Vol. 4, April 2026

THE CAPITOL ECONOMICS JOURNAL

A Publication of the Undergraduate Economics Society at The George Washington University

Capitol Economics Journal

Volume IV, Spring 2026

The CEJ is published annually and all publication policies and processes are conducted in accordance with international standards of academic ethics and integrity. The CEJ accepts and publishes original, undergraduate research papers relating to all fields of economics. Consideration for each year's edition opens at the beginning of the fall semester and closes one month prior to the end of the semester.

Read the journal online:

https://www.gwuues.org/publications

Contact: ues.gwu@gmail.com

ISSN: 2771-7348

George Washington University Undergraduate Economics Society
University Student Center
800 21st Street NW
Washington, D.C. 20052

| Contents

| Letter from the Editor

Dear Reader,

This is the fourth edition of the Capitol Economics Journal and the second since its revival in 2024. The following publication is the culmination of a year's worth of work by our writers and editorial staff. The Capitol Economics Journal provides an opportunity for economics students at George Washington University to publish their senior theses. The journal furthers the mission of the Undergraduate Economics Society by promoting high-quality, groundbreaking undergraduate economic research.

The papers chosen for publication represent the best of what George Washington University undergraduates have to offer in research prowess. Over two semesters, seniors developed econometric models to address a range of economic questions. Our writers have continued to refine their work alongside a strong team of associate editors and faculty in the economics department. The selected papers feature research on the labor impacts of California's cap-and-trade program, stock market reactions to FTC Antitrust lawsuits, and the effect of the 2008 Financial Crisis on the cosmetics industry.

I would like to extend my immense gratitude to everyone who was involved in the publication process, and my deepest congratulations to the published writers. Our associate editors, Adrian, Ashka, Grace, Satya, and Sneha, were a dedicated team and indispensable to the production of this journal. To the Executive Board of the Undergraduate Economics Society, thank you for assisting with every step of this process. The Capitol Economics Journal would not be possible without the help of every single person involved. This is the collective product of a devoted and passionate team. Their work has guaranteed that this journal will continue to publish student research for many years to come.

Finally, thank you, the reader, for exploring our journal. Your curiosity sustains our research and gives our work purpose. Through research, we can better understand the human experience, challenge prior notions, and develop better solutions to the great economic challenges of our time. By engaging with our research, you are reaffirming a great tradition of empirical work and the importance of its results.

Thank You,

Aidan J. Cullers
Editor-in-Chief
Capitol Economics Review

FTC Antitrust Lawsuits and Stock Market Reactions: Evidence from 2020–2025

Meghan R. Bankapur
George Washington University

Abstract

This paper examines how financial markets react to Federal Trade Commission enforcement actions from 2020 to 2025. Using a standard event-study framework, I estimate cumulative abnormal returns (CARs) for 40 publicly traded defendants. Cross-sectional regressions control for case type and industry fixed effects. Average CARs are economically small and statistically indistinguishable from zero across all windows, with mean CARs of approximately −0.3 percent in the (−1,+1) window and −0.2 percent in the (−3,+3) window. These findings suggest that investors largely anticipate modern antitrust enforcement or view it as routine. Sectoral results reveal modest heterogeneity: energy firms exhibit small but positive CARs in longer windows, averaging roughly 5 percent in the (−10,+10) window, consistent with interpretations of enforcement as reducing regulatory uncertainty or signaling more predictable competitive conditions. The evidence points to muted short-term market effects of antitrust actions, consistent with a regulatory environment in which enforcement is expected and efficiently priced. **JEL Codes:** G14, L41, G38

1 Introduction

Antitrust enforcement is a central tool policymakers use to promote competition, limit excessive market power, and shape firm behavior. Because enforcement actions can alter expected profitability, legal exposure, and future pricing freedom, financial markets should react to enforcement 'surprises' when such information becomes public. Stock-price responses therefore offer a real-time signal of how investors interpret enforcement, whether as a costly regulatory intervention or as a stabilizing action that strengthens long-run competitive conditions. Yet despite the recent increase in high-profile FTC cases, relatively little is known about how modern antitrust enforcement affects firm valuations in today's fast-moving information environment.

