Pontifícia Universidade Católica
do Rio de Janeiro

Samuel Efraim

Essays in Macroeconomics and Financial Econometrics

Tese de Doutorado

Thesis presented to the Programa de Pós-graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia.

Advisor: Prof. Carlos Viana de Carvalho

Rio de Janeiro
Setembro 2025

Pontifícia Universidade Católica
do Rio de Janeiro

Samuel Efraim

Essays in Macroeconomics and Financial Econometrics

Thesis presented to the Programa de Pós-graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia. Approved by the Examination Committee.

Prof. Carlos Viana de Carvalho
Advisor
Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Prof. Marcelo Cunha Medeiros
Department of Economics – UIUC

Prof. Eduardo Zilberman
Department of Economics – PUC-Rio

Felipe Farah Schwartzman
Senior Economist – Richmond Fed

Prof. Miguel Maria Charters de Oliveira Bandeira da Silva
Department of Economics – Insper

Rio de Janeiro, Setembro 29th, 2025

Samuel Efraim

BA in Economics from University of São Paulo in 2021.

Bibliographic data

Efraim, Samuel

Essays in Macroeconomics and Financial Econometrics / Samuel Efraim; advisor: Carlos Viana de Carvalho. – Rio de janeiro: PUC-Rio, Departamento de Economia, 2025.

v., 156 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia.

Inclui bibliografia

1. Economia – Teses. 2. Finanças – Teses. 3. Curva de Phillips;. 4. Heterogeneidade em rigidez de preço;. 5. Volatilidade Realizada;. 6. Normalidade.. I. Carvalho, Carlos. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. III. Título.

CDD: 330

To my family and friends.

Acknowledgments

I am deeply grateful to everyone who supported me along this journey. First and foremost, I thank my family. My mother, Andrea, and my father, Bruno, stood by me at every step—without them, pursuing a PhD would not even have been possible. I am also thankful to my sisters, Lana and Lais, and to my grandmother, Marilena, for their constant encouragement and unwavering belief in me.

I am especially grateful to Rosalia, my partner and the best thing this PhD brought me. Her support made this work possible and made the journey truly worthwhile.

I also want to thank all the teachers and professors who shaped my path, from school to my undergraduate years at USP and my PhD at PUC-Rio. In particular, I am deeply grateful to my advisor Carlos Carvalho and to Marcelo Medeiros, for their guidance, expertise, and friendship, which were fundamental to this work.

I would also like to thank the Department of Economics at PUC-Rio, as well as CAPES, CNPq, and FAPERJ for their financial support. *This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) — Finance Code 001.*

Abstract

Efraim, Samuel; Carvalho, Carlos (Advisor). **Essays in Macroeconomics and Financial Econometrics**. Rio de Janeiro, 2025. 156p. Tese de doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

This dissertation consists of three essays on macroeconomics and financial econometrics. In the first essay, we evaluate each category within the Personal Consumption Expenditures (PCE) basket separately to construct a cyclically sensitive inflation index. Building on Stock and Watson's Common Activity Index (CAI), we show that our category-by-category approach – augmented with a broader set of items – yields superior out-of-sample performance, when evaluating the correlation between our measure and the CAI. We also do an exercise using regional data that provides interesting insights into the dynamics of inflation for different items. In the second essay, we investigate sectoral heterogeneity in price stickiness by deriving both sectoral and aggregate Phillips Curves from a New Keynesian framework. Using PCE data for fifteen sectors, we demonstrate that ignoring heterogeneity biases aggregate slope estimates and that our sector-level price adjustment frequencies closely align with micro-data evidence, thereby bridging the gap between micro price behavior and macroeconomic models. Finally, the third essay examines the properties of realized volatility estimators using high-frequency data on nearly 11,000 U.S. stocks from 2003 to 2020. We compare a variety of measures – addressing microstructure noise and jumps – and find that simpler estimators (e.g., five-minute realized volatility) often outperform more complex methods for illiquid stocks, highlighting important implications for risk management and volatility forecasting.

Keywords

Phillips curve; Heterogeneity in price stickiness; Realized Volatility; Normality.

Resumo

Efraim, Samuel; Carvalho, Carlos. **Ensaios em Macroeconomia e Econometria Financeira**. Rio de Janeiro, 2025. 156p. Tese de Doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Esta dissertação é composta por três ensaios em macroeconomia e econometria financeira. No primeiro ensaio, avaliamos separadamente cada categoria da cesta de *Personal Consumption Expenditures* (PCE) para construir um índice de inflação ciclicamente sensível. Com base no *Common Activity Index* (CAI) de Stock e Watson, mostramos que nossa abordagem categoria a categoria — ampliada por um conjunto mais abrangente de itens — apresenta desempenho superior fora da amostra, ao avaliar a correlação entre nossa medida e o CAI. Também realizamos um exercício com dados regionais que fornece insights interessantes sobre a dinâmica da inflação para diferentes itens.
No segundo ensaio, investigamos a heterogeneidade setorial na rigidez de preços, derivando Curvas de Phillips setoriais e agregada a partir de um arcabouço Novo-Keynesiano. Utilizando dados do PCE para quinze setores, demonstramos que ignorar a heterogeneidade gera viés nas estimativas da inclinação agregada e que nossas frequências de ajuste de preços em nível setorial se alinham de perto às evidências de microdados, reduzindo a distância entre o comportamento microeconômico de preços e os modelos macroeconômicos.
Por fim, o terceiro ensaio examina as propriedades de estimadores de volatilidade realizada usando dados de alta frequência de quase 11.000 ações norte-americanas no período de 2003 a 2020. Comparamos uma variedade de medidas — que tratam de ruído de microestrutura e saltos — e constatamos que estimadores mais simples (por exemplo, a volatilidade realizada em intervalos de cinco minutos) frequentemente superam métodos mais complexos para ações com baixa liquidez, destacando implicações importantes para gestão de risco e previsão de volatilidade.

Palavras-chave

Curva de Phillips; Heterogeneidade em rigidez de preço; Volatilidade Realizada; Normalidade.

Table of contents

List of figures

List of tables

1
Cyclically Sensitive Inflation Components

Abstract: This paper evaluates individual categories within the inflation basket to produce a cyclically sensitive inflation index. Our work is similar to Stock & Watson (2020) in trying to create such a measure. We also take advantage of the filter shown in their paper to be the most effective in excluding undesirable frequency noises. However, unlike their paper that attributes index weights based on the coefficients from the regression of a measure of economic activity that combines several different metrics, the Common Activity Index (CAI), on a partition of categories of the Personal Consumption Expenditures (PCE) simultaneously, we regress the CAI on each category separately. We include in our index the items that exhibit a positive and significant coefficient and keep the relative weight of the selected items proportionally to the original PCE weights. We show that, our measure exhibits a higher correlation with CAI when using out-of-sample data and presents more stability selecting and excluding items. Therefore, our index is potentially a better fit for applications that aim to use the same selection for multiple periods or to preserve the relative importance of the selected items, since our method maintains a more clear intuition and closer relation with the observed inflation. Furthermore, we explore a more granular partition of the PCE, allowing the creation of a more comprehensive aggregate measure. Finally, we repeat the exercise using regional data to get insights on how the itemized regional Phillips-Curve relates to the national one.

Keywords: Phillips curve; Heterogeneity in price stickiness.

1.1 Introduction

Understanding the relationship between economic slack and inflation is a cornerstone of effective monetary policy. Policymakers must navigate the complexities of inflation dynamics to stabilize the economy, especially in times of significant economic disturbances. The recent rise in inflation in the United States has underscored the importance of accurately gauging which components of inflation are truly cyclical, that is, which items systematically co-move with economic activity rather than reflecting supply shocks, measurement error, or idiosyncratic relative price changes.

Stock & Watson (2020) made a substantial contribution to this agenda by proposing a cyclically sensitive inflation index (CSI). Their approach applies non-negative least squares (NNLS) to filtered inflation rates of PCE categories, using a filtered measure of cyclical activity, the Common Activity Index (CAI), as the reference cyclical signal. A key insight in Stock & Watson (2020) is that the empirical stability of the Phillips relation depends not only on the selection of cyclically sensitive components, but also on the treatment of activity: the Phillips relation becomes "resilient" precisely when economic slack is proxied by the CAI and when gaps in conventional slack measures are replaced by bandpass-filtered or year-over-year changes in real activity. In other words, Stock & Watson (2020) provide direct evidence that the stability of the Phillips relationship depends jointly on using their cyclically sensitive inflation index and on proxying slack with the CAI under appropriate filtering choices.

Their methodology is designed primarily to maximize the cyclical variation in the constructed index, delivering an indicator of the response of inflation to cyclical tightness that can be computed in real time and used for monitoring and policy discussion. As Stock & Watson (2020) emphasize, the CSI is not intended to be an inflation target because it does not measure the overall (share-weighted) cost of living. This distinction matters because a separate and increasingly important objective in the literature is to construct a measure of inflation that is both (i) more tightly linked to slack and (ii) less contaminated by supply disturbances, so that it can serve as a cleaner inflation object for Phillips-curve estimation and related empirical exercises. For instance, Barnichon & Shapiro (2024) estimate Phillips curves using narrower inflation measures, including the San Francisco Fed's "cyclical core PCE inflation" (Shapiro *et al.*, 2020), precisely to mitigate endogeneity concerns that become more acute in recent samples.

This paper builds on Stock & Watson (2020) and Shapiro *et al.* (2020) and proposes a unified approach. Like Stock & Watson (2020), we take seriously the evidence that the CAI and filtering choices are central for a stable Phillips relation. Like Shapiro *et al.* (2020), we focus on identifying the inflation components that move systematically with slack. Our core methodological difference relative to Stock & Watson (2020) is that we evaluate cyclical sensitivity at the level of individual items and use this item-by-item evidence to construct a cyclically sensitive inflation measure that remains interpretable as a cost-of-living index *over the cyclically sensitive components within the PCE basket*. Concretely, rather than generating an index whose weights are chosen to maximize cyclical comovement (and which therefore need not preserve a cost-of-living interpretation), we maintain the economic meaning of PCE expenditure shares *within the cyclically sensitive basket*. This yields an inflation measure that retains a clear cost-of-living interpretation, while focusing on components that are empirically linked to slack.

Our motivation for departing from the simultaneous selection in Stock & Watson (2020) is interpretability and selection sharpness. Because their selection is conducted jointly through NNLS weights, the resulting CSI does not yield a sharp item-level conclusion about cyclicality: items may receive small but positive weights due to the non-negativity constraint, and weights can be influenced by cross-category correlations. Consequently, it is difficult to interpret a small CSI weight as evidence that a category is acyclical *per se*. An item-by-item procedure instead delivers a transparent classification and allows us to characterize heterogeneity in cyclical sensitivity across PCE categories.

A second key contribution of our paper is to emphasize and test *stability*. Although stability is not the primary objective of Stock & Watson (2020), they conduct several stability analyses of their selection and weighting procedure. We adopt and extend this perspective in a setting that directly connects to Shapiro *et al.* (2020). Specifically, we compare the stability of our index against the CSI along multiple dimensions, and we find that our approach delivers (i) greater stability of selection and weights and (ii) a higher out-of-sample correlation with cyclical activity measures. Importantly, Shapiro *et al.* (2020) does not focus on stability diagnostics of selection, so our stability analysis also fills a gap in that strand of the literature. In the Appendix, we provide a systematic comparison between our index and Shapiro *et al.* (2020), trying to isolate the roles of the CAI and alternative filtering choices, and clarifying how these design decisions map into differences in stability and out-of-sample performance.

Empirically, we replicate Stock & Watson (2020)'s approach on a more disaggregated set of 75 PCE categories, compared with the more aggregated set used in their main analysis, which allows a direct comparison of robustness and performance at a much finer level of disaggregation. We also perform rolling-window exercises and out-of-sample evaluations to quantify the stability gains of our method relative to the CSI. Our findings underscore that a simpler, item-by-item approach can deliver a more stable cyclically sensitive inflation measure while preserving a cost-of-living interpretation within the cyclically sensitive basket. In the Appendix, we also report a parallel comparison using the same level of aggregation (17 PCE categories) as in Stock & Watson (2020), with qualitatively similar conclusions.

We further extend our analysis by exploiting regional variation, a strategy increasingly adopted in recent studies (e.g., Fitzgerald *et al.* (2024); McLeay & Tenreyro (2020); Hazell *et al.* (2022)). Using disaggregated inflation data from the Bureau of Labor Statistics for 25 regions and 21 PCE categories, combined with corresponding regional unemployment rates, we obtain a larger cross-sectional sample and the ability to include time fixed effects. These fixed effects can potentially help address common endogeneities, such as nationwide monetary policy shocks, by purging shocks shared across all regions. This regional design provides an additional lens on the cyclical behavior of inflation components and offers complementary evidence to our national time-series analysis.

The structure of this paper is as follows: Section 1.2 presents the national-data analysis. Subsection 1.2.1 describes the data sources and preprocessing steps, Subsection 1.2.2 details our methodology and its connection to Stock & Watson (2020) and Shapiro *et al.* (2020), and Subsection 1.2.3 presents the national results and the construction of our cyclically sensitive cost-of-living inflation measure. Section 1.3 presents the regional analysis: Subsection 1.3.1 describes the regional data, Subsection 1.3.2 outlines the regional methodology, and Subsection 1.3.3 reports the regional results. Finally, Section 1.4 concludes.

1.2 National

1.2.1 Data

Our analysis utilizes a comprehensive dataset that includes detailed price categories within the Personal Consumption Expenditures (PCE) as well as various measures of economic activity. This section outlines the sources, preprocessing steps, and key characteristics of the data used in constructing our cyclically sensitive inflation index.

1.2.1.1 Personal Consumption Expenditures (PCE)

The PCE data is obtained from the Bureau of Economic Analysis (BEA). The dataset includes quarterly price indices for a broad range of goods and services consumed by households. These indices are classified into major categories such as food, housing, apparel, transportation, medical care, recreation, education, and communication, among others. Each category is further divided into subcategories, providing a granular view of consumer spending patterns.

For our analysis, we focus on the quarterly price indices at the most detailed level available, covering 75 categories from the period of January 1989 to December 2023. This choice of 75 categories represents the maximum level of detail that allows us to extend our analysis back to 1989. Including more items would significantly reduce the historical coverage of the data. The data is seasonally adjusted and expressed in terms of year-over-year percentage changes to account for inflation dynamics over time. The detailed level of categorization allows us to analyze the cyclicality of individual items within the PCE basket.

1.2.1.2 Composite Index of Real Cyclical Activity (CAI)

To measure economic activity, we use the Composite Index of Real Cyclical Activity (CAI) developed by Stock & Watson (2020). The CAI is a composite measure that combines several key economic indicators, capturing the overall state of economic activity. The indicators included in the CAI are:

- Gross Domestic Output (GDO)
- Capacity Utilization Rate
- Establishment Employment
- Overall Employment-to-Population Ratio
- Employment-to-Population Ratio for Ages 25-54
- Unemployment Rate
- Short-term Unemployment Rate

These indicators are standardized, lagged, and then bandpass filtered to isolate the cyclical component. The CAI is calculated as the first principal component of these filtered series, providing a single, robust measure of cyclical economic activity. The data for these indicators is sourced from various federal agencies, including the Bureau of Labor Statistics (BLS) and the Federal Reserve.

1.2.1.3
Comparison with Stock and Watson

While Stock and Watson used only 17 categories in their analysis, our dataset includes 75 categories. This broader scope allows for a more detailed and comprehensive analysis of the cyclicality of inflation components. However, including a larger number of categories presents challenges for Stock and Watson's method, which uses a regression model that includes all items simultaneously. With a larger number of items (N) and the same number of time periods (T), their method is at a disadvantage due to increased complexity and potential multicollinearity.

In the appendix, we provide a detailed comparison of our results with those obtained using Stock and Watson's method, applied to the same 17 categories used in their paper. This comparison demonstrates the advantages of our item-by-item approach, which continues to perform better in terms of out-of-sample tests and capturing the nuanced inflationary dynamics associated with economic slack.

This dataset provides the foundation for our item-by-item analysis of the cyclicality of inflation components. The granularity and comprehensiveness of the data enable a detailed examination of inflation dynamics, facilitating the construction of a robust cyclically sensitive inflation index.

1.2.2
Methodology

This section outlines the methodology employed to construct our cyclically sensitive inflation index, contrasting it with the methodologies used by SW. We detail our item-by-item analysis approach, the use of the Composite Index of Real Cyclical Activity (CAI), and the advantages of our method.

1.2.2.1
Our Approach

Our approach involves analyzing the cyclicality of each item in the Personal Consumption Expenditures (PCE) basket individually. This item-by-item analysis allows us to identify the most cyclically sensitive components and create a more granular and comprehensive cyclically sensitive inflation index. The steps involved are as follows:

1. **Data Preparation:** We start with quarterly price indices for 75 detailed cate-

gories within the PCE, covering the period from January 1989 to December 2023. The data is seasonally adjusted and expressed in terms of year-over-year percentage changes.

2. **Calculation of Cyclicality:** For each item, we calculate its cyclicality by regressing the year-over-year percentage change in its price on the Composite Index of Real Cyclical Activity (CAI), developed by Stock and Watson. The CAI is a robust measure of cyclical economic activity, capturing the overall state of the economy.

3. **Filtering and Standardization:** The CAI is constructed by standardizing and bandpass filtering seven key economic indicators: Gross Domestic Output (GDO), capacity utilization rate, establishment employment, overall employment-to-population ratio, employment-to-population ratio for ages 25-54, the unemployment rate, and the short-term unemployment rate. These indicators are sourced from various federal agencies and combined using principal component analysis.

4. **Regression Analysis:** We perform a separate regression for each of the 75 PCE items, using the CAI as the explanatory variable. This allows us to quantify the cyclicality of each item independently.

5. **Aggregation:** The individual cyclicality measures are then aggregated to construct our cyclically sensitive inflation index. This index provides a comprehensive measure of how inflation responds to economic slack.

Formally, for each item i, we estimate

$$\Delta\pi_{i,t} = \hat{\alpha}_i + \hat{\beta}_i \, \mathrm{CAI}_t + \epsilon_{i,t},$$

where $\Delta\pi_{i,t}$ is the year-over-year percent change in the price of item i at time t, CAI_t is the Composite Index of Real Cyclical Activity, and $\epsilon_{i,t}$ is an error term.

We classify an item as cyclically sensitive if it exhibits a statistically significant positive relationship with cyclical activity, consistent with the Phillips-curve mechanism: specifically, we select item i when $\widehat{\beta}_i > 0$ and the corresponding p-value is below 0.05.

With the selected items we rebalance the PCE weights to sum to one and aggregate this items.

1.2.2.2
Stock and Watson's Approach

Stock and Watson (2020) developed a cyclically sensitive inflation index by including all PCE items simultaneously in a regression model. Their approach involves the following steps:

1. **Data Preparation:** They used a smaller subset of 17 PCE categories in their original analysis.
2. **Simultaneous Regression:** The CAI is regressed by all items filtered PCE price indexes in a single regression model.
3. **Estimation:** The coefficients obtained from this regression indicate the sensitivity of the CAI to each each category's filtered inflation rate. This is performed using a non-negative least squares (NNLS) method to ensure all coefficients are non-negative.
4. **Rebalancing:** After estimation, the coefficients (β_i) are rebalanced to sum to one, effectively addressing the weights assigned to each item in the aggregate measure.

Their model is specified as:

$$\mathrm{CAI}_t = \beta_0 + \sum_{i=1}^{75} w_i \Delta\pi_{i,t} + u_t$$

where w_i are the weights assigned to each item, subject to $0 \leq w_i \leq 1$ and $\sum_{i=1}^{75} w_i = 1$. The use of non negative least squares (NNLS) helps in ensuring non-negative weights, but when applied to a larger dataset, the method faces challenges such as increased complexity and potential multicollinearity.

1.2.2.3
Advantages of Our Method

Our item-by-item approach offers several advantages over the methodology used by Stock and Watson:

- **Granularity:** By analyzing each PCE item individually, we can better capture the heterogeneity in cyclicality across different price categories.
- **Comprehensiveness:** Our method includes 75 PCE categories, providing a more detailed and comprehensive view of inflation dynamics.

- **Robustness:** Our approach is robust to the inclusion of a large number of items, avoiding the pitfalls of multicollinearity and overfitting associated with simultaneous regression models.
- **Out-of-Sample Performance:** Rolling window analyses show that our method performs better in out-of-sample tests, providing more reliable estimates of cyclically sensitive inflation.

In summary, our methodology provides a detailed and robust measure of cyclically sensitive inflation, leveraging the strengths of the CAI and offering a comprehensive view of how inflation responds to economic slack.

The Federal Reserve Bank of San Francisco also constructs cyclically sensitive inflation measures by evaluating each item individually, much like we do. The key differences are that they rely on raw inflation and the unemployment rate, whereas we apply a filtered inflation series and use the Common Activity Index (CAI) index instead of unemployment alone. In the appendix, we provide a detailed comparison of both approaches and show that our item-selection procedure produces more stable results.

1.2.3 Results

In this section, we present the findings of our analysis, comparing the cyclicality of inflation components using our item-by-item approach and the method proposed by Stock and Watson (2020). We begin by discussing the item selection process for both methods, using the 75 price categories, followed by a comparison of the resulting cyclically sensitive inflation indices, and conclude with an evaluation of out-of-sample performance.

1.2.3.1 Item Selection: Stock and Watson's Method

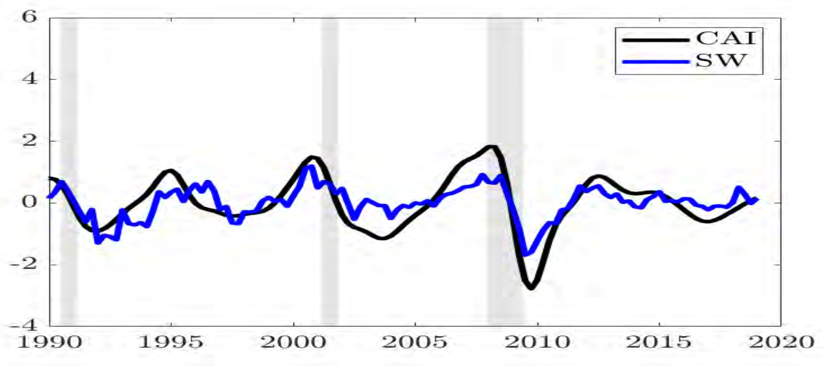
Stock and Watson's approach involves a simultaneous regression of all items to determine their cyclicality. The following table summarizes the items selected using their method in the entire sample, along with the weight of each item in the their selection and the weight of each item in the PCE. In this table we only report an item if Stock and Watson's Method attributed a weight $w_i \geq 0.001$. We then present a graphic comparing this index with the CAI. We can see that although their selection doesn't directly select items the arbitrary rule $w_i \geq 0.001$ already encompass 99.98% of their index.

Item	**PCE**	**S&W**
Glassware, tableware, and household utensils	0.38%	2.88%
Video, audio, ...	1.86%	0.59%
Sporting equipment, supplies, guns, and ammunition	0.53%	3.03%
Recreational books	0.24%	9.79%
Musical instruments	0.05%	2.08%
Jewelry and watches	0.58%	0.21%
Educational books	0.09%	6.10%
Luggage and similar personal items	0.22%	1.16%
Food and nonalcoholic beverages ...	6.69%	6.89%
Pharmaceutical and other medical products	3.10%	1.89%
Recreational items	1.26%	1.14%
Personal care products	1.02%	1.14%
Tobacco	0.89%	1.97%
Magazines, newspapers, and stationery	0.52%	3.58%
Imputed rental of owner-occupied nonfarm housing	12.39%	6.02%
Group housing	0.01%	21.92%
Electricity	1.61%	0.41%
Natural gas	0.63%	0.15%
Paramedical services	2.44%	0.88%
Other motor vehicle services	0.66%	1.83%
Membership clubs, sports centers, parks, theaters, and museums	1.41%	4.00%
Other recreational services	0.40%	7.34%
Financial services furnished without payment	2.08%	1.13%
Financial service charges, fees, and commissions	2.45%	0.25%
Net household insurance	0.07%	5.62%
Postal and delivery services	0.14%	0.53%
Nursery, elementary, and secondary schools	0.35%	3.82%
Professional and other services	1.59%	0.44%
Foreign travel by U.S. residents	1.03%	2.63%
Final consumption expenditures ...	2.92%	0.43%
Total	**47.61%**	**99.98%**

Notes: Items selected by the Stock–Watson procedure using the **full sample**. “PCE” reports expenditure shares in the PCE. “S&W” reports weights assigned within the Stock–Watson selected basket; these sum to approximately 100% (rounding error).

Table 1.1: Selected Items: Stock and Watson Method (Whole Sample)

Figure 1.1: Stock and Watson's method



Notes: This figure shows the CAI and a measure of inflation using Stock & Watson (2020) method.

1.2.3.2
Item Selection: Our Method

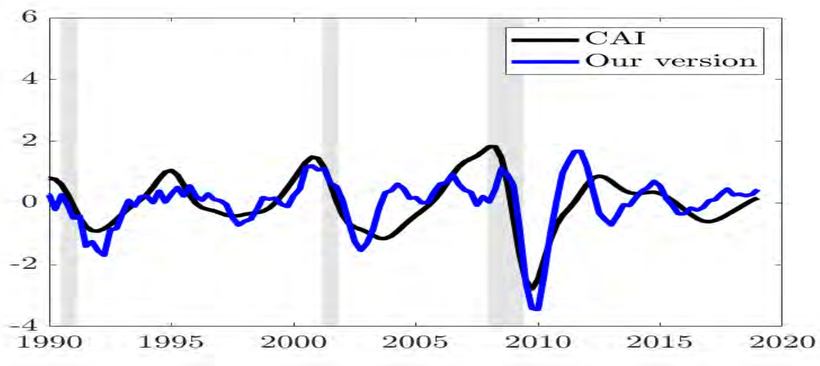
Our item-by-item approach evaluates each PCE category independently to determine its cyclicality. The following table summarizes the items selected using our method in the entire sample, along with the weight of each item in the PCE. We then present a graphic comparing this index with the CAI. In the tables using our method we only present one weight since we only adjust PCE weights in our index.

Item	PCE weights
Motor vehicle parts and accessories	0.53%
Sporting equipment, supplies, guns, and ammunition	0.53%
Recreational books	0.24%
Educational books	0.09%
Luggage and similar personal items	0.22%
Food and nonalcoholic beverages purchased for off-premises consumption	6.69%
Alcoholic beverages purchased for off-premises consumption	1.03%
Recreational items	1.26%
Household supplies	1.05%
Rental of tenant-occupied nonfarm housing	3.41%
Imputed rental of owner-occupied nonfarm housing	12.39%
Group housing	0.01%
Electricity	1.61%
Paramedical services	2.44%
Nursing homes	1.43%
Motor vehicle maintenance and repair	1.35%
Other motor vehicle services	0.66%
Membership clubs, sports centers, parks, theaters, and museums	1.41%
Other recreational services	0.40%
Purchased meals and beverages	5.06%
Nursery, elementary, and secondary schools	0.35%
Personal care and clothing services	1.12%
Household maintenance	0.64%
Total	**43.92%**

Notes: This table lists the items selected by **our method** using the **full sample**. “PCE weights” are expenditure shares in the PCE. The last row reports the total PCE weight covered by the selected subset.

Table 1.2: Selected Items: Our Method (Whole Sample)

Figure 1.2: Our method



Notes: This figure shows the CAI and a measure of inflation using our method.

1.2.3.3
Categories selected in one measure and not in the other

We now try to focus on the difference of the indexes created using our method and theirs. The 2 tables below present items that were selected by only of the two methods, along with the coefficient β and the p-value we found for each item. The selection process using Stock and Watson's method does not necessarily identify the most cyclically sensitive categories. For example, Tobacco is included in their index despite being considered anti-cyclical when analyzed individually. Conversely, Rental of tenant-occupied nonfarm housing, which is cyclically sensitive, is excluded from their selection, likely due to its strong covariance with other housing measures that are selected instead. This does not indicate a flaw in their methodology, as their goal is to produce an index that is collectively the most cyclically sensitive. Nonetheless, it highlights differences in item selection. We will now compare the performance of these selections.

Item	β	**p-value**	**PCE Weight**
Motor vehicle parts and accessories	0.670	0.000	0.53%
Alcoholic beverages ...	0.581	0.016	1.03%
Household supplies	0.701	0.000	1.05%
Rental of tenant-occupied nonfarm housing	0.573	0.000	3.41%
Nursing homes	0.552	0.000	1.43%
Motor vehicle maintenance and repair	0.437	0.000	1.35%
Purchased meals and beverages	0.539	0.000	5.06%
Personal care and clothing services	0.469	0.000	1.12%
Household maintenance	0.467	0.001	0.64%
Total			**15.62%**

Notes: This table lists items selected by **our method** but not selected by the Stock–Watson basket when using the **full sample**. Reported β and p-values come from item-level cyclicality regressions used in the selection stage (higher β indicates more procyclical inflation for that category). The last column reports each item's PCE weight; the final row sums the PCE weight of items in this set.

Table 1.3: Items Selected by Our Method but Not by Stock–Watson (Whole Sample)

Item	β	p-value	PCE	S&W
Glassware, tableware, and household utensils	0.369	0.2199	0.38%	2.88%
Video, audio, photographic …	0.477	0.069	1.86%	0.59%
Musical instruments	0.314	0.152	0.05%	2.08%
Jewelry and watches	0.693	0.085	0.58%	0.21%
Pharmaceutical and other medical products	-0.141	0.418	3.10%	1.89%
Personal care products	0.231	0.085	1.02%	1.14%
Tobacco	-1.563	0.043	0.89%	1.97%
Magazines, newspapers, and stationery	0.223	0.258	0.52%	3.58%
Natural gas	2.556	0.2128	0.63%	0.15%
Financial services furnished without payment	-2.185	0.014	2.08%	1.13%
Financial service charges, fees, and commissions	-1.286	0.020	2.45%	0.25%
Net household insurance	0.062	0.881	0.07%	5.62%
Postal and delivery services	0.290	0.608	0.14%	0.53%
Professional and other services	0.153	0.118	1.59%	0.44%
Foreign travel by U.S. residents	0.564	0.336	1.03%	2.63%
Final consumption expenditures of NPISHs	0.351	0.264	2.92%	0.43%
Total			**19.31%**	**25.52%**

Notes: This table lists items included in the Stock–Watson basket but not selected by **our method** when using the **full sample**. Reported β and p-values come from the same item-level cyclicality regressions used in our selection stage. "PCE" is the expenditure share in the PCE, while "S&W" is the weight assigned by Stock–Watson within their selected basket (weights sum to approximately 100% within their selection, up to rounding).

Table 1.4: Items Selected by Stock–Watson but Not by Our Method (Whole Sample)

Taken together, the tables 1.3 and 1.4 show that the two selection differ in economically meaningful ways: roughly one third of our index weight is allocated to categories that receive less than 0.1% weight each in Stock and Watson's CSI, while our rule excludes a set of categories that jointly account for about 25% of their CSI weight.

1.2.3.4
In-sample

With our selections done, we now regress our estimated core inflation $\Delta\pi_t^*$ on the CAI, since we are first analyzing in-sample it is clear that their index will present a higher correlation with CAI, since that is the way it is designed. The results presented are for the following regresion:

$$\Delta\pi_t^* = \alpha + \kappa \text{CAI}_t + \epsilon_{i,t}$$

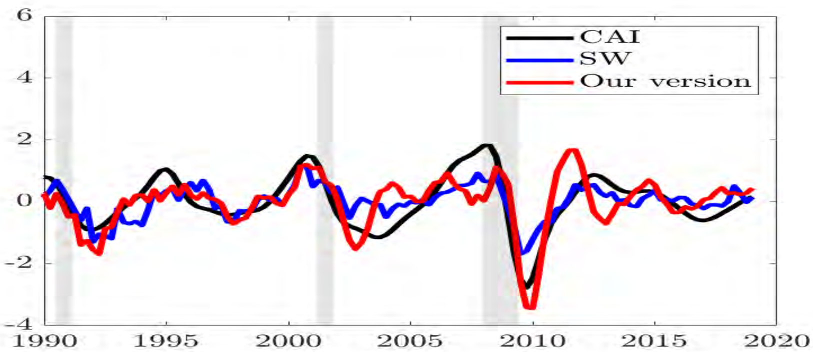
Notice that when testing the correlation of the defined indexes we are using κ opposed to the β_i used in the selection process.

	S&W Index	**Our Index**
α	-0.0113	-0.0134
	(0.0243)	(0.0568)
κ	0.5181***	0.6648***
	(0.0284)	(0.0666)
N	121	121
R^2	0.73604	0.45567

Notes: This table reports in-sample regressions of each method's inflation index on the Chicago Fed National Activity Index (CAI). κ is the cyclicality coefficient (slope on CAI). Standard errors are in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.5: In-Sample Comparison: Inflation Index on CAI (Whole Sample)

Figure 1.3: Inflation Index Comparison Over Time: S&W vs Our Method vs Benchmark



Notes: This figure plots over time the inflation indexes produced by the Stock–Watson method and by our method, together with the benchmark measure used in the comparison. All series are constructed on the full sample and shown at the frequency used in the main analysis.

We can observe in Table 1.5 that both methods lead to high R^2, with Stock and Watson's method achieving a higher R^2 by construction. While our method demonstrates a higher slope, assessing their performance requires examining out-of-sample results. It is important to note that the R^2 achieved by both measures exceeds that of the 17 items originally used in Stock and Watson's paper, showing how using more disaggregated data can be informative.

A crucial feature desired in our inflation index is that any item excluded from the index must itself be acyclic. If we were to choose only the single item whose inflation rate is most highly correlated with the CAI, we might obtain an apparently steep Phillips curve and a high R^2, but in doing so we would omit many genuinely cycle-sensitive goods—exactly the opposite of our goal. To guard against this, we compare the excluded items under both our method and Stock–Watson's. Since the Stock–Watson weights are continuous rather than binary, we construct two discrete analogues for comparison:

- **SW1:** $\max(\text{PCE weight} - \text{SW weight}, 0)$ for each item, then renormalize so the remaining positive weights sum to one.
- **SW2:** include only those items to which the Stock–Watson algorithm assigns less than 0.1 % weight.

The first measure considers that if the weight attributed by SW is smaller then the PCE weight, it is not selected, the second one consider the item not to be selected like before if assigns less than 0.1% weight. With both their "non-selections" and ours in hand we re-normalize weights to sum to 1. We run the same regressions as before in those 3 measures of inflation.

