D.14: Impulse responses from local projections for shocks in entrepreneur subjects.

Figure 0-D.15: Forecast errors by shock session for entrepreneurs

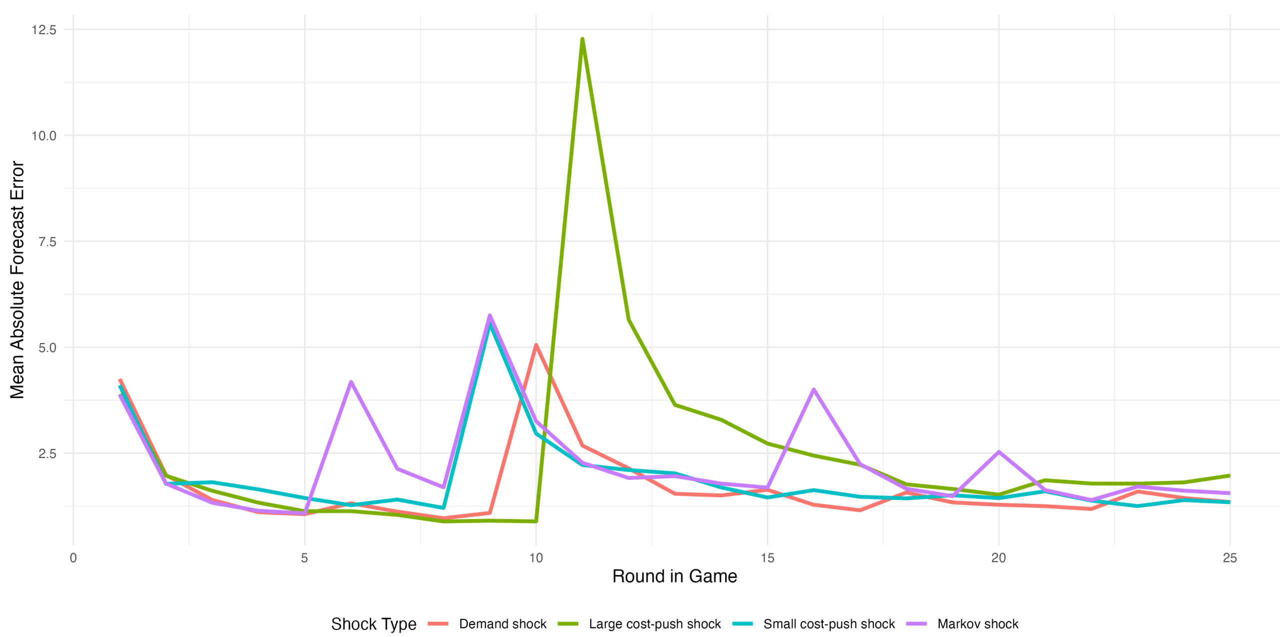


Figure 0-D.16: Individual price changes and average price change across treatments and sessions, for entrepreneurs