This paper examines whether Federal Trade Commission (FTC) enforcement announcements between 2020 and 2025 generate abnormal stock returns for publicly traded defendants. The question matters both for antitrust policy and financial economics. Under the semi-strong form of the Efficient Market Hypothesis (EMH), security prices should incorporate all publicly available information. If enforcement actions are anticipated, price adjustments may occur before the announcement. If they contain new information, abnormal returns should appear within narrow event windows around the disclosure. Modern enforcement also occurs in an environment with faster information dissemination, digital media coverage, and heightened regulatory scrutiny, making it unclear whether older empirical patterns still hold. Abnormal returns (ARs) measure the deviation of a firm's realized return from its expected return predicted by a market model. Cumulative abnormal returns (CARs) aggregate these deviations over a specified event window and capture the total valuation effect of new information. An enforcement action is considered "anticipated" if investigations, media reporting, regulatory filings, or procedural developments provide investors with advance signals regarding likely legal intervention. In such cases, price adjustments may occur prior to the formal announcement date.

Using a standard event-study framework, I estimate cumulative abnormal returns (CARs) for 40 publicly traded defendants across four event windows surrounding each FTC announcement. The sample spans diverse industries, including technology, health care, consumer goods, and energy. I then estimate cross-sectional regressions with industry fixed effects to test whether case type, such as merger or non-merger enforcement, systematically predicts CARs, while also examining sectoral heterogeneity and potential information leakage prior to announcements.

The remainder of the paper proceeds as follows. Section 1.1 reviews the relevant literature and situates the contribution within prior event-study research. Section 1.2 outlines the Efficient Market Hypothesis and its implications for regulatory announcements. Section 2 describes the empirical framework, Section 3 presents the data, Section 4 reports results and robustness checks, and Section 5 concludes.

1.1 Antitrust Enforcement and Stock-Market Reactions

A substantial empirical literature studies how antitrust enforcement affects firm valuations using event-study methods. Classic work examines how mergers, enforcement challenges, and collusion investigations generate abnormal stock returns. Eckbo (1983) analyzes 113 horizontal mergers challenged by U.S. antitrust authorities between 1963 and 1978 and finds combined bidder–target abnormal returns of approximately −1.6 percent, consistent with expectations of reduced post-merger market power. Combined bidder–target abnormal returns refer to the value-weighted abnormal returns of both merging firms, interpreted as the market's assessment of expected merger synergies or reductions in market power. A -1.6 percent response suggests that investors anticipated reduced monopoly rents following enforcement challenges.Extending this, Eckbo (1985) shows that rival firms do not earn significant gains from challenged mergers, casting doubt on traditional monopoly-power arguments predicting positive spillovers.

In a related analysis, Stillman (1983) studies 11 DOJ merger challenges and finds no significant rival effects, suggesting that enforcement interventions do not systematically shift competitive rents. Criminal antitrust actions generate clearer market reactions: Bosch and Eckard (1991) examine 37 price-fixing indictments and document stock losses of -2 to -3 percent, reflecting expected penalties and reputational costs. Karpoff, Lee, and Vendrzyk (1999) show that fraud and procurement-related investigations can produce even larger CARs, often exceeding -10 percent, highlighting that allegations of misconduct destroy firm value.

More recent research shifts attention to merger remedies and rival-firm responses. Duso, Gugler, and Yurtoglu (2010) analyze 151 European mergers from 1990 to 2002 and find that remedies imposed by competition authorities correlate with negative CARs of -0.8 to -1.2 percent for merging firms but positive CARs for rivals, suggesting reallocation of market power. Clougherty and Duso (2013) deepen this logic by using rival reactions to infer competitive effects, concluding that positive rival CARs serve as indirect evidence that enforcement strengthens market competition.

Methodologically, MacKinlay (1997) provides the framework that underpins this empirical approach, outlining standard practices for estimation-window selection, event-window definition, and abnormal return aggregation. These studies collectively suggest that antitrust enforcement can affect firm valuations, but effects vary substantially by era, enforcement type, and industry structure. This motivates re-examining market responses in the modern, information-rich environment of 2020–2025.

1.2 Efficient Market Hypothesis and Information Transmission

The event-study approach used in this paper relies on the semi-strong form of the Efficient Market Hypothesis (EMH), which states that security prices fully and immediately reflect all publicly available information. Fama (1970) formalized this idea, arguing that abnormal returns arise only when announcements contain information not yet incorporated into prices. Under this view, enforcement announcements should generate measurable stock-price reactions only if they provide new, value-relevant information to investors.

Brown (1968), who show that firms' earnings news is rapidly impounded into stock prices, and Fama, Fisher, Jensen, and Roll (1969), who document near-instantaneous adjustment to stock splits. Jensen (1978) interpreted these findings as evidence that markets process information efficiently enough that predictable, risk-adjusted abnormal profits are infeasible. Later critiques, including Hall (1983) and Malkiel (2003), emphasize that efficiency depends on frictions, disclosure timing, and investor attention, factors especially relevant for regulatory announcements. Critics argue that market efficiency may be limited by investor inattention, information-processing frictions, or strategic disclosure timing. Regulatory announcements may be complex and subject to interpretation, potentially delaying price adjustment. These considerations imply that muted CARs may reflect informational frictions rather than pure anticipation.