	SW1	**SW2**	**Our model**
α	0.0218	0.0218	0.0228
	(0.0781)	(0.0782)	(0.0775)
κ	0.0422	0.0018	-0.0590
	(0.0689)	(0.0593)	(0.0393)
N	121	121	121
R^2	0.00314	0.00001	0.0186

Notes: This table reports regressions of **non-selected (excluded)** inflation indexes on the CAI. For Stock–Watson, we construct two discrete analogues of the excluded set: **SW1** assigns each item weight $\max(\text{PCE weight} - \text{SW weight}, 0)$ and renormalizes positive weights to sum to one; **SW2** treats as excluded any item receiving less than 0.1% Stock–Watson weight, then renormalizes excluded-item weights to sum to one. **Our model** uses the items excluded by our selection rule, with their PCE weights renormalized to sum to one. κ is the cyclicality coefficient (slope on CAI). Standard errors are in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table 1.6: Cyclicality of Excluded Items: Stock–Watson Analogues vs Our Method

As shown in table 1.6, the acyclicality of the items we don't select is more pronounced, as evidenced by its negative slope.

1.2.3.5
Out-of-sample

To assess out-of-sample performance, we divide the sample into two periods: 1989-2004 and 2005-2020. The first period is used to select items, and the second period is used to test how cyclical the inflation with those items is. We also reverse the process, using the second period for selection and the first for testing. The category selection using each half of the sample can be found in the appendix 1.G.

	1989–2004		2005–2020	
	SW	**Our model**	**SW**	**Our model**
α	-0.0389	-0.0895	0.1538*	0.0777
	(0.0583)	(0.0542)	(0.0835)	(0.1064)
κ	0.4392***	0.6616***	0.1938**	0.5941***
	(0.0836)	(0.0778)	(0.0849)	(0.1082)
N	60	60	61	61
R^2	0.32233	0.55518	0.081175	0.33816

Notes: Out-of-sample evaluation splits the sample into two periods (1989–2004 and 2005–2020). For each column pair, the inflation index is constructed using the items selected in the *other* subsample and then tested by regressing the resulting inflation index on the CAI within the displayed subsample. κ is the cyclicality coefficient (slope on CAI). Standard errors are in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table 1.7: Out-of-Sample Results: Selection and Testing Across Subsamples

The R^2 of our method is higher in both partitions of the sample, as is the slope of the regressions. This suggests that our method performs significantly better in out-of-sample scenarios, which is particularly important for policymakers who rely on accurate and reliable measures for forecasting and decision-making.

1.2.3.6
Rolling window

We now use a rolling window for the time sample to asses how robust this selection really is. We present the percentage of time in which the selected items in the whole sample is selected using a 15 year rolling window. The percentage of selection in the rolling window for each category can be found in the appendix 1.H.

Selected Items		
	Our Method	**S&W Method**
% Selection across rolling windows	81.21	57.90
Std. dev. across items	0.1983	0.2968

Non-Selected Items		
	Our Method	**S&W Method**
% Selection across rolling windows	4.19	16.31
Std. dev. across items	0.0849	0.2111

Notes: Items are first classified as **Selected** or **Non-Selected** using the **whole-sample** selection for each method. For each item, we compute the fraction of rolling windows in which it is selected. The table reports (i) the mean of these fractions within each group (Selected vs Non-Selected) and (ii) the standard deviation of these fractions across items within the group. Higher values for "Selected Items" and lower values for "Non-Selected Items" indicate greater stability and fewer classification reversals across rolling windows.

Table 1.8: Rolling-Window Stability Relative to Whole-Sample Classification

Now we can see a different advantage of our method, this time not in creating a cyclically sensitive inflation index but in assessing which items are individually cyclical. In the above tables, we observe that items selected in the whole sample are selected at a much higher rate using our method compared to theirs. Conversely, items not selected in the whole sample are more rarely selected by us. This indicates that our method is more consistent in identifying truly cyclically sensitive items and excluding those that are not, providing a clearer and more accurate assessment of individual item cyclicality.

1.3
Regional Exercise

Using regional data has become more important for assessing the Phillips curve as Fitzgerald *et al.* (2024);McLeay & Tenreyro (2020) show how it can be beneficial for its identification. Regional data, other then having more variation in our sample for each period creates the possibility of using time fixed effects. The use of this fixed-effect means that anything that affects all regions the same way are removed, potentially removing all effects from monetary policy, making it easier to identify the Phillips curve.

Hazell *et al.* (2022) make the argument that using time fixed effects can also remove long term inflation expectation due to every region sharing a monetary union. Their paper does make an argument that not every item has the same dynamics of the Phillips curve in the regional level, their idea is that tradable goods are priced only depending on national measures of slack. This means that if we were to estimate a regional Phillips curve using time fixed effects and their theory was correct we would virtually find a slope of 0 for every tradable good.

Because of both the possibility of better identification and the fact that we can test the Hazell *et al.* (2022) hypothesis, we believe there is a major importance in making the same test we just did using regional data with and without fixed effects.

This tests can give us important information if using regional data makes it easier to identify the Phillips curve for each item, as well as a better understanding of the dynamics of regional inflation may in fact be different from the aggregated one.

1.3.1
Data

The data we use for this part all comes from the BLS, both regional prices and unemployment, our regional price data only starts in 1991 and is semi-annual.

The BLS does not report many product categories at a disaggregated level, so we can only analyze eight items compressing 100% of CPI. To expand this set, we use the Handbook to generate additional categories as follows:

- As an example, take the BLS category "Food and beverages", which breaks down into "Food at home," "Food away from home," and "Alcoholic beverages." If we have the indices and weights for the first two sub-categories, we can algebraically extract the index for "Alcoholic beverages":

  where all weights w are defined using the December value in a pivot month. In non-pivot months, we update each weight by

$$w_{\text{new}} = \left(\frac{index_{\text{new}}}{index_{\text{pivot}}}\right) \times w_{\text{pivot}}.$$

$$w_{\text{alc,new}} = w_{\text{FB,pivot}} \times \frac{index_{\text{FB,new}}}{index_{\text{FB,pivot}}} - \sum_{i \in \{\text{home,away}\}} w_{i,\text{pivot}} \times \frac{index_{i,\text{new}}}{index_{i,\text{pivot}}}.$$

- We then derive the "Alcoholic beverages" index by setting its first pivot-month value to 100 and computing

$$index_{\text{alc,new}} = index_{\text{alc,pivot}} \times \frac{w_{\text{alc,new}}}{w_{\text{alc,pivot}}}.$$

Our only challenge is that for some cities we do not have the December pivot index; in a few cities we have only November or a second-half average. Nonetheless, we apply this procedure to a limited set of items that correspond to a small share of the BLS, and the approximations prove very accurate. The appendix 1.F presents a regression comparing the official "Alcoholic beverages" index with our estimated values and the items we had to inter and their corresponding weights.

1.3.2 Methodology

We now compare the 3 selections, trying to stay as close as possible to the previously presented method. The first selection method still uses national data, but the point of doing this is to have a fair comparison for the selections using regional data and to really understand what is different between them. The difference of what we do this time and in the first section is that now we are using unemployment ratio instead of CAI, because we don't have data to create a regional CAI, and using CPI data, not PCE.

For each item we do the following regression:

$$\Delta\pi_{i,t} = \alpha_i + \beta_i \Delta \text{U}_t + \epsilon_{i,t}$$

for each inflation measure we select items that present $\beta < 0$ and p-value< 0.05. Since we are using unemployment instead of the CAI is natural that we now select items with negative coefficients (instead of positive, like previosuly). We use two selection methods using regional data, one being the Pooled method, not using any fixed effects, as follows:

$$\Delta\pi_{i,t,j} = \alpha_i + \beta_i \Delta \text{U}_{t,j} + \epsilon_{i,t,j}$$

And the other method includes a time fixed effect:

$$\Delta\pi_{i,t,j} = \alpha_t + \alpha_j + \beta_i \Delta \text{U}_{t,j} + \epsilon_{i,t,j}$$

Since our data only starts in 1991 we make 3 selections, one with the whole sample (1991-2019), one with 1991-2005 and one with 2006-2019 to allow us to do out-of-sample testing.

1.3.3 Results

The selections of each method are:

Fixed Effects	Pooled	Aggregated
Medical care commodities	Alcoholic beverages	Electricity
Rent of primary residence	Electricity	Food at home
	Food at home	Food away from home
	Food away from home	Household furnishings and operations
	Fuel oil and other fuels	Lodging away from home
	Gasoline (all types)	Recreation
	Household furnishings and operations	Rent of primary residence
	Lodging away from home	Utility (piped) gas service
	Medical care services	
	Other motor fuels	
	Public transportation	
	Recreation	
	Rent of primary residence	
	Utility (piped) gas service	
Total weight	**50.457%**	**38.005%**

Notes: This table lists the product categories selected when the selection step is performed using the **full sample**, separately for each specification: **Fixed Effects** (regional panel with region fixed effects), **Pooled** (regional panel without fixed effects), and **Aggregated** (national/aggregate series). "Total weight" reports the sum of CPI expenditure weights covered by the selected categories within each specification.

Table 1.9: Products Selected by Specification (Whole Sample)

As we can see the Pooled and aggregated versions have similar selections, with the Pooled version selecting a few more items. This probably comes from the fact that we are dealing with more data, having a better chance of finding statistical significance. The version using fixed effects, on the other hand, selects fewer items. That would go against the theories that using fixed effects would improve identification and more inline with items being priced nationally.

A problem of using this variety of methodologies is that testing each selection has to be made in each one of the scenarios. That is, since we have 3 different indexes, generated using different methods, we have to test each index in each of the specifications in which they were selected.

In each of the following tables the first column presents the result with respect of the selections made using fixed effects, the second column, the ones made using pooled data; the third column, the ones using aggregated data and the final two columns are using Headline inflation and core inflation (inflation ex-food and energy) respectively. All of this regressions are made in-sample, that is, the period used to make the selection is the same as the one used for testing. The results of the tests using out-of-sample selection can be found in the appendix 1.E.

This table tests each of the selections in the specification using fixed effects :

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.612^{***}	-0.045	-0.099	-0.162^{***}	-0.303^{***}
	(0.099)	(0.071)	(0.079)	(0.054)	(0.059)
N	1,302	1,302	1,302	1,302	1,302
R^2	0.030	0.0003	0.001	0.007	0.021

Notes: This table reports **in-sample** estimates using the **full sample**, where the **testing regression is run in the regional panel with region fixed effects**. Each column uses a different inflation index as the regressor of interest: **FE** is the index built under the fixed-effects specification; **POOL** is built from the regional panel without fixed effects; **NAC** is built from national/aggregated data; **HL** uses the filtered unemployment specification; and **CORE** is the benchmark core measure. κ is the cyclicality coefficient (slope on the activity variable used in the test regression). Standard errors are in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table 1.10: In-Sample Results (Whole Sample): Testing Using Fixed Effects

We can see that the fixed effects method achieves a really high, significant coefficient, with a higher R^2 then any other selection when testing in this specification.

We now present the results of each selection using national data :

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.464^{***}	-0.792^{***}	-0.779^{***}	-0.533^{***}	-0.227^{***}
	(0.081)	(0.246)	(0.133)	(0.176)	(0.065)
α	-0.089	-0.182	-0.125	-0.152	-0.108^{*}
	(0.073)	(0.222)	(0.119)	(0.158)	(0.059)
N	58	58	58	58	58
R^2	0.369	0.156	0.381	0.141	0.178

Notes: This table reports **in-sample** estimates using the **full sample**, where the **testing regression is run on national (aggregated) data**. Each column uses a different inflation index as the regressor of interest: **FE** is the index built under the fixed-effects specification; **POOL** is built from the regional panel without fixed effects; **NAC** is built from national/aggregated data; **HL** uses the filtered unemployment specification; and **CORE** is the benchmark core measure. κ is the cyclicality coefficient (slope on the activity variable used in the test regression). Standard errors are in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table 1.11: In-Sample Results (Whole Sample): Testing Using National Aggregated Data

And the results of each selection using the pooled data specification :

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.484^{***}	-0.686^{***}	-0.676^{***}	-0.465^{***}	-0.273^{***}
	(0.047)	(0.055)	(0.044)	(0.040)	(0.028)
α	-0.064	-0.169^{***}	-0.131^{***}	-0.120^{***}	-0.106^{***}
	(0.048)	(0.056)	(0.045)	(0.041)	(0.029)
N	1,302	1,302	1,302	1,302	1,302
R^2	0.075	0.106	0.156	0.096	0.066

Notes: This table reports **in-sample** estimates using the **full sample**, where the **testing regression is run on pooled regional data without region fixed effects**. Each column uses a different inflation index as the regressor of interest: **FE** is the index built under the fixed-effects specification; **POOL** is built from the pooled regional panel; **NAC** is built from national/aggregated data; **HL** uses the filtered unemployment specification; and **CORE** is the benchmark core measure. κ is the cyclicality coefficient (slope on the activity variable used in the test regression). Standard errors are in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.12: In-Sample Results (Whole Sample): Testing Using Pooled Regional Data (No Fixed Effects)

We now test a basket made with items that are not selected by each specification, it's clear that we don't just want our selection to be cyclically sensitive, we want that the items that are not selected to not be sensitive.

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.024	-0.166^{*}	-0.113	-0.162^{***}	-0.303^{***}
	(0.059)	(0.097)	(0.083)	(0.054)	(0.059)
N	1,302	1,302	1,302	1,302	1,302
R^2	0.0001	0.002	0.002	0.007	0.021

Notes: This table repeats the in-sample test in the **regional panel with region fixed effects**, but the inflation measure in each column is constructed from **non-selected items only**. For each specification (FE, POOL, NAC, HL, CORE), we take the set of items classified as *non-selected* by that specification's whole-sample selection rule, re-normalize their weights to sum to one, and build the corresponding inflation index. The coefficient κ measures cyclicality (slope on the activity variable in the testing regression). Standard errors are in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table 1.13: In-Sample Results (Whole Sample): Testing Using Fixed Effects with the *Non-Selected* Items Index

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.467^{***}	-0.120	-0.274	-0.533^{***}	-0.227^{***}
	(0.142)	(0.077)	(0.179)	(0.176)	(0.065)
α	-0.160	-0.124^{*}	-0.168	-0.152	-0.108^{*}
	(0.128)	(0.070)	(0.162)	(0.158)	(0.059)
N	58	58	58	58	58
R^2	0.162	0.041	0.040	0.141	0.178

Notes: This table reports **in-sample** estimates using the **full sample**, where the **testing regression is run on national (aggregated) data** and the inflation measure in each column is constructed from **non-selected items only**. For each specification (FE, POOL, NAC, HL, CORE), we take the set of items classified as *non-selected* by that specification's whole-sample selection rule, re-normalize their weights to sum to one, and build the corresponding inflation index. The coefficient κ measures cyclicality (slope on the activity variable in the testing regression). Standard errors are in parentheses. *p$<$0.10, **p$<$0.05, ***p$<$0.01.

Table 1.14: In-Sample Results (Whole Sample): Testing Using National Aggregated Data with the *Non-Selected* Items Index

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.411^{***}	0.137^{***}	-0.158^{**}	-0.465^{***}	-0.273^{***}
	(0.046)	(0.047)	(0.070)	(0.040)	(0.028)
α	-0.152^{***}	-0.098^{**}	-0.141^{**}	-0.120^{***}	-0.106^{***}
	(0.047)	(0.048)	(0.071)	(0.041)	(0.029)
N	1,302	1,302	1,302	1,302	1,302
R^2	0.058	0.006	0.004	0.096	0.066

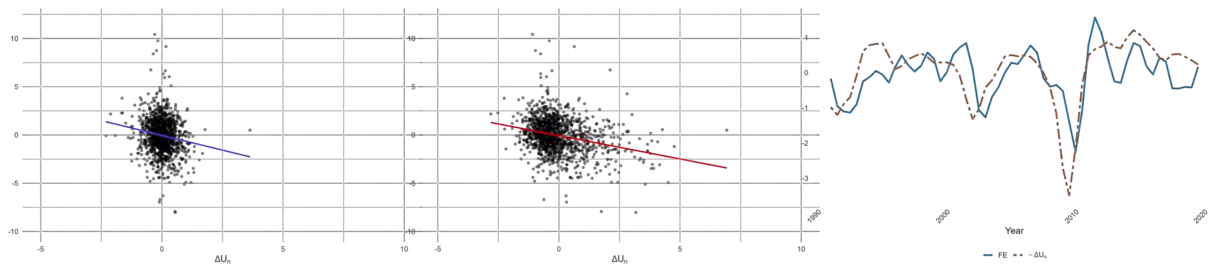
Notes: This table reports **in-sample** estimates using the **full sample**, where the **testing regression is run on pooled regional data without region fixed effects** and the inflation measure in each column is constructed from **non-selected items only**. For each specification (FE, POOL, NAC, HL, CORE), we take the set of items classified as *non-selected* by that specification's whole-sample selection rule, re-normalize their weights to sum to one, and build the corresponding inflation index. The coefficient κ measures cyclicality (slope on the activity variable in the testing regression). Standard errors are in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.15: In-Sample Results (Whole Sample): Testing Using Pooled Regional Data (No Fixed Effects) with the *Non-Selected* Items Index

The pooled data method and aggregated data method yield to a similar selection, so they perform well when tested on the other environment. The pooled version yield a steeper Philips curve on both scenarios but a smaller R^2. When we look at the tests of not selected items it is clear that the selection using aggregated data doesn't select all of the cyclically sensitive items. On the other hand neither of these selections performs well when using fixed effects, the fixed effects version is also not effective in other scenarios but it's really effective on its own.

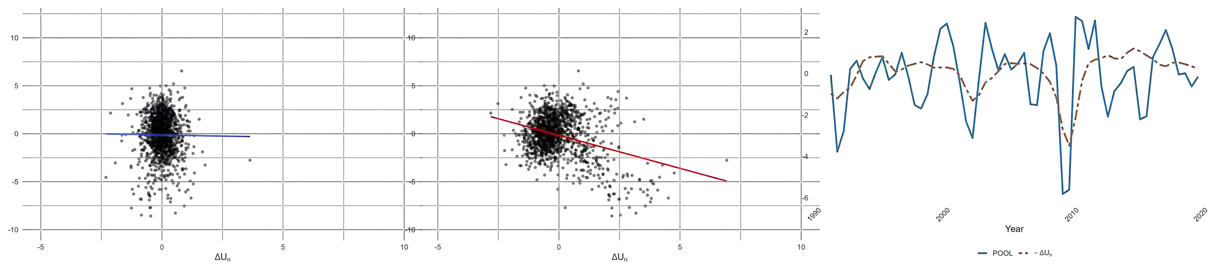
We now present the graphs of each the tested inflation measures in the three specifications. The graphs on the left column are plots of filtered unemployment and filtered inflation index after controlling for fixed effects, the middle column are plots of filtered unemployment and filtered inflation index and the right column are plots are of the filtered inflation index and filtered unemployment nationally. Each line of the graph represents one of the selection methods for the construction of the inflation index.

Figure 1.4: Fixed effects



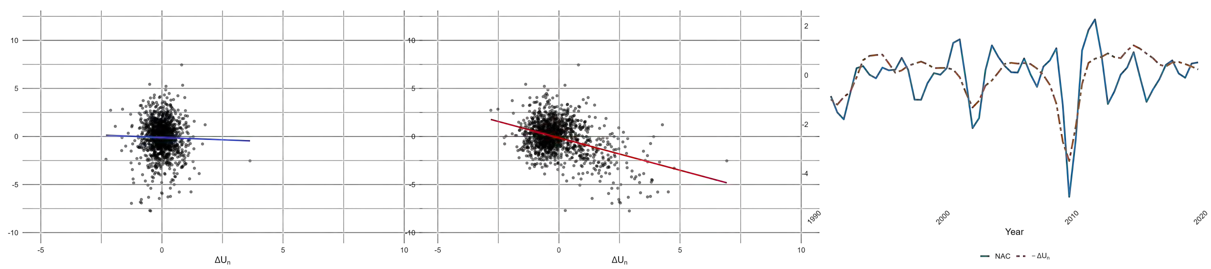
(a) FE (b) Pooled (c) National

Figure 1.5: Pooled data



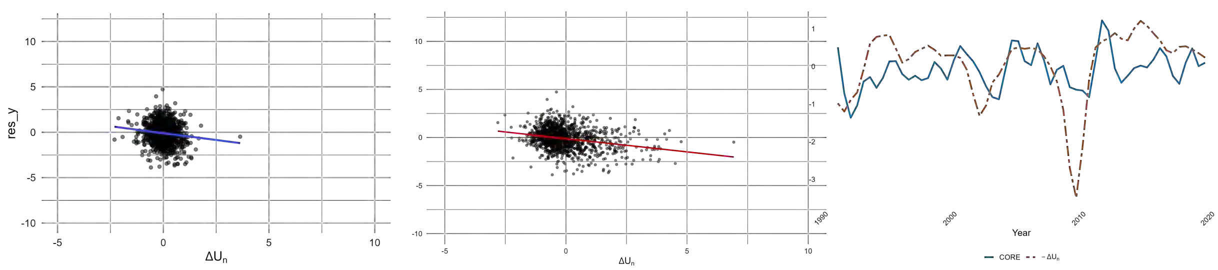
(a) FE (b) Pooled (c) National

Figure 1.6: Aggregated data



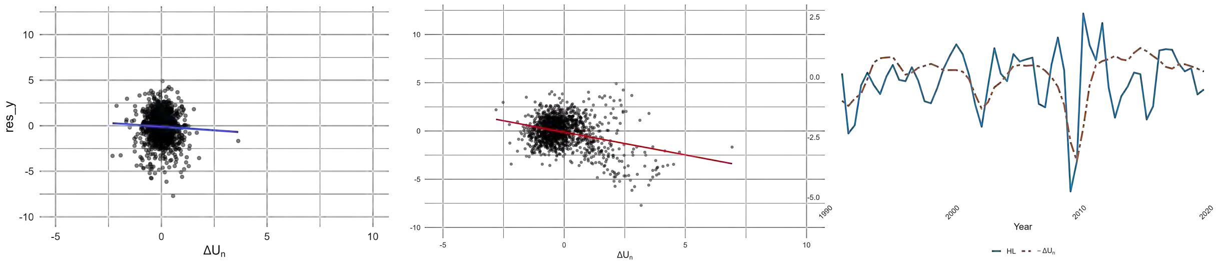
(a) FE (b) Pooled (c) National

Figure 1.7: Core inflation



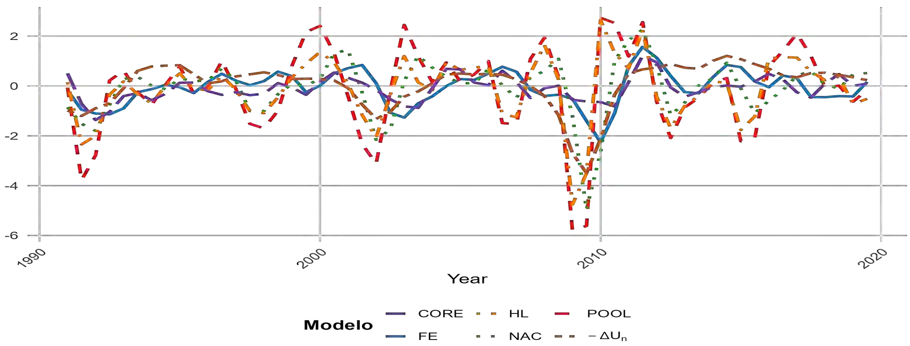
(a) FE (b) Pooled (c) National

Figure 1.8: Headline inflation



(a) FE (b) Pooled (c) National

Figure 1.9: Comparing



If the regional Phillips has the same specification of the aggregated Phillips curve, using time fixed effects would be just removing endogeneities shared by each region, one would expect to actually be able to select more items when using it, we in fact select much items, this would make sense if some items have their prices set nationally. Another interesting fact we can notice is that, the papers that defend that using fixed effects help identify the Phillips curve show a way steeper κ when using it, but when we use all items except for shelter as the inflation measure this is no longer the case, with the pooled version having a steeper κ. This would lead us to believe that some specific items are driving this identification, with probably some more interesting dynamics to be discovered.

To address this we use the following Phillips curve specifications, using the exact specification of McLeay & Tenreyro (2020). With the first equation using fixed effects and the second one using pooled data.

$$\pi_{t,j} = \alpha_t + \alpha_j + \kappa \mathrm{U}_{t,j} + \gamma_1 \pi_{t-1,j} + \gamma_2 E[\pi_{t+1,j}] + \epsilon_{t,j}$$

$$\pi_{i,t,j} = \alpha + \kappa_i \mathrm{U}_{t,j} + \gamma_1 \pi_{i,t-1,j} + \gamma_2 E[\pi_{i,t+1,j}] + \epsilon_{i,t,j}$$

In Table 1.16, we estimate the Phillips curve using **headline inflation** as the dependent variable. Column (1) uses regional **fixed effects** (FE), while column (2) reports the **pooled** specification without fixed effects (POOL).

	FE	POOL
	(1)	(2)
κ	-0.352^{***}	-0.241^{***}
	(0.111)	(0.050)
γ_1	-0.044	-0.031
	(0.028)	(0.027)
γ_2	0.657	1.350^{***}
	(0.733)	(0.215)
Observations	1,378	1,378
R^2	0.010	0.039

Notes: This table reports Phillips-curve estimates with **headline inflation** as the dependent variable. Column (1) includes **regional fixed effects** (FE); column (2) is a **pooled** specification without fixed effects (POOL). The coefficient κ is the Phillips-curve slope. Standard errors are in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.16: Headline Inflation: Pooled vs. Fixed-Effects Phillips Curve

Table 1.17 repeats the exercise using **all-items less shelter inflation** as the inflation measure. As before, column (1) includes regional fixed effects and column (2) reports the pooled specification.

	FE	POOL
	(1)	(2)
κ	-0.111	-0.125^{***}
	(0.098)	(0.046)
γ_1	-0.063^{**}	-0.063^{**}
	(0.028)	(0.027)
γ_2	0.736	2.181^{***}
	(0.648)	(0.204)
Observations	1,378	1,378
R^2	0.006	0.078

Notes: This table reports Phillips-curve estimates with **all-items less shelter inflation** as the dependent variable. Column (1) includes **regional fixed effects** (FE); column (2) is a **pooled** specification without fixed effects (POOL). The coefficient κ is the Phillips-curve slope. Standard errors are in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.17: All-Items Less Shelter Inflation: Pooled vs. Fixed-Effects Phillips Curve

In fact not only in our analysis housing seems to be a special case, in Hazell *et al.* (2022) paper they do not consider a heterogeneous Phillips curve but when they estimate the regional Philips curve with non-tradables items including shelter they get a coefficient four times higher then when using non-tradables items and not including housing.

1.4 Conclusion

In this paper, we have demonstrated the advantages of our item-by-item approach in constructing a cyclically sensitive inflation index and in assessing individual item cyclicality. Although our method is simple, it outperforms both Stock and Watson's approach and other existing methods in several key aspects.

First, our method does not face the problem of attributing weights in a way that controls for the inflation of different items, which can lead to the selection of non-cyclical items and the exclusion of truly cyclical ones. This issue, while not present in Stock and Watson's original level of aggregation, becomes apparent when extending their technique to a larger dataset. By testing each product individually, our method ensures a more accurate selection of cyclically sensitive items.

Second, although Stock and Watson's method achieves a higher R^2 in-sample regressions what is true by construction, our method consistently shows a steeper κ and R^2 in out-of-sample tests. This indicates that our approach provides more reliable and accurate predictions of inflation sensitivity, which is crucial for policymakers.

Furthermore, when comparing item selection, our method shows greater consistency. Items that are selected in the whole sample are chosen at a much higher rate when using a rolling window method of selection with our approach, and items not selected in the whole sample are rarely selected by us. This consistency underscores the robustness of our method in identifying truly cyclically sensitive items.

In conclusion, our item-by-item approach not only provides a more granular and comprehensive measure of cyclically sensitive inflation but also demonstrates superior performance in out-of-sample scenarios. This makes it a valuable tool for policymakers who require accurate and reliable measures for forecasting and decision-making. Our method's simplicity and effectiveness highlight its potential for broader application in economic analysis and policy formulation.

The regional exercise shows how our method can be applied using regional data and how important it is to have a better understanding of the different dynamics we observe in inflation using fixed effects with regional data.

1.A
Comparison with the FED SF method

FED SF method

$$\pi_{it} = \alpha_i + \beta(U_t - U_t^n) + \varepsilon_{it}$$

π_{it}: Inflation of item i in time t

U_t: unemployment rate in time t

U_t^n:CBO's natural unemployment rate in time t

ε_{it}: error term

Items are selected when:

$\beta < 0$

and

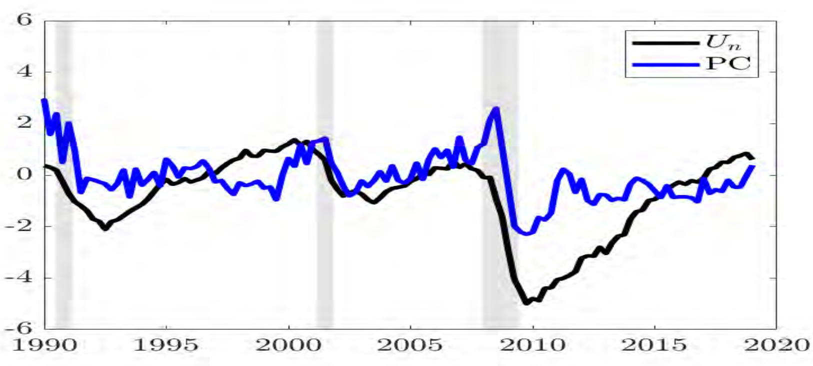
p-value< 0.05

Item	PCE weight
Furniture and furnishings	1.41%
Therapeutic appliances and equipment	0.51%
Household supplies	1.05%
Rental of tenant-occupied nonfarm housing	3.41%
Imputed rental of owner-occupied nonfarm housing	12.39%
Group housing	0.01%
Natural gas	0.63%
Dental services	1.03%
Paramedical services	2.44%
Nursing homes	1.43%
Motor vehicle maintenance and repair	1.35%
Membership clubs, sports centers, parks, theaters, and museums	1.41%
Other recreational services	0.40%
Purchased meals and beverages	5.06%
Nursery, elementary, and secondary schools	0.35%
Professional and other services	1.59%
Personal care and clothing services	1.12%
Social services and religious activities	1.31%
Household maintenance	0.64%
Final consumption expenditures of nonprofit institutions serving households	2.92%
Total	**40.46%**

Notes: Items selected using the **Federal Reserve's** "PC" method estimated on the **full sample**. Reported weights are headline PCE expenditure shares.

Table 1.18: Selected Items (Full Sample): Federal Reserve (PC) Method

Figure 1.10: Federal Reserve (PC) core inflation index and the unemployment gap over time



Notes: This figure plots, over time, the core inflation index constructed using the Federal Reserve's PC selection method and the unemployment gap.

Item	PCE weight
Motor vehicle parts and accessories	0.53%
Sporting equipment, supplies, guns, and ammunition	0.53%
Recreational books	0.24%
Educational books	0.09%
Luggage and similar personal items	0.22%
Food and nonalcoholic beverages purchased for off-premises consumption	6.69%
Alcoholic beverages purchased for off-premises consumption	1.03%
Recreational items	1.26%
Electricity	1.61%
Other motor vehicle services	0.66%
Total	**12.86%**

Notes: List of categories selected by **our method** but **not** selected by the **Federal Reserve's** PC method. PCE weights are headline PCE expenditure shares.

Table 1.19: Items Selected by Our Method but Not by the Federal Reserve (PC) Method

Item	PCE weight
Furniture and furnishings	1.41%
Therapeutic appliances and equipment	0.51%
Natural gas	0.63%
Dental services	1.03%
Professional and other services	1.59%
Social services and religious activities	1.31%
Final consumption expenditures of nonprofit institutions serving households	2.92%
Total	**9.40%**

Notes: List of categories selected by the **Federal Reserve's** PC method but **not** selected by **our method**. PCE weights are headline PCE expenditure shares.

Table 1.20: Items Selected by the Federal Reserve (PC) Method but Not by Our Method

	ΔPC	**Our model**
α	-0.0063	-0.0134
	(0.0459)	(0.0568)
β	0.4512***	0.6648***
	(0.0539)	(0.0666)
N	121	121
R^2	0.37098	0.45567

Notes: OLS regression of the inflation measure on the cyclical activity indicator (CAI): $\pi_t = \alpha + \beta CAI_t + \varepsilon_t$. Columns compare the inflation measure constructed using the Federal Reserve's PC selection method (ΔPC) versus our method. Standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.21: In-sample Phillips-curve regressions using the CAI

	ΔPC	**Our model**
α	-0.0058	-0.0101
	(0.0362)	(0.0549)
β	0.5708***	0.6813***
	(0.0418)	(0.0635)
N	121	121
R^2	0.60992	0.49151

Notes: OLS regression of the inflation measure on the filtered unemployment gap u_t^{gap}: $\pi_t = \alpha + \beta u_t^{gap} + \varepsilon_t$. Columns compare the inflation measure constructed using the Federal Reserve's PC selection method (ΔPC) versus our method. Standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.22: In-sample Phillips-curve regressions using the filtered unemployment gap

	1989–2004		2005–2020	
	PC	**Our model**	**PC**	**Our model**
α	-0.1604***	-0.0895	0.0828	0.0777
	(0.0486)	(0.0542)	(0.0841)	(0.1064)
β	0.5714***	0.6616***	0.3965***	0.5941***
	(0.0698)	(0.0778)	(0.0855)	(0.1082)
N	60	60	61	61
R^2	0.53600	0.55518	0.26691	0.33816

Notes: Split-sample (out-of-sample) OLS regressions of inflation on the CAI: $\pi_t = \alpha + \beta\, CAI_t + \varepsilon_t$. "PC" uses the Federal Reserve's PC-based inflation measure; "Our model" uses our inflation measure. Each subperiod regression uses T observations within the indicated dates. Standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.23: Out-of-sample Phillips-curve regressions using the CAI (split-sample)

	1989–2004		2005–2020	
	PC	**Our model**	**PC**	**Our model**
α	-0.0975*	-0.0213	0.0308	0.0155
	(0.0517)	(0.0633)	(0.0633)	(0.0951)
β	0.6061***	0.6275***	0.5540***	0.6626***
	(0.0829)	(0.1015)	(0.0603)	(0.0907)
N	60	60	61	61
R^2	0.47963	0.39700	0.58858	0.47520

Notes: Split-sample (out-of-sample) OLS regressions of inflation on the filtered unemployment gap: $\pi_t = \alpha + \beta\, u_t^{gap} + \varepsilon_t$. "PC" uses the Federal Reserve's PC-based inflation measure; "Our model" uses our inflation measure. Each subperiod regression uses T observations within the indicated dates. Standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.24: Out-of-sample Phillips-curve regressions using the filtered unemployment gap (split-sample)

The method used by the Federal Reserve Bank of San Francisco also checks products one by-one. But it has the disadvantages of using only gap unemployment instead of CAI and of not using the year on year filter proposed by Stock and Watson. This filter has two roles, it eliminates an I(1) trend in inflation and marginal cost and somewhat smothers both high-frequency variance (noise) and very low frequency-variance. Their method is closer to ours than Stock and Watson's. The rolling window table shows that we tend to select items more consistently. When using half the sample, their method only selects 9 items (compared

Selected Items		
	Our Method	**FED's Method**
% Selection across rolling windows	81.21	66.61
Std Dev	0.1983	0.2773

Non-Selected Items		
	Our Method	**FED's Method**
% Selection across rolling windows	4.19	5.01
Std Dev	0.0849	0.0977

Notes: Items are classified as *selected* vs. *non-selected* using the **full-sample** selection for each method. "% of selection" is the average fraction of rolling windows in which an item is selected. "Std Dev" is the standard deviation of the rolling-window selection indicator across items within each group.