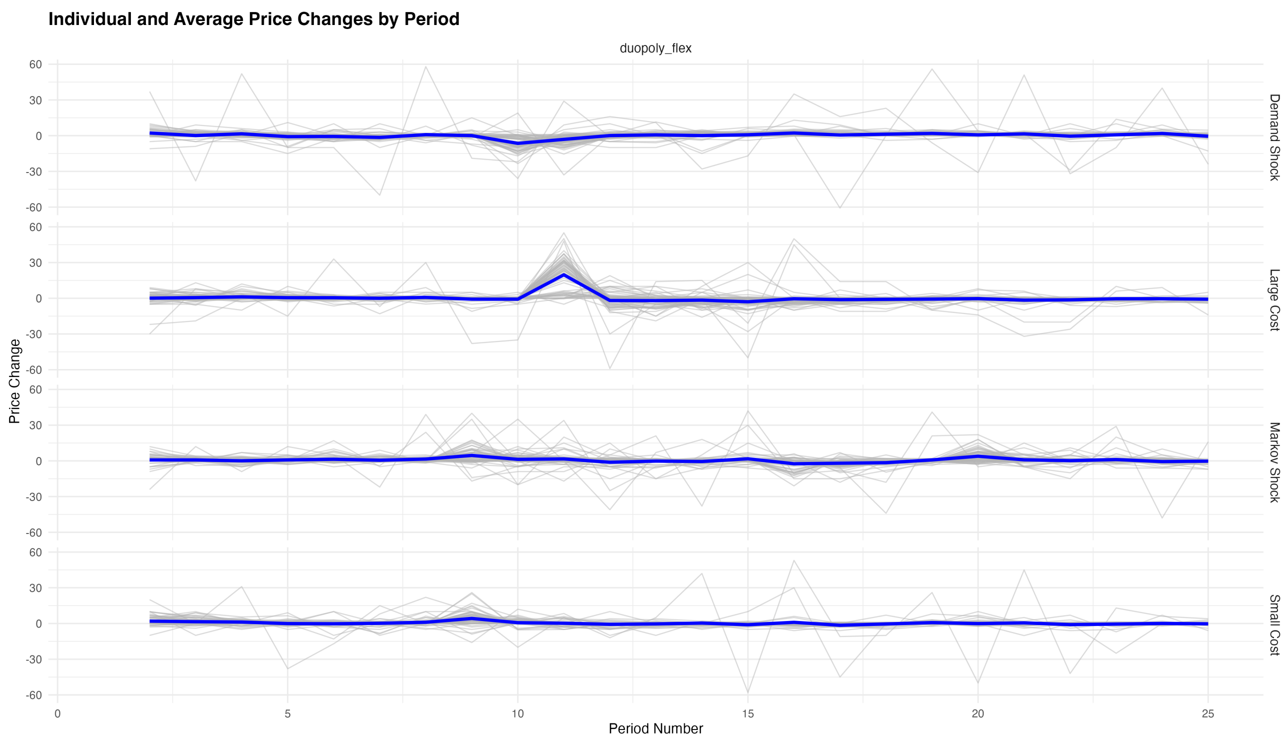
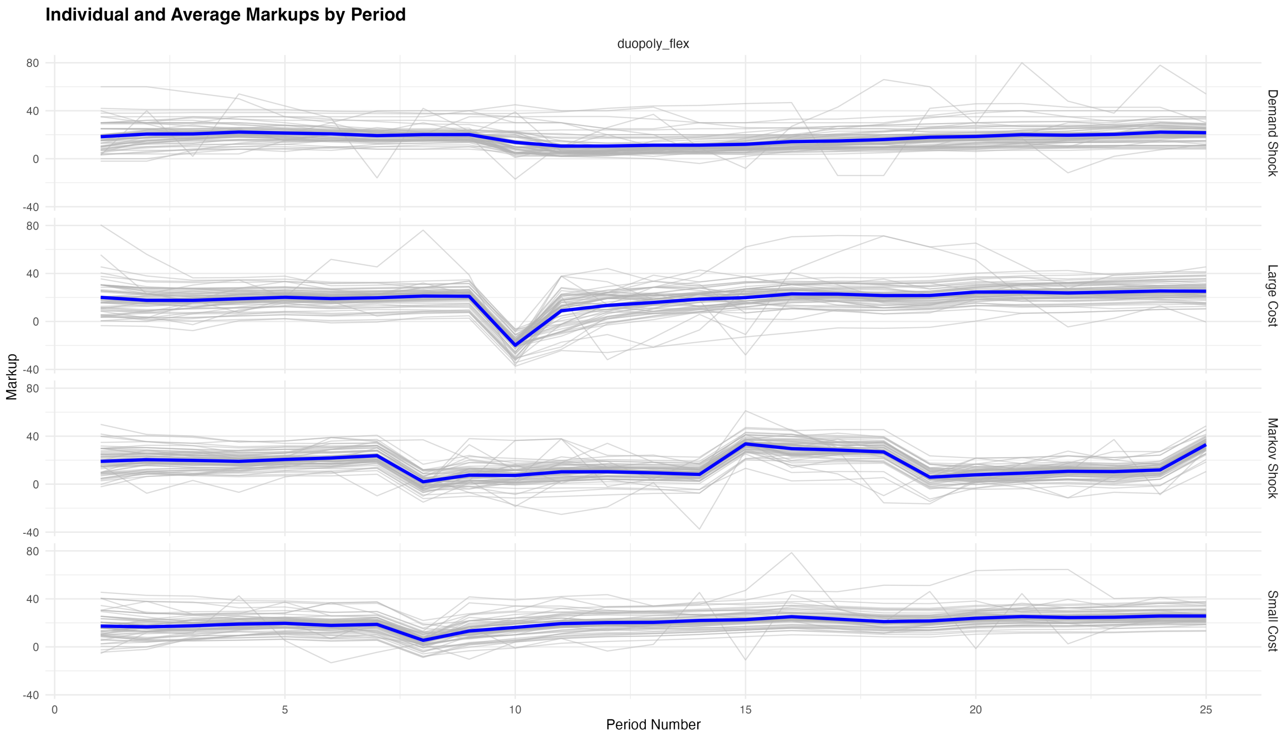


Figure 0-D.17: Individual markups and average markup across treatments and sessions, for entrepreneurs



Bibliography

Abraham, K. G. and Wachter, M. (1987). Help-Wanted Advertising, Job Vacancies, and Unemployment. *Brookings Papers on Economic Activity*, 1987(1):207–248.