Event studies serve as practical tests of EMH. If markets are informationally efficient, price reactions should appear precisely when enforcement information becomes public. Conversely, muted or statistically insignificant cumulative abnormal returns (CARs) suggest that the announcement was anticipated or economically unimportant. Information-leakage tests, such as examining pre-announcement CARs, correspond directly to EMH principles: significant lead returns indicate that markets incorporated information earlier, while insignificant leads support the timing of the event definition. Importantly, a finding of insignificant CARs does not imply that the EMH fails. Rather, it may indicate that information was incorporated prior to the announcement date. Distinguishing between information leakage and informational inefficiency requires examining pre-event returns.

This framework guides the empirical analysis below. By measuring abnormal returns around FTC enforcement announcements, I evaluate not only the economic impact of regulatory actions but also the extent to which markets anticipate and price such actions, providing insight into the informational environment surrounding modern antitrust policy.

2 Model Description

2.1 Event-Study Framework

The market model specification follows MacKinlay (1997), who demonstrates that it improves statistical power relative to mean-adjusted returns by controlling for systematic market risk. This design is grounded in the semi-strong form of the Efficient Market Hypothesis (Fama, 1970; Fama, 1991), which states that asset prices adjust fully and immediately to publicly available information. Under this framework, enforcement announcements should generate abnormal returns only if they contain new information that investors had not already incorporated into prices. If markets anticipate enforcement, the price reaction will occur earlier; if markets do not, the reaction should be concentrated within a narrow event window. Symmetric windows are standard in event-study design to capture both immediate reaction and minor delays in information processing (MacKinlay, 1997). Starting at t=0 alone risks missing anticipatory or slightly lagged responses. Because windows overlap mechanically, statistical inference is conducted separately for each specification. To estimate abnormal returns, I model daily stock returns $\mathrm{R_{it}}$ using the market model:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \tag{1}$$

where $\mathrm{R_{it}}$ is the return on a broad market index and ϵ_{it} is the abnormal component of returns. Parameters α_i and β_i are estimated over an estimation window of 250 trading days ending 30 days before the announcement date (t ϵ [− 250, − 30]). The event window captures the short-term reaction surrounding the announcement.

2.2 Abnormal and Cumulative Abnormal Returns

Abnormal returns (AR) represent deviations from expected market-related performance:

$$AR_{it} = R_{it} - (\hat{\alpha}_i + \hat{\beta}_i R_{mt}) \tag{2}$$

The disturbance term it represents the abnormal component of returns. Abnormal returns ($\mathrm{AR_{it}}$) are thus the realized residuals from the market model during the event window. Cumulative abnormal returns (CARs) for firm i over event window W are defined as:

$$CAR_i(W) = \sum_{t \in W} AR_{it} \tag{3}$$

CARs quantify the total price response to new information released during the event window. These are the dependent variables in the cross-sectional analysis. I compute CARs for each window using daily data for 40 publicly traded defendants and aggregate them into an event-level dataset. The choice of symmetric windows [− 1, + 1], [− 3, + 3], [− 5, + 5] and [− 10, + 10] reflects the assumption, implicit in the EMH, that markets react within a short period around the time new information becomes public. If enforcement conveys meaningful new information, CARs should deviate from zero; if not, CARs should remain statistically indistinguishable from zero.

2.3 Cross-Sectional Regressions

To test whether enforcement type and industry characteristics explain variation in CARs, I estimate the following specification:

$$CAR_i(W) = \gamma_0 + \gamma_1 CaseType_i + \delta' IndustryFE_i + u_i \tag{4}$$

where $\mathrm{CaseType_i}$ indicates the type of FTC action (merger, non-merger, or neutral) and $\mathrm{IndustryFE_i}$ denotes a set of industry fixed effects. Standard errors are heteroskedasticity-robust. For robustness, I also re-estimate the model after trimming extreme CARs (±0.20) and include pre-event "lead" windows to test for information leakage.

2.4 Identification Considerations

Because the validity of an event study depends on correctly identifying the moment information reaches the market, I also examine whether enforcement news leaks prior to the announcement. Under the semi-strong EMH, significant abnormal returns before the focal date would signal that markets incorporated enforcement expectations early. To assess this, I test for abnormal returns in pre-event "lead" windows and evaluate whether removing firm-days with confounding disclosures affects the results. If CARs remain stable and pre-event returns are indistinguishable from zero, this supports the interpretation that the event date captures the relevant information arrival.