Table 1.25: Rolling-window selection frequency by method

to 20 in the full sample), while our method selects 17 items (compared to 23 in the full sample).

1.B
Two samples out of sample comparison

Item	PCE weight	S&W weight
Net purchases of used motor vehicles	0.99%	0.94%
Motor vehicle parts and accessories	0.53%	4.47%
Furniture and furnishings	1.41%	2.86%
Tools and equipment for house and garden	0.19%	3.44%
Video, audio, photographic, and information processing equipment and media	1.86%	3.15%
Sports and recreational vehicles	0.37%	5.50%
Recreational books	0.24%	5.60%
Jewelry and watches	0.58%	0.42%
Telephone and related communication equipment	0.16%	4.06%
Motor vehicle fuels, lubricants, and fluids	2.21%	0.18%
Pharmaceutical and other medical products	3.10%	0.28%
Recreational items	1.26%	1.15%
Household supplies	1.05%	4.25%
Personal care products	1.02%	1.85%
Tobacco	0.89%	1.44%
Magazines, newspapers, and stationery	0.52%	2.29%
Imputed rental of owner-occupied nonfarm housing	12.39%	14.75%
Group housing	0.01%	16.49%
Water supply and sanitation	0.73%	0.38%
Electricity	1.61%	1.04%
Natural gas	0.63%	0.17%
Physician services	3.92%	3.69%
Dental services	1.03%	4.39%
Other motor vehicle services	0.66%	3.03%
Ground transportation	0.34%	0.60%
Air transportation	0.62%	0.32%
Water transportation	0.02%	1.02%
Membership clubs, sports centers, parks, theaters, and museums	1.41%	3.03%
Accommodations	0.92%	1.72%
Net household insurance	0.07%	2.25%
Telecommunication services	1.48%	1.55%
Commercial and vocational schools	0.37%	0.76%
Professional and other services	1.59%	2.78%
Total	**44.18%**	**99.85%**

Notes: Items selected using Stock and Watson's procedure estimated on the **first half** of the sample. "PCE weight" is the expenditure share in headline PCE. "S&W weight" reports the corresponding Stock–Watson selection weight (sums to 100% up to rounding).

Table 1.26: Selected Items (First Half of Sample): Stock and Watson Method

Item	PCE weight
Motor vehicle parts and accessories	0.53%
Recreational books	0.24%
Jewelry and watches	0.58%
Therapeutic appliances and equipment	0.51%
Educational books	0.09%
Luggage and similar personal items	0.22%
Telephone and related communication equipment	0.16%
Food and nonalcoholic beverages purchased for off-premises consumption	6.69%
Recreational items	1.26%
Household supplies	1.05%
Rental of tenant-occupied nonfarm housing	3.41%
Imputed rental of owner-occupied nonfarm housing	12.39%
Group housing	0.01%
Electricity	1.61%
Nursing homes	1.43%
Motor vehicle maintenance and repair	1.35%
Other motor vehicle services	0.66%
Gambling	1.08%
Other recreational services	0.40%
Purchased meals and beverages	5.06%
Professional and other services	1.59%
Personal care and clothing services	1.12%
Total	**41.44%**

Notes: Items selected using **our method** estimated on the **first half** of the sample. Reported weights are headline PCE expenditure shares.

Table 1.27: Selected Items (First Half of Sample): Our Method

Item	PCE weight
Household supplies	1.05%
Rental of tenant-occupied nonfarm housing	3.41%
Imputed rental of owner-occupied nonfarm housing	12.39%
Group housing	0.01%
Nursing homes	1.43%
Other recreational services	0.40%
Purchased meals and beverages	5.06%
Internet access	0.35%
Final consumption expenditures of nonprofit institutions serving households	2.92%
Total	**27.02%**

Notes: Items selected using the **Federal Reserve's** "PC" (persistent components) method estimated on the **first half** of the sample. Reported weights are headline PCE expenditure shares.

Table 1.28: Selected Items (First Half of Sample): Federal Reserve (PC) Method

Item	PCE weight	S&W weight
Glassware, tableware, and household utensils	0.27%	1.60%
Sporting equipment, supplies, guns, and ammunition	0.48%	3.13%
Recreational books	0.15%	9.18%
Musical instruments	0.04%	2.21%
Educational books	0.05%	7.71%
Luggage and similar personal items	0.21%	2.94%
Food and nonalcoholic beverages purchased for off-premises consumption	6.23%	4.34%
Alcoholic beverages purchased for off-premises consumption	1.09%	7.87%
Food produced and consumed on farms	0.00%	0.24%
Pharmaceutical and other medical products	3.29%	0.49%
Recreational items	1.30%	3.32%
Tobacco	0.80%	2.95%
Rental of tenant-occupied nonfarm housing	3.48%	22.55%
Electricity	1.30%	0.34%
Paramedical services	2.66%	2.50%
Membership clubs, sports centers, parks, theaters, and museums	1.56%	0.96%
Other recreational services	0.48%	7.86%
Financial services furnished without payment	2.49%	1.80%
Net household insurance	0.09%	3.09%
Net health insurance	1.64%	1.04%
Net motor vehicle and other transportation insurance	0.56%	2.19%
Internet access	0.56%	0.62%
Nursery, elementary, and secondary schools	0.34%	7.88%
Foreign travel by U.S. residents	1.24%	3.19%
Total	**30.31%**	**100.00%**

Notes: Items selected using Stock and Watson's procedure estimated on the **second half** of the sample. "PCE weight" is the expenditure share in headline PCE. "S&W weight" reports the corresponding Stock–Watson selection weight.

Table 1.29: Selected Items (Second Half of Sample): Stock and Watson Method

Item	PCE weight
Motor vehicle parts and accessories	0.59%
Sporting equipment, supplies, guns, and ammunition	0.48%
Recreational books	0.15%
Educational books	0.05%
Food and nonalcoholic beverages purchased for off-premises consumption	6.23%
Alcoholic beverages purchased for off-premises consumption	1.09%
Recreational items	1.30%
Rental of tenant-occupied nonfarm housing	3.48%
Imputed rental of owner-occupied nonfarm housing	11.23%
Group housing	0.02%
Dental services	0.95%
Paramedical services	2.66%
Motor vehicle maintenance and repair	1.24%
Other motor vehicle services	0.75%
Membership clubs, sports centers, parks, theaters, and museums	1.56%
Other recreational services	0.48%
Purchased meals and beverages	5.45%
Nursery, elementary, and secondary schools	0.34%
Personal care and clothing services	1.13%
Household maintenance	0.62%
Total	**39.80%**

Notes: Items selected using **our method** estimated on the **second half** of the sample. Reported weights are headline PCE expenditure shares.

Table 1.30: Selected Items (Second Half of Sample): Our Method

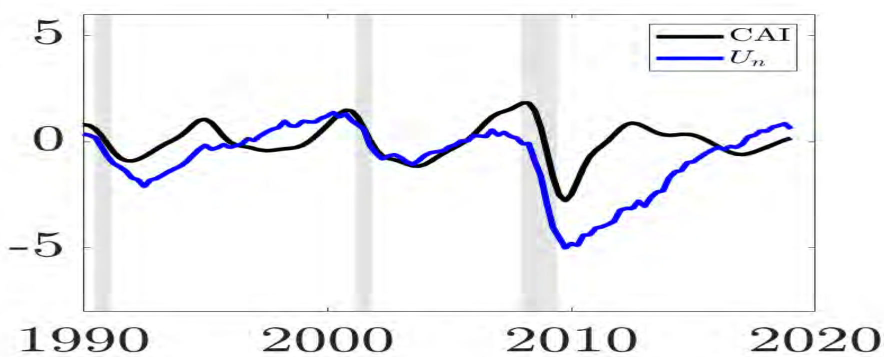
Item	PCE weight
Rental of tenant-occupied nonfarm housing	3.48%
Imputed rental of owner-occupied nonfarm housing	11.23%
Group housing	0.02%
Dental services	0.95%
Paramedical services	2.66%
Nursing homes	1.35%
Motor vehicle maintenance and repair	1.24%
Membership clubs, sports centers, parks, theaters, and museums	1.56%
Other recreational services	0.48%
Purchased meals and beverages	5.45%
Nursery, elementary, and secondary schools	0.34%
Personal care and clothing services	1.13%
Social services and religious activities	1.50%
Household maintenance	0.62%
Final consumption expenditures of nonprofit institutions serving households (NPISHs)	2.83%
Total	**34.84%**

Notes: Items selected using the **Federal Reserve's** "PC" method estimated on the **second half** of the sample. Reported weights are headline PCE expenditure shares.

Table 1.31: Selected Items (Second Half of Sample): Federal Reserve (PC) Method

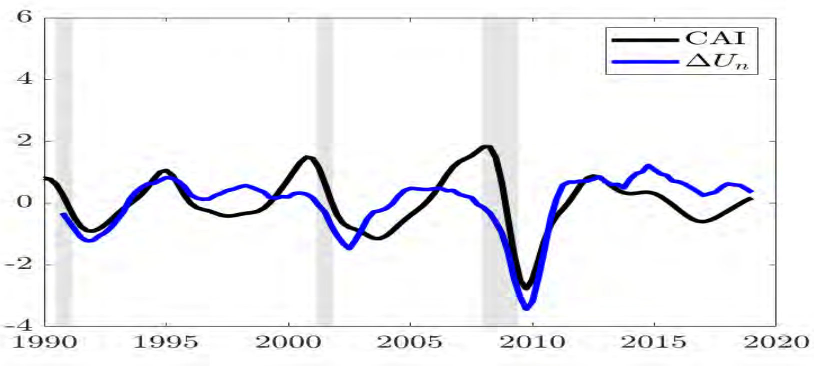
1.C
Difference from FED comes from filter or CAI usage?

Figure 1.11: CAI and the unemployment gap over time



Notes: This figure plots the cyclical activity indicator (CAI) and the unemployment gap through time. It motivates the comparison between using CAI versus unemployment-based measures as the cyclical regressor in the selection step.

Figure 1.12: CAI and the filtered unemployment measure over time



Notes: This figure plots the CAI together with the filtered unemployment measure used in the alternative selection exercise. The goal is to assess whether differences relative to the Federal Reserve's PC method are driven by the choice of cyclical regressor (CAI vs unemployment) or by the filtering of unemployment.

Item	**CAI β**	**CAI p-value**	**Un β**	**Un p-value**	**% Weight**
Furniture and ...	0.1391	0.4605	0.4122	0.0253	1.41
Household appliances	0.5263	0.1717	1.0812	0.0039	0.44
Therapeutic...	0.1027	0.3895	0.2923	0.0121	0.51
Personal care...	0.2310	0.0851	0.4070	0.0018	1.02
Dental services	0.1712	0.0702	0.2192	0.0182	1.03
Gambling	0.2892	0.0613	0.3975	0.0087	1.08
Professional and ...	0.1533	0.1188	0.2323	0.0158	1.59
Social services ...	0.0845	0.2708	0.2989	0.0000	1.31
Final non-prof...	0.3514	0.2644	1.1294	0.0002	2.92
Total					**11.31**

Notes: This table lists items that enter the selected basket under our selection rule when the cyclical regressor is (filtered) unemployment, but not when the regressor is the CAI. Reported coefficients and p-values come from item-level regressions used in the selection step. "% Weight" is the PCE expenditure share of the item.

Table 1.32: Items selected using (filtered) unemployment but not using CAI (Our method)

Item	CAI β	CAI p-value	Un β	Un p-value	% Weight
Recreational books	0.8922	0.0001	0.4094	0.0822	0.24
Educational books	0.6816	0.0022	0.2109	0.3447	0.09
Luggage and ...	1.3091	0.0086	0.6514	0.1900	0.22
Alcoholic beverages ...	0.5809	0.0160	0.3773	0.1155	1.03
Total					**1.58**

Notes: This table lists items that enter the selected basket under our selection rule when the cyclical regressor is the CAI, but not when the regressor is (filtered) unemployment. Reported coefficients and p-values come from item-level regressions used in the selection step. "% Weight" is the PCE expenditure share of the item.

Table 1.33: Items selected using CAI but not using (filtered) unemployment (Our method)

	Filtered unemployment	CAI
α	-0.01567	-0.0134
	(0.0545)	(0.0568)
β	0.5739***	0.6648***
	(0.0639)	(0.0666)
N	121	121
R^2	0.404	0.455

Notes: OLS regressions of our inflation measure on alternative cyclical regressors: $\pi_t^{our} = \alpha + \beta x_t + \varepsilon_t$, where x_t is either the filtered unemployment measure or the CAI. Standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.34: Inflation regressions under alternative cyclical regressors (Our method)

1.D
Using SW aggregation

In their paper, Stock and Watson use a more aggregated set of categories, totaling only 17 items. We therefore re-run our selection and forecasting exercise restricting attention to the same 17 items used in their aggregation. Our method continues to outperform the Stock–Watson measure out-of-sample. Importantly, the Stock–Watson selection also mitigates the earlier issue of selecting items that appear acyclical under our approach. Using the full sample, our method does not select any item that Stock and Watson do not select, whereas the converse is not true.

Item	β	**p-value**	**Weight**
Gas & electric utilities	1.1081	0.1208	3.46%
Other services	0.2182	0.0563	2.80%
Final consumption expenditures of nonprofit institutions serving households (NPISH)	0.5193	0.0844	15.56%
Total			**21.82%**

Notes: This table reports categories that are included in the Stock–Watson selected basket but excluded by our selection rule when both are implemented using the Stock–Watson 17-item aggregation. The coefficient and p-value refer to the item-level cyclicality regression used in the selection step. "Weight" is the expenditure share within the Stock–Watson aggregation.

Table 1.35: Items selected by Stock–Watson but not by our method (SW aggregation, full sample)

We do not select any item that Stock and Watson do not select.

	SW	**Our method**
α	-0.0254	-0.0108
	(0.0432)	(0.0680)
β	0.4747***	0.5727***
	(0.0513)	(0.0808)
N	141	141
R^2	0.38084	0.26530

Notes: OLS regression of the (method-specific) inflation measure on the cyclical regressor used in this section. The "SW" column uses the Stock–Watson inflation measure computed from the 17-item aggregation; "Our method" uses our inflation measure constructed on the same 17-item aggregation. Standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.36: In-sample Phillips-curve regression: Stock–Watson vs our method (SW aggregation)

	1989–2004		2005–2020	
	SW	**Our method**	**SW**	**Our method**
α	-0.1494**	-0.1028	0.0040	0.0043
	(0.0710)	(0.1035)	(0.0834)	(0.0752)
β	0.1613	0.4639***	0.3676***	0.3415***
	(0.1111)	(0.1621)	(0.0763)	(0.0688)
N	60	60	61	61
R^2	0.03506	0.12376	0.28212	0.29453

Notes: Split-sample OLS regressions of the (method-specific) inflation measure on the cyclical regressor used in this section. The "SW" column uses the Stock–Watson inflation measure computed from the 17-item aggregation; "Our method" uses our inflation measure constructed on the same 17-item aggregation. Standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.37: Split-sample (out-of-sample) regression: Stock–Watson vs our method (SW aggregation)

	Our aggregation	**Stock–Watson aggregation**
α	-0.0134	-0.0193
	(0.0568)	(0.0705)
β	0.6648***	0.6580***
	(0.0666)	(0.0826)
N	121	121
R^2	0.45500	0.34700

Notes: Both columns implement **our** selection and inflation construction, varying only the level of aggregation used to define the item universe. "Our aggregation" refers to our baseline (more disaggregated) universe; "Stock–Watson aggregation" restricts the universe to the 17-item aggregation used in Stock and Watson. Standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.38: Our method under alternative aggregations (our original vs Stock–Watson)

1.E
Regional exercise, Two samples out of sample comparison

	Fixed Effects	Pooled	Aggregated
	Alcoholic beverages	Alcoholic beverages	Food at home
	Medical care commodities	Electricity	Food away from home
	Rent of primary residence	Food at home	Household furnishings and operations
		Food away from home	Lodging away from home
		Fuel oil and other fuels	Rent of primary residence
		Gasoline (all types)	
		Household furnishings and operations	
		Lodging away from home	
		Medical care services	
		Other private transportation	
		Public transportation	
		Rent of primary residence	
		Utility (piped) gas service	
Weight	**10.480%**	**55.776%**	**29.073%**

Notes: This table reports the set of products selected in the **first half of the sample** under three specifications: **Fixed Effects**, **Pooled**, and **Aggregated**. "Weight" reports the total expenditure weight of the selected basket under each specification.

Table 1.39: Products Selected I (First-half selection)

	Fixed Effects	Pooled	Aggregated
	Rent of primary residence	Alcoholic beverages	Alcoholic beverages
		Electricity	Food at home
		Food at home	Food away from home
		Food away from home	Lodging away from home
		Fuel oil and other fuels	Recreation
		Household furnishings and operations	Rent of primary residence
		Lodging away from home	
		Medical care services	
		Other motor fuels	
		Public transportation	
		Recreation	
		Rent of primary residence	
		Utility (piped) gas service	
Weight	**7.862%**	**47.646%**	**31.226%**

Notes: This table reports the set of products selected in the **second half of the sample** under three specifications: **Fixed Effects**, **Pooled**, and **Aggregated**. "Weight" reports the total expenditure weight of the selected basket under each specification.

Table 1.40: Products Selected II (Second-half selection)

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.603^{***}	-0.150	-0.117	-0.244^{***}	-0.330^{***}
	(0.131)	(0.121)	(0.114)	(0.083)	(0.092)
N	618	629	618	618	618
R^2	0.036	0.003	0.002	0.015	0.022

Notes: This table reports regressions of inflation on the cyclical regressor for the regional exercise, evaluated on the 1991–2005 subsample (first half). The inflation indexes are constructed using baskets selected in the **second half of the sample** ("newsel"). Columns FE/POOL/NAC correspond to indexes built under the Fixed Effects, Pooled, and NAC specifications, respectively; HL and CORE are benchmark measures. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table 1.41: Regional exercise: second-half selection, evaluated on 1991–2005

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.471^{***}	-0.002	-0.015	-0.099	-0.286^{***}
	(0.124)	(0.088)	(0.087)	(0.073)	(0.080)
N	684	684	684	684	684
R^2	0.022	0.00000	0.0001	0.003	0.020

Notes: This table reports regressions of inflation on the cyclical regressor for the regional exercise, evaluated on the 2006–2019 subsample (second half). The inflation indexes are constructed using baskets selected in the **first half of the sample** ("oldsel"). Columns FE/POOL/NAC correspond to indexes built under the Fixed Effects, Pooled, and NAC specifications, respectively; HL and CORE are benchmark measures. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table 1.42: Regional exercise: first-half selection, evaluated on 2006–2019

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.458^{***}	-0.712^{***}	-0.809^{***}	-0.689^{**}	-0.230
	(0.156)	(0.226)	(0.201)	(0.252)	(0.139)
α	-0.100	-0.165	-0.145	-0.177	-0.223^{**}
	(0.099)	(0.143)	(0.128)	(0.160)	(0.088)
N	28	28	28	28	28
R^2	0.250	0.277	0.384	0.223	0.095

Notes: This table reports the regional exercise under the **NAC specification**, which uses **aggregated data**. The regression is evaluated on the 1991–2005 subsample (first half), while the inflation indexes are constructed using baskets selected in the **second half of the sample** ("newsel"). Columns FE/POOL/NAC correspond to alternative index constructions (Fixed Effects, Pooled, and NAC); HL and CORE are benchmark measures. Standard errors in parentheses. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 1.43: Regional exercise: second-half selection, evaluated on 1991–2005 (NAC specification; aggregated data)

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.459^{***}	-0.488	-0.744^{***}	-0.480^{*}	-0.214^{***}
	(0.091)	(0.305)	(0.179)	(0.251)	(0.070)
α	-0.043	-0.157	-0.116	-0.119	0.001
	(0.100)	(0.333)	(0.195)	(0.274)	(0.076)
N	30	30	30	30	30
R^2	0.474	0.084	0.383	0.116	0.252

Notes: This table reports the regional exercise under the **NAC specification**, which uses **aggregated data**. The regression is evaluated on the 2006–2019 subsample (second half), while the inflation indexes are constructed using baskets selected in the **first half of the sample** ("oldsel"). Columns FE/POOL/NAC correspond to alternative index constructions (Fixed Effects, Pooled, and NAC); HL and CORE are benchmark measures. Standard errors in parentheses. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 1.44: Regional exercise: first-half selection, evaluated on 2006–2019 (NAC specification; aggregated data)

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.446^{***}	-0.627^{***}	-0.572^{***}	-0.597^{***}	-0.317^{***}
	(0.080)	(0.075)	(0.073)	(0.061)	(0.054)
α	-0.165^{***}	-0.179^{***}	-0.138^{**}	-0.175^{***}	-0.242^{***}
	(0.061)	(0.058)	(0.056)	(0.047)	(0.041)
N	618	629	618	618	618
R^2	0.049	0.099	0.092	0.134	0.053

Notes: This table reports the regional exercise under the **Pooled specification**, which uses **pooled regional data (without fixed effects)**. The regression is evaluated on the 1991–2005 subsample (first half), while the inflation indexes are constructed using baskets selected in the **second half of the sample** ("newsel"). Columns FE/POOL/NAC correspond to alternative index constructions (Fixed Effects, Pooled, and NAC); HL and CORE are benchmark measures. Standard errors in parentheses. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 1.45: Regional exercise: second-half selection, evaluated on 1991–2005 (Pooled specification; pooled regional data)

	FE	POOL	NAC	HL	CORE
	(1)	(2)	(3)	(4)	(5)
κ	-0.456^{***}	-0.458^{***}	-0.672^{***}	-0.412^{***}	-0.245^{***}
	(0.052)	(0.069)	(0.050)	(0.053)	(0.034)
α	0.062	-0.110	-0.135^{**}	-0.060	0.021
	(0.063)	(0.084)	(0.060)	(0.065)	(0.041)
N	684	684	684	684	684
R^2	0.101	0.060	0.211	0.080	0.072

Notes: This table reports the regional exercise under the **Pooled specification**, which uses **pooled regional data (without fixed effects)**. The regression is evaluated on the 2006–2019 subsample (second half), while the inflation indexes are constructed using baskets selected in the **first half of the sample** ("oldsel"). Columns FE/POOL/NAC correspond to alternative index constructions (Fixed Effects, Pooled, and NAC); HL and CORE are benchmark measures. Standard errors in parentheses. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 1.46: Regional exercise: first-half selection, evaluated on 2006–2019 (Pooled specification; pooled regional data)

1.F
Items inference

We now present the inference of our method for assessing inflation of categories not directly presented by BLS. We first present the weight of each category and the weight of what we are inferring. Then we present a table of the regression of the inflation we get from using our method for alcoholic beverages and the values presented by BLS for each region that we use.

	Item	Weight (%)
1	Rent of primary residence	7.862%
2	Owners' equivalent rent of primary residence	23.044%
3	Lodging away from home	2.410%
4	Alcoholic beverages	1.038%
5	Apparel	2.663%
6	Education and communication	6.810%
7	Water and sewer and trash collection services	1.107%
8	Fuel oil and other fuels	0.145%
9	Food at home	7.772%
10	Food away from home	6.347%
11	Gasoline (all types)	2.811%
12	Household furnishings and operations	4.682%
13	Medical care commodities	1.580%
14	Medical care services	7.289%
15	Other goods and services	3.159%
16	Recreation	5.797%
17	Other private transportation	11.180%
18	Public transportation	1.105%
19	Electricity	2.425%
20	Utility (piped) gas service	0.710%
21	Other motor fuels	0.064%
	Total (all items)	**100.000%**
	Total inferred (red-highlighted items)	**24.88%**

Notes: We report expenditure weights for the set of CPI major categories used in the regional exercise. Items highlighted in red are categories for which regional inflation is not directly reported by the BLS at the level required for our analysis; we therefore *infer* their inflation using our method. The final row reports the total weight of the inferred subset.

Table 1.47: Items and weights

	Place	**Coefficient**	**R^2**
1	Atlanta-Sandy Springs-Roswell, GA	1.00	1.00
2	Boston-Cambridge-Newton, MA-NH	1.00	1.00
3	Chicago-Naperville-Elgin, IL-IN-WI	1.00	1.00
4	Cincinnati-Hamilton, OH-KY-IN	1.05	0.98
5	Cleveland-Akron, OH	1.00	1.00
6	Dallas-Fort Worth-Arlington, TX	1.00	1.00
7	Denver-Aurora-Lakewood, CO	1.00	0.98
8	Detroit-Warren-Dearborn, MI	1.00	1.00
9	Houston-The Woodlands-Sugar Land, TX	1.00	1.00
10	Kansas City, MO-KS	1.12	0.96
11	Los Angeles-Long Beach-Anaheim, CA	1.00	1.00
12	Miami-Fort Lauderdale-West Palm Beach, FL	0.99	1.00
13	Milwaukee-Racine, WI	1.03	0.98
14	Minneapolis-St.Paul-Bloomington, MN-WI	1.00	1.00
15	New York-Newark-Jersey City, NY-NJ-PA	1.00	1.00
16	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	1.00	1.00
17	Phoenix-Mesa-Scottsdale, AZ	1.00	0.95
18	Pittsburgh, PA	1.01	0.99
19	Portland-Salem, OR-WA	0.92	0.98
20	San Diego-Carlsbad, CA	1.00	0.99
21	San Francisco-Oakland-Hayward, CA	1.00	1.00
22	Seattle-Tacoma-Bellevue, WA	1.00	1.00
23	St. Louis, MO-IL	1.01	0.98
24	Tampa-St. Petersburg-Clearwater, FL	0.98	1.00
25	Urban Alaska	0.99	0.92
26	Urban Hawaii	1.03	0.98
27	Washington-Baltimore, DC-MD-VA-WV	1.01	1.00

Notes: For each metropolitan area, we regress the official inflation rate for *Alcoholic beverages* on the inflation series for *Alcoholic beverages* implied by our method using the set of categories available in the city-level CPI. The reported coefficient is the estimated slope from this city-specific regression, and R^2 is the corresponding goodness-of-fit. Coefficients close to one indicate that the implied series closely matches the BLS series in that city.

Table 1.48: Inferring Alcoholic Beverages Inflation from City-Level CPI

1.G
PCE categories

Glassware, tableware, and household utensils	New motor vehicles	Household appliances
Sporting equipment, supplies, guns, and ammunition	Net purchases of used motor vehicles	Tools and equipment for house and garden
Recreational books	Motor vehicle parts and accessories	Video, audio, photographic, and information processing equipment and media
Musical instruments	Furniture and furnishings	Sports and recreational vehicles
Educational books	Jewelry and watches	Therapeutic appliances and equipment
Luggage and similar personal items	Telephone and related communication equipment	Women's and girls' clothing
Food and nonalcoholic beverages purchased for off-premises consumption	Men's and boys' clothing	Children's and infants' clothing
Alcoholic beverages purchased for off-premises consumption	Other clothing materials and footwear	Motor vehicle fuels, lubricants, and fluids
Food produced and consumed on farms	Fuel oil and other fuels	Household supplies
Pharmaceutical and other medical products	Personal care products	Magazines, newspapers, and stationery
Recreational items	Imputed rental of owner-occupied nonfarm housing	Rental value of farm dwellings
Tobacco	Group housing	Water supply and sanitation
Rental of tenant-occupied nonfarm housing	Natural gas	Physician services
Electricity	Dental services	Hospitals
Paramedical services	Nursing homes	Motor vehicle maintenance and repair
Membership clubs, sports centers, parks, theaters, and museums	Other motor vehicle services	Ground transportation
Other recreational services	Air transportation	Water transportation
Financial services furnished without payment	Audio-video, photographic, and information processing equipment and media	Gambling
Net household insurance	Purchased meals and beverages	Food furnished to employees (including military)
Net health insurance	Accommodations	Financial service charges, fees, and commissions
Net motor vehicle and other transportation insurance	Life insurance	Telecommunication services
Internet access	Postal and delivery services	Higher education
Nursery, elementary, and secondary schools	Commercial and vocational schools	Professional and other services
Foreign travel by U.S. residents	Personal care and clothing services	Social services and religious activities
Final consumption expenditures of nonprofit institutions serving households	Household maintenance	Less: Expenditures in the United States by nonresidents

Notes: This table lists the 75 disaggregated PCE categories used in the rolling-window selection exercises and in the full-sample classification comparisons across methods.

Table 1.49: PCE item universe used in the analysis (75 categories)

1.H
Rolling window comparison

Item	Share of windows selected
Glassware, tableware, and household utensils	0.2419
Sporting equipment, supplies, guns, and ammunition	0.7258
Recreational books	0.9839
Musical instruments	0.6452
Educational books	0.9516
Luggage and similar personal items	0.6613
Food and nonalcoholic beverages purchased for off-premises consumption	0.7258
Alcoholic beverages purchased for off-premises consumption	0.7258
Food produced and consumed on farms	0.4032
Pharmaceutical and other medical products	0.3065
Recreational items	0.4355
Tobacco	0.9839
Rental of tenant-occupied nonfarm housing	0.0484
Electricity	0.0968
Paramedical services	0.4194
Membership clubs, sports centers, parks, theaters, and museums	0.5968
Other recreational services	0.3871
Financial services furnished without payment	0.9194
Net household insurance	0.8871
Net health insurance	0.1452
Net motor vehicle and other transportation insurance	0.7903
Internet access	0.4355
Nursery, elementary, and secondary schools	0.7258
Foreign travel by U.S. residents	0.9355

Notes: Each entry reports the fraction of rolling windows in which the item is selected, restricting attention to items classified as *selected* under the Stock–Watson method (based on the full-sample classification used in your analysis). Values lie in $[0, 1]$ (e.g., 0.2419 means the item is selected in 24.19% of rolling windows).

Table 1.50: Rolling-window selection frequency (Selected items): Stock and Watson method

Item	Share of windows selected
Furniture and furnishings	0.2903
Therapeutic appliances and equipment	0.2742
Household supplies	0.3226
Rental of tenant-occupied nonfarm housing	1.0000
Imputed rental of owner-occupied nonfarm housing	0.9677
Group housing	1.0000
Natural gas	0.0484
Dental services	0.6290
Paramedical services	0.5484
Nursing homes	0.9839
Motor vehicle maintenance and repair	0.5323
Membership clubs, sports centers, parks, theaters, and museums	0.6935
Other recreational services	0.9032
Purchased meals and beverages	0.9032
Nursery, elementary, and secondary schools	0.6452
Professional and other services	0.6452
Personal care and clothing services	0.4839
Social services and religious activities	0.8387
Household maintenance	0.7742
Final consumption expenditures of nonprofit institutions serving households	0.8387

Notes: Federal Reserve "PC" method. Each entry reports the fraction of rolling windows in which the item is selected, restricting attention to items classified as *selected* under the PC method (based on the full-sample classification used in your analysis). Values lie in $[0, 1]$.

Table 1.51: Rolling-window selection frequency (Selected items): Federal Reserve (PC) method

Item	Share of windows selected
Motor vehicle parts and accessories	1.0000
Sporting equipment, supplies, guns, and ammunition	0.6290
Recreational books	0.9516
Educational books	1.0000
Luggage and similar personal items	1.0000
Food and nonalcoholic beverages purchased for off-premises consumption	0.9839
Alcoholic beverages purchased for off-premises consumption	0.6774
Recreational items	1.0000
Household supplies	0.8871
Rental of tenant-occupied nonfarm housing	1.0000
Imputed rental of owner-occupied nonfarm housing	1.0000
Group housing	1.0000
Electricity	0.6290
Paramedical services	0.6613
Nursing homes	0.8226
Motor vehicle maintenance and repair	0.9194
Other motor vehicle services	0.8871
Membership clubs, sports centers, parks, theaters, and museums	0.5968
Other recreational services	1.0000
Purchased meals and beverages	1.0000
Nursery, elementary, and secondary schools	0.4355
Personal care and clothing services	0.8226
Household maintenance	0.8548

Notes: Each entry reports the fraction of rolling windows in which the item is selected, restricting attention to items classified as *selected* under our method (based on the full-sample classification used in your analysis). Values lie in $[0, 1]$.

Table 1.52: Rolling-window selection frequency (Selected items): Our method

Item	Share	Item	Share
New motor vehicles	0.0161	Household appliances	0.0484
Net purchases of used motor vehicles	0.0323	Tools and equipment for house and garden	0.0806
Motor vehicle parts and accessories	0.2258	Video, audio, photographic, and information processing equipment and media	0.2097
Furniture and furnishings	0.0161	Sports and recreational vehicles	0.1452
Jewelry and watches	0.3548	Therapeutic appliances and equipment	0.0161
Telephone and related communication equipment	0.2419	Women's and girls' clothing	0.0323
Men's and boys' clothing	0.0645	Children's and infants' clothing	0.5000
Other clothing materials and footwear	0	Motor vehicle fuels, lubricants, and fluids	0.0323
Fuel oil and other fuels	0.0484	Household supplies	0.3065
Personal care products	0.8065	Magazines, newspapers, and stationery	0.1935
Imputed rental of owner-occupied nonfarm housing	0.4194	Rental value of farm dwellings	0
Group housing	0.9355	Water supply and sanitation	0.0161
Natural gas	0.1613	Physician services	0.2097
Dental services	0.0968	Hospitals	0
Nursing homes	0.5645	Motor vehicle maintenance and repair	0.3065
Other motor vehicle services	0.8226	Ground transportation	0.2903
Air transportation	0.3065	Water transportation	0.0161
Audio-video, photographic, and information processing equipment and media	0.0645	Gambling	0
Purchased meals and beverages	0	Food furnished to employees (including military)	0
Accommodations	0	Financial service charges, fees, and commissions	0.0161
Life insurance	0.0323	Telecommunication services	0.6129
Postal and delivery services	0.2742	Higher education	0.8548
Commercial and vocational schools	0	Professional and other services	0.1935
Personal care and clothing services	0.4677	Social services and religious activities	0
Household maintenance	0.0161	Less: Expenditures in the United States by nonresidents	0
Final consumption expenditures of nonprofit institutions serving households	0.4839		

Notes: Non-selected items under the Stock–Watson method (based on the full-sample classification used in your analysis). "Share" is the fraction of rolling windows in which the item is selected; values lie in $[0, 1]$.