Abuallail, I. and Vu, V. (2022). Race Alongside the Machines: Occupational Digitalization Trends in Canada, 2006-2021. *Brookfield Institute for Innovation+ Entrepreneurship*.

Alstadheim, R., Bjørnland, H. C., and Maih, J. (2013). Do central banks respond to exchange rate movements? a markov-switching structural investigation.

Aruoba, S. B., Fernández, A., Guzmán, D., Pastén, E., and Saffie, F. (2024). Pricing under distress. Technical report, National Bureau of Economic Research.

Assenza, T., Grazzini, J., Hommes, C., and Massaro, D. (2015). Pq strategies in monopolistic competition: Some insights from the lab. *Journal of Economic Dynamics and Control*, 50:62–77.

Barth, J. R. and Bennett, J. T. (1974). The role of money in the canadian economy: An empirical test. *The Canadian Journal of Economics/Revue canadienne d'Economique*, 7(2):306–311.

Beine, M., Bos, C. S., and Coulombe, S. (2012). Does the canadian economy suffer from dutch disease? *Resource and Energy Economics*, 34(4):468–492.

Bernanke, B. S. and Blinder, A. S. (1988). Credit, money, and aggregate demand.

Bernanke, B. S. and Blinder, A. S. (1992). The federal funds rate and the channels of monetary transmission. *The American economic review*, 82(4):901–921.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1):387–422.

Cajner, T., Ratner, D., et al. (2016). A Cautionary Note on the Help Wanted Online Data. *FEDS Notes, Board of Governors of the Federal Reserve System*.

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

Carnevale, A. P., Jayasundera, T., and Repnikov, D. (2014). Understanding Online Job Ads Data. *Georgetown University, Center on Education and the Workforce, Technical Report (April)*.

Cavallo, A. (2018a). More amazon effects: online competition and pricing behaviors. Technical report, National Bureau of Economic Research.

Cavallo, A. (2018b). Scraped data and sticky prices. *Review of Economics and Statistics*, 100(1):105–119.

Cavallo, A., Lippi, F., and Miyahara, K. (2024). Large shocks travel fast. *American Economic Review: Insights*, 6(4):558–574.

Champagne, J. and Sekkel, R. (2018). Changes in monetary regimes and the identification of monetary policy shocks: Narrative evidence from canada. *Journal of Monetary Economics*, 99:72–87.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1996). Identification and the effects of monetary policy shocks. *Financial factors in economic stabilization and growth*, 36.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of macroeconomics*, 1:65–148.

Claus, E. and Dungey, M. (2016). Can monetary policy surprises affect the term structure? *Journal of Macroeconomics*, 47:68–83.

Coibion, O. (2012). Are the effects of monetary policy shocks big or small? *American Economic Journal: Macroeconomics*, 4(2):1–32.

Coibion, O., Gorodnichenko, Y., and Kumar, S. (2018). How Do Firms Form Their Expectations? New Survey Evidence. *American Economic Review*, 108(9):2671–2713.

Dedola, L., Gautier, E., Osbat, C., and Santoro, S. (2025). Price stickiness in the euro area. *Research Handbook on Inflation*, pages 265–292.

Evans, D., Mason, C., Chen, H., and Reeson, A. (2023). An Algorithm for Predicting Job Vacancies Using Online Job Postings in Australia. *Nature, Humanities and Social Sciences Communications*, 10(1):1–9.

Fehr, E. and Tyran, J.-R. (2008). Limited rationality and strategic interaction: the impact of the strategic environment on nominal inertia. *Econometrica*, 76(2):353–394.

Fortin-Gagnon, O., Leroux, M., Stevanovic, D., and Surprenant, S. (2022). A large canadian database for macroeconomic analysis. *Canadian Journal of Economics/Revue canadienne d'économique*, 55(4):1799–1833.

Gagliardone, L., Gertler, M., Lenzu, S., and Tielens, J. (2025). Micro and macro cost-price dynamics in normal times and during inflation surges. Technical report, National Bureau of Economic Research.