2.5 Model Intuition

The intuition behind this model is straightforward. The event-study isolates how markets react when new regulatory information becomes public. Under the semi-strong Efficient Market Hypothesis, prices should adjust immediately to unexpected enforcement announcements, so any abnormal return reflects information that investors had not previously incorporated. If investors interpret FTC enforcement as harmful, because it raises legal costs, limits market power, or signals future restrictions, the estimated coefficient on the relevant case type will be negative, reflecting price declines. Conversely, a positive coefficient would indicate that investors view enforcement as beneficial, perhaps by reducing uncertainty, preventing monopolistic inefficiency, or creating opportunities for competitors.

Industry fixed effects capture systematic differences across sectors: for instance, energy or healthcare may exhibit stronger reactions if regulation meaningfully affects profitability in those markets. Robust standard errors ensure that inference is not distorted by heteroskedasticity or small-sample noise. Altogether, this framework links legal and regulatory actions to real-time changes in firm valuation and, combined with the EMH, allows us to infer how capital markets perceive the economic consequences and informational content of antitrust policy.

3 Data Description

The dataset consists of 40 publicly traded firms subject to Federal Trade Commission (FTC) enforcement actions between 2020 and 2025. Each observation corresponds to a single event, the first public announcement of an enforcement action, matched to the stock returns of the affected firm. Information on enforcement type (merger, non-merger, or neutral) and industry classification was collected from the FTC case docket and cross-checked with company filings. Daily stock prices were obtained from Google Finance, and market returns were proxied using the S&P 500 Index. Market returns are calculated using daily adjusted closing prices of the S&P 500 Index, ensuring dividends and splits are reflected.

To compute cumulative abnormal returns (CARs), I use an estimation window of 250 trading days ending 30 days before each event and four symmetric event windows: [−1,+1], [−3,+3], [−5,+5], and [−10,+10] days. These windows capture both immediate and slightly delayed market reactions. The resulting variables, CAR_m1_p1, CAR_m3_p3, CAR_m5_p5, and CAR_m10_p10, serve as the dependent variables in the regression analysis.

Descriptive statistics for these variables are presented in Table 1, which reports mean and sample size by industry. Average CARs range between −0.3 percent and +0.5 percent across industries in the (−1,+1) window, suggesting that markets do not respond strongly to enforcement announcements on average. Figure 1 further illustrates this pattern: median CARs across case types remain near zero, with only modest dispersion. This pattern foreshadows the regression results, where industry differences, rather than case type, account for most of the variation in returns.

Table 1. Summary Statistics by Industry					
Industry	**N**	**Mean CAR (-1,+1)**	**Mean CAR (-3,+3)**	**Mean CAR (-5,+5)**	**Mean CAR (-10,+10)**
Consumer Disc.	4	-0.026	0.001	-0.030	-0.081
Consumer Staples	7	0.022	0.029	0.020	0.042
Energy	3	0.049	0.116	0.060	0.092
Financials	1	−0.0011585	−0.0052217	−0.000598	−0.0021033
Health Care	15	−0.0204941	−0.0185762	−0.0247572	−0.0443073
Industrials	2	0.027	−0.013424	−0.006401	−0.0127482
Information Tech.	3	−0.0442929	0.032	−0.0185762	−0.0926488
Materials	3	0.030	−0.009693	0.005	−0.0054588
Real Estate	1	−0.0006816	−0.0185509	0.012	−0.0702878
Utilities	1	−0.0077777	0.012	0.015	0.001
Total	40	-0.003	0.005	-0.009	-0.020

Table 1: Note: This table reports mean cumulative abnormal returns (CARs) by industry across four event windows. CARs are calculated using an event-study framework described in Section 2. Source: Author's calculations using FTC case data and Google Finance.

3.1 Data Sources and Cleaning

Information on enforcement actions was compiled from the Federal Trade Commission's (FTC) online case database, which reports the date, type, and industry of each action. I identified publicly traded defendants by cross-referencing case names with company filings and ticker symbols from Google Finance and Yahoo Finance. For firms with multiple actions during the sample period, I kept the first event to avoid overlapping announcement windows. Some industries have small sample sizes (e.g., n=1), so mean CARs should be interpreted cautiously. Cross-sectional regression analysis accounts for this imbalance through fixed effects rather than simple mean comparison. For example, materials firms exhibit a 3 percent mean CAR in the (−1,+1) window. While this magnitude appears economically meaningful, it reflects only three observations and is not statistically distinguishable from zero in regression analysis.