Table 1.53: Rolling-window selection frequency (Non-selected items): Stock and Watson method

Item	Share	Item	Share
New motor vehicles	0	Tools and equipment for house and garden	0
Net purchases of used motor vehicles	0	Video, audio, photographic, and information processing equipment and media	0
Motor vehicle parts and accessories	0	Sporting equipment, supplies, guns, and ammunition	0
Household appliances	0.0806	Sports and recreational vehicles	0
Glassware, tableware, and household utensils	0	Recreational books	0.1452
Musical instruments	0	Jewelry and watches	0
Educational books	0	Luggage and similar personal items	0
Telephone and related communication equipment	0	Food and nonalcoholic beverages purchased for off-premises consumption	0.1613
Alcoholic beverages purchased for off-premises consumption	0.5000	Food produced and consumed on farms	0
Women's and girls' clothing	0	Men's and boys' clothing	0
Children's and infants' clothing	0	Other clothing materials and footwear	0
Motor vehicle fuels, lubricants, and fluids	0.0161	Fuel oil and other fuels	0.0323
Pharmaceutical and other medical products	0.0161	Recreational items	0.0161
Personal care products	0.2419	Tobacco	0.1290
Magazines, newspapers, and stationery	0	Rental value of farm dwellings	0.1129
Water supply and sanitation	0	Electricity	0.3710
Physician services	0	Hospitals	0
Other motor vehicle services	0	Ground transportation	0
Air transportation	0	Water transportation	0
Audio-video, photographic, and information processing equipment and media	0.0161	Gambling	0.0806
Food furnished to employees (including military)	0	Accommodations	0.0323
Financial services furnished without payment	0	Financial service charges, fees, and commissions	0
Life insurance	0	Net household insurance	0.1452
Net health insurance	0	Net motor vehicle and other transportation insurance	0.0484
Telecommunication services	0	Postal and delivery services	0
Internet access	0	Higher education	0.1774
Commercial and vocational schools	0	Foreign travel by U.S. residents	0.0323
Less: Expenditures in the United States by nonresidents	0		

Notes: Non-selected items under the Federal Reserve's PC method (based on the full-sample classification used in your analysis). "Share" is the fraction of rolling windows in which the item is selected; values lie in $[0, 1]$.

Table 1.54: Rolling-window selection frequency (Non-selected items): Federal Reserve (PC) method

Item	Share	Item	Share
New motor vehicles	0	Household appliances	0.1774
Net purchases of used motor vehicles	0	Glassware, tableware, and household utensils	0
Furniture and furnishings	0	Tools and equipment for house and garden	0
Video, audio, photographic, and information processing equipment and media	0	Sports and recreational vehicles	0
Musical instruments	0.2742	Jewelry and watches	0.3387
Therapeutic appliances and equipment	0.0484	Telephone and related communication equipment	0.0968
Food produced and consumed on farms	0	Women's and girls' clothing	0
Men's and boys' clothing	0	Children's and infants' clothing	0
Other clothing materials and footwear	0	Motor vehicle fuels, lubricants, and fluids	0
Fuel oil and other fuels	0.0323	Pharmaceutical and other medical products	0
Personal care products	0	Tobacco	0
Magazines, newspapers, and stationery	0	Rental value of farm dwellings	0
Water supply and sanitation	0	Natural gas	0.0323
Physician services	0	Dental services	0
Hospitals	0	Ground transportation	0
Air transportation	0	Water transportation	0
Audio-video, photographic, and information processing equipment and media	0	Gambling	0.2581
Food furnished to employees (including military)	0	Accommodations	0
Financial services furnished without payment	0	Financial service charges, fees, and commissions	0
Life insurance	0	Net household insurance	0
Net health insurance	0	Net motor vehicle and other transportation insurance	0
Telecommunication services	0	Postal and delivery services	0.2903
Internet access	0	Higher education	0.0645
Commercial and vocational schools	0	Professional and other services	0.1129
Social services and religious activities	0	Foreign travel by U.S. residents	0.0645
Less: Expenditures in the United States by nonresidents	0	Final consumption expenditures of nonprofit institutions serving households	0.0968

Notes: Non-selected items under our method (based on the full-sample classification used in your analysis). "Share" is the fraction of rolling windows in which the item is selected; values lie in $[0, 1]$.

Table 1.55: Rolling-window selection frequency (Non-selected items): Our method

2
Sectoral Inflation and the Phillips Curve

Carlos Carvalho[1,4] **Marcus Castro**[2] **Samuel Efraim**[1] **Jae Won Lee**[3] **Woong Yong Park**[3]

[1] PUC–Rio

[2] LSE

[3] Seoul National University

[4] Kapitalo Investimentos

This paper investigates the role of sectoral heterogeneity in price stickiness for the estimation of the Phillips Curve. Building on existing New Keynesian models with sector-specific pricing frictions, we derive sectoral and aggregate Phillips Curves that incorporate relative heterogeneity. While the theoretical implications of such models are well established, our main contribution is to take these predictions seriously in an empirical context. Using PCE data for 15 sectors, we estimate both sectoral and aggregate Phillips Curves, showing that ignoring heterogeneity leads to biased estimates of the aggregate slope. Importantly, the sector-level estimates imply price adjustment frequencies that closely match micro-data evidence, providing validation for our empirical approach and helping to connect micro-level price data with the structure of standard macroeconomic models.

Keywords: GARCH; HAR; Realized Volatility; Value at Risk; Forecasting.

2.1 Introduction

The relationship between inflation and economic slack, as captured by the Phillips Curve, lies at the core of macroeconomic theory and monetary policy transmission. Formulated in its modern version through the development of New Keynesian models during the 1980s and 1990s, the New Keynesian Phillips Curve (NKPC) remains the dominant framework for modeling inflation dynamics (Galí, 2015; Woodford, 2004). However, its early theoretical success contrasts with a variety of empirical difficulties, sparking an extensive and ongoing debate in the literature.

Several studies have documented significant challenges in estimating the NKPC, including conflicting results on slope coefficients (Nason & Smith, 2008a,b; Mavroeidis *et al.*, 2014), instability across time and model specifications (Stock & Watson, 2007; Albuquerque & Baumann, 2017; Luengo-Prado *et al.*, 2018; Jordà & Nechio, 2018; Jordi & Gambetti, 2019), concerns over robustness (Dufour *et al.*, 2010) and even existence (Hooper *et al.*, 2020). These problems are often attributed to factors such as the endogeneity of regressors, the contamination of slack variables by supply-side cost-push shocks and the weakness of available instruments.

Most of the empirical literature relies on benchmark specifications derived from canonical New Keynesian models such as Woodford (2004) and Galí (2015), or on extensions that incorporate backward-looking behavior through price indexation (Christiano *et al.*, 2005), staggered wage contracts (Fuhrer & Moore, 1995), or rule-of-thumb behavior (Galí & Gertler, 1999; Gali *et al.*, 2005). Additional terms also arise in models with investment frictions (Woodford, 2005; Carvalho & Nechio, 2016) or international trade linkages (Monacelli, 2007; Zaniboni & Zaniboni, 2008).

In this paper, we take seriously the implications of heterogeneity in price stickiness across sectors to address these empirical shortcomings. While theoretical contributions such as Carvalho (2006) and Aoki (2001) have shown how heterogeneous price-setting behavior can reshape the aggregate Phillips Curve, this insight has not yet been fully integrated into empirical estimation. We show that ignoring sectoral differences in price rigidity leads to biased slope estimates and distorts the inferred degree of nominal rigidity in the economy.

We estimate the multi-sector New Keynesian Phillips Curve by stacking the fifteen sectoral equations into a system and applying a two-step system GMM estimator. Our instrument set comprises sectoral demand shocks, aggregate demand shocks, defensive expenditure shocks and high-frequency monetary policy surprises (around FOMC announcements), all chosen to isolate the demand-driven variation in economic slack. To assess stability over time, we split the sample into two sub-periods (1975–2000 vs. 2001–2024) and compare slope estimates across these windows. We also re-estimate a homogeneous (aggregate) Phillips Curve as a baseline; this comparison illustrates the bias induced by ignoring sectoral heterogeneity.

Departing from previous studies that estimate sectoral Phillips Curves independently—such as Imbs *et al.* (2011), Griffa (2025) or Schwartzman (2006), our approach jointly estimates an integrated system of sectoral equations. This allows us to estimate common parameters across sectors—such as the discount factor and preference elasticities—enhancing empirical discipline and improving alignment with the underlying theoretical structure.

Our empirical analysis, based on PCE price and quantity data for 15 consumption sectors in the U.S. economy from 1975 to 2024, reveals that accounting for sectoral heterogeneity is critical for identifying a stable and economically meaningful aggregate Phillips Curve. Moreover, our sector-level estimates imply price adjustment frequencies that closely match those documented in micro-level price data—providing external validation for our estimation strategy and strengthening the connection between macroeconomic theory and micro-data evidence.

While most of the empirical literature estimates the NKPC using a proxy for real marginal costs—most commonly the labor share—we instead rely on the output gap as the measure of economic slack. This decision is partly driven by data limitations, as reliable marginal cost measures are not available at the sectoral level. However, using the output gap poses challenges of its own. Following the influential critique by Galí & Gertler (1999), many studies have shown that regressions using the output gap often yield weak or even negative slope estimates, primarily due to the presence of supply-side shocks that bias the relationship. To address this issue, we implement an instrumental variable strategy that isolates the demand-driven component of fluctuations. Our findings demonstrate that, when heterogeneity is accounted for and valid instruments are employed, the output gap can deliver robust and theoretically consistent estimates of the NKPC slope.

Our estimates show that incorporating sectoral heterogeneity significantly increases the estimated slope of the aggregate Phillips Curve compared to standard specifications. The results are remarkably robust across identification strategies, whether key structural parameters—such as the elasticity of intertemporal substitution and the elasticity of labor supply—are calibrated or estimated. Moreover, the estimated sectoral degrees of stickiness exhibit strong correlations (around 0.7) with micro-level measures of price-change frequency presented in Nakamura & Steinsson (2008), providing further empirical support for the heterogeneous NKPC framework and its capacity to reconcile disaggregated evidence with aggregate inflation dynamics.

The remainder of the paper is structured as follows: Section 2.2 presents the basics of the theoretical framework and illustrates the problem of estimating the aggregated NKPC without considering heterogeneity in price stickiness, Section 2.3 details the data, Section 2.4 describes the empirical strategy, Section 2.5 reports the main findings and robustness exercises and finally Section 2.6 concludes.

2.2 Model

Our framework is a multi-sector extension of the standard New Keynesian model closely related to the one in Carvalho *et al.* (2021). It features: (i) multiple sectors that differ in the degree of nominal rigidity and are subject to both sector-specific demand and supply shocks; and (ii) sector-specific labor markets, allowing for heterogeneity in wage-setting and employment dynamics across sectors.

The economy is composed of $K = 15$ sectors, corresponding to the first level of disaggregation in the Personal Consumption Expenditures (PCE) classification. Each sector $k \in \{1, 2, \ldots, K\}$ contains a continuum of firms indexed by $i \in [0, 1]$. Every firm produces a differentiated consumption good and belongs to a unique sector. We denote firm i that operates in sector k as "firm ik."

Let $\mathcal{I}_k$ denote the set of firms assigned to sector k, so that the union of all sectoral firm sets covers the entire economy: $\bigcup_{k=1}^{K} \mathcal{I}_k = [0, 1]$. The measure of firms in sector k is denoted by n_k.

Firms within each sector face sector-specific Calvo-style price-setting frictions, and their production technologies may vary in response to idiosyncratic shocks. Labor is specific to each sector, and wages are determined in decentralized markets, potentially leading to sectoral wage differentials.

2.2.1 Representative household

The representative consumer derives utility from a composite consumption good, supplies different types of labor to firms in different sectors and has access to a complete set of state-contingent claims. Subject to the budget constraint presented below, they maximize

$$E_0\left[\sum_{t=0}^{\infty}\beta^t\Gamma_t\Big(\frac{C_t^{1-\sigma}}{1-\sigma}\;-\;\sum_{k=1}^{K}\omega_k\,\frac{H_{k,t}^{1+\varphi}}{1+\varphi}\Big)\right] \tag{2.1}$$

where C_t denotes the household's consumption of the composite good, and $H_{k,t}$ denotes the hours of labor services supplied to sector k. Labor is fully mobile within each sector, but immobile across sectors. The parameters β, φ, and $\{\omega_k\}_{k=1}^{K}$ are, respectively, the discount factor, the inverse of the (Frisch) elasticity of labor supply, and the relative disutilities of supplying hours to sector k. Lastly, Γ_t denotes the aggregate preference shock. The flow budget constraint of the household is given by

$$P_t\,C_t\;+\;E_t\big[Q_{t,t+1}\,B_{t+1}\big]\;=\;B_t\;+\;\sum_{k=1}^{K}W_{k,t}\,H_{k,t}\;+\;\sum_{k=1}^{K}\int_{I_k}\Pi_{k,t}(i)\,\mathrm{d}i. \tag{2.2}$$

The aggregate consumption composite is defined as

$$C_t=\left(\sum_{k=1}^{K}\left(n_kD_{k,t}\right)^{1/\eta}C_{k,t}{}^{(\eta-1)/\eta}\right)^{\eta/(\eta-1)},$$

with

$$\sum_{k=1}^{K}n_kD_{k,t}=1.$$

Aggregate price index is given by

$$P_t=\left(\sum_{k=1}^{K}\left(n_kD_{k,t}\right)P_{k,t}^{1-\eta}\right)^{1/(1-\eta)}.$$

Sectoral consumption index:

$$C_{k,t}=\left(\left(\frac{1}{n_k}\right)^{1/\theta}\int_{\mathcal{I}_k}C_{k,t}(i)^{(\theta-1)/\theta}di\right)^{\theta/(\theta-1)}$$

Sectoral price index:

$$P_{k,t} = \left(\frac{1}{n_k} \int_{\mathcal{I}_k} P_{k,t}(i)^{1-\theta} di \right)^{1/(1-\theta)}.$$

where θ denotes the within-sector elasticity of substitution between consumption varieties.

Given the aggregate consumption good, C_t, and the price levels, $P_{k,t}$ and P_t, the optimal sectoral goods minimize the total expenditure $P_t C_t$. The demands for sectoral goods are then obtained as

$$C_{k,t} = n_k D_{k,t} \left(\frac{P_{k,t}}{P_t} \right)^{-\eta} C_t$$

Given $C_{k,t}$, the optimal demand for type-i good $C_{k,t}(i)$ is

$$C_{k,t}(i) = \frac{1}{n_k} \left(\frac{P_{k,t}(i)}{P_{k,t}} \right)^{-\theta} C_{k,t}.$$

The remaining FOCs are

$$Q_{t,t+1} = \beta \left(\frac{\Gamma_t}{\Gamma_{t+1}} \right) \left(\frac{C_t}{C_{t+1}} \right)^{-\sigma} \left(\frac{P_t}{P_{t+1}} \right),$$

$$\frac{W_{k,t}(i)}{P_t} = \omega_k H_{k,t}(i)^{\varphi} C_t^{\sigma}.$$

2.2.2 Firms

Firms use sector-specific labor to produce according to the following technology:

$$Y_{k,t}\,(i) = A_t A_{k,t} H_{k,t}(i). \tag{2.3}$$

Where $Y_{k,t}$ is the production of firm ik, A_t is economy-wide productivity, $A_{k,t}$ is sector-specific productivity, $H_{k,t}(i)$ denotes hours of labor that firm ik employs.

The firm's nominal profit is given by

$$\Pi_{k,t}(i) = P_{k,t}(i) Y_{k,t}(i) - (1 - \tau_{k,t}) \left[W_{k,t}(i) H_{k,t}(i) \right],$$

where $\tau_{k,t}$ is a time-varying subsidy to production costs. Real profit is:

$$\Pi_{k,t}^{real}(i) \equiv \frac{P_{k,t}(i)}{P_t} Y_{k,t}(i) - (1-\tau_{k,t})\,[\frac{W_{k,t}(i)}{P_t} H_{k,t}(i)]$$

from the production function 2.3, we can substitute $H_{k,t}(i)$ and get to

$$\Pi_{k,t}^{real}(i) \equiv \frac{P_{k,t}(i)}{P_t} Y_{k,t}(i) - (1-\tau_{k,t})\,[\frac{W_{k,t}(i)Y_{k,t}(i)}{P_t A_t A_{k,t}}] \tag{2.4}$$

Prices are sticky as in Calvo (1983): a firm in sector k adjusts its price with probability $1-\alpha_k$ each period. Thus, the sectoral price level $P_{k,t}$ evolves according to

$$\begin{aligned} P_{k,t} &= \left[\frac{1}{n_k}\int_{\mathcal{I}_{k,t}^*} P_{k,t}^{*1-\theta}\,di + \frac{1}{n_k}\int_{\mathcal{I}_k-\mathcal{I}_{k,t}^*} P_{k,t-1}(i)^{1-\theta}\,di\right]^{\frac{1}{1-\theta}} \\ &= \left[(1-\alpha_k)P_{k,t}^{*1-\theta} + \alpha_k P_{k,t-1}^{1-\theta}\right]^{\frac{1}{1-\theta}}. \end{aligned} \tag{2.5}$$

Reoptimizing firms maximize expected discounted profits:

$$\max_{P_{k,t}(i)} E_t \sum_{s=0}^{\infty} \alpha^s Q_{t,t+s}^{real} \Pi_{k,t+s}^{real}(i),$$

where $Q_{t,t+s}^{real} = \beta\left(\frac{\Gamma_t}{\Gamma_{t+s}}\right)\left(\frac{C_t}{C_{t+s}}\right)^{-\sigma}$ is the real stochastic discount factor between time t and $t+s$, and $\Pi_{k,t+s}^{real}(i)$ is the real profit at time $t+s$ under the condition that the firm's price has not been reoptimized since time t, in which case the demand function is given by

$$Y_{k,t+s}(i) = D_{k,t+s} Y_{t+s}\left(\frac{P_{k,t}(i)}{P_{k,t+s}}\right)^{-\theta}\left(\frac{P_{k,t+s}}{P_{t+s}}\right)^{-\eta}.$$

Let us rewrite the firm's objective function:

$$\max_{P_{k,t}(i)} E_t \sum_{s=0}^{\infty} \alpha^s Q_{t,t+s}^{real}\left[\frac{P_{k,t}(i)}{P_{t+s}} - (1-\tau_{k,t+s})\underbrace{\left(\frac{W_{k,t+s}(i)}{P_{t+s}A_{t+s}A_{k,t+s}}\right)}_{\equiv MC_{k,t+s}(i)}\right] Y_{k,t+s}(i),$$

where $MC_{k,t+s}(i)$ is firm ik's real marginal costs (in the absence of subsidy) in time $t+s$ when the firm has not reoptimzed since time t.

The first order condition is given by

$$0 = E_t \sum_{s=0}^{\infty} \alpha^s Q_{t,t+s}^{real} \left(\frac{P_{k,t}^*}{P_{k,t+s}} \right)^{-\theta} \left(\frac{P_{k,t+s}}{P_{t+s}} \right)^{-\eta} Y_{t+s} D_{k,t+s} \left[\frac{P_{k,t}^*}{P_{t+s}} - \tilde{U}_{k,t+s} MC_{k,t+s}(i) \right] \quad (2.6)$$

where $\tilde{U}_{k,t+s}$ are economy wide cost push shocks.

Loglinearizing the FOC gives:

$$E_t \sum_{s=0}^{\infty} \alpha_k^s \beta^s \left\{ p_{k,t}^* \right\} = E_t \sum_{s=0}^{\infty} \alpha_k^s \beta^s \left\{ mc_{k,t+s}(i) + p_{t+s} + \tilde{u}_{k,t+s} \right\},$$

$$h_{k,t}(i) = y_{k,t}(i) - a_{k,t} - a_t$$

$$c_t = y_t$$

$$h_{k,t}(i) = -\theta \left(p_{k,t}(i) - p_{k,t} \right) + y_{k,t} - a_{k,t} - a_t$$

The real marginal costs are (taking into account price rigidity):

$$\begin{aligned} mc_{k,t+s}(i) &= \left(w_{k,t+s}(i) - p_{t+s} \right) - a_{k,t+s} - a_{t+s} \\ &= \left(\varphi h_{k,t+s}(i) + \sigma c_{t+s} \right) - a_{k,t+s} - a_{t+s} \\ &= \left(\varphi \left(-\theta \left(p_{k,t+s}(i) - p_{k,t+s} \right) + y_{k,t+s} - a_{k,t+s} - a_{t+s} \right) + \sigma y_{t+s} \right) - a_{k,t+s} - a_{t+s} \\ &= -\varphi\theta \left(p_{k,t+s}(i) - p_{k,t+s} \right) + \varphi y_{k,t+s} - (1+\varphi)(a_{k,t+s} - a_{t+s}) + \sigma y_{t+s}. \end{aligned}$$

Replacing the FOC:

$$E_t \sum_{s=0}^{\infty} \alpha_k^s \beta^s \left\{ p_{k,t}^* \right\} = E_t \sum_{s=0}^{\infty} \alpha_k^s \beta^s \Big[-\varphi\theta \left(p_{k,t+s}(i) - p_{k,t+s} \right) + \varphi y_{k,t+s}$$

$$+\sigma y_{t+s} + p_{t+s} - (1+\varphi) \left(a_{k,t+s} - a_{t+s} \right) + \tilde{u}_{k,t+s} \Big],$$

notice that $p_{k,t+s}(i) = p^*_{k,t}$, so:

$$(1+\varphi\theta)E_t \sum_{s=0}^{\infty} \alpha_k^s \beta^s \left\{p^*_{k,t}\right\} = E_t \sum_{s=0}^{\infty} \alpha_k^s \beta^s \Big[\varphi\theta p_{k,t+s} + \varphi y_{k,t+s}$$

$$+\sigma y_{t+s} + p_{t+s} - (1+\varphi)\left(a_{k,t+s} - a_{t+s}\right) + \tilde{u}_{k,t+s}\Big],$$

$$\frac{(1+\varphi\theta)}{1-\alpha_k\beta}\left\{p^*_{k,t}\right\} = E_t \sum_{s=0}^{\infty} \alpha_k^s \beta^s \{\varphi\theta\left(p_{k,t+s}\right) + \varphi y_{k,t+s} + \sigma y_{t+s}$$

$$+p_{t+s} - (1+\varphi)(a_{k,t+s} - a_{t+s}) + \tilde{u}_{k,t+s}\},$$

$$p^*_{k,t} = (1-\alpha_k\beta)p_{k,t} + \alpha_k\beta\mathbb{E}_t[p^*_{k,t+1}] + \frac{1-\alpha_k\beta}{1+\varphi\theta}\left\{\varphi y_{k,t} + \sigma y_t - p^r_t - (1+\varphi)(a_{k,t} - a_t) + \tilde{u}_{k,t}\right\}$$

adding and subtracting $\alpha_k\beta\mathbb{E}_t[p_{k,t+1}]$:

$$p^*_{k,t} - p_{k,t} = (\alpha_k\beta)\mathbb{E}_t[\pi_{k,t+1}] + \alpha_k\beta\mathbb{E}_t[p^*_{k,t+1} - p_{k,t+1}]$$

$$+\frac{1-\alpha_k\beta}{1+\varphi\theta}\left\{\varphi y_{k,t} + \sigma y_t - p^r_t - (1+\varphi)(a_{k,t} - a_t) + \tilde{u}_{k,t}\right\} \tag{2.7}$$

From the sectoral price index, we have:

$$p^*_{k,t} - p_{k,t} = \frac{\alpha_k}{1-\alpha_k}\left(\pi_{k,t}\right). \tag{2.8}$$

substituting 2.8 into 2.2.2 we get:

$$\frac{\alpha_k}{1-\alpha_k}(\pi_{k,t}) = \frac{\alpha_k \beta}{1-\alpha_k}\mathbb{E}_t\left[(\pi_{k,t+1})\right] + \frac{1-\alpha_k\beta}{1+\varphi\theta}\left\{\varphi y_{k,t} + \sigma y_t - p_t^r - (1+\varphi)(a_{k,t}-a_t) + \tilde{u}_{k,t}\right\}$$

$$\pi_{k,t} = \beta\mathbb{E}_t[\pi_{k,t+1}] + \frac{(1-\alpha_k\beta)(1-\alpha_k)}{(1+\varphi\theta)(\alpha_k)}\left\{\varphi y_{k,t} + \sigma y_t - p_t^r - (1+\varphi)(a_{k,t}-a_t) + \tilde{u}_{k,t}\right\}$$

adding and subtracting φy_t we finally arrive at our Phillips curve:

$$\pi_{k,t} = \beta\mathbb{E}_t[\pi_{k,t+1}] + \frac{(1-\alpha_k\beta)(1-\alpha_k)}{(1+\varphi\theta)(\alpha_k)}\left\{\varphi y_{k,t}^r + (\sigma+\varphi)y_t - p_t^r - (1+\varphi)(a_{k,t}-a_t) + \tilde{u}_{k,t}\right\}. \tag{2.9}$$

To better visualize the Sectoral NKPC we change notation and get to:

$$\pi_{k,t} = \beta\mathbb{E}_t\pi_{k,t+1} + \psi_k\Theta_y y_t + \psi_k q_{k,t} + v_{k,t} \tag{2.10}$$

with

$$\psi_k = \frac{(1-\alpha_k)(1-\alpha_k\beta)}{\alpha_k},$$

$$\Theta_y = \frac{\sigma+\varphi}{1+\theta\varphi},$$

$$q_{k,t} = \frac{\varphi}{1+\theta\varphi}y_{k,t}^R - \frac{1}{1+\theta\varphi}p_{k,t}^R,$$

$$v_{k,t} = -\psi_k\frac{1+\varphi}{1+\theta\varphi}\tilde{a}_{k,t} + \psi_k\frac{1}{1+\theta\varphi}\tilde{u}_{k,t},$$

where:

$$\tilde{a}_{k,t} = a_t + a_{k,t},$$

$$\tilde{u}_{k,t} = u_t + u_{k,t}.$$

To get to the aggregated NKPC we use the definition:

$$\pi_t = \sum_{k=1}^{K} n_k \pi_{k,t}$$

This leads us to:

$$\pi_t = \beta\mathbb{E}_t\pi_{t+1} + \left(\sum_{k=1}^{K} n_k\psi_k\right)\Theta_y y_t + \sum_{k=1}^{K} n_k\psi_k q_{k,t} + v_t \tag{2.11}$$

For a complete version of the model and the equilibrium one should refer to Carvalho *et al.* (2021).

2.3 Data

We use sectoral data from the Personal Consumption Expenditures (PCE) series published by the U.S. Bureau of Economic Analysis (BEA). Specifically, we focus on the first level of disaggregation of the PCE, which divides total consumption into 15 broad sectors. This level of aggregation strikes a balance between capturing meaningful heterogeneity in price-setting behavior and keeping the number of moments in our GMM estimation tractable. Using more granular item-level data, while in principle desirable, would substantially increase the dimensionality of the estimation problem and exacerbate the weak instrument concerns already noted in the Phillips Curve literature.

Our dataset spans the period from 1975 to 2024. For each sector, we collect data on nominal and real expenditures, allowing us to compute sectoral price indices and quantities. These series are used to construct sector-specific inflation rates and output.

To identify demand shocks, we rely on two external instruments. The first is the series of high-frequency monetary policy shocks constructed by Bauer & Swanson (2023), which captures unexpected changes in policy rates around FOMC announcements. The second is the series for U.S. defense expenditures, obtained from the BEA, which provides an alternative source of variation plausibly orthogonal to sectoral demand conditions. These instruments have been widely used in recent empirical work to address endogeneity concerns in Phillips Curve estimation. We discuss their construction and implementation in greater detail in Section 2.4.

To create our benchmarks for sectoral frequency of price changes we use the data from Nakamura & Steinsson (2008).

All series are seasonally adjusted and we apply a Hodrick-Prescott (HP) filter to prices and quantities to remove low-frequency trends. Variables are expressed in logarithmic differences when appropriate.

2.4 Estimation

2.4.1 The bias

As shown in Section 2.2 the model leads us to the aggregated Phillips Curve:

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \left(\sum_{k=1}^{K} n_k \psi_k \right) \Theta_y y_t + \sum_{k=1}^{K} n_k \psi_k q_{k,t} + v_t$$

The heterogeneous stickness in price makes it so the term $q_{k,t}$ is present in the NKPC. It is very easy to observe this point, by construction:

$$\sum_{k=1}^{K} n_k q_{k,t} = 0$$

so if the stickness is the same for all sectors we get to:

$$\psi_k = \psi \; \forall k \quad \Longrightarrow \quad \sum_{k=1}^{K} n_k \psi_k q_{k,t} = 0, \quad \forall \, \psi \in \mathbb{R}.$$

and

$$\pi_t = \sum_{k=1}^{K} n_k \beta \mathbb{E}_t \pi_{k,t+1} + \sum_{k=1}^{K} n_k \psi_k \Theta_y y_{k,t} + \sum_{k=1}^{K} n_k v_{k,t} \tag{2.12}$$

Which leads us to:

$$\pi_t = \beta E_t \pi_{t+1} + \psi \Theta_y y_t + v_{k,t} \tag{2.13}$$

This means that when we estimate the Phillips curve without taking into account this heterogeneity there is an actual omitted variable bias that leads us to estimate a flatter slope for the Phillips curve. To understand why, one can think of a illustrative example with only two sectors: one of them has flexible prices and the other one has sticky prices. In a contractionary (expansionary) shock to the whole economy, the flexible sector decreases (increases) its prices while the sticky sector doesn't. As a consequence, demands shifts make the relative demand of the flexible sector to raise (lower). Meanwhile, relative demand of the sticky sector goes in the opposite direction. Since ψ_k is higher for the sector with flexible prices, $\sum_{k=1}^{K} n_k \psi_k q_{k,t}$ increases (decreases).

Summarizing:

- Contractionary shock ($y_t \downarrow$)
  - $y_{k,t}^R > 0$ (flexible sector) and $y_{k,t}^R < 0$ (sticky sector)
  - $p_{k,t}^R < 0$ (flexible sector) and $p_{k,t}^R > 0$ (sticky sector)
  - $\Longrightarrow q_t \uparrow$

– $\text{Cov}(y_t, q_t) < 0$ and $\text{Cov}(\pi_t, q_t) > 0 \implies$ downward bias

It's important to note that this bias doesn't arise from using output instead of marginal cost in the NKPC. The potential problem of using output is that the supply shocks a_t and $a_{k,t}$ are both present in $v_{k,t}$ and has no connection with the aforementioned bias.

2.4.2 Empirical approach

As demonstrated in Subsection 2.4.1, estimating the NKPC while ignoring heterogeneity of price stickness leads to a bias. This is not corrected with the most common instruments, usually related with demand shocks, as they would also correlate with the omitted term. Because of this, to properly estimate $\sum_{k=1}^{K} n_k \psi_k \Theta_y$ (the effect of an increase in output in every sector) we need to estimate the NKPC sector by sector.

The SNKPC was derived previously as:

$$\pi_{k,t} = \beta \mathbb{E}_t \pi_{k,t+1} + \psi_k \Theta_y y_t + \psi_k q_{k,t} + v_{k,t}$$

with:

$$v_{k,t} = \frac{\psi_k}{(1+\theta\varphi)} \left(-a_t - a_{k,t} + \tilde{u}_{k,t}\right)$$

$$\psi_k = \frac{(1-\alpha_k)(1-\alpha_k\beta)}{\alpha_k}$$

$$\Theta_y = \frac{\sigma+\varphi}{1+\theta\varphi}$$

$$q_{k,t} = \frac{\varphi}{1+\theta\varphi} y_{k,t}^R - \frac{1}{1+\theta\varphi} p_{k,t}^R$$

Considering:

$$\eta_{k,t+1} = \pi_{k,t+1} - \mathbb{E}_t \pi_{k,t+1}$$

Using the equations of $\eta_{k,t+1}$, $q_{k,t}$ and Θ_y, the SNKPC can be rewritten as:

$$\pi_{k,t} = \beta \pi_{k,t+1} + \frac{\psi_k(\sigma+\varphi)}{1+\theta\varphi} y_t + \frac{\psi_k \varphi}{1+\theta\varphi} y_{k,t}^R - \frac{\psi_k}{1+\theta\varphi} p_{k,t}^R + v_{k,t} - \beta \eta_{k,t+1}$$

Defining:

$$\delta_k = \frac{\psi_k}{1+\theta\varphi},$$

we can write:

$$\pi_{k,t} = \beta\pi_{k,t+1} + \delta_k\left(\sigma+\varphi\right)y_t + \delta_k\varphi y^R_{k,t} - \delta_k p^R_{k,t} + v_{k,t} - \beta\eta_{k,t+1}$$

Since β,φ and θ are common across all sectors, we estimate the SNKPC simultaneously in a system GMM using for each sector the moments:

$$\mathbb{E}_t\left[\left(\beta\pi_{k,t+1} + \delta_k\left(\sigma+\varphi\right)y_t + \delta_k\varphi y^R_{k,t} - \delta_k p^R_{k,t} - \pi_{k,t}\right)z_{k,t}\right] = 0. \tag{2.14}$$

with $z_{k,t}$ being a instrument. For that instrument to be valid we need:

$$\mathbb{E}_t[(v_{k,t} - \beta\eta_{k,t+1})z_{k,t}] = 0$$

We will discuss the instruments used in the estimation in the next subsection. With δ_k estimated, we can recover the slope of the aggregated Phillips curve and the sectoral frequency of price adjustment using:

$$\kappa \equiv \left(\sum_{k=1}^{K} n_k \frac{\psi_k}{1+\theta\varphi}\right)(\sigma+\varphi)$$

$$\alpha_k = \frac{1+\beta+\psi_k - \sqrt{(1+\beta+\psi_k)^2 - 4\beta}}{2\beta}$$

In the GMM we restrict the coefficients estimated to be positive in conform with macroeconomic theory.

2.4.3 Instruments

We rely on four external instruments in our estimation strategy to address potential endogeneity in the Phillips Curve and to achieve identification of sectoral and aggregate relationships: monetary policy shocks, military expenditures, sectoral demand shocks, and aggregate demand shocks.

The first two instruments are standard in the literature. Monetary policy shocks are constructed using the high-frequency series of unexpected changes in policy rates around Federal Open Market Committee (FOMC) announcements, following the approach in Bauer & Swanson (2023). These shocks capture exogenous variation in monetary policy by focusing on movements that were not anticipated by financial markets. Since anticipated policy changes should already be incorporated into private agents' decisions, these high-frequency surprises are plausibly orthogonal to the contemporaneous state of the economy and are valid instruments for identifying demand-driven fluctuations.