Genakos, C. and Pagliero, M. (2022). Competition and pass-through: evidence from isolated markets. *American Economic Journal: Applied Economics*, 14(4):35–57.

Golosov, M. and Lucas Jr, R. E. (2007). Menu costs and phillips curves. *Journal of Political Economy*, 115(2):171–199.

Gorodnichenko, Y. and Peter, K. S. (2007). Public Sector Pay and Corruption: Measuring Bribery from Micro Data. *Journal of Public Economics*, 91(5-6):963–991.

Granger, C. W. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics*, 2(2):111–120.

Gürkaynak, R. S., Sack, B., and Swanson, E. (2005). The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models. *American economic review*, 95(1):425–436.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):190–195.

Hassan, C. A. U., Khan, M. S., and Shah, M. A. (2018). Comparison of machine learning algorithms in data classification. In *2018 24th International Conference on Automation and Computing (ICAC)*, pages 1–6. IEEE.

Hong, G. H. and Li, N. (2017). Market structure and cost pass-through in retail. *Review of Economics and Statistics*, 99(1):151–166.

Huber, P. J. et al. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–233. Berkeley, CA: University of California Press.

Jijkoun, V. (2016). Online Job Postings Have Many Duplicates. But How Can You Detect Them If They Are Not Exact Copies of Each Other.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, 95(1):161–182.

Juurikkala, T., Karas, A., and Solanko, L. (2011). The role of banks in monetary policy transmission: Empirical evidence from russia. *Review of international economics*, 19(1):109–121.

Keerthana, B., Reddy, A. R., and Tiwari, A. (2021). Accurate Prediction of Fake Job Offers Using Machine Learning. In *Machine Intelligence and Soft Computing: Proceedings of ICMISC 2020*, pages 101–112. Springer.

Kim, S.-H., Moon, S., and Velasco, C. (2017). Delayed overshooting: Is it an'80s puzzle? *Journal of Political Economy*, 125(5):1570–1598.

Klenow, P. J. and Kryvtsov, O. (2008). State-dependent or time-dependent pricing: Does it matter for recent us inflation? *The Quarterly Journal of Economics*, 123(3):863–904.

Klenow, P. J. and Malin, B. A. (2010). Microeconomic evidence on price-setting. In *Handbook of monetary economics*, volume 3, pages 231–284. Elsevier.

Koeppl, T. V., Kronick, J. M., and McNeil, J. (2024). Using functional shocks to assess conventional and unconventional monetary policy in canada. *Canadian Journal of Economics/Revue canadienne d'économique*.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621.

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of monetary economics*, 47(3):523–544.

Labour Market Information Council (LMIC) (2020). How Representative Are Online Job Postings? Blog post.

LeBlanc, J. D., Civelli, A., Deck, C., and Bregu, K. (2016). State dependent price setting rules under implicit thresholds: An experiment. *Journal of Economic Dynamics and Control*, 68:17–44.

Magnani, J., Gorry, A., and Oprea, R. (2016). Time and state dependence in an ss decision experiment. *American Economic Journal: Macroeconomics*, 8(1):285–310.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

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

Montag, H. and Villar Vallenas, D. (2025). Post-pandemic price flexibility in the us: evidence and implications for price setting models.

Nakamura, E. and Steinsson, J. (2008). Five facts about prices: A reevaluation of menu cost models. *The Quarterly Journal of Economics*, 123(4):1415–1464.

Noussair, C. N., Pfajfar, D., and Zsiros, J. (2015). Pricing decisions in an experimental dynamic stochastic general equilibrium economy. *Journal of Economic Behavior & Organization*, 109:188–202.

Osborne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research, and Evaluation*, 20(1).

Osisanwo, F., Akinsola, J., Awodele, O., Hinmikaiye, J., Olakanmi, O., Akinjobi, J., et al. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3):128–138.

Parry, R. T. (1998). Conducting monetary policy in a global economy. *Business Economics*, pages 14–19.

Pissarides, C. (1986). Unemployment and Vacancies in Britain. *Economic Policy*, 1(3):499–541.

Ragan, C. (2005). The exchange rate and canadian inflation targeting. Technical report, Bank of Canada.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. *Handbook of macroeconomics*, 2:71–162.

Rigobon, R. and Sack, B. (2004). The impact of monetary policy on asset prices. *Journal of monetary economics*, 51(8):1553–1575.

Ritz, R. A. (2024). Does competition increase pass-through? *The RAND Journal of Economics*, 55(1):140–165.