Stock price data were collected using Google Finance formulas in Google Sheets, pulling daily adjusted close prices for each firm and the S&P 500 market index. I verified that all tickers had continuous trading data over the estimation and event windows; any with missing prices were excluded. Industry labels were standardized to match the ten-sector classification used in the regressions. Duplicates and inconsistent case identifiers were removed in Stata after merging the event-level dataset with case-level descriptors. The final dataset contains 40 unique events covering firms across ten industries.

4 Results

4.1 Main Estimates

Table 2 reports cross-sectional regressions of cumulative abnormal returns (CARs) across four symmetric event windows surrounding FTC enforcement announcements. Across all specifications, coefficients on the case-type indicators are small in magnitude and statistically indistinguishable from zero. This pattern indicates that investors do not systematically differentiate between merger and non-merger enforcement in the short run. R2 values range from 0.18 to 0.25, suggesting that case characteristics and industry fixed effects explain only a modest share of the variation in event-window returns.

Industry effects, however, exhibit mild heterogeneity. Energy-sector CARs reach approximately 9 percent in the (−10,+10) window, though estimates remain imprecise due to small sample size. This hints that investors may view enforcement in this sector as reducing uncertainty, constraining dominant firm behavior, or stabilizing future competitive conditions. Other industries show estimates close to zero with no consistent direction. These patterns mirror the descriptive evidence in Table 1, where mean CARs cluster near zero for most industries but are somewhat higher and more variable for energy and consumer staples. Figure 2 presents mean CARs by industry and provides complementary evidence of modest, sector-specific heterogeneity. While most industries exhibit small mean responses, energy stands out with more positive average returns, echoing the regression results.

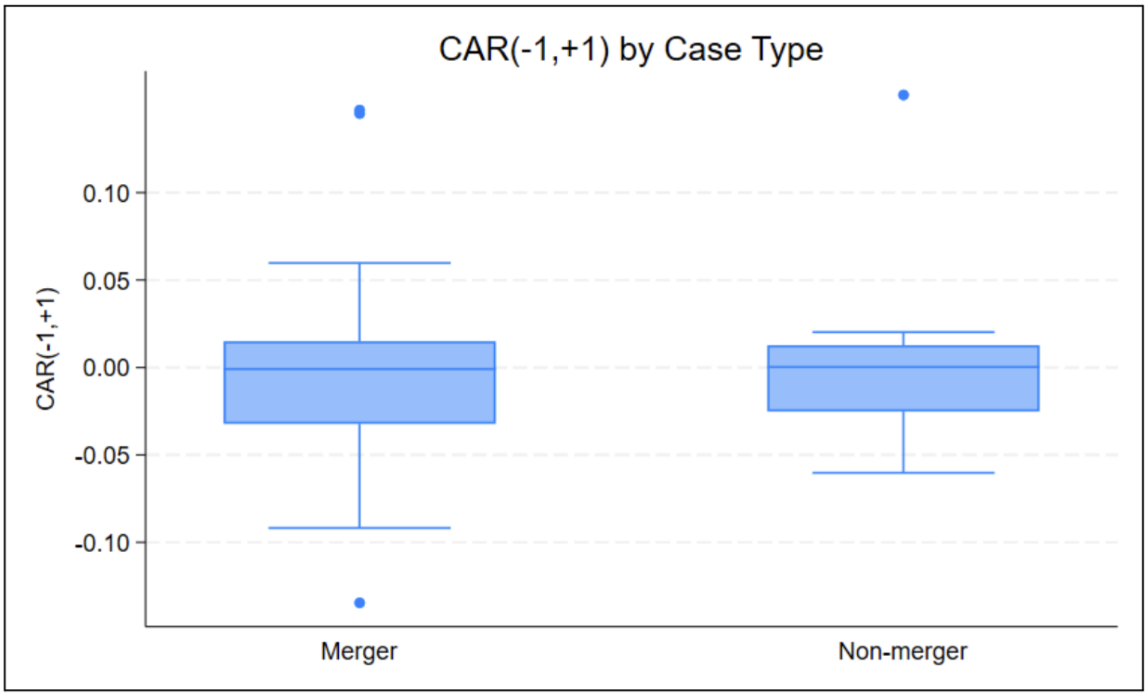

Figure 1: CAR(-1,+1) by Case Type. Boxplots show cumulative abnormal returns surrounding FTC enforcement announcements for merger and non-merger cases. Source: Author's calculations