Military expenditures provide a second source of exogenous variation. We use the three-year difference in U.S. defense spending, obtained from the Bureau of Economic Analysis (BEA). Defense expenditures in the United States are largely determined by political and geopolitical considerations rather than by the cyclical state of the economy, making them a plausible instrument. Moreover, as a sizable component of aggregate demand, military spending affects both overall economic activity and relative sectoral demands, thereby serving as a valid instrument for both aggregate and sectoral dynamics.

The remaining two instruments are constructed using a structural vector autoregression (SVAR) model with sign restrictions, following the methodology of Rubio-Ramirez *et al.* (2010). In particular, we estimate a simplified two-variable SVAR of output and prices:

$$\mathbf{X}_t = \mathbf{fi}_1 \mathbf{X}_{t-1} + \mathbf{fi}_2 \mathbf{X}_{t-2} + \mathbf{u}_t, \quad \mathbf{X}_t = \begin{pmatrix} y_t \\ p_t \end{pmatrix}, \quad \mathbf{u}_t = \begin{pmatrix} u_{y,t} \\ u_{p,t} \end{pmatrix}.$$

We impose sign restrictions on the impulse responses to separate demand and supply shocks, requiring demand shocks to produce a positive co-movement between output and prices, and supply shocks to generate an inverse relationship:

$$\begin{array}{c|cc} & y_t & p_t \\ \hline \varepsilon^D & + & + \\ \varepsilon^S & + & - \end{array}.$$

Following Rubio-Ramirez *et al.* (2010), we adopt a Bayesian approach and draw random orthonormal matrices to rotate the reduced-form residuals. For each draw, we retain the candidate shock series if the sign restrictions on impulse responses are satisfied. We iterate this procedure until obtaining 10,000 admissible draws, allowing us to construct a posterior distribution of structural shocks consistent with our identifying assumptions.

The median of the demand shock series is then used as an instrument for sectoral demand, while a weighted sum of sectoral demand shocks serves as the instrument for aggregate demand. This approach ensures that both instruments reflect orthogonal demand-driven fluctuations and are robust to alternative identification schemes.

Notice that the goal of this instrument is to remove supply shocks from the output. If the instrument is not able to do that the slope we would find in the NKPC would be smaller than the real one, so we can treat the slope we find as a lower bound for the real coefficient.

2.5 Empirical Results

For each specification we present two results, one with every parameter except from θ (would not be possible) estimated, and one with β,φ and σ calibrated to 0.99, 1 and 1, respectively. In the Appendix we also present the results with different calibrations. The data on monetary policy shocks is available starting in 1990, so we first present results using data from 1990-2024. Later, for robustness check and to test for sample bias, we remove this instrument and estimate the results with two sub-samples, one from 1975-2000 and the other for 2001-2024.

We first present the list of items in Table 2.1 so one can better understand the coefficients that are specific to industries like δ_k and α_k.

The correlation with micro data infrequencies is compared to the results presented by Nakamura & Steinsson (2008).

	Category
1	Motor vehicles and parts
2	Furnishings and durable household equipment
3	Recreational goods and vehicles
4	Other durable goods
5	Food and beverages purchased for off-premises consumption
6	Clothing and footwear
7	Gasoline and other energy goods
8	Other nondurable goods
9	Housing and utilities
10	Healthcare
11	Transportation services
12	Recreation services
13	Food services and accommodations
14	Financial services and insurance
15	Other services

Table 2.1: Items

2.5.1 Uncalibrated

	Estimate	Std. Error	T-value	P-Value
δ_1	0.1153	0.0437	2.6353	0.0084
δ_2	0.0357	0.0132	2.7099	0.0067
δ_3	0.0218	0.0082	2.6483	0.0081
δ_4	0.0113	0.0038	2.9642	0.0030
δ_5	0.0000	0.0023	0.0000	1.0000
δ_6	0.0200	0.0075	2.6777	0.0074
δ_7	0.1567	0.0653	2.3987	0.0165
δ_8	0.0061	0.0019	3.2489	0.0012
δ_9	0.0039	0.0016	2.4109	0.0159
δ_{10}	0.0035	0.0012	2.9937	0.0028
δ_{11}	0.0208	0.0083	2.4940	0.0126
δ_{12}	0.0028	0.0013	2.2533	0.0242
δ_{13}	0.0000	0.0004	0.0000	1.0000
δ_{14}	0.0000	0.0020	0.0000	1.0000
δ_{15}	0.0000	0.0005	0.0000	1.0000
β	0.7659	0.0263	29.1101	0.0000
φ	1.4398	0.5168	2.7857	0.0053
σ	4.6110	1.8191	2.5347	0.0113

Table 2.2: GMM Results, not calibrated

κ
0.2075^{***}
(0.0171)

Table 2.3: Slope of the Phillips curve, not calibrated

	$\theta = 3$	$\theta = 4$	$\theta = 5$	$\theta = 8$	$\theta = 10$
Correlation with benchmarks	0.6862	0.6815	0.6770	0.6646	0.6571

Table 2.4: Correlation with micro-data, not calibrated

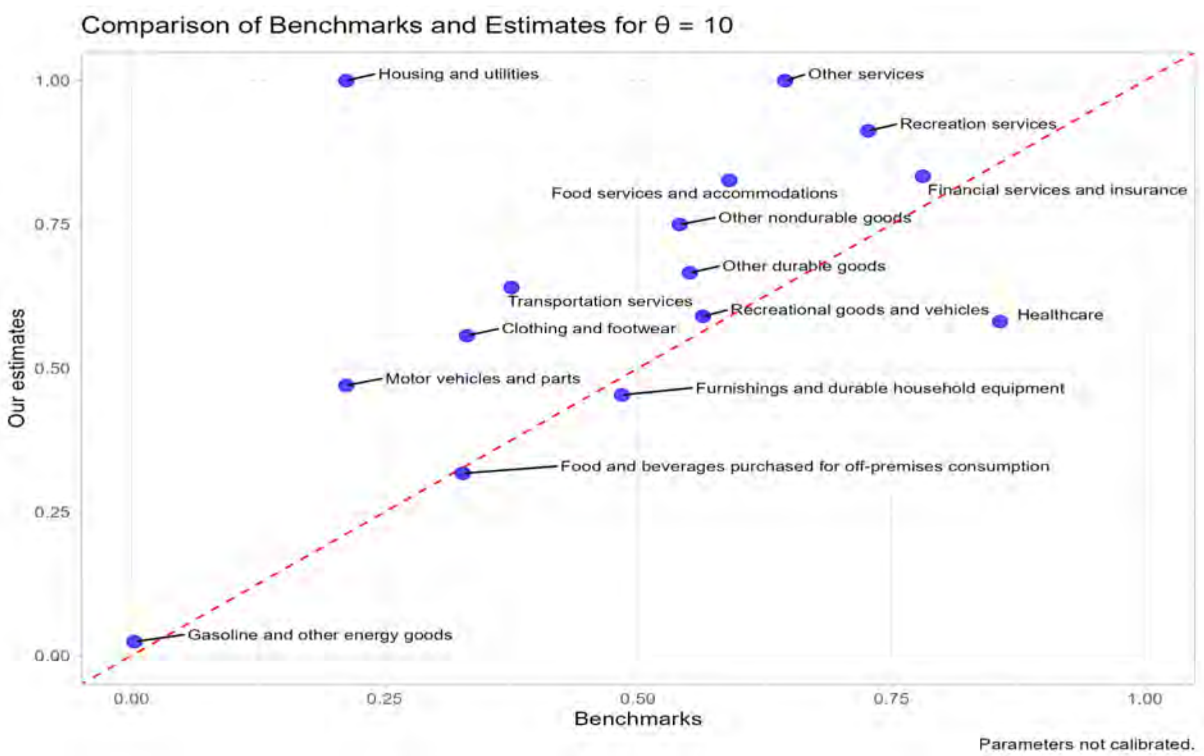
As we show in Section 2.4, we are actually estimating:

$$\delta_k = \frac{\psi_k}{1+\theta\varphi}$$

and

$$\kappa \equiv \left(\sum_{k=1}^{K} n_k \frac{\psi_k}{1+\theta\varphi} \right) (\sigma + \varphi)$$

This means that changing our calibration of θ doesn't affect the estimation of $delta_k$ or κ and only have an effect on the calculated infrequencies of each sector. Below, on table x we present the results of the infrequencies we find and in graph comparing them with the benchmarks.



As we can see, even with no calibration we find results closely matching the infrequencies found in papers using micro-data and a reasonable value for φ, the Frisch elasticity. The value found for σ would indicate a low EIS $\equiv \frac{1}{\sigma}$ compared to calibrated models designed to match growth and fluctuations facts, but the values found in empirical studies using aggregate consumption data typically find the EIS to be close to zero, as shown by Guvenen (2006). The result found for β, however, is low compared to the literature, typically required to be closer to one.

Since we are closely relating our empirical analyses with the NKPC derived in the model, we also estimate it calibrating the common parameters across sectors.

2.5.2
Results calibrated $\beta = 0.99$, $\varphi = 1$,$\sigma = 1$

	Estimate	Std. Error	T-value	P-Value
δ_1	0.2423	0.0072	33.5744	0.0000
δ_2	0.1224	0.0046	26.3951	0.0000
δ_3	0.0568	0.0030	18.8604	0.0000
δ_4	0.0470	0.0020	23.3480	0.0000
δ_5	0.0134	0.0060	2.2319	0.0256
δ_6	0.0430	0.0029	14.9073	0.0000
δ_7	0.7573	0.0584	12.9689	0.0000
δ_8	0.0474	0.0025	19.2879	0.0000
δ_9	0.0313	0.0030	10.5442	0.0000
δ_{10}	0.0249	0.0015	16.3452	0.0000
δ_{11}	0.0288	0.0026	11.2328	0.0000
δ_{12}	0.0096	0.0010	9.4770	0.0000
δ_{13}	0.0154	0.0009	16.4227	0.0000
δ_{14}	0.0000	0.0083	0.0000	1.0000
δ_{15}	0.0000	0.0016	0.0000	1.0000

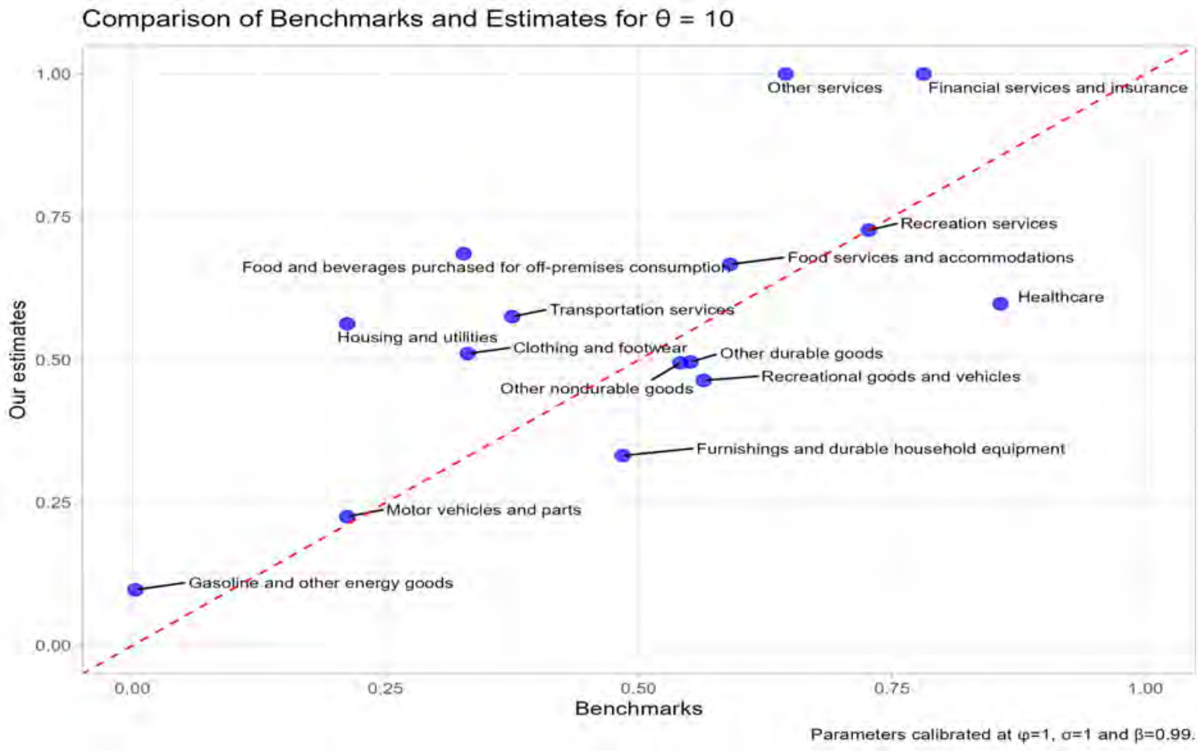
Table 2.5: GMM Results, calibrated

κ
0.2740^{***}
(0.0152)

Table 2.6: Slope of the Phillips curve, calibrated

	$\theta = 3$	$\theta = 4$	$\theta = 5$	$\theta = 8$	$\theta = 10$
Correlation with benchmarks	0.7108	0.7063	0.7018	0.6890	0.6813

Table 2.7: Correlation with micro-data, calibrated



As we can see, the results when calibrating β, φ and σ yield an even closer correlation with the microdata and a steeper NKPC slope.

To address how much of our coefficient is influenced by the instruments and how much is by the specification and data, we compare the results presented with different specification, like the NKPC considering homogeneous price stickness across sectors but using sectoral data and using aggregated data.

For each of this specification, including the one we present our main results, we test it using and not using instruments. Additionally, for this exercise we use β, φ and σ calibrated as it would be very hard to compare results of κ if those parameters were to be different in each specification.

In Table 2.8 columns I and II are standard NKPC using aggregated data, like this:

$$\mathbb{E}_t[(\pi_t - 0.99\pi_{t+1} - \kappa y_t)z_t] = 0 \tag{2.15}$$

In column I with z_t being y_t and π_{t+1}, so, no instruments but respecting that residuals should not correlate with either of the variables in the regression. In column II the instruments z_t are: monetary policy shocks, defense expending and the aggregate demand shock we generated.

For columns III and IV

$$\mathbb{E}_t \left[\left(0.99\pi_{k,t+1} + 2\delta y_t + \delta y^R_{k,t} - \delta p^R_{k,t} - \pi_{k,t} \right) z_{k,t} \right] = 0. \tag{2.16}$$

and with that $\kappa = 2\delta$, in III we use the regressors as the instruments and in IV we use the same instruments as in the main specification.

In Columns V and VI we use the specification 2.14 used before, only with the parameters calibrated.

	I	II	III	IV	V	VI
κ	0.0389	0.0698	0.0042***	0.0277***	0.0125***	0.2740***
	(0.0955)	(0.2514)	(0.0001)	(0.0001)	(0.0040)	(0.0152)

Standard errors in parentheses, *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$, I Aggregate GMM without instruments, II Aggregate GMM with instruments, III Homogeneous model without instruments. IV Homogeneous calibrated instrument, V Heterogeneous model without instruments, VI Heterogeneous instrument.

Table 2.8: Aggregate GMM

As we can see in Table 2.8 both the instruments and the specification seem to matter for the estimation of the Phillips curve.

2.5.3 Sub-sample

To investigate uncertainties involving the sample, a typical practice has been to divide the sample into two or more periods, e.g. Gali and Gertler (1999) and Clarida, Gali, and Gertler (2000).

To do so we drop the monetary policy shocks instrument. This reduces the number of moments we have and allows us to expand our sample from 1990-2024 to 1975-2024. We split this in two samples, 1975-2000 and 2001-2024.

We again use β, φ and σ calibrated as it would be very hard to compare results of κ if those parameters were to be different in each specification. In the Appendix we present the results for the not calibrated case.

	1975–1999	2000–2024
κ	0.3293***	0.2395***
	(0.0483)	(0.0196)
Correlation with benchmarks	0.3326	0.6778

Table 2.9: Comparing, calibrated

We can see significant high values for the slope of the Phillips curve in both specifications. It is in accordance to literature that the slope is reducing with time as this has been stated in many papers. It is also expected that we would find a smaller correlation with the benchmarks infrequencies in the older period since the analysis in Nakamura & Steinsson (2008) is done using data from 1998-2005.

2.6 Conclusion

This paper has shown that accounting for sectoral heterogeneity in price stickiness is critical for both the theoretical and empirical assessment of the New Keynesian Phillips Curve. Building on multi-sector Calvo-style models, we derive a system of sectoral Phillips Curves and an aggregate Phillips Curve that explicitly incorporate relative differences in nominal rigidity. Ignoring these differences generates an omitted-variable bias—driving down ordinary estimates of the aggregate slope—and leads to misleading inferences about the degree of nominal rigidity in the economy.

To address these challenges, we develop a joint GMM estimation strategy that exploits cross-sectoral variation in PCE price and quantity data (1975–2024) and employs external instruments (high-frequency monetary surprises, defense spending, and SVAR-based demand shocks). By estimating all 15 sectoral equations simultaneously, we discipline common parameters (discount factor, labor supply elasticity) with economic-theoretic cross-equation restrictions, improving identification and consistency with the underlying model.

Our empirical results reveal that once heterogeneity is accounted for, the estimated aggregate Phillips slope rises markedly—from near zero in homogeneous specifications to roughly 0.27 under calibrated parameters—and remains statistically robust across alternative calibrations and subsamples. Moreover, the implied sectoral frequencies of price adjustment (the α_k's) display correlations of about 0.7 with micro-level menu-cost evidence (Nakamura & Steinsson (2008)), providing external validation for our approach. Finally, sub-sample analyses show a slight time-variation in the slope, consistent with changes in underlying frictions documented elsewhere.

These findings underscore the importance of integrating disaggregated price-setting behavior into aggregate inflation models. For policy, our results suggest that central banks relying on a single, homogeneous Phillips Curve may understate the true sensitivity of inflation to economic slack. Future research could extend this framework by incorporating firm-level micro data, allowing for endogenous sectoral linkages, or exploring the role of marginal-cost proxies at the sectoral level.

2.A Appendix

	1975–1999	2000–2024
κ	0.5500***	0.3302***
	(0.0865)	(0.0215)
β	0.8013***	0.9744***
	(0.0256)	(0.0076)
φ	4.0777***	0.1938***
	(2.5293)	(0.2379)
σ	11.6153***	7.3453***
	(6.3995)	(2.2453)
Correlation with benchmarks	0.5519	0.6889

Table 2.10: Comparing, not calibrated

Infrequencies	Benchmark	$\theta = 3$	$\theta = 4$	$\theta = 5$	$\theta = 8$	$\theta = 10$
α_1	0.212	0.3884	0.3503	0.3198	0.2552	0.2256
		(0.0051)	(0.0050)	(0.0049)	(0.0045)	(0.0043)
α_2	0.484	0.5051	0.4674	0.4362	0.3664	0.3325
		(0.0063)	(0.0065)	(0.0065)	(0.0065)	(0.0063)
α_3	0.564	0.6261	0.5928	0.5644	0.4981	0.4641
		(0.0077)	(0.0081)	(0.0084)	(0.0089)	(0.0091)
α_4	0.551	0.6530	0.6212	0.5939	0.5296	0.4962
		(0.0059)	(0.0063)	(0.0065)	(0.0070)	(0.0072)
α_5	0.327	0.7974	0.7760	0.7572	0.7109	0.6857
		(0.0411)	(0.0447)	(0.0477)	(0.0546)	(0.0580)
α_6	0.331	0.6650	0.6339	0.6072	0.5438	0.5109
		(0.0091)	(0.0096)	(0.0101)	(0.0109)	(0.0112)
α_7	0.003	0.2077	0.1786	0.1568	0.1151	0.0978
		(0.0105)	(0.0096)	(0.0088)	(0.0070)	(0.0062)
α_8	0.541	0.6518	0.6199	0.5925	0.5281	0.4947
		(0.0072)	(0.0076)	(0.0079)	(0.0085)	(0.0087)
α_9	0.212	0.7063	0.6778	0.6532	0.5941	0.5628
		(0.0117)	(0.0125)	(0.0131)	(0.0145)	(0.0151)
α_{10}	0.857	0.7332	0.7066	0.6836	0.6277	0.5979
		(0.0070)	(0.0075)	(0.0080)	(0.0089)	(0.0093)
α_{11}	0.375	0.7162	0.6884	0.6643	0.6064	0.5756
		(0.0107)	(0.0115)	(0.0121)	(0.0134)	(0.0139)
α_{12}	0.727	0.8261	0.8073	0.7907	0.7495	0.7268
		(0.0085)	(0.0093)	(0.0100)	(0.0115)	(0.0123)
α_{13}	0.590	0.7839	0.7614	0.7417	0.6932	0.6669
		(0.0059)	(0.0064)	(0.0068)	(0.0077)	(0.0082)
α_{14}	0.781	1.0000	1.0000	1.0000	1.0000	1.0000
		(3.3306)	(4.1632)	(4.9957)	(7.4931)	(9.1578)
α_{15}	0.645	1.0000	1.0000	1.0000	1.0000	1.0000
		(0.6246)	(0.7808)	(0.9369)	(1.4053)	(1.7175)

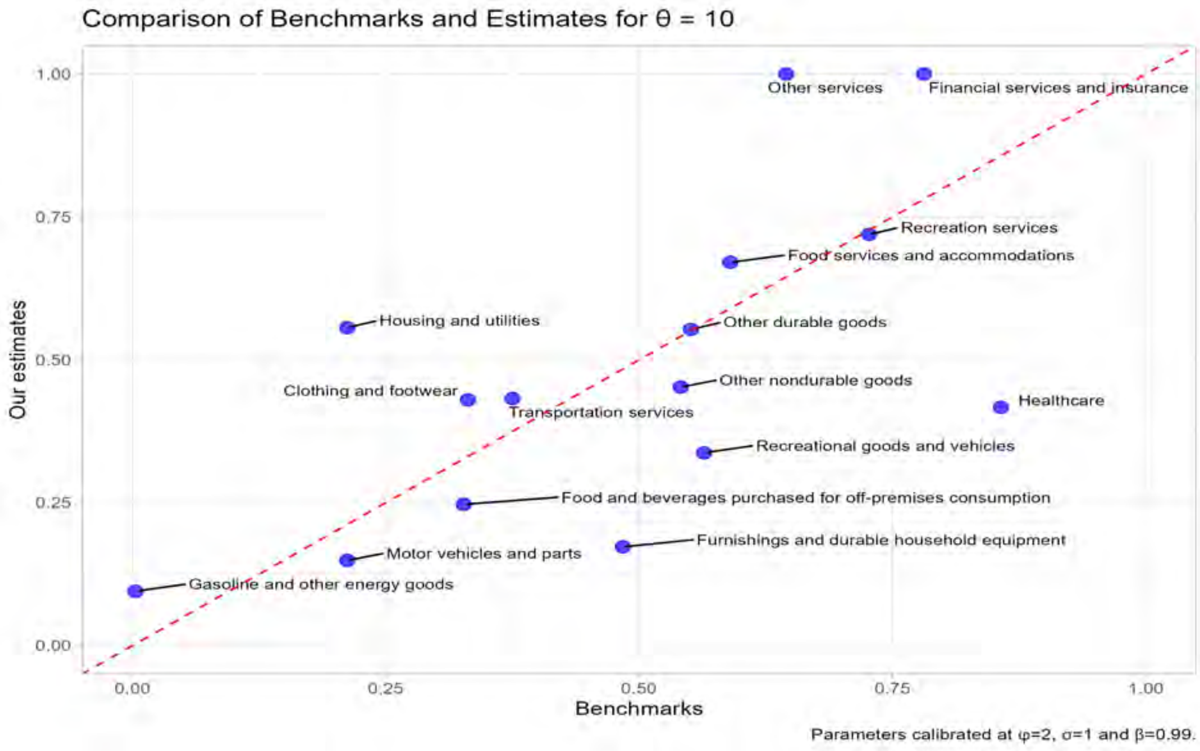
Table 2.11: Infrequencies, calibrated

Infrequencies	Benchmark	$\theta = 3$	$\theta = 4$	$\theta = 5$	$\theta = 8$	$\theta = 10$
α_1	0.212	0.5012	0.4553	0.4183	0.3391	0.3022
		(0.3972)	(0.3189)	(0.2584)	(0.1352)	(0.0802)
α_2	0.484	0.7070	0.6688	0.6363	0.5606	0.5218
		(0.8253)	(0.7331)	(0.6607)	(0.5080)	(0.4365)
α_3	0.564	0.7775	0.7449	0.7166	0.6488	0.6130
		(1.0248)	(0.9275)	(0.8508)	(0.6886)	(0.6122)
α_4	0.551	0.8527	0.8278	0.8057	0.7509	0.7209
		(1.3063)	(1.2040)	(1.1227)	(0.9492)	(0.8669)
α_5	0.327	1.0000	1.0000	1.0000	1.0000	1.0000
		(2.4463)	(2.4445)	(2.4427)	(2.4380)	(2.4352)
α_6	0.331	0.7885	0.7568	0.7293	0.6631	0.6279
		(1.0614)	(0.9634)	(0.8861)	(0.7225)	(0.6454)
α_7	0.003	0.4424	0.3966	0.3604	0.2850	0.2509
		(0.3034)	(0.2305)	(0.1749)	(0.0673)	(0.0337)
α_8	0.541	0.9050	0.8869	0.8704	0.8283	0.8046
		(1.5765)	(1.4747)	(1.3920)	(1.2117)	(1.1251)
α_9	0.212	0.9329	0.9190	0.9062	0.8728	0.8535
		(1.7554)	(1.6580)	(1.5769)	(1.3948)	(1.3053)
α_{10}	0.857	0.9387	0.9257	0.9138	0.8824	0.8641
		(1.8020)	(1.7072)	(1.6277)	(1.4477)	(1.3586)
α_{11}	0.375	0.7832	0.7511	0.7232	0.6562	0.6207
		(1.0391)	(0.9410)	(0.8637)	(0.7001)	(0.6230)
α_{12}	0.727	0.9478	0.9364	0.9259	0.8978	0.8814
		(1.8710)	(1.7798)	(1.7023)	(1.5240)	(1.4343)
α_{13}	0.590	1.0000	1.0000	1.0000	1.0000	1.0000
		(2.4558)	(2.4564)	(2.4570)	(2.4587)	(2.4599)
α_{14}	0.781	1.0000	1.0000	1.0000	1.0000	1.0000
		(2.4610)	(2.4631)	(2.4653)	(2.4722)	(2.4771)
α_{15}	0.645	1.0000	1.0000	1.0000	1.0000	1.0000
		(2.4528)	(2.4526)	(2.4523)	(2.4517)	(2.4513)

Table 2.12: Infrequencies, not calibrated

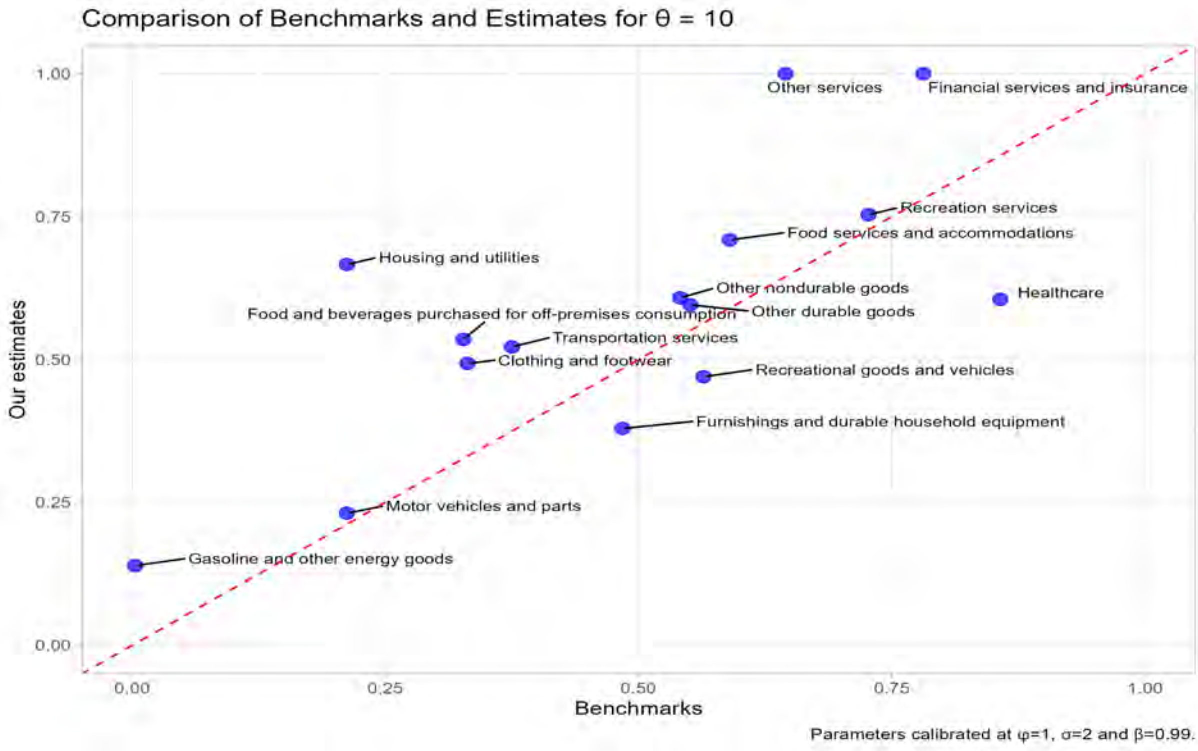
	$\theta = 3$	$\theta = 4$	$\theta = 5$	$\theta = 8$	$\theta = 10$
κ	0.2830***	0.2830***	0.2830***	0.2830***	0.2830***
	(0.0161)	(0.0161)	(0.0161)	(0.0161)	(0.0161)
Correlation with benchmarks	0.7039	0.6977	0.6917	0.6752	0.6656

Table 2.13: Slope of the Phillips Curve and correlations ($\varphi = 2$ and $\sigma = 1$)



	$\theta = 3$	$\theta = 4$	$\theta = 5$	$\theta = 8$	$\theta = 10$
κ	0.2345***	0.2345***	0.2345***	0.2345***	0.2345***
	(0.0172)	(0.0172)	(0.0172)	(0.0172)	(0.0172)
Correlation with benchmarks	0.6186	0.6108	0.6035	0.5841	0.5728

Table 2.14: Slope of the Phillips Curve and correlations ($\varphi = 1$ and $\sigma = 2$)



Comparison of Benchmarks and Estimates for θ = 10
Our estimates
Benchmarks
0.00
0.25
0.50
0.75
1.00
Other services
Financial services and insurance
Recreation services
Food services and accommodations
Housing and utilities
Other nondurable goods
Healthcare
Food and beverages purchased for off-premises consumption
Other durable goods
Transportation services
Clothing and footwear
Recreational goods and vehicles
Furnishings and durable household equipment
Motor vehicles and parts
Gasoline and other energy goods
Parameters calibrated at φ=1, σ=2 and β=0.99.

3
Properties of Realized Measures of Variance

Samuel Efraim[1] **Rosalia Kjaer**[1] **Marcelo Medeiros**[1,2]
[1] PUC–Rio
[2] UIUC

Abstract: This study investigates the characteristics of volatility in financial asset returns, a critical variable for financial decision-making across various domains, including risk management, portfolio optimization, and option pricing. Since volatility cannot be directly observed, it necessitates estimation, which has led to the development of various estimators, especially with the rise of high-frequency data. These estimators are designed to address challenges such as microstructure noise and discontinuities in asset prices. Despite numerous studies comparing different measures, the literature lacks consensus on the optimal volatility estimator, with much research primarily focusing on highly liquid equities. This paper utilizes high-frequency data from nearly 11,000 stocks spanning the years 2003 to 2020, providing a comprehensive analysis across both NASDAQ and NYSE, thus extending beyond the S&P 500 benchmark. The study examines the distributional properties of daily returns across various estimators, exploring how stock-specific attributes like liquidity and book-to-market ratios influence the distribution of daily returns. The findings suggest that stocks with lower liquidity display non-normal, fatter-tailed return distributions, and that simpler volatility measures—such as the 5-minute realized volatility—often outperform their more complex counterparts in evaluating the performance of illiquid stocks. These findings have significant implications for risk analysis and modeling future volatility forecasts.
Keywords: Volatility; Realized Measures; Normality; Asset Return.

3.1 Introduction

Estimating and predicting the volatility of financial asset returns is of paramount importance in the field of financial economics. Accurate measurement of volatility is crucial for effective portfolio construction, robust risk assessment, option pricing, alongside various other financial applications. Given that volatility is not directly observable in asset markets, the current methodological emphasis is on estimating it through high-frequency data; refer to McAleer & Medeiros (2008) and Barndorff-Nielsen & Shephard (2006b) for comprehensive reviews. Both academics and market practitioners acknowledge the intricate characteristics of asset returns, which has spurred considerable research on their distributional properties.

In this paper, we revisit the work of Andersen *et al.* (2001) nearly 25 years later, emphasizing the need for updated analyses in this evolving field. Unlike most studies that concentrate solely on S&P 500 stocks, we utilize a comprehensive cross-section of nearly 11,000 stocks from the Trades-and-Quotes (TAQ) database over 20 years, spanning data from 2003 to 2020, to examine the distributional properties of daily returns alongside several realized measures of daily volatility. By analyzing this broader range of assets, we advance the understanding of volatility, thereby refining existing knowledge and addressing a more relevant time frame.

The main takeaways from this paper are as follows. First, we confirm the early findings of Andersen *et al.* (2001) only for the most liquid stocks. We investigate whether firm- and stock-specific characteristics are correlated with the normality of standardized returns across various volatility estimators. Second, we find strong evidence that the standardized returns of less liquid stocks deviate from normality, displaying fatter tails than typically associated with a normal distribution. The use of various volatility measures does not change this conclusion. Third, we notice that a simple estimator, such as the 5-minute realized volatility, outperforms more complex volatility measures designed to address issues like jumps and microstructure noise when applied to illiquid stocks. Finally, we employ decision tree methods to identify potential cut-off points based on firm- and stock-specific features where stocks deviate from normality. These decision trees are also utilized to compare the normality of standardized returns across different volatility measures.

Illiquid stocks often exhibit higher expected returns, as noted by Amihud (2002), and can significantly improve portfolio optimization. However, for this enhancement to be effective, it is crucial to comprehend how the distribution of illiquid stocks may diverge from standard assumptions. While further research is needed in this area, this paper illustrates that presuming normality in the distribution of these stocks is inappropriate.

Over the past two decades, the availability of a much larger cross-section of high-frequency intraday data has enabled the development of new volatility estimators designed to correct potential biases within the standard Realized Variance (RV) estimator. A significant challenge stems from microstructure noise, which can be affected by factors such as the bid-ask spread. This noise introduces bias into the RV estimator when finer time intervals are used. However, completely discarding this information would compromise the principles of continuous-time modeling. To tackle this issue, estimators like the kernel-based estimator and Two-Time-Scale Realized Variance (TSRV) have been created to reduce noise in high-frequency data. Furthermore, the occurrence of price jumps can complicate volatility estimation further, introducing additional biases that must be considered in robust volatility modeling.