Table 2. Cross-Sectional Regressions				
	CAR (-1,+1)	CAR (-3,+3)	CAR (-5,+5)	CAR (-10,+10)
Neutral (n)	–0.0081	0.0133	–0.0474	–0.0509
	-0.0146	-0.0207	-0.0305	-0.0520
Non-Merger (nm)	0.0313	0.0355	0.0327	0.0416
	-0.0273	-0.0253	-0.0352	-0.0463
Industry FE	Yes	Yes	Yes	Yes
Observations	40	40	40	40
R^2	0.2490	0.2490	0.2000	0.1790

Table 2: Cross-Sectional Regressions of Cumulative Abnormal Returns. Note: Table reports OLS estimates from cross-sectional regressions of cumulative abnormal returns (CARs) surrounding FTC enforcement announcements. The dependent variable is CAR over the indicated event window. Independent variables include indicators for neutral and non-merger enforcement actions; merger cases serve as the omitted reference category. Industry fixed effects are included in all specifications. Heteroskedasticity-robust standard errors are reported in parentheses. R^2 values correspond to each regression. Source: Author's calculations.

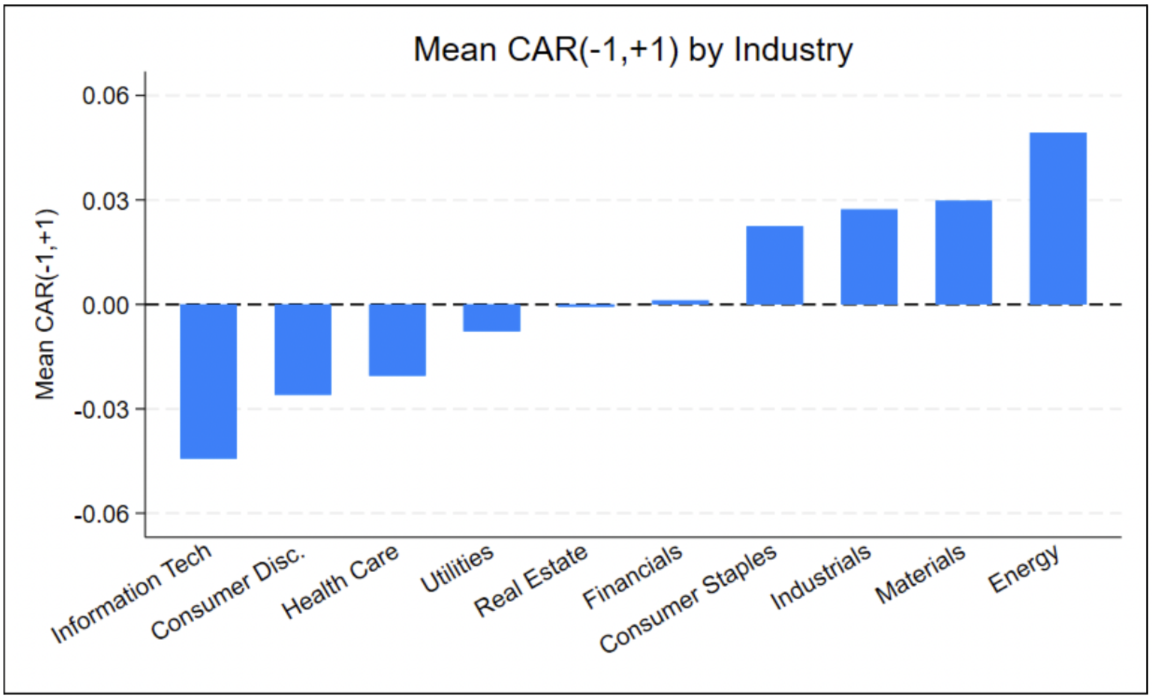

Figure 2: Mean CAR(-1,+1) by industry. Bars represent average cumulative abnormal returns across industries following FTC enforcement announcements. Source: Author's calculations

4.2 Interpretation Through the Lens of Market Efficiency

Viewed through the semi-strong form of the Efficient Market Hypothesis, the results suggest that FTC enforcement announcements often convey little new information to investors. If markets anticipate enforcement, because investigations are public, procedural timelines are predictable, or media coverage reduces uncertainty, prices may adjust prior to the announcement. In this scenario, CARs around the event date will appear statistically flat even if enforcement matters economically. Alternatively, investors may interpret enforcement actions as routine rather than transformative, especially in industries accustomed to regulatory oversight. These interpretations align with recent work emphasizing rapid information diffusion and high investor attention in modern disclosure environments.

4.3 Robustness Checks

Trimming Outliers

To assess sensitivity to extreme observations, I re-estimate the regressions after excluding firms with CARs exceeding ± 0.20 in the $(-1,+1)$ window. Results remain qualitatively unchanged: case-type coefficients stay small and insignificant, and the relative strength of energy-sector estimates persists in the longer windows. Five observations were trimmed for exceeding ± 0.20 in the $(-1,+1)$ window.

Pre-Announcement Leakage

I next evaluate whether markets incorporated enforcement information ahead of time by constructing CARs over wider lead windows, including $(-180,-2)$ and $(-360,-2)$. The 180- and 360-day lead windows are constructed to test for extended pre-event drift and are independent of the 250-day estimation window used for parameter estimation. Across all windows, average pre-event CARs exhibit no systematic trends or significant deviations from zero. This provides no evidence of anticipatory trading or information leakage and supports the assumption that the announcement date captures the release of value-relevant information.

Alternative Event Windows

Finally, comparing results across the four primary windows reveals no meaningful differences: CAR patterns remain flat across case types, with modest, industry-specific variation. The stability of these patterns across window lengths strengthens the conclusion that enforcement announcements do not generate large or systematic abnormal returns for defendants.

Limitations

Several limitations warrant mention. First, the sample size is modest (n=40), limiting statistical power. Second, event-date identification may not perfectly capture initial information arrival. Third, enforcement actions vary in economic magnitude, and aggregation may obscure heterogeneous effects. Finally, the analysis focuses only on defendants and does not incorporate rival-firm spillovers.

5 Conclusion

This paper examines how financial markets respond to Federal Trade Commission (FTC) enforcement actions between 2020 and 2025. Using a standard event-study design, I estimate cumulative abnormal returns (CARs) for 40 publicly traded defendants across multiple event windows and assess how these reactions vary by case type and industry. The analysis offers new evidence on the valuation effects of modern antitrust enforcement during a period of heightened regulatory scrutiny and rapid information dissemination.

The results suggest that investors largely anticipate or discount FTC actions.

Average CARs are close to zero in all windows, and case-type coefficients are statistically insignificant, implying that enforcement announcements contain limited new information for markets or are perceived as routine regulatory events. Sectoral heterogeneity, however, reveals that energy firms experience modest positive CARs in longer windows, consistent with interpretations of enforcement as reducing uncertainty or stabilizing future competitive dynamics. Viewed through the semi-strong form of the Efficient Market Hypothesis, these muted reactions indicate that markets either incorporated enforcement expectations earlier or judged the announcements to have minimal valuation consequences.

Future work can examine mechanisms underlying this heterogeneity. Two promising extensions involve testing whether media saliency amplifies investor attention and incorporating rival-firm data to identify spillover effects. Even if the semi-strong EMH does not hold perfectly, the absence of strong abnormal returns suggests that enforcement announcements do not generate immediate wealth transfers for publicly traded defendants, an important insight for policymakers evaluating the financial impact of antitrust intervention. Together, these analyses may clarify whether enforcement redistributes value across competitors or signals broader changes in industry conduct. Overall, the evidence points to limited short-term financial impacts of antitrust policy, highlighting a regulatory environment in which enforcement is generally expected and efficiently priced.

References

[1] Ball, Ray, and Philip Brown. 1968. "An Empirical Evaluation of Accounting Income Numbers." *Journal of Accounting Research* 6 (2): 159–178.

[2] Bosch, Jean, and E. Woodrow Eckard. 1991. "The Profitability of Price Fixing: Evidence from Stock Market Reaction to Federal Indictments." *Review of Economics and Statistics* 73 (2): 309–317.

[3] Clougherty, Joseph A., and Thomas Duso. 2013. "Using Rival Effects to Identify Synergies and Improve Merger Typologies." *Strategic Organization* 11 (4): 310–335.

[4] Duso, Tomaso, Klaus Gugler, and Burcin B. Yurtoglu. 2010. "Is the Event Study Methodology Useful for Merger Analysis? A Comparison of Stock Market and Accounting Data." *International Review of Law and Economics* 30 (2): 186–192.

[5] Eckbo, B. Espen. 1983. "Horizontal Mergers, Collusion, and Stockholder Wealth." *Journal of Financial Economics* 11 (1–4): 241–273.

[6] Eckbo, B. Espen. 1985. "Mergers and the Market Concentration Doctrine: Evidence from the Capital Market." *Journal of Business* 58 (3): 325–349.

[7] Fama, Eugene F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* 25 (2): 383–417.

[8] Fama, Eugene F. 1991. "Efficient Capital Markets: II." *Journal of Finance* 46 (5): 1575–1617.