Comparative studies of these estimators have emerged naturally. Liu *et al.* (2015) and Arnerić & Žigman (2022), among others, have investigated the relative merits of various estimators. Further comparisons are presented in Hansen & Lunde (2005). Despite these extensive analyses, no consensus has been reached, and differing variance measures continue to be employed by researchers and market participants. However, much of the literature focuses on the most liquid stocks, and we are unaware of recent papers that extend the analysis conducted by Andersen *et al.* (2001) to a broader range of stocks across various time periods and volatility measures. It is crucial to provide a thorough and precise econometric description of financial databases. Our study aims to fill this gap.

Since our research emphasizes delivering information and insights regarding volatility across a broader spectrum of stocks, it is essential to incorporate recent modeling approaches and the characteristics of a more intricate dynamic, as outlined previously. In this context, we introduce and examine additional estimators beyond RV.

This article is organized as follows: In Section 3.2 presents comprehensive descriptive statistics for our dataset. In Section 3.3, we conduct an analysis of the normality properties of returns along with their relationships to firm characteristics.

3.2
Data and Descriptive Analysis

The data utilized in this study are derived from the Trades-and-Quotes (TAQ) database, administered by the Center for Research in Security Prices (CRSP). The TAQ database is widely recognized for its comprehensive coverage and high precision. Our sample encompasses the period from 2003 to 2020, incorporating data from all 10,875 available tickers, which includes stock prices from both the New York Stock Exchange (NYSE) and NASDAQ. A significant advantage of the TAQ database is its provision of high-frequency data with a temporal resolution of up to one second, facilitating in-depth analyses of intraday stock market dynamics. It is crucial to acknowledge that not all stocks are consistently represented in every year within this time frame.

To ensure robust liquidity pruning and maintain the integrity of our analysis and results, we excluded from our sample any tickers that reported on fewer than 100 days and had liquidity (measured by the percentage of daily trades) below 2%. In other words, we excluded stocks that had trades in less than 2% of the time periods, which in our case consist of 5-second intervals.

For each firm, we compute six distinct realized measures as delineated in the Appendix. In particular, we focus on the Kernel realized variance (Kernel) of Barndorff-Nielsen *et al.* (2008), the Robust Two Time Scale Realized Variance (RT-SRV) introduced by Boudt & Zhang (2015), and the MedRV estimator proposed by Andersen *et al.* (2012), alongside intervals of 1, 5, and 15 minutes.

Table 3.1 provides descriptive statistics for all estimators utilized in our analysis. The left panel illustrates statistics related to daily stock returns, while the right panel focuses on various daily variance estimators. These statistics are reported over time to demonstrate their evolution throughout the sample period. A noteworthy observation is the adjustment introduced by robust variance estimators: the MedRV and RTSRV estimators are, on average, consistently lower than the RV measures, regardless of the sample size considered.

	Daily return statistics				Estimators daily averages					
Year	Count	Mean	Median	% Zero	Kernel	MedRV	RTSRV	RV 1m	RV 5m	RV 15m
Panel A: 2003–2008										
2003	5803	0.158	0	6.894	0.353	0.120	0.071	0.401	0.337	0.276
2004	5583	0.030	0	5.964	0.228	0.075	0.053	0.252	0.211	0.174
2005	5602	-0.039	0	6.327	0.186	0.059	0.044	0.208	0.174	0.146
2006	5434	0.025	0	6.413	0.151	0.048	0.039	0.166	0.139	0.119
2007	5465	-0.075	0	6.310	0.271	0.248	0.043	0.353	0.271	0.228
2008	5136	-0.275	0	7.000	0.659	0.207	0.137	0.769	0.636	0.543
Panel B: 2009–2015										
2009	4804	-0.015	0	7.731	0.551	0.165	0.111	0.626	0.528	0.459
2010	4643	-0.017	0	7.649	0.196	0.062	0.045	0.213	0.182	0.160
2011	4453	-0.122	0	7.820	0.200	0.057	0.051	0.218	0.188	0.165
2012	4286	0.011	0	8.621	0.175	0.045	0.036	0.193	0.165	0.143
2013	4226	0.042	0	8.958	0.142	0.037	0.032	0.154	0.130	0.111
2014	4240	-0.058	0	8.947	0.188	0.053	0.047	0.217	0.180	0.155
2015	4293	-0.088	0	9.183	0.259	0.074	0.058	0.301	0.251	0.217
Panel C: 2016–2020										
2016	4283	-0.043	0	9.638	0.323	0.100	0.072	0.378	0.287	0.223
2017	4239	-0.068	0	11.581	0.271	0.130	0.079	0.306	0.211	0.157
2018	4248	-0.185	0	11.182	0.353	0.152	0.099	0.388	0.273	0.203
2019	4250	-0.066	0	10.417	0.377	0.163	0.102	0.419	0.285	0.207
2020	4406	-0.149	0	9.663	0.616	0.353	0.199	0.629	0.442	0.339

The table displays the summary statistics for the *full sample* detailed by year for returns (on the left side) and for quadratic variance estimators (on the right side). The table reports the number of tickers in that year (Count), the unconditional mean and median of daily returns (Mean and Median), the average daily percentage of zero returns (% Zeros), and the daily averages of the Kernel, Robust Two Time Scale RV, Median RV, RV with 1-minute, 5-minute, and 15-minute intervals (Kernel, RTSRV, MedRV, RV 1m, RV 5m, and RV 15m) estimators.

Table 3.1: Daily return and realized measures descriptive statistics

To align our research with emerging trends highlighted in the current literature, which predominantly focuses on larger firms, we specifically identified the top 10% of the largest companies each year based on their monthly market capitalization. Subsequently, we present the same statistical analyses in Table 3.15 as shown in the preceding table. In addition, we highlight several stylized facts that have been well-documented in prior studies, including a lower average daily return coupled with a significantly diminished average percentage of days exhibiting zero returns. Furthermore, the variances exhibit a consistent trend, suggesting lower levels across all employed estimators.

	Daily return statistics				Estimators daily averages					
Year	Count	Mean	Median	% Zero	RV 1m	RV 5m	RV 15m	Kernel	RTSRV	MedRV
Panel A: 2003–2008										
2003	529	0.064	0	2.080	0.040	0.042	0.022	0.039	0.034	0.032
2004	512	0.020	0	1.404	0.026	0.027	0.015	0.025	0.023	0.022
2005	505	-0.027	0	1.284	0.023	0.023	0.014	0.023	0.021	0.020
2006	500	0.017	0	1.276	0.025	0.027	0.015	0.025	0.023	0.022
2007	497	-0.028	0	0.894	0.035	0.040	0.020	0.034	0.032	0.030
2008	470	-0.158	0	0.673	0.166	0.172	0.101	0.164	0.151	0.141
Panel B: 2009–2015										
2009	439	0.087	0	0.999	0.084	0.104	0.052	0.083	0.075	0.072
2010	420	0.044	0	1.092	0.039	0.041	0.019	0.032	0.029	0.028
2011	403	-0.022	0	0.808	0.036	0.040	0.024	0.035	0.033	0.032
2012	390	0.048	0	1.184	0.022	0.024	0.012	0.021	0.019	0.018
2013	384	0.072	0	1.119	0.018	0.015	0.010	0.017	0.016	0.015
2014	389	0.008	0	0.994	0.022	0.021	0.011	0.020	0.018	0.017
2015	392	-0.007	0	1.519	0.029	0.027	0.014	0.026	0.022	0.021
Panel B: 2016–2020										
2016	381	0.064	0	0.957	0.030	0.029	0.016	0.027	0.024	0.023
2017	376	0.024	0	1.122	0.019	0.019	0.008	0.017	0.014	0.013
2018	375	-0.072	0	0.596	0.034	0.035	0.018	0.030	0.026	0.024
2019	375	0.055	0	0.679	0.027	0.029	0.012	0.023	0.020	0.019
2020	373	0.021	0	0.357	0.096	0.123	0.050	0.082	0.074	0.070

The table displays the summary statistics for the *10% largest companies* (measured in monthly capitalization) detailed by year for returns (on the left side) and for quadratic variance estimators (on the right side). The table reports the number of tickers in that year (Count), the unconditional mean and median of daily returns (Mean and Median), the average daily percentage of zero returns (% Zeros), and the daily averages of the Kernel, Robust Two Time Scale RV, Median RV, RV with 1-minute, 5-minute, and 15-minute intervals (Kernel, RTSRV, MedRV, RV 1m, RV 5m, and RV 15m) estimators.

Table 3.2:
Daily return statistics and estimators averages - Top 10% Largest Firms

3.2.1 Jumps and Asset Liquidity

One relevant question to consider is which factors are causing the lack of normality observed in a large number of firms. One potential candidate is illiquidity. We can assess the lack of liquidity in several ways. The first measure could be the average percentage of zero returns in a day. This measure is significant as it directly highlights the buying and selling activity of the asset. However, it may be ineffective for distributions with few but well-separated non-zero returns. It is essential to distinguish between well-separated and non-separated returns, as many robust measures for jumps tend to overcorrect for well-separated returns, treating all instances as jumps.

This issue may stem from the Bi-power variation estimator, which will only yield non-zero values for distributions with at least two consecutive non-zero returns—a straightforward conclusion drawn from equation 3.13 in Appendix A. This is vital because, for instance, if we have 50% non-zero returns consistently interspersed with zero returns, the estimator may incorrectly interpret everything as jumps since the threshold established by the power will be nonexistent. Consequently, we will use the count of zero bi-power values daily as a pertinent measure of liquidity.

In tabel 3.3, we present the values for the complete sample on the left side and for the sample containing the top 10% of the largest companies on the right side. The table includes the 10th, 25th, 50th, 75th, and 90th percentiles, along with the mean for each variable of interest. The first block displays statistics for the number of days in the year when the bi-power RV estimator is zero. The second block details the number of jumps in the year for each distribution. The third block indicates the percentage of days when returns are zero, while the last block shows the percentage of days without any trades.

The table shows that the two measures are not entirely consistent. For the BP zeros statistic, we see that the potential for overcorrection is completely eliminated by the end, contrasting with the initial year. For the larger companies, we did not include this in the table because, with the exception of 8 tickers, all others are equal to zero. Regarding the percentage of zeros measure, it is also interesting to note a drastic decrease for both the sample of the largest companies and the total sample.

Furthermore, it is intrinsic to our research to add a variable to control for jumps directly. The strategy for jump identification we used is derived from Barndorff-Nielsen & Shephard (2006a), that also uses bi-power variation. We chose to represent this by the number of days with jumps in that distribution. In other words, a value of 100 would indicate that out of 252 trading days, 100 days experienced a jump event. The percentile statistics for these jump counts are also presented in table 3.3.

We will also use more standard measures of liquidity, such as the percentage of returns equal to zero and the percentage of intervals without trades on that day. These numbers can also be seen in table 3.3. In it, we can observe a convergence towards values that may indicate greater generalized liquidity.

Year	All Data						Top 10%					
	10%	25%	50%	75%	90%	Mean	10%	25%	50%	75%	90%	Mean
						Zeros BP						
2003	0.0	0.0	1.0	59.0	120.0	32.7	-	-	-	-	-	-
2011	0.0	0.0	0.0	7.0	102.0	20.6	-	-	-	-	-	-
2020	0.0	0.0	0.0	0.0	6.0	5.5	-	-	-	-	-	-
						Jumps						
2003	7.0	29.0	112.0	171.0	200.0	106.3	84.8	94.0	108.0	128.0	145.0	111.6
2011	11.0	68.0	103.0	164.0	203.0	108.8	69.8	77.0	87.0	96.0	106.2	86.5
2020	36.0	78.0	94.0	133.0	175.8	102.8	72.0	80.0	86.0	94.0	101.0	86.4
						% Zeros						
2003	40.4	61.1	83.8	95.7	98.3	75.9	17.5	23.9	33.6	41.5	48.4	33.7
2011	17.0	33.0	63.2	89.6	97.8	60.3	18.3	24.6	34.4	43.2	51.3	35.2
2020	5.5	13.1	34.5	64.7	86.2	40.2	1.4	2.3	3.6	5.6	8.1	4.4
						% No Trade						
2003	77.7	90.0	97.3	99.4	99.8.3	92.1	75.6	88.3	96.1	99.1	99.7	91.0
2011	21.8	75.5	91.9	98.4	99.6	83.6	49.2	71.6	88.7	96.9	99.3	81.1
2020	37.8	37.8	81.8	93.0	97.8	74.3	34.8	58.7	77.9	90.4	96.9	71.5

This table displays the values for the complete sample on the left side and the sample with the top 10% largest companies on the right side. The table shows the percentile statistics for the 10th, 25th, 50th, 75th, and 90th percentiles, as well as the mean for each variable of interest. The first block shows the statistics for the number of days in the year that the bi-power RV estimator is zero, the second block shows the number of jumps in the year for each distribution, the third block shows the percentage of days when returns are zero, and the last block shows the percentage of the day that did not have any trade.

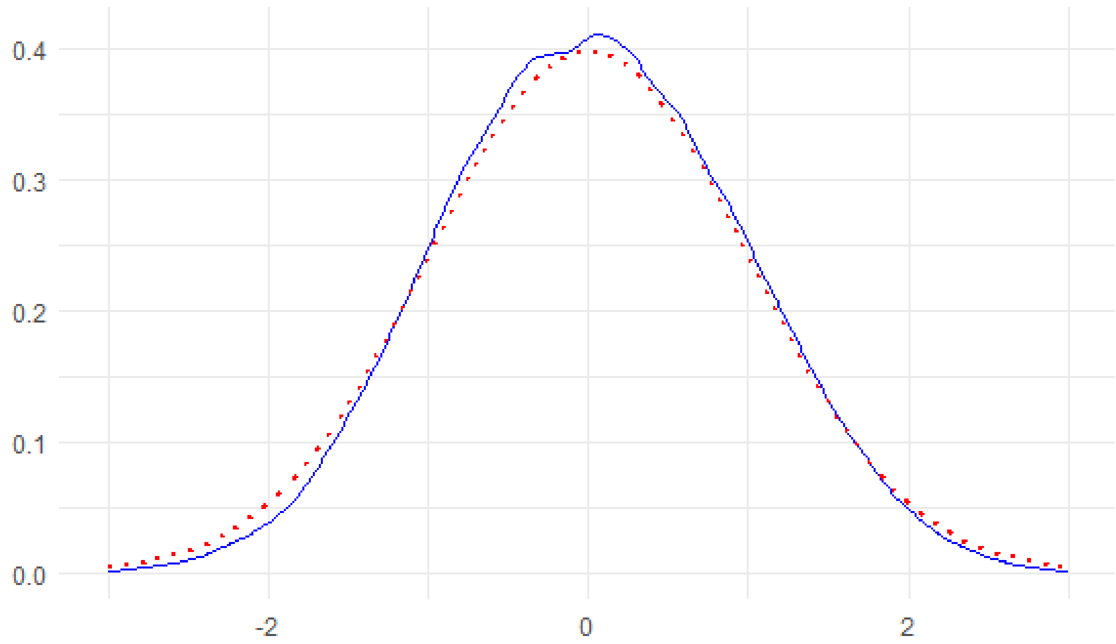
Table 3.3: Jumps and asset liquidity

3.3 Are daily standardized returns normal?

In this section, we present a comprehensive analysis of the distributional properties of standardized returns constructed using either the 5-minute realized variance measure or the realized kernel estimators. The Appendix reports results concerning the other realized measures. We also show the distributional properties of the realized measures.

We begin by presenting our findings on the distribution of standardized returns, focusing initially on the largest companies, with the aim of aligning our results with prior research conducted by Andersen *et al.* (2001). For this initial analysis, we selected the top 10% of the largest firms in 2020. Figure 3.1 shows the density distribution of their standardized returns, represented by the blue line. The red dotted line overlaid represents the theoretical standard normal distribution. We observe that the two distributions exhibit considerable similarity, indicating that the standardized returns of these large companies in 2020 closely approximate a Gaussian distribution. The figure also presents the moments of the distribution, calculated unconditionally with respect to the ticker. We find initial evidence that this distribution closely approximates a normal distribution.

Figure 3.1: Distribution of standardized returns for the top 10% largest firms in 2020.



Mean: 0.23   Standard deviation: 0.941   Skewness: 0.014   Kurtosis: 2.85

The figure shows the distribution of standardized daily returns using the 5-minute RV estimator for the top 10% largest firms in 2020 (blue line) and the standard normal distribution (red dashed line).

Table 3.4 reports the empirical distribution of the mean, standard deviation, asymmetry, and kurtosis across the cross-section of all firms. It displays the 10th, 25th, 50th, 75th, and 90th percentiles, as well as the mean and standard deviation of the first four moments of the unconditional distributions of standardized daily returns using either the 5-minute RV estimator or the realized kernel estimator. [1] From the numbers in Table 3.4, we observe a subtle deviation from the normal distribution. For instance, kurtosis varies from 2.2 at the 10th percentile to 3.6 at the 90th percentile—indicating heavier tails as we progress along the spectrum. Interestingly, skewness remains relatively aligned with the Gaussian benchmark,

[1] It is essential to recognize that a stock in the 90th percentile for kurtosis may not hold the same position regarding the mean as each statistical moment functions independently, with percentiles computed separately.

with both the median and mean hovering close to zero. Yet, despite this apparent balance, the standard deviation remains subdued, failing to reach one even at the 90th percentile.

	5-minute RV				Realized Kernel			
	Mean	St. dev.	Asymmetry	Kurtosis	Mean	St. dev.	Asymmetry	Kurtosis
0.10	−0.110	0.534	−0.213	2.239	−0.108	0.571	−0.364	2.676
0.25	−0.049	0.697	−0.091	2.656	−0.050	0.679	−0.147	2.943
0.50	0.000	0.825	0.024	2.952	0.003	0.795	0.023	3.318
0.75	0.051	0.916	0.143	3.270	0.053	0.902	0.191	4.037
0.90	0.105	0.979	0.266	3.672	0.106	1.001	0.398	5.829
Mean	0.000	0.770	0.027	3.152	0.001	0.796	0.018	4.108
St. dev.	0.104	0.234	0.307	7.090	0.103	0.459	0.543	4.693

This table displays the percentile statistics, specifically the 10th, 25th, 50th, 75th, and 90th percentiles, as well as the mean and standard deviation of the first four moments of the unconditional distributions of standardized daily returns using the 5-minute RV or the realized-kernel estimators.

Table 3.4: Empirical distribution of descriptive statistics

To provide a more robust analysis of the distributional properties of standardized returns, we consider several tests for normality. The first one is the Jarque-Bera (JB) (Bera & Jarque (1980)) test, arguably the most commonly used, and will serve as the primary tool in our analysis. However, as noted in Bontemps & Meddahi (2005), the JB test can exhibit over-rejection in finite samples. We will also display three additional tests as robustness measures against potential distortions from the Jarque-Bera test. We consider the Cramer-von Mises, Kolmogorov-Smirnov, and Anderson-Darling tests.

In Table 3.5, we report the values for the 10th, 25th, 50th, 75th, and 90th percentiles, along with the mean and standard deviation of the p-values for the null hypothesis of normality derived from the Jarque-Bera (JB), Cramer-von Mises (CvM), Anderson-Darling (AD), and Kolmogorov-Smirnov (KS) tests for the standardized returns (Anderson & Darling (1952), Smirnov (1948), Sim & Yap (2011)). We find that approximately 25% of the data demonstrates significant deviations from normality at the 5% significance level based on the results from the first three tests. The results are robust to the choice of the realized measure used.

In the Appendix, we present the results using the jump-robust realized measures. The general findings remain similar, regardless of the choice of realized measure.

Furthermore, the same analysis was conducted for the logarithm of the standard deviation of all estimators cited in this paper, another value that, theoretically, should be normally distributed. The tables related to its moments are provided in Appendix B.

	5-minute RV				Realized Kernel			
	JB	CvM	KS	AD	JB	CvM	KS	ADb
0.10	0.010	0.001	0.001	0.158	0.000	0.000	0.000	0.076
0.25	0.099	0.050	0.042	0.510	0.001	0.017	0.023	0.356
0.50	0.362	0.265	0.249	0.791	0.182	0.186	0.190	0.711
0.75	0.657	0.586	0.563	0.934	0.561	0.516	0.513	0.909
0.90	0.857	0.823	0.798	0.982	0.812	0.791	0.771	0.976

This table displays the percentile statistics, specifically the 10th, 25th, 50th, 75th, and 90th percentiles, as well as the mean and standard deviation of the p-values for the null hypothesis of normality from the Jarque-Bera (JB), Cramer-von Mises (CvM), Anderson-Darling (AD), and Kolmogorov-Smirnov (KS) tests for the standardized, using RV 5 minutes, daily returns' distributions.

Table 3.5: p-values for normality tests

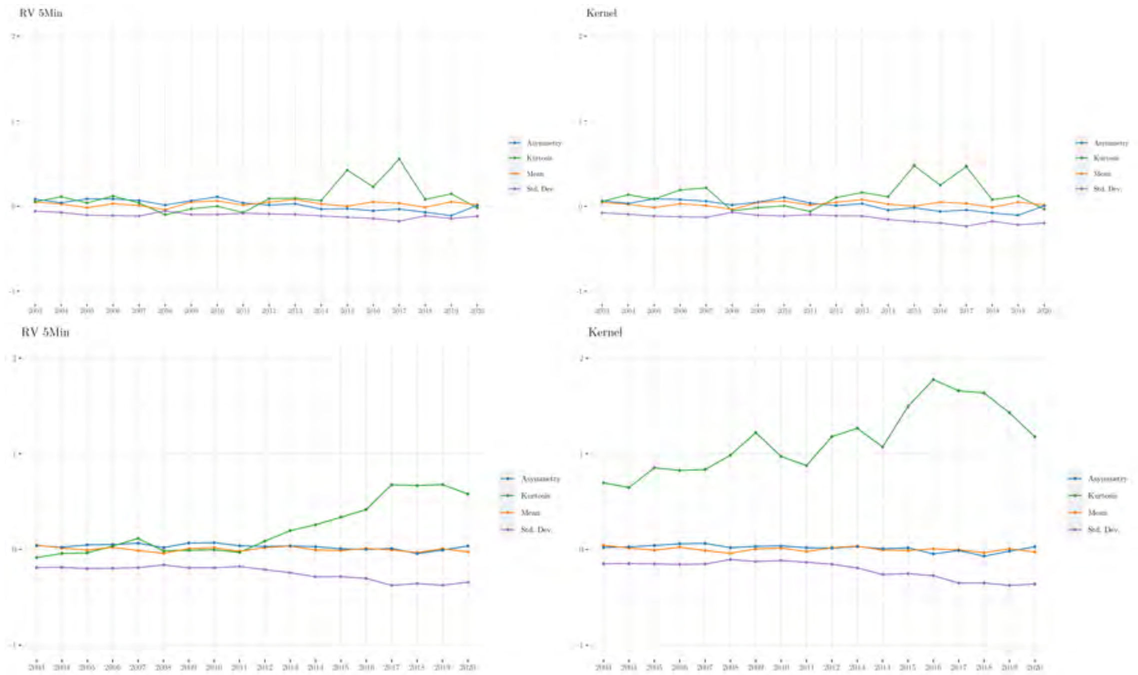
3.3.1 Normality through time

An intriguing aspect to examine is how the concept of normality may have evolved over the years. As markets expanded, with more participants and companies entering the stock exchange, the dynamics may have shifted, potentially leading standardized returns to converge toward a particular distribution.

Figure 3.10 presents the average moments over time, for the 5-minute and realized kernel estimators. The measures are displayed as deviations from their Gaussian counterparts. For instance, the graphs depict kurtosis minus 3. In the upper panel of Figure 3.10, which shows the top 10% largest companies, we observe relative stability over the years, with values fluctuating around zero for both estimators. An exception occurs between 2015 and 2017, where kurtosis shows a notable spike. These patterns are somewhat expected, as we assume that the markets for the largest companies are well-developed and liquid enough.

In contrast, the lower panel of Figure 3.10 illustrates a more pronounced divergence from the Gaussian benchmarks, particularly in terms of standard deviation and kurtosis for both estimators. Notably, the standard deviation exhibits a downward trajectory over the years, whereas kurtosis reveals a general upward trend, although this trend appears to have reversed in more recent years.

Figure 3.2: Evolution of average moments over time for various estimators.



This figure shows the unconditional mean of the first four moments (mean, standard deviation, asymmetry, and kurtosis) of *the 10% largest companies* distributions for each year for the Kernel and RV 5-minute estimators. The top panel refers to the top 10% firms, and the lower panel to all firms.

3.3.2
The sources of (non) normality

To explore which firm characteristics relate to a lack of normality, we employ two strategies. First, we conduct a series of panel regressions where the dependent variable is the p-value of the JB test, the coefficient of asymmetry, or the coefficient of kurtosis of the standardized returns. As explanatory variables, we include time fixed effects, fundamental factors generally linked to the distribution of premiums Fama & French (2015), and liquidity measures.

3.3.2.1
Linear Model

We estimate versions of the following linear panel model:

$$Y_{it} = \gamma_t + \beta' X_{it} + \varepsilon_{it}, \tag{3.1}$$

where Y_{it} may represent the p-value of the JB test, the coefficient of asymmetry, or the coefficient of kurtosis for firm i, averaged over year t; γ_t is a year fixed effect; X_{it} is a vector containing explanatory variables; and ε_{it} is the zero-mean error term. We consider the following regressors in X_{it}: percentage of days with bipower variace equals to zero (Zeros BP); percentage of days with jumps; percentage of no trade; size (measures in market capitalization); book-to-market ration (B/M); price-to-earnings ratio (P/E), dividend yield, and finally price/cash flow.

The variation in the R-squared was investigated using the forward selection algorithm, aiming to maximize it. From these results (shown in Appendix D), we can observe the recurrence of some anomalies more than others. Consequently, we selected characteristics that emerged as particularly relevant for this category. Note that this initial analysis served only as a basis for exploring the following correlations.

In Tables 3.6 and 3.7, we see the results of the regressions with the chosen explanatory variables. In all regressions, time fixed effects were added and each variable was demeaned.

	Dependent variable:					
	RV 5Min			Kernel 5s		
	JB P-Val	Asymmetry	Kurtosis	JB P-Val	Asymmetry	Kurtosis
% Trade	−0.05560**	0.05588***	−0.53937***	0.11605***	0.07385***	−0.76724***
	(0.02187)	(0.01389)	(0.10386)	(0.02349)	(0.01643)	(0.10247)
Jumps	−0.00028*	0.00028***	0.00083	−0.00055***	0.00030**	−0.00035
	(0.00016)	(0.00010)	(0.00076)	(0.00017)	(0.00012)	(0.00075)
Size	−0.02094	0.11989***	0.32345	−0.17442***	0.15813***	0.57179**
	(0.05826)	(0.03701)	(0.27666)	(0.06257)	(0.04378)	(0.27296)
Constant	0.46306***	0.07180***	2.82997***	0.46116***	0.06615***	3.05669***
	(0.01243)	(0.00790)	(0.05903)	(0.01335)	(0.00934)	(0.05824)
Observations	7,569	7,569	7,569	7,569	7,569	7,569
R^2	0.00961	0.06942	0.01428	0.01366	0.07930	0.02016
Adjusted R^2	0.00699	0.06695	0.01167	0.01104	0.07686	0.01756
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 3.6: RV 5Min vs Kernel - Top 10%

Comparing the two tables, size does not seem offset the liquidity impact in all the variables. On the contrary, for the RV p-value, size seems to play a secondary part.

Additionally, the R^2 of the regressions with the p-value and kurtosis as the dependent variable is higher compared to its counterpart with only a subset of the sample. This might again indicate the existence of a more elaborate explanatory structure when considering the entire universe of stocks.

	Dependent variable:					
	RV 5Min			Kernel 5s		
	JB P-Val	Asymmetry	Kurtosis	JB P-Val	Asymmetry	Kurtosis
% Trade	0.03239***	−0.04514***	−0.99718***	0.46908***	−0.05331***	−3.83210***
	(0.00831)	(0.00540)	(0.02438)	(0.00830)	(0.01378)	(0.12649)
Jumps	−0.00068***	0.00023***	0.00014	0.00023***	0.00015***	−0.00689***
	(0.00003)	(0.00002)	(0.00009)	(0.00003)	(0.00005)	(0.00048)
Size	−0.04238	0.17307***	1.39763***	−0.82158***	0.23459***	6.10766***
	(0.05344)	(0.03468)	(0.15668)	(0.05335)	(0.08860)	(0.81304)
Zeros BP	−0.58813***	−0.04811***	−1.27616***	−0.36533***	−0.10633***	3.12651***
	(0.01160)	(0.00753)	(0.03401)	(0.01158)	(0.01923)	(0.17650)
Constant	0.39863***	0.04573***	2.89144***	0.30048***	0.03830***	3.77924***
	(0.00448)	(0.00291)	(0.01313)	(0.00447)	(0.00743)	(0.06816)
Observations	68,861	68,861	68,861	68,861	68,861	68,861
R^2	0.07568	0.02605	0.09657	0.14957	0.00810	0.06587
Adjusted R^2	0.07540	0.02575	0.09630	0.14931	0.00779	0.06558
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 3.7: RV 5Min vs Kernel - Full Sample

3.3.2.2
Variable Correlation Through Time

Given that our analysis spans 20 years, it is also important to understand if these correlations between normality and explanatory variables change over the years. In other words, whether the variables related to liquidity change their correlation with aspects of normality over time. The regression in equation 3.1 is just like the one in equation 3.2, but now we let the variables' coefficients vary over time.

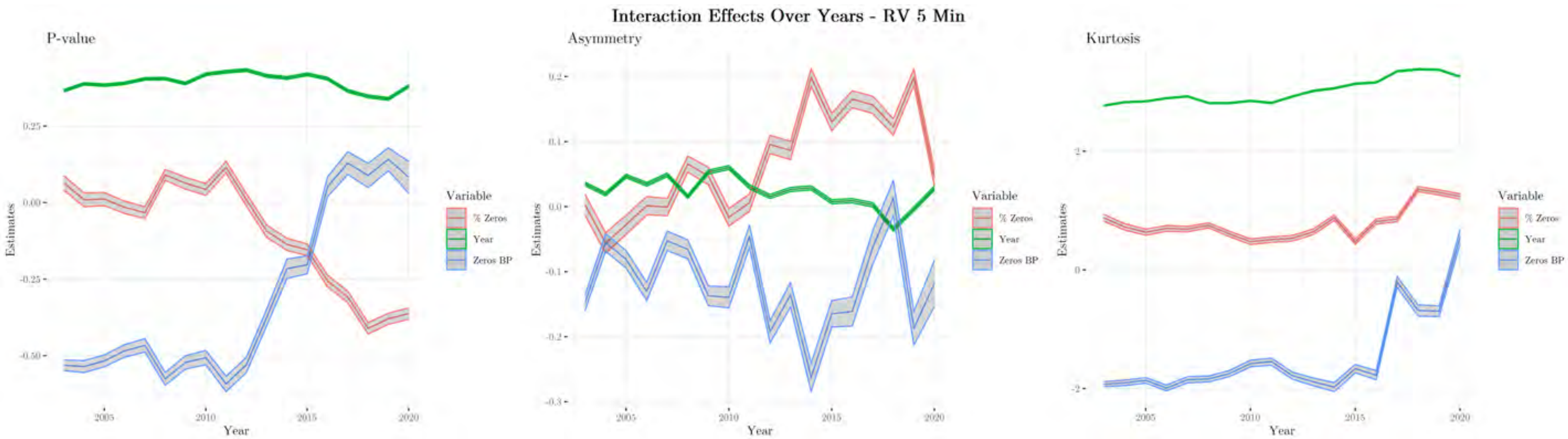
$$y_{it}^{M} = \sum_{t} \beta_{0t} * \mathbf{I}_t + \sum_{t} \sum_{c \in C} (\beta_{tc}^{M} x_{it,c} * \mathbf{I}_t) + \epsilon_{it}. \tag{3.2}$$

M is related to the dependent variables, including the p-value for the JB test, Asymmetry, and Kurtosis; β_{0t} denotes the time fixed effects for 2003 to 2020, $\mathbf{I}_t$ is the indicator variable for year, ε_{it} is the gaussian error and c pertains to the set C of chosen independent variables.

In Figure 3.3, we can see the evolution of the correlations for the variables of daily zero percentage mean and the annual percentage at which the bipower variation estimator returns zero. At the beginning of the sample, zeros BP seems to dominate the daily zero percentage mean in relation to the p-value of the Jarque-Bera normality test. In 2015, this relationship reverses. This can be explained by the increasing correlation of the daily zero percentage mean with skewness, thus moving away from the Gaussian counterpart. In the graph related to kurtosis, we can see a trend of the correlations of Zeros BP tending towards zero, meaning that while it previously had a negative correlation, it is now increasingly less correlated with this aspect.

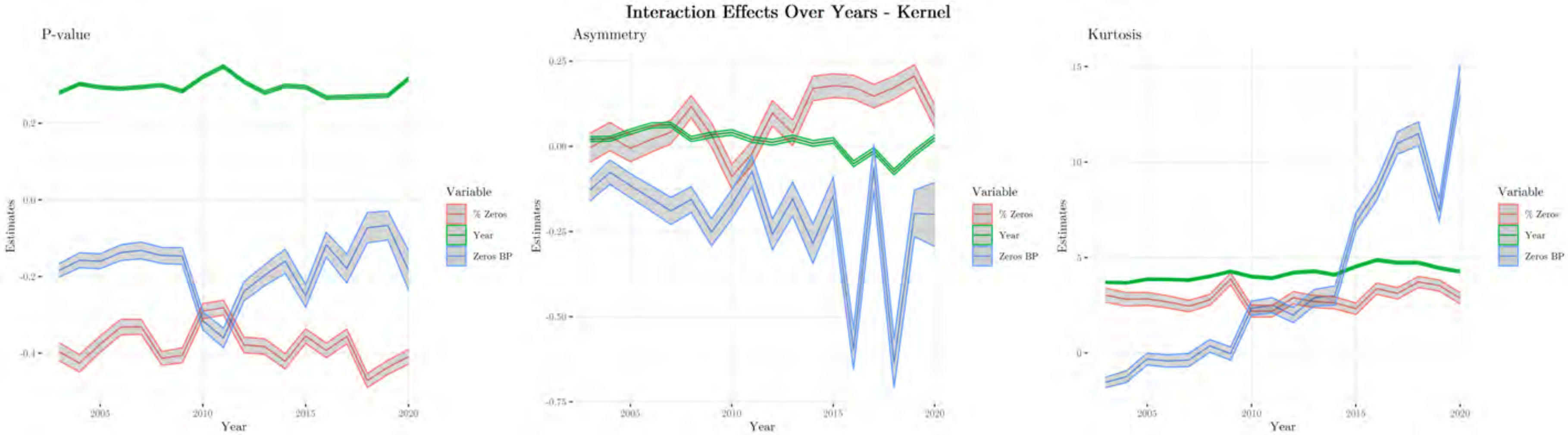
In Figure 3.4, we can observe the same coefficients, now applied to the distributions using the kernel estimator. In the first column, related to the p-value of the Jarque-Bera test, we see both liquidity measures jointly contributing to a lower probability of normality. Another trend that the graphs seem to present is a distancing of kurtosis from its Gaussian counterpart regarding the correlation with Zeros BP: in this variable, it is possible to see an increasing trend in the correlation of Zeros BP with higher kurtosis.

Figure 3.3: Evolution of variable correlation - RV 5 Min.



This figure shows the the coefficients from the interaction between years and the variables % *Zeros* and *Zeros BP* besides the year fixed effects for the regression with P-value (first column), Asymmetry (second column) and Kurtosis (third column) as dependent value.

Figure 3.4: Evolution of of variable correlation - Kernel.



This figure shows the the coefficients from the interaction between years and the variables *% Zeros* and *Zeros BP* besides the year fixed effects for the regression with P-value (first column), Asymmetry (second column) and Kurtosis (third column) as dependent value.

3.3.2.3
Tree Model

In the previous section, it was possible to observe some correlations between the characteristics of stocks and companies and the behavior of standardized return distributions using the 5-minute RV estimator and the kernel estimator. With this in mind, one can wonder if some subsets of the sample might be better fitted for one estimator than for another. Building on this question, we constructed several decision and regression trees to try to gain some insights and details on the topic.

Decision trees are supervised learning algorithms used for classification and regression tasks. They work by recursively splitting the data into subsets based on the values of input features, aiming to improve the homogeneity of the target variable within each subset. The structure of a decision tree resembles a flowchart, with internal nodes representing tests on features, branches representing the outcomes of these tests, and leaf nodes representing the target variable or class label.

The typical steps of a decision tree algorithm are as follows:

1. Select the best feature f to split the data D based on a criterion. In this case, Gini Impurity.
2. Split the data D into subsets $D_1, D_2, \ldots, D_n$ based on the selected feature.
3. Recursively apply this process to each subset until a stopping condition is met. In this case, the Complexity Parameter.

The Complexity Parameter algorithm stops splitting when the cost complexity of adding another split is not worthwhile. The cost complexity α is defined as:

$$R_\alpha(T) = R(T) + \alpha|T|$$

where $R(T)$ is the resubstitution error of the tree T, α is the complexity parameter, and $|T|$ is the number of terminal nodes in the tree. The tree grows until any further splitting does not decrease the overall error by a factor of α.

The Gini Impurity for a feature f is given by:

$$G(D) = 1 - \sum_{k=1}^{K} p_k^2$$

where $G(D)$ is the Gini impurity of the dataset D, K is the number of classes, and p_k is the probability of selecting an element of class k in the dataset. When incorporating a custom cost matrix C, the Gini impurity can be adjusted as:

$$G_C(D) = \sum_{i=1}^{K}\sum_{j=1}^{K} p_i p_j C_{ij}$$

where C_{ij} represents the cost of classifying an element of class i as class j.

Regression trees, on the other hand, used for predicting continuous values, operate similarly to decision trees but use criteria like Mean Squared Error (MSE) to determine the best splits. The goal is to minimize the variance within each subset. The MSE for a split is calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2$$

where y_i are the observed values and $\hat{y}$ is the predicted value for subset i.

With the aim of better visualization, we decided to transform the continuous p-value variable into classes. We created four classes: we call it 0 when the p-value is less than 1%, indicating that the distribution is likely not normal. We call it 1 when the p-value is between 1% and 5%, meaning it may not be normal, but it's a grey area. We call it 2 when the p-value is between 5% and 10%, meaning it may be normal, but it's also a grey area. Finally, we call it 3 for the class where the distributions have a p-value greater than 10%, indicating that they are likely Gaussian.

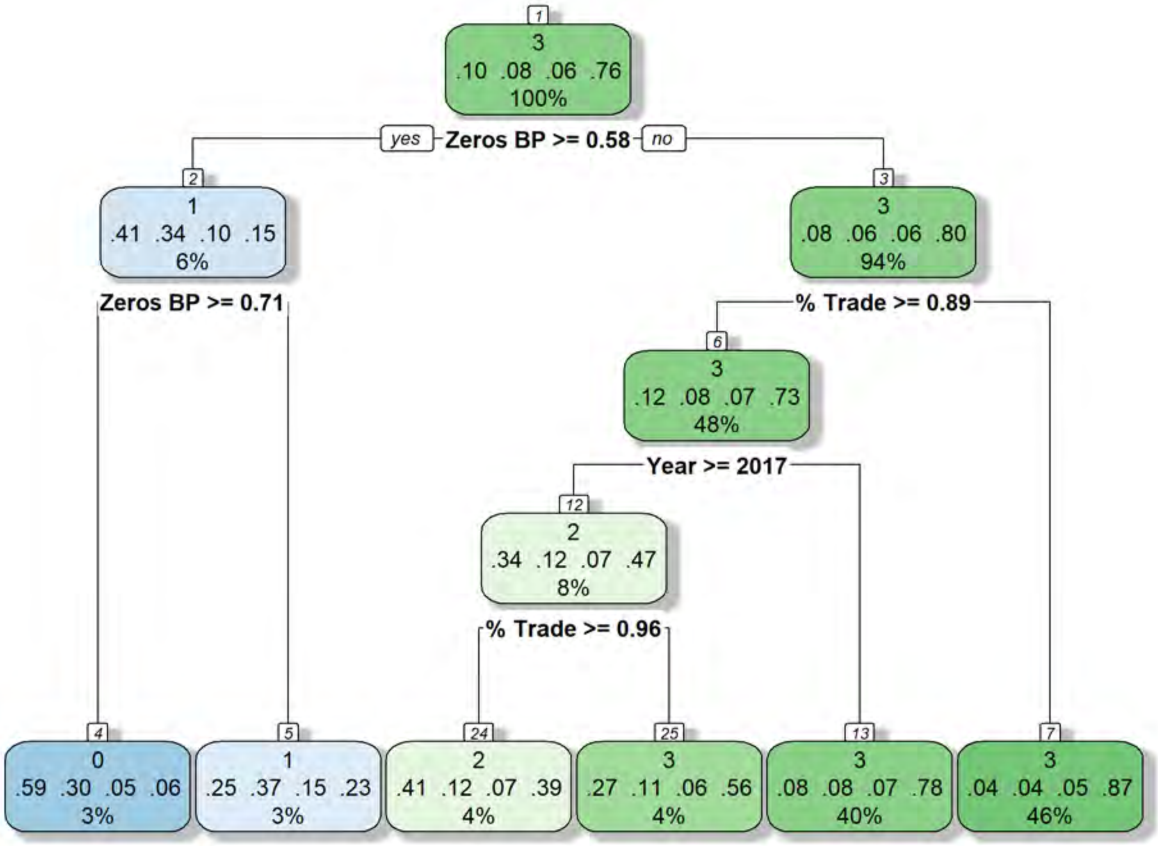
It is possible to observe that classifying something as 0 when it belongs to class 1 is much "better" than classifying it as class 2, for example. Additionally, in our case, we opted to assign a higher penalty for cases that were misclassified in relation to being normal. That is, if the distribution according to the Jarque-Bera normality reject the null hypothesis of normality but was classified in our decision tree as class 3 (probrably normal), it is worse than the opposite. Thus, an ad hoc cost matrix was constructed:

$$\mathbf{C} = \begin{pmatrix} 0 & 2 & 3 & 10 \\ 2 & 0 & 1 & 8 \\ 3 & 1 & 0 & 6 \\ 10 & 8 & 6 & 0 \end{pmatrix} \tag{3.3}$$

Here, the rows refer to the chosen class, and the columns to the actual class. For instance, the element in row 1 and column 1 represents the penalty for choosing class 0 when it is actually class 0, which is no penalty at all. Conversely, the element in row 1 and column 2 represents the penalty for choosing class 0 when it is actually class 1. As described in the previous paragraph, classifying something as 3 when it is class 2 costs 6 (as shown in the element of row 4 and column 3), but classifying something as 0 when it is actually class 1 costs 2.

In Figures 4 and 5, we present the decision trees related to the p-values of the Jarque-Bera normality test using the 5-minute RV estimator (Figure 4) and the kernel estimator (Figure 5). In these figures, we can observe a pattern of decision splits that seems to repeat. The first most important factor is the percentage of days in that distribution where the estimated value by bi-power Variation is equal to zero. Following this, we see a split concerning the percentage of the day with no trades at all. This appears to be closely intertwined with the liquidity of the asset.

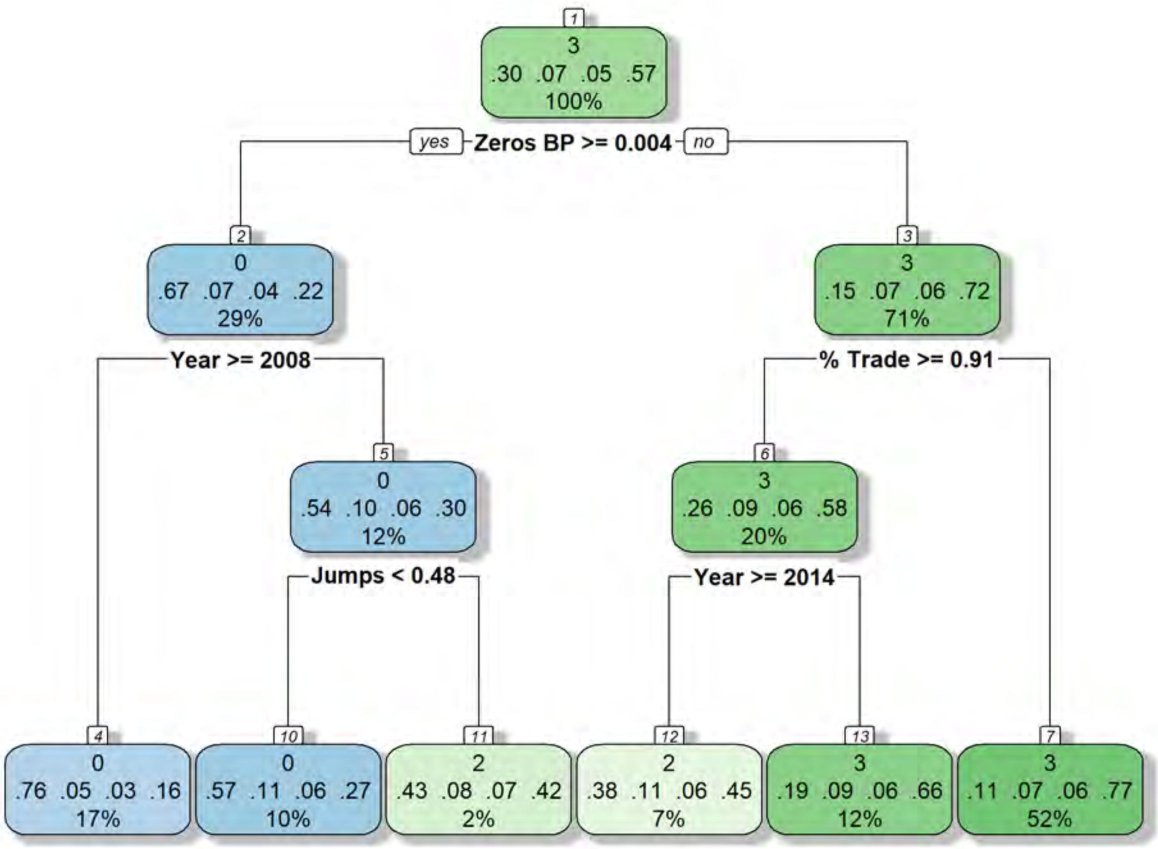
Figure 3.5: Decision Tree - RV 5 minutes.



The figure displays the results of the *decision tree* where the dependent variable is the p-value of the Jarque-Bera normality test for the standardized returns' distributions using the *5 Minutes RV*. The colored squares represent the nodes, and the smaller square above each node indicates the corresponding node number. Inside each colored square, the first line shows the class assigned to that node. The second line displays the percentage distribution of each class (0, 1, 2, 3). The third line indicates the percentage of the total sample that fits into that node. The splitting criterion for each node (excluding the final nodes, which are the nodes in the last row) is shown below the node. Observations that meet the criterion go to the left, while those that do not go to the right. The colors also have their significance: blue represents nodes assigned to class 0 and 1, and green otherwise.

In both figures, it is possible to clearly observe the zones where we can be more confident about the normality of the distributions. Additionally, the liquidity threshold for these zones appears to be less stringent when analyzing the 5-minute RV estimator. For this estimator, having less than approximately 60% of the days with zero BP is sufficient to ensure a more securely normal distribution. On the other hand, for the kernel estimator, a threshold above 0.4% significantly influences the probability that the distribution we are dealing with is statistically not normal.

Figure 3.6: Decision Tree - Kernel.



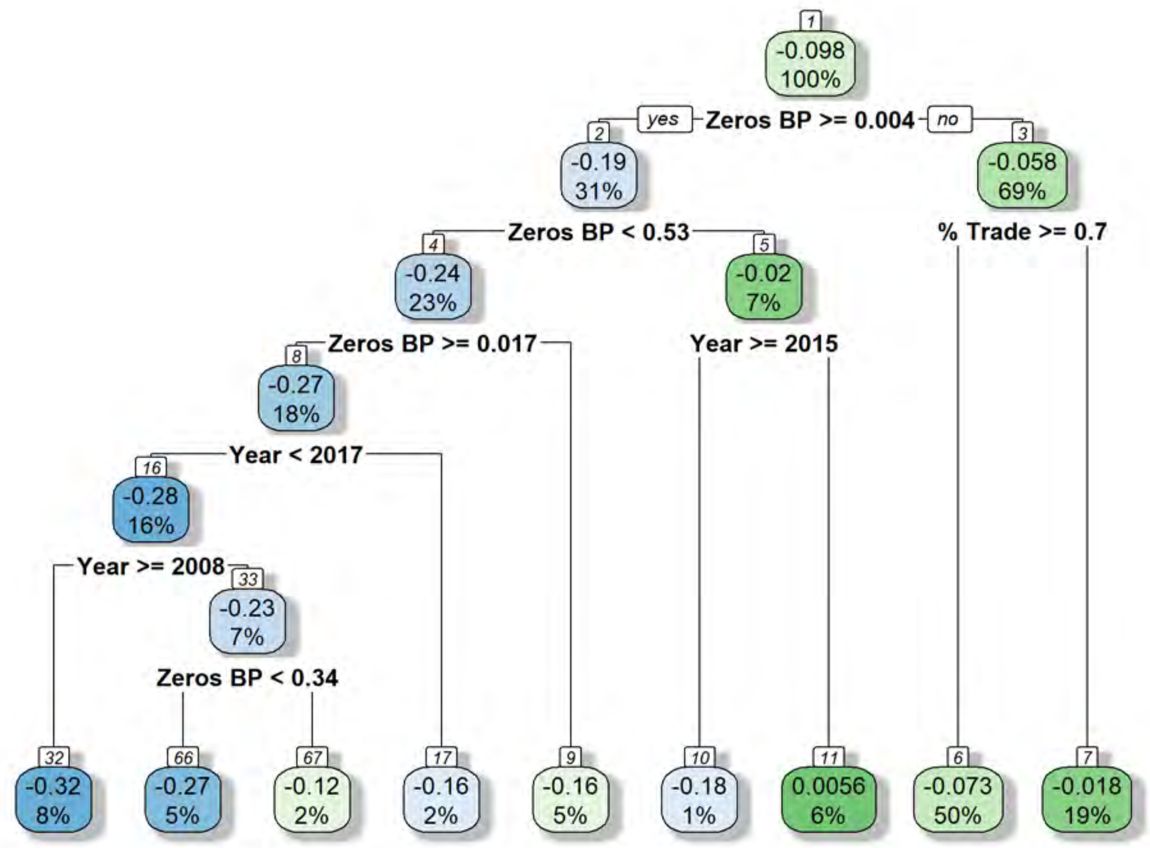
The figure displays the results of the *decision tree* where the dependent variable is the p-value of the Jarque-Bera normality test for the standardized returns' distributions using the *Kernel estimator*. The colored squares represent the nodes, and the smaller square above each node indicates the corresponding node number. Inside each colored square, the first line shows the class assigned to that node. The second line displays the percentage distribution of each class (0, 1, 2, 3). The third line indicates the percentage of the total sample that fits into that node. The splitting criterion for each node (excluding the final nodes, which are the nodes in the last row) is shown below the node. Observations that meet the criterion go to the left, while those that do not go to the right. The colors also have their significance: blue represents nodes assigned to class 0 and 1, and green otherwise.

Next, we will analyze the difference between the p-values of the distributions using the kernel estimator and the 5-minute RV estimator. More specifically,

$$\text{Diff} = \text{p-value}_{\text{Kernel}} - \text{p-value}_{\text{RV 5m}} \tag{3.4}$$

This analysis can be interesting to observe regions where the estimators agree and regions where one estimator might 'perform better' towards normality than the other. In Figure 6, we present the regression tree that always returns the mean difference of the p-values for each node. In this image, we can see that the nodes consistently indicate either a certain degree of similarity between the p-values (seen on the right side of the tree) or a significant negative bias (left side). Again, the decision criteria are based on the same variables as the previous trees: Zeros BP and % Trade.

Figure 3.7: Regression Tree - P-value differences.



The figure displays the results of the *regression tree* where the dependent variable is the *differences of the p-values* of the Jarque-Bera normality test for the standardized returns' distributions using the 5 Minutes RV and the Kernel estimator. The colored squares represent the nodes, and the smaller square above each node indicates the corresponding node number. Inside each colored square, the first line shows the mean value for that node. The second line indicates the percentage of the total sample that fits into that node. The splitting criterion for each node (excluding the final nodes, which are the nodes in the last row) is shown below the node. Observations that meet the criterion go to the left, while those that do not go to the right. The colors also have their significance: blue represents nodes with negative values and green represents nodes with positive values. The color strengthens with the number's absolute value.

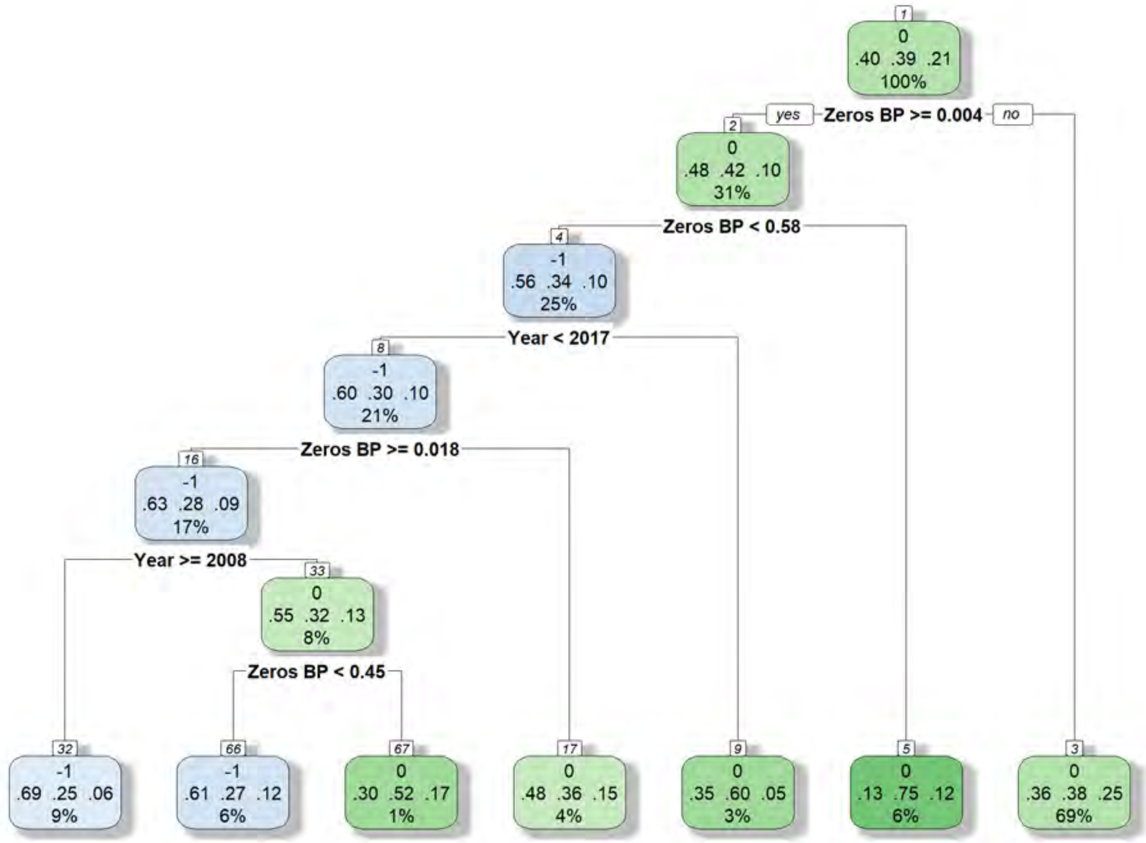
With the aim of better visualizing these two zones, we constructed three classes: -1 if the difference is less than -0.1, 0 if it is between -0.1 and 0.1, and 1 if it is greater than 0.1. Note that -1 represents those distributions with higher p-values for standardized returns using the 5-minute RV estimator, and 1 for those standardized with the kernel estimator. We also constructed an ad hoc cost matrix:

$$\mathbf{C}_2 = \begin{pmatrix} 0 & 1 & 3 \\ 1 & 0 & 1 \\ 3 & 1 & 0 \end{pmatrix} \tag{3.5}$$

This matrix was constructed in such a way that assigning class 0 costs less. Thus, the other classes would be used more sparingly.

In Figure 7, we again observe the decision criteria tied to the variable Zeros BP. An interesting point is the absence of terminal nodes predicting class 1. Additionally, the first split goes to the right when more than 4% of the days have the bi-power Variation estimator returning zero, but the second split goes the other way, influencing the probability of a higher p-value when standardized with the 5-minute RV when less than 58% of the days have Zeros BP. This might indicate counter-evidence that only the most liquid stocks can return statistically normal distributions.

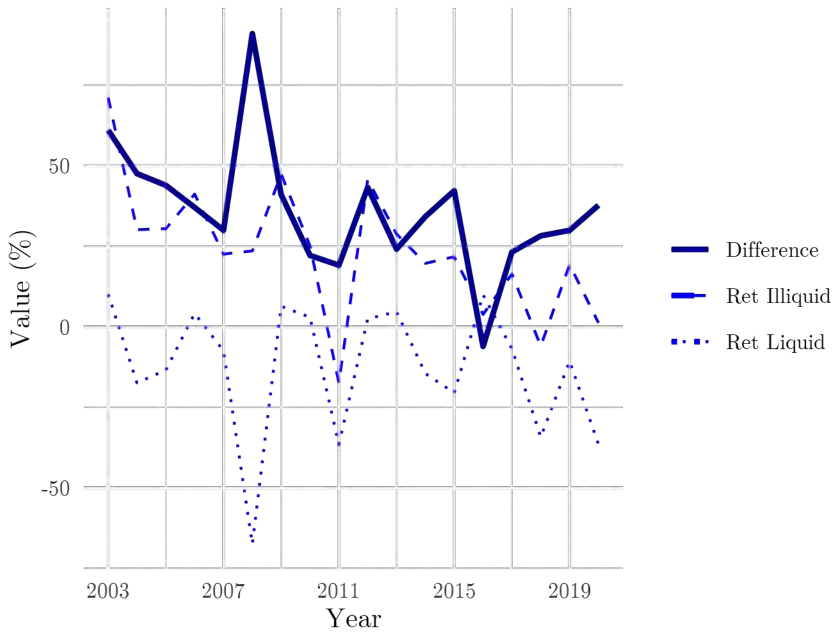
Figure 3.8: Decision Tree - P-value differences.



The figure displays the results of the *decision tree* where the dependent variable is the differences of the p-values of the Jarque-Bera normality test for the standardized returns' distributions using the 5 Minutes RV and the Kernel estimator. The colored squares represent the nodes, and the smaller square above each node indicates the corresponding node number. Inside each colored square, the first line shows the class assigned to that node. The second line displays the percentage distribution of each class (-1, 0, 1). The third line indicates the percentage of the total sample that fits into that node. The splitting criterion for each node (excluding the final nodes, which are the nodes in the last row) is shown below the node. Observations that meet the criterion go to the left, while those that do not go to the right. The colors also have their significance: blue represents nodes assigned to class -1, and green represents nodes assigned to class 0.

The findings in this section are particularly interesting, especially regarding the results from the first decision tree in Figure 6. With these splits in mind, we selected the stocks that would fall into nodes 2 and 7 and calculated the annual return of an equally weighted portfolio[2]. In Figure 10 below, we can consistently observe that the portfolio return for stocks with "non-Gaussian" standardized returns is higher than that of stocks with so-called Gaussian standardized returns.

Figure 3.9: Difference in returns - Liquid vs Illiquid.



The chart presents the percentage returns of equally weighted portfolios. The dotted line represents stocks in Node 7 in Figure 6, while the dashed line corresponds to stocks in Node 2 of the same figure. Lastly, the solid bold line illustrates the difference in returns between these two portfolios.

3.4 Conclusion

The findings from Andersen *et al.* (2001) that standardized returns follow a normal distribution process mainly hold for stocks over the years. Specifically, when considering the largest stocks, a Gaussian relationship is clearly observable in standardized returns. However, when expanding the samples to include all stocks, this relationship deteriorates. We can especially observe that less liquid stocks tend to exhibit more non-normal behavior. This deviation from normality is most noticeable in an excess of kurtosis, or "heavy tails." A deeper understanding of the impact of these heavy tails on volatility forecasting and portfolio optimization remains an area for future research, as well as whether all stocks should be represented by the same data-generating process.

[2]For the portfolio corresponding to the final node 7, due to the large number of stocks considered, a random sample of 100 stocks per year was used.

Another interesting point is that more robust measures of realized volatility show even less normal standardized returns than a simpler measure like the 5-minute RV. This is intriguing since, in theory, these estimators should be more robust to complex dynamics. This result is particularly important for less liquid stocks. It would also be valuable to further investigate why this occurs and whether this suggests that the 5-minute RV should primarily be used for less liquid stocks, where the difference in normality between estimators is more significant.

3.A
Theoretical Assumptions

The current section formalizes the theoretical assumptions regarding the univariate stock prices. The exploration of continuous time processes began with the assumption of a diffusion model for the logarithmic prices:

$$dp(t) = \mu(t)\,dt + \sigma(t)\,dW(t), \quad 0 \leq t \leq T, \tag{3.6}$$

where $p(t)$ is the logarithmic price, $\mu(t)$ is a predictable drift component, $\sigma(t)$ is the instantaneous volatility, and $W(t)$ is a Wiener process. An additional assumption is that the Wiener process $W(t)$ and the process governing the time variation in $\sigma(t)$ are independent, implying the absence of a leverage effect. This means that shocks to the price process do not systematically affect future volatility. The time index T is typically normalized to represent one trading day.

For the model presented in Equation (3.6), the use of intraday returns to obtain an estimate of the process's quadratic variation is common in the literature. The realized variance (RV) estimator is defined as the squared sum of discretely sampled returns.

Formally, we first define the quadratic variation of the cumulative return, $r(t) \equiv p(t) - p(0)$,

$$[r, r]_t = \int_0^t \sigma^2(s)\,ds, \tag{3.7}$$

which is a natural representation of the total *ex post* variability of the process in (3.6). In this initial model, the quadratic variation equals the integrated variance (IV), $IV \equiv \int_0^t \sigma^2(s)\,ds$. Subsequently, the RV estimator is obtained from the discretely sampled Δ-period return, $r_{t,\Delta} \equiv p(t) - p(t - \Delta)$, as

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^{1/\Delta} r^2_{t+j,\Delta}, \tag{3.8}$$

where $1/\Delta$ defines the sampling frequency of the intraday returns. The realized variance estimator converges to the quadratic variation between trading days $t + 1$ and t for $\Delta \to 0$

$$RV_{t+1}(\Delta) \to [r, r]_{t+1} - [r, r]_t = \int_t^{t+1} \sigma^2(s)ds. \tag{3.9}$$

This result implies that by increasing the sampling frequency, we can treat the quadratic variation of the process defined in (3.6) as observable. Under the conditions of no microstructure noise and as $\Delta \to 0$, the realized variance is a consistent estimator of the integrated variance.

In this simple case, the *ex post* distribution of daily returns is conditionally Gaussian:

$$r(t)|\mathcal{F}_t \sim \mathrm{N}\left(\int_0^t \mu(s)ds, \int_0^t \sigma^2(s)ds\right), \tag{3.10}$$

where $\mathcal{F}_t$ represents the σ-algebra generated by the sample paths of $\mu(t)$ and $\sigma(t)$, $\mathcal{F}_t \equiv \{\mu(\tau), \sigma(\tau)\}_{\tau=t-1}^{\tau=t}$. Intuitively, $\mathcal{F}_t$ encompasses all information available up to time t regarding the drift and volatility processes.

The model in Equation (3.6) can be extended to incorporate jumps, which represent occasional discontinuities in the price paths:

$$dp(t) = \mu(t)\,dt + \sigma(t)\,dW(t) + \kappa(t)\,dq(t), \quad 0 \le t \le T, \tag{3.11}$$

where the added terms are $dq(t)$, a counting process with $dq(t) = 1$ if a jump occurs at time t and $dq(t) = 0$ otherwise, with jump intensity $\lambda(t)$, which may be time-varying, and $\kappa(t)$ is the corresponding jump size. The inclusion of jumps is motivated by the need to capture unexpected economic events that can cause abrupt movements in stock prices, such as earnings announcements or macroeconomic shocks.

In the presence of jumps, the quadratic variation includes a term associated with the cumulative sum of squared jump sizes:

$$[r,r]_t = \int_0^t \sigma^2(s)\,ds + \sum_{0 \le s \le t} \kappa^2(s). \tag{3.12}$$

In this setup, the RV remains a consistent estimator of the quadratic variation of the processes but provides a biased estimate of the IV.

The proper removal of jumps is of interest due to its non-diversifiable nature Pan (2002). Some of the estimators compared in this work, presented later in this section, are jump robust estimators of the integrated variance, but the identification of the intraday returns containing jump occurrences can also provide interesting insights about the assets' dynamics. We also adopt the jump identification procedure devised by Andersen *et al.* (2007), which we briefly review in the following paragraphs, in order to characterize the variation in jump frequency through time and across the multiple stocks in our dataset. The work of Johnson *et al.* (2022) provides a deeper analysis of jump dynamics.

The procedure is based on the standard realized bi-power variation measure proposed by Barndorff-Nielsen & Shephard (2004)

$$BV_{t+1}(\Delta) \equiv \mu_1^{-2} \sum_{j=2}^{1/\Delta} |r_{t+j*\Delta,\Delta}||r_{t+(j-1)*\Delta,\Delta}|, \tag{3.13}$$

where $\mu_1 \equiv \sqrt{2/\pi}$. The BV is a consistent estimator of the integrated variance even in the presence of jumps.

$$BV_{t+1}(\Delta) \to \int_t^{t+1} \sigma^2(s)ds. \tag{3.14}$$

Then, relying on the result that in the absence of jumps the returns sampled at very high frequency are approximately Gaussian, and further assuming that the volatility is constant within the trading day, the authors propose the following jump detection rule

$$\kappa_s(\Delta) = r_{t+s*\Delta,\Delta} * \mathrm{I}|[r_{t+s*\Delta,\Delta}| > \Phi_{1-\beta/2} * \sqrt{\Delta * BV_{t+1}(\Delta)}], \quad s = 1, 2, \ldots, \frac{1}{\Delta}, \tag{3.15}$$

where $\Phi_{1-\beta/2}$ is the critical value for the standard normal distribution. The value $\beta = 1 - (1-\alpha)^{\Delta}$ defines the $(1-\beta)$ confidence interval for a randomly drawn return in a pure diffusion process, which is approximately distributed as $\mathrm{N}(0, \Delta * BV_{t+1})$. α is the size of the jump test at the daily level, for which a value of $\alpha = 10^{-5}$ is claimed to present satisfactory performance, which we employ in our analysis as well. The reader is referred to the original work Andersen *et al.* (2007) for a detailed presentation.

The presence of microstructure noise, arising from factors such as the bid-ask spread and price discreteness, alters the limiting behavior of the RV estimator as $\Delta \to 0$.

$$RV_{t+1}(\Delta) \to \infty, \tag{3.16}$$

as demonstrated by Bandi & Russell (2008). Although biased and divergent, the same work shows how to determine the optimal sampling frequency, $1/\Delta$, that balances the bias-variance trade-off.

The absence of leverage, i.e. correlation between $W(t)$ and the process generating time variation in $\sigma(t)$, is one of the assumptions for the calculation of the optimal sampling frequency for the RV estimator Bandi & Russell (2008) . Even in the case of no microstructure noise, an asymmetric relationship between the innovations in the returns and $\sigma(t)$, affects the distribution of *ex post* daily returns. Some authors Ané & Geman (2000) Zhou (1998) Andersen *et al.* (2007) propose calculating returns in "financial" time τ^*. In this case, returns are computed over intervals of quadratic variation larger than a certain threshold

$$t_k = \inf_{t>0}([r,r]_t - [r,r]_{t_{k-1}} > \tau^*), \quad k = 0, 1, \ldots, \tag{3.17}$$

$$R_k \equiv p(t_k) - p(t_{k-1}), \quad k = 0, 1, \ldots, \tag{3.18}$$

where τ^* indicates the defined threshold. Then, the following result holds

$$R_k/\sqrt{\tau^*} \sim \text{i.i.d. N}(0,1), \quad k = 0, 1, 2, \ldots \tag{3.19}$$

3.A.1 Robust Two Time Scale Realized Variance

The Robust Two Time Scaled Realized Variance (RTSRV) is designed to mitigate the impact of market microstructure noise, such as measurement errors and the bid-ask spread, which can distort volatility estimates at finer time scales. It is particularly useful for high-frequency data, where these factors can significantly distort calculations of integrated variance and covariance. Otherwise, it can overcorrect.

The main equation of RTSRV is:

$$\text{RTSRV} = \left(1 - \frac{\bar{n}}{n}\right)^{-1} \left(\{X,X\}_T^{(\text{avg})*} - \frac{\bar{n}}{n}\{X,X\}_T^{(\text{all})*}\right) \tag{3.20}$$

Where

$$\{X,X\}_T^{(\text{avg})*} = \frac{\frac{c_\eta^*}{K}\sum_{i=1}^{n-K+1}(X_{t_{i+K}} - X_{t_i})^2 I_K^X(i;\eta)}{\left(\frac{1}{n-K+1}\sum_{i=1}^{n-K+1} I_K^X(i;\eta)\right)} \tag{3.21}$$

$$\{X,X\}_T^{(\text{all})*} = \frac{c_\eta^*\sum_{i=1}^{n}(X_{t_{i+1}} - X_{t_i})^2 I_1^X(i;\eta)}{\left(\frac{1}{n}\sum_{i=1}^{n} I_1^X(i;\eta)\right)} \tag{3.22}$$

$$I_K^X(i;\eta) = \begin{cases} & \text{if } \frac{(X_{t_{i+K}} - X_{t_i})^2}{\int_{t_i}^{t_{i+K}}\sigma_s^2 ds + 2\sigma_\epsilon^2} \leq \eta \\ 0 & \text{otherwise} \end{cases} \tag{3.23}$$

$$c_\eta = \frac{1}{F_{\chi_3^2}(\eta)} \tag{3.24}$$

Here, η is set at 9 to truncate returns that are larger than 3 standard deviations away from the mean of the normal distribution. Thereby, $c_\eta^* = 1.027$. The standard value for K is such that the "slow" subintervals are equal to 5 minutes. Although K can be optimized with respect to some parameters, we choose to stick with the literature. Nevertheless, one can refer to the Appendix for the "RSTRV Optimal K" results.

3.A.2
Median Realized Variance

The Median Realized Variance (MedRV) estimator is a robust form of estimating realized variance that is less sensitive to outliers compared to the traditional realized variance estimator. The MedRV of subsamples can be calculated as follows:

$$\text{MedRV} = \frac{\pi}{6 - 4\sqrt{3} + \pi} \sum_{k=1}^{n} \left(\text{Median}_{i=1,\ldots,m} |r_{k,i}|\right)^2$$

where $r_{k,i}$ are the returns within the subinterval k. In this paper, we set n such that the subinterval is 5 seconds and $m = 3$.

This formula adjusts the median of the absolute returns within each subinterval squared, multiplied by a normalization constant. The MedRV estimator is particularly useful in scenarios where the data exhibit jumps, as the median is less influenced by extreme values than the mean.

3.A.3
Kernel

This group of estimators was introduced by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) Barndorff-Nielsen *et al.* (2008) and was formulated to process high-frequency noisy data, being robust to some frictions.

$$K(X) = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \quad \gamma_h = \sum_{j=|h|+1}^{n} x_j x_{j-|h|} \tag{3.25}$$

where $k(x)$ is the chosen kernel weight function. In our analysis, we use the Parzen kernel, because it is guaranteed to produce a non-negative number. The Parzen kernel function is given by:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \le x \le \frac{1}{2} \\ 2(1-x)^3 & \frac{1}{2} \le x \le 1 \\ 0 & x > 1 \end{cases} \tag{3.26}$$

Here, x is not calculated trivially. For a more detailed explanation, the reader may refer to **?**.

3.B
Moments Distributions - Logarithm Standard Deviation

Table 3.8:
Moments Distributions - Logarithm Standard Deviation

	0.10	0.25	0.50	0.75	0.90	Mean	Std. Dev.
			Andersen et al.				
Mean	−1.28	−0.25	0.40	0.87	1.30	0.20	1.03
Std. Dev.	0.26	0.32	0.50	1.29	1.85	0.84	0.68
Skewness	−1.70	−1.07	−0.29	0.37	0.75	−0.45	1.18
Kurtosis	2.57	3.09	3.89	5.39	8.23	5.51	8.22
p-val JB	0.00	0.00	0.00	0.05	0.43	0.10	0.21
p-val CvM	0.00	0.00	0.00	0.09	0.39	0.10	0.20
p-val KS	0.00	0.00	0.01	0.13	0.45	0.12	0.21
p-val AD	0.00	0.02	0.17	0.54	0.83	0.30	0.32
			RV - 5 Minutes				
Mean	0.05	0.43	0.85	1.33	1.78	0.88	0.70
Std. Dev.	0.29	0.33	0.41	0.61	1.02	0.54	0.33
Skewness	−1.46	−0.51	0.13	0.55	0.89	−0.09	0.99
Kurtosis	2.79	3.20	3.87	5.22	8.39	5.10	4.30
p-val JB	0.00	0.00	0.00	0.08	0.49	0.11	0.23
p-val CvM	0.00	0.00	0.01	0.18	0.52	0.14	0.24
p-val KS	0.00	0.00	0.03	0.23	0.56	0.16	0.24
p-val AD	0.01	0.06	0.29	0.67	0.90	0.38	0.33
			RV - 15 Minutes				
Mean	−0.02	0.35	0.76	1.23	1.65	0.78	0.69
Std. Dev.	0.33	0.37	0.44	0.63	1.03	0.57	0.32
Skewness	−1.47	−0.50	0.06	0.41	0.70	−0.17	0.92
Kurtosis	2.81	3.19	3.77	4.89	7.91	4.93	4.12
p-val JB	0.00	0.00	0.00	0.15	0.57	0.14	0.25
p-val CvM	0.00	0.00	0.03	0.26	0.61	0.18	0.26
p-val KS	0.00	0.00	0.06	0.32	0.64	0.20	0.26
p-val AD	0.01	0.11	0.41	0.76	0.93	0.44	0.34
			Median				
Mean	−0.89	−0.27	0.21	0.69	1.14	0.17	0.83
Std. Dev.	0.32	0.40	0.57	1.10	1.58	0.79	0.53

(continued)

	0.10	0.25	0.50	0.75	0.90	Mean	Std. Dev.
Skewness	−1.60	−0.91	−0.05	0.47	0.81	−0.26	1.09
Kurtosis	2.68	3.19	3.93	5.44	8.16	5.25	7.01
p-val JB	0.00	0.00	0.00	0.07	0.47	0.11	0.22
p-val CvM	0.00	0.00	0.01	0.13	0.45	0.12	0.22
p-val KS	0.00	0.00	0.02	0.18	0.50	0.14	0.23
p-val AD	0.00	0.05	0.25	0.60	0.87	0.34	0.32
			Kernel				
Mean	0.11	0.51	0.94	1.41	1.85	0.95	0.71
Std. Dev.	0.25	0.30	0.40	0.65	1.13	0.55	0.39
Skewness	−1.90	−0.70	0.04	0.53	0.90	−0.23	1.19
Kurtosis	2.78	3.22	4.00	5.75	10.39	5.87	6.41
p-val JB	0.00	0.00	0.00	0.05	0.45	0.10	0.22
p-val CvM	0.00	0.00	0.00	0.14	0.49	0.13	0.23
p-val KS	0.00	0.00	0.02	0.19	0.53	0.14	0.23
p-val AD	0.00	0.03	0.23	0.62	0.88	0.34	0.33

3.C
More Moments Distributions

	0.10	0.25	0.50	0.75	0.90	Mean	Std. Dev.
Mean	−0.101	−0.048	0.004	0.053	0.103	0.002	0.091
Std. Dev.	0.535	0.653	0.775	0.875	0.952	0.759	0.164
Skewness	−0.241	−0.107	0.029	0.167	0.306	0.031	0.255
Kurtosis	2.363	2.776	3.106	3.491	3.975	3.187	1.137
p-val JB	0.002	0.044	0.286	0.616	0.838	0.352	0.314
p-val CvM	0.001	0.032	0.214	0.534	0.801	0.309	0.301
p-val KS	0.001	0.031	0.208	0.524	0.779	0.302	0.295
p-val AD	0.149	0.445	0.746	0.917	0.978	0.655	0.303

This table displays the percentile statistics, specifically the 10th, 25th, 50th, 75th, and 90th percentiles, as well as the mean and standard deviation of the first four moments of the unconditional distributions of standardized daily returns using the RV 1 minute estimator and for the the p-values for the null hypothesis of normality from the Jarque-Bera (JB), Cramer-von Mises (CvM), Anderson-Darling (AD), and Kolmogorov-Smirnov (KS) tests for these distributions.

Table 3.9:
Moments Distributions - RV 1 minute

	0.10	0.25	0.50	0.75	0.90	Mean	Std. Dev.
Mean	−0.127	−0.060	0.003	0.060	0.116	−0.001	0.106
Std. Dev.	0.690	0.792	0.890	0.960	1.011	0.868	0.129
Skewness	−0.196	−0.090	0.017	0.125	0.231	0.017	0.201
Kurtosis	2.159	2.488	2.710	2.954	3.244	2.737	0.900
p-val JB	0.024	0.132	0.347	0.616	0.824	0.388	0.289
p-val CvM	0.001	0.053	0.261	0.569	0.814	0.335	0.301
p-val KS	0.000	0.044	0.245	0.551	0.789	0.322	0.296
p-val AD	0.134	0.518	0.787	0.930	0.981	0.683	0.302

This table displays the percentile statistics, specifically the 10th, 25th, 50th, 75th, and 90th percentiles, as well as the mean and standard deviation of the first four moments of the unconditional distributions of standardized daily returns using the RV 15 minutes estimator and for the the p-values for the null hypothesis of normality from the Jarque-Bera (JB), Cramer-von Mises (CvM), Anderson-Darling (AD), and Kolmogorov-Smirnov (KS) tests for these distributions.

Table 3.10:
Moments Distributions - RV 15 minutes

	0.10	0.25	0.50	0.75	0.90	Mean	Std. Dev.
Mean	−0.096	−0.045	0.004	0.054	0.108	0.006	0.096
Std. Dev.	0.747	0.838	0.922	0.987	1.041	0.908	0.124
Skewness	−0.210	−0.095	0.012	0.120	0.236	0.012	0.200
Kurtosis	1.791	2.602	2.953	3.244	3.558	2.865	0.686
p-val JB	0.004	0.056	0.314	0.642	0.850	0.370	0.317
p-val CvM	0.000	0.025	0.262	0.598	0.833	0.338	0.315
p-val KS	0.000	0.031	0.268	0.589	0.814	0.336	0.309
p-val AD	0.025	0.412	0.785	0.935	0.983	0.647	0.344

This table displays the percentile statistics, specifically the 10th, 25th, 50th, 75th, and 90th percentiles, as well as the mean and standard deviation of the first four moments of the unconditional distributions of standardized daily returns using the estimator from Andersen et al. and for the the p-values for the null hypothesis of normality from the Jarque-Bera (JB), Cramer-von Mises (CvM), Anderson-Darling (AD), and Kolmogorov-Smirnov (KS) tests for these distributions.

Table 3.12:
Moments Distributions - Andersen et al.

	0.10	0.25	0.50	0.75	0.90	Mean	Std. Dev.
Mean	-1.943	-0.135	0.005	0.135	1.907	6.439×10^{8}	1.812×10^{11}
Std. Dev.	0.893	1.157	1.802	12.819	103.688	2.906×10^{10}	1.103×10^{12}
Skewness	-4.569	-0.359	0.011	0.372	4.199	-0.051	4.362
Kurtosis	2.896	3.280	4.684	28.311	91.711	28.416	49.404
p-val JB	0.000	0.000	0.000	0.275	0.686	0.173	0.284
p-val CvM	0.000	0.000	0.005	0.281	0.667	0.177	0.274
p-val KS	0.000	0.000	0.014	0.309	0.659	0.185	0.272
p-val AD	0.000	0.000	0.261	0.792	0.949	0.383	0.389

This table displays the percentile statistics, specifically the 10th, 25th, 50th, 75th, and 90th percentiles, as well as the mean and standard deviation of the first four moments of the unconditional distributions of standardized daily returns using the Median RV estimator and for the the p-values for the null hypothesis of normality from the Jarque-Bera (JB), Cramer-von Mises (CvM), Anderson-Darling (AD), and Kolmogorov-Smirnov (KS) tests for these distributions.

Table 3.11:
Moments Distributions - Median

	0.10	0.25	0.50	0.75	0.90	Mean [a]	Std. Dev.
Mean	-31.911	-0.163	0.012	0.146	43.223	2.00	43.40
Std. Dev.	1.026	1.133	1.346	233.119	2539.762	430.95	1190.9
Skewness	-6.416	-0.435	0.041	0.474	6.062	-0.042	5.224
Kurtosis	2.793	3.141	4.983	47.229	127.166	37.496	58.585
p-val JB	0.000	0.000	0.000	0.328	0.705	0.187	0.290
p-val CvM	0.000	0.000	0.007	0.348	0.720	0.200	0.289
p-val KS	0.000	0.000	0.015	0.364	0.702	0.204	0.286
p-val AD	0.000	0.000	0.287	0.833	0.961	0.398	0.405

This table displays the percentile statistics, specifically the 10th, 25th, 50th, 75th, and 90th percentiles, as well as the mean and standard deviation of the first four moments of the unconditional distributions of standardized daily returns using the RSTRV estimator and for the the p-values for the null hypothesis of normality from the Jarque-Bera (JB), Cramer-von Mises (CvM), Anderson-Darling (AD), and Kolmogorov-Smirnov (KS) tests for these distributions.

Table 3.13:
Moments Distributions - RTSRV

[a] We chose to filter the results by removing the 5% most extreme values, which we will consider as outliers in this context.

	0.10	0.25	0.50	0.75	0.90	Mean	Std. Dev.
Mean	-63.787	-0.479	0.013	0.270	78.894	4.9×10^{10}	2.6×10^{13}
Std. Dev.	1.043	1.169	5.626	472.321	3708.7	3.8×10^{12}	1.2×10^{14}
Skewness	-7.904	-1.475	0.036	1.340	7.686	-0.022	5.9
Kurtosis	2.879	3.350	12.320	67.031	151.6	46.5	63.9
p-val JB	0.000	0.000	0.000	0.174	0.637	0.148	0.269
p-val CvM	0.000	0.000	0.000	0.223	0.642	0.157	0.268
p-val KS	0.000	0.000	0.000	0.241	0.633	0.161	0.266
p-val AD	0.000	0.000	0.004	0.742	0.942	0.319	0.392

This table displays the percentile statistics, specifically the 10th, 25th, 50th, 75th, and 90th percentiles, as well as the mean and standard deviation of the first four moments of the unconditional distributions of standardized daily returns using the K-Optimized RTSRV estimator and for the the p-values for the null hypothesis of normality from the Jarque-Bera (JB), Cramer-von Mises (CvM), Anderson-Darling (AD), and Kolmogorov-Smirnov (KS) tests for these distributions.

Table 3.14:
Moments Distributions - RTSRV Optimal K

3.D
Sources of Variation

Sources of Variation - RV 5Min - Top 10

	(1)	(2)	(3)	(4)	(5)
JB P - Value					
R^2	0.007	0.0091	0.0098	0.0101	0.0104
Unc. Var		0.0021	0.0007	0.0003	0.0003
Year FE	✓	✓	✓	✓	✓
% Trade		✓	✓	✓	✓
Div. Yield			✓	✓	✓
B/M				✓	✓
Other					✓
Asymmetry					
R^2	0.062	0.073	0.081	0.089	0.092
Unc. Var		0.011	0.008	0.008	0.003
Year FE	✓	✓	✓	✓	✓
B/M		✓	✓	✓	✓
P/E			✓	✓	✓
% Trade				✓	✓
Other					✓
Kurtosis					
R^2	0.010	0.016	0.021	0.027	0.029
Unc. Var		0.006	0.005	0.006	0.002
Year FE	✓	✓	✓	✓	✓
B/M		✓	✓	✓	✓
% Trade			✓	✓	✓
P/E				✓	✓
Other					✓

This table illustrates the R^2 values and the R^2 increments for the equations regarding the standardized return distributions using the *5-minute RV* estimator for *all companies*. Each block displays the four variables selected by the forward selection algorithm. In the first block, the dependent variable is the p-value of the Jarque-Bera test. In the second block, it is the skewness, and in the third block, the kurtosis. The first row of each block corresponds to the R^2 value, while the second row shows the increment in R^2 contributed by the inclusion of each variable.

Sources of Variation - Kernel - Top 10

	(1)	(2)	(3)	(4)	(5)
JB P - Value					
R^2	0.008	0.011	0.012	0.013	0.014
Unc. Var		0.003	0.001	0.001	0.001
Year FE	✓	✓	✓	✓	✓
% Trade		✓	✓	✓	✓
Jumps			✓	✓	✓
Size				✓	✓
Other					✓
Asymmetry					
R^2	0.07	0.081	0.088	0.092	0.097
Unc. Var		0.011	0.007	0.004	0.005
Year FE	✓	✓	✓	✓	✓
B/M		✓	✓	✓	✓
% Trade			✓	✓	✓
P/E				✓	✓
Other					✓
Kurtosis					
R^2	0.012	0.019	0.027	0.033	0.036
Unc. Var		0.007	0.008	0.006	0.003
Year FE	✓	✓	✓	✓	✓
% Trade		✓	✓	✓	✓
B/M			✓	✓	✓
P/E				✓	✓
Other					✓

This table illustrates the R^2 values and the R^2 increments for the equations regarding the standardized return distributions using the *Kernel* estimator for *all companies*. Each block displays the four variables selected by the forward selection algorithm. In the first block, the dependent variable is the p-value of the Jarque-Bera test. In the second block, it is the skewness, and in the third block, the kurtosis. The first row of each block corresponds to the R^2 value, while the second row shows the increment in R^2 contributed by the inclusion of each variable.

Sources of Variation - RV 5Min - Total

	(1)	(2)	(3)	(4)	(5)
JB P - Value					
R^2	0.020	0.031	0.040	0.041	0.042
Unc. Var		0.011	0.009	0.001	0.001
% Trade	✓	✓	✓	✓	✓
Jumps		✓	✓	✓	✓
Year FE			✓	✓	✓
Size				✓	✓
Other					✓
Asymmetry					
R^2	0.014	0.024	0.028	0.030	0.032
Unc. Var		0.01	0.004	0.002	0.002
Year FE	✓	✓	✓	✓	✓
Jumps		✓	✓	✓	✓
P/E			✓	✓	✓
Div. Yield				✓	✓
Other					✓
Kurtosis					
R^2	0.031	0.062	0.085	0.098	0.110
Unc. Var		0.031	0.023	0.013	0.012
Year FE	✓	✓	✓	✓	✓
Jumps		✓	✓	✓	✓
B/M			✓	✓	✓
P/E				✓	✓
Other					✓

This table illustrates the R^2 values and the R^2 increments for the equations regarding the standardized return distributions using the *5-minute RV* estimator for *only the 10% largerst companies*. Each block displays the four variables selected by the forward selection algorithm. In the first block, the dependent variable is the p-value of the Jarque-Bera test. In the second block, it is the skewness, and in the third block, the kurtosis. The first row of each block corresponds to the R^2 value, while the second row shows the increment in R^2 contributed by the inclusion of each variable.

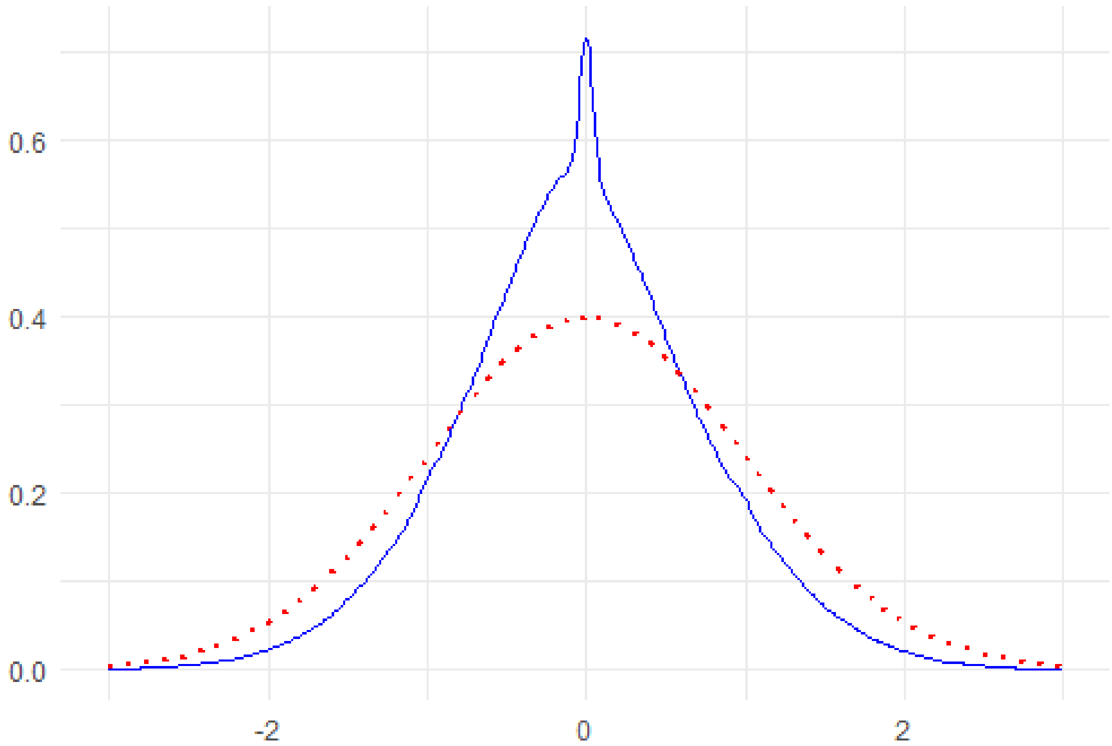
Sources of Variation - Kernel - Total

	(1)	(2)	(3)	(4)	(5)
JB P - Value					
R^2	0.094	0.127	0.132	0.137	0.138
Unc. Var		0.033	0.005	0.005	0.001
% Trade	✓	✓	✓	✓	✓
Jumps		✓	✓	✓	✓
Size			✓	✓	✓
Year FE				✓	✓
Other					✓
Asymmetry					
R^2	0.005	0.007	0.0080	0.0083	0.0085
Unc. Var		0.002	0.001	0.0001	0.0002
Year FE	✓	✓	✓	✓	✓
Jumps		✓	✓	✓	✓
Div. Yield			✓	✓	✓
P/E				✓	✓
Other					✓
Kurtosis					
R^2	0.027	0.055	0.059	0.061	0.061
Unc. Var		0.028	0.002	0.002	0.00
% Trade	✓	✓	✓	✓	✓
Jumps		✓	✓	✓	✓
Year FE			✓	✓	✓
Size				✓	✓
Other					✓

The table illustrates the R^2 values and the R^2 increments for the equations regarding the standardized return distributions using the *Kernel* estimator for *only the 10% largerst companies*. Each block displays the four variables selected by the forward selection algorithm. In the first block, the dependent variable is the p-value of the Jarque-Bera test. In the second block, it is the skewness, and in the third block, the kurtosis. The first row of each block corresponds to the R^2 value, while the second row shows the increment in R^2 contributed by the inclusion of each variable.

3.E
90% Smallest Firms

Figure 3.10: Distribution of standardized returns for the top 90% smallest firms in 2020.



Mean: -0.06 Standard deviation: 0.56 Skewness: 0.06 Kurtosis: 3.9
The figure shows the distribution of standardized daily returns using the 5-minute RV estimator for the top 90% smallest firms in 2020 (blue line) and the standard normal distribution (red dashed line).

	Daily return statistics				Estimators daily averages					
Year	Count	Mean	Median	% Zero	RV 1m	RV 5m	RV 15m	Kernel	RTSRV	MedRV
Panel A: 2003–2008										
2003	4755	0.17	0	6.29	0.438	0.368	0.301	0.385	0.078	0.128
2004	4607	0.03	0	4.77	0.280	0.234	0.193	0.253	0.058	0.078
2005	4544	-0.04	0	5.16	0.226	0.190	0.160	0.203	0.047	0.062
2006	4491	0.02	0	4.63	0.182	0.152	0.130	0.165	0.042	0.051
2007	4468	-0.08	0	4.30	0.228	0.194	0.155	0.197	0.046	0.063
2008	4422	-0.30	0	4.10	0.845	0.697	0.595	0.719	0.142	0.215
Panel B: 2009–2015										
2009	3951	-0.029	0	4.72	0.705	0.594	0.517	0.618	0.120	0.175
2010	3772	-0.025	0	4.17	0.231	0.198	0.174	0.211	0.049	0.065
2011	3621	-0.142	0	3.67	0.246	0.212	0.186	0.223	0.055	0.060
2012	3507	0.003	0	4.03	0.219	0.188	0.162	0.198	0.039	0.048
2013	3454	0.041	0	3.79	0.175	0.148	0.126	0.160	0.035	0.040
2014	3498	-0.007	0	3.35	0.250	0.208	0.179	0.214	0.052	0.058
2015	3521	-0.001	0	3.55	0.304	0.248	0.209	0.271	0.064	0.078
Panel B: 2016–2020										
2016	3422	-0.005	0	3.92	0.402	0.296	0.240	0.342	0.078	0.109
2017	3379	-0.084	0	5.77	0.354	0.244	0.181	0.315	0.091	0.150
2018	3374	-0.216	0	4.9	0.439	0.304	0.224	0.399	0.114	0.176
2019	3373	-0.085	0	3.9	0.490	0.331	0.240	0.439	0.119	0.188
2020	3348	-0.184	0	3.01	0.715	0.499	0.382	0.700	0.225	0.393

The table displays the summary statistics for the *90% smallest companies* (measured in monthly capitalization) detailed by year for returns (on the left side) and for quadratic variance estimators (on the right side). The table reports the number of tickers in that year (Count), the unconditional mean and median of daily returns (Mean and Median), the average daily percentage of zero returns (% Zeros), and the daily averages of the Kernel, Robust Two Time Scale RV, Median RV, RV with 1-minute, 5-minute, and 15-minute intervals (Kernel, RTSRV, MedRV, RV 1m, RV 5m, and RV 15m) estimators.

Table 3.15:
Daily return statistics and estimators averages - Top 90% Smallest Firms

Bibliography

Albuquerque, Bruno, & Baumann, Ursel. 2017. Will US inflation awake from the dead? The role of slack and non-linearities in the Phillips curve. *Journal of Policy Modeling*, **39**(2), 247–271.

Amihud, Yakov. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, **5**(1), 31–56.

Andersen, Torben G, Bollerslev, Tim, Diebold, Francis X, & Ebens, Heiko. 2001. The distribution of realized stock return volatility. *Journal of financial economics*, **61**(1), 43–76.

Andersen, Torben G, Bollerslev, Tim, & Dobrev, Dobrislav. 2007. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications. *Journal of Econometrics*, **138**(1), 125–180.

Andersen, Torben G, Dobrev, Dobrislav, & Schaumburg, Ernst. 2012. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, **169**(1), 75–93.

Anderson, Theodore W, & Darling, Donald A. 1952. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, 193–212.

Ané, Thierry, & Geman, Hélyette. 2000. Order flow, transaction clock, and normality of asset returns. *The Journal of Finance*, **55**(5), 2259–2284.

Aoki, Kosuke. 2001. Optimal monetary policy responses to relative-price changes. *Journal of monetary economics*, **48**(1), 55–80.

Arnerić, Josip, & Žigman, Ante Čuljak, Maria and. 2022. Is Jump Robust Two Times Scaled Estimator Superior among Realized Volatility Competitors? *Mathematics*, **10**(12), 2124.

Bandi, Federico M, & Russell, Jeffrey R. 2008. Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies*, **75**(2), 339–369.

Barndorff-Nielsen, Ole E, & Shephard, Neil. 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, **2**(1), 1–37.

Barndorff-Nielsen, Ole E, & Shephard, Neil. 2006a. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics*, **4**(1), 1–30.

Barndorff-Nielsen, Ole E, & Shephard, Neil. 2006b. Variation, jumps and high frequency data in financial econometrics. *Advanced in Economics and Econometrics. Theory and Applications*.

Barndorff-Nielsen, Ole E, Hansen, Peter R, Lunde, Asger, & Shephard, Neil. 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, **76**(6), 1481–1536.

Barnichon, Régis, & Shapiro, Adam Hale. 2024. Phillips meets Beveridge. *Journal of Monetary Economics*, **148**, 103660. Inflation in the COVID Era and Beyond.

Bauer, Michael D, & Swanson, Eric T. 2023. A reassessment of monetary policy surprises and high-frequency identification. *NBER Macroeconomics Annual*, **37**(1), 87–155.

Bera, Anil K, & Jarque, Carlos M. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, **6**(3), 255–259.

Bontemps, Christian, & Meddahi, Nour. 2005. Testing normality: a GMM approach. *Journal of Econometrics*, **124**(1), 149–186.

Boudt, Kris, & Zhang, Jin. 2015. Jump robust two time scale covariance estimation and realized volatility budgets. *Quantitative Finance*, **15**(6), 1041–1054.

Carvalho, Carlos. 2006. Heterogeneity in price stickiness and the real effects of monetary shocks. *Frontiers in Macroeconomics*, **2**(1).

Carvalho, Carlos, & Nechio, Fernanda. 2016. Factor specificity and real rigidities. *Review of Economic Dynamics*, **22**, 208–222.

Carvalho, Carlos, Lee, Jae Won, & Park, Woong Yong. 2021. Sectoral price facts in a sticky-price model. *American Economic Journal: Macroeconomics*, **13**(1), 216–256.

Christiano, Lawrence J, Eichenbaum, Martin, & Evans, Charles L. 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, **113**(1), 1–45.

Dufour, Jean-Marie, Khalaf, Lynda, & Kichian, Maral. 2010. On the precision of Calvo parameter estimates in structural NKPC models. *Journal of Economic Dynamics and Control*, **34**(9), 1582–1595.

Fama, Eugene F., & French, Kenneth R. 2015. A five-factor asset pricing model. *Journal of Financial Economics*, **116**(1), 1–22.

Fitzgerald, Terry, Jones, Callum, Kulish, Mariano, & Nicolini, Juan Pablo. 2024. Is There a Stable Relationship between Unemployment and Future Inflation? *American Economic Journal: Macroeconomics*, **16**(4), 114–42.

Fuhrer, Jeff, & Moore, George. 1995. Inflation persistence. *The Quarterly Journal of Economics*, **110**(1), 127–159.

Gali, Jordi, Gertler, Mark, & Lopez-Salido, J David. 2005. Robustness of the estimates of the hybrid New Keynesian Phillips curve. *Journal of Monetary Economics*, **52**(6), 1107–1118.

Galí, Jordi. 2015. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press.

Galí, Jordi, & Gertler, Mark. 1999. Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, **44**(2), 195–222.

Griffa, Cristina. 2025. *Sectoral Phillips Curves: The Role of Expectations and Production Networks in Price Setting*. Manuscript, January 2025.

Guvenen, Fatih. 2006. Reconciling conflicting evidence on the elasticity of intertemporal substitution: A macroeconomic perspective. *Journal of Monetary Economics*, **53**(7), 1451–1472.

Hansen, Peter R, & Lunde, Asger. 2005. A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of applied econometrics*, **20**(7), 873–889.

Hazell, Jonathon, Herreno, Juan, Nakamura, Emi, & Steinsson, Jón. 2022. The slope of the Phillips Curve: evidence from US states. *The Quarterly Journal of Economics*, **137**(3), 1299–1344.

Hooper, Peter, Mishkin, Frederic S, & Sufi, Amir. 2020. Prospects for inflation in a high pressure economy: Is the Phillips curve dead or is it just hibernating? *Research in Economics*, **74**(1), 26–62.

Imbs, Jean, Jondeau, Eric, & Pelgrin, Florian. 2011. Sectoral Phillips curves and the aggregate Phillips curve. *Journal of Monetary Economics*, **58**(4), 328–344.

Johnson, James A, Medeiros, Marcelo C, & Paye, Bradley S. 2022. Jumps in stock prices: New insights from old data. *Journal of Financial Markets*, **60**, 100708.

Jordà, Òscar, & Nechio, Fernanda. 2018. Inflation Globally. *Federal Reserve Bank of San Francisco*.

Jordi, GalÃ, & Gambetti, Luca. 2019. *Has the US Wage Phillips Curve Flattened? A Semi-Structural Exploration*. Tech. rept. CEPR Discussion Papers.

Liu, Lily Y., Patton, Andrew J., & Sheppard, Kevin. 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, **187**(1), 293–311.

Luengo-Prado, María José, Rao, Nikhil, & Sheremirov, Viacheslav. 2018. Sectoral inflation and the Phillips curve: What has changed since the Great Recession? *Economics Letters*, **172**, 63–68.

Mavroeidis, Sophocles, Plagborg-Møller, Mikkel, & Stock, James H. 2014. Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve. *Journal of Economic Literature*, **52**(1), 124–88.

McAleer, Michael, & Medeiros, Marcelo C. 2008. Realized volatility: A review. *Econometric reviews*, **27**(1-3), 10–45.

McLeay, Michael, & Tenreyro, Silvana. 2020. Optimal inflation and the identification of the Phillips curve. *NBER macroeconomics annual*, **34**(1), 199–255.

Monacelli, Tommaso. 2007. Comment on" Globalization and Inflation Dynamics: A Structural Perspective". *Pages 579–590 of: International Dimensions of Monetary Policy*. University of Chicago Press.

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

Nason, James M, & Smith, Gregor W. 2008a. Identifying the new Keynesian Phillips curve. *Journal of Applied Econometrics*, **23**(5), 525–551.

Nason, James M., & Smith, Gregor W. 2008b. The New Keynesian Phillips Curve: Lessons from Single-Equation Econometric Estimation. *Federal Reserve Bank of Richmond Economic Quarterly*, **94**(4), 361–395. Available at https://ssrn.com/abstract=2187862.

Pan, Jun. 2002. The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, **63**(1), 3–50.

Rubio-Ramirez, Juan F, Waggoner, Daniel F, & Zha, Tao. 2010. Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies*, **77**(2), 665–696.

Schwartzman, Felipe Farah. 2006. Estimativa de curva de Phillips para o Brasil com preços desagregados. *Economia aplicada*, **10**, 137–155.

Shapiro, Adam Hale, *et al.* 2020. A simple framework to monitor inflation. Federal Reserve Bank of San Francisco.

Sim, Chiaw Hock, & Yap, Bee Wah. 2011. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, **81**(12), 2141–2155.

Smirnov, Nickolay. 1948. Table for estimating the goodness of fit of empirical distributions. *The annals of mathematical statistics*, **19**(2), 279–281.

Stock, James H, & Watson, Mark W. 2007. Why has US inflation become harder to forecast? *Journal of Money, Credit and banking*, **39**, 3–33.

Stock, James H, & Watson, Mark W. 2020. Slack and cyclically sensitive inflation. *Journal of Money, Credit and Banking*, **52**(S2), 393–428.

Woodford, Michael. 2004. Inflation targeting and optimal monetary policy. *Review-Federal Reserve Bank of Saint Louis*, **86**(4), 15–42.

Woodford, Michael. 2005. *Firm-specific capital and the New-Keynesian Phillips curve*.

Zaniboni, Nicola, & Zaniboni, N. 2008. Globalization and the phillips curve. *V Unpublished manuscript*.

Zhou, Bin. 1998. F-consistency, De-volatilization and normalization of high-frequency financial data. *Nonlinear Modelling of High Frequency Financial Time Series*, 109–123.