[9] Fama, Eugene F., Lawrence Fisher, Michael C. Jensen, and Richard Roll. 1969. "The Adjustment of Stock Prices to New Information." *International Economic Review* 10 (1): 1–21.

[10] Hall, Robert E. 1983. "The Importance of Lifetime Jobs in the U.S. Economy." *American Economic Review* 73 (4): 716–724.

[11] Jensen, Michael C. 1978. "Some Anomalous Evidence Regarding Market Efficiency." *Journal of Financial Economics* 6 (2–3): 95–101.

[12] Karpoff, Jonathan M., Daniel S. Lee, and Valaria P. Vendrzyk. 1999. "Defense Contracting Fraud, Penalties, and Contractor Influence." *Journal of Political Economy* 107 (6): 1067–1095.

[13] MacKinlay, A. Craig. 1997. "Event Studies in Economics and Finance." *Journal of Economic Literature* 35 (1): 13–39.

[14] Malkiel, Burton G. 2003. "The Efficient Market Hypothesis and Its Critics." *Journal of Economic Perspectives* 17 (1): 59–82.

[15] Stillman, Robert. 1983. "Examining Antitrust Policy toward Horizontal Mergers." *Journal of Financial Economics* 11 (1–4): 225–240.

[16] U.S. Federal Trade Commission. 2020–2025. "FTC Cases and Proceedings Database." https://www.ftc.gov/enforcement/cases-proceedings. Accessed November 2025.

Estimating Demand for Cosmetics, Perfume, and Bath Preparations Before, During, After the Global Financial Crisis

Chesapeake Dowdy

George Washington University

Abstract

This paper utilizes an Almost Ideal Demand System (AIDS) to construct demand for Cosmetics, Perfume, and Bath Preparations and related goods from the Consumption Expenditure Survey and track how demand changed with the Global Financial Crisis. The results can be applied to test the "Lipstick Effect," a theory that consumption of lipstick and other cosmetics increases during recessions. Cosmetics were found to be a normal good before and after the recession, with expenditure elasticities of demand 1.37 and 0.97, but became inferior during with an expenditure elasticity of demand -2.77. Jewelry became a much stronger luxury good during the recession with an expenditure elasticity of demand of 8.22 compared with 1.85 beforehand. Substitutability and complementarity between cosmetics and jewelry or dresses evolved with the crisis as well. **JEL Codes:** D12, E21, L66

1 Introduction

Recessions are known to change consumption in many ways, namely the reduction in consumption of luxuries that have higher income elasticities of demand resulting from increased unemployment and lower household income during recessionary periods. Given the severity of the Global Financial Crisis, however, there is potential that it changed underlying consumption and savings behaviors, such as elasticities, especially with the destruction of wealth that results from a financial crisis (Chakrabarti, et al. 2010). For example, Attanasio, et al. (2022) found that car purchases decreased during the Global Financial Crisis to a greater extent than would have occurred during a "normal" recession. The estimation of demand can provide insight into how consumption patterns, such as price or income elasticity of demand and substitution between goods, change during or after a recession as severe as the Global Financial Crisis. Cosmetics provide an especially interesting case study because there are conflicting opinions as to whether they fall under the category of a "luxury" or a "necessity." Estimating the demand for cosmetics alone and when compared to other goods typically consumed by women such as dresses or jewelry can provide valuable insight into the consumption patterns of a large portion of the adult population.

There is substantial literature using the Almost Ideal Demand System (AIDS) to estimate demand for goods, mainly in the field of agricultural economics. Additionally, there is some literature estimating certain demand elasticities for cosmetics products, mainly to test the "Lipstick Effect" hypothesis that demand for cosmetics increases during recessions. However, this paper presents the first use of the AIDS model to estimate demand for cosmetics. The paper is organized as follows. The next section examines how demand estimation has evolved over time and different applications of the AIDS model. Section 3 introduces the Consumer Expenditure Survey and Consumer Price Index data used. Section 4 specifies the equations used for an Almost Ideal Demand System. Section 5 presents the results, and Section 6 applies these results to test the "Lipstick Effect" hypothesis. Section 7 discusses potential limitations to the methodology. Finally, Section 8 provides a brief summary and conclusions.

2 Literature Review

Estimating demand has been an empirical focus for much of modern economics with many applications from understanding overall market functioning to evaluating the impacts of trade policy or designing optimal taxation systems. Early exercises in demand estimation built upon the theoretical framework of the laws of supply and demand to assign values to elasticities of demand. Pigou (1910) estimated elasticity of demand for food and clothes of workmen. Lehfeldt (1914) used an exponential model based on population growth to model demand for imported wheat in the UK. Moore (1922) used a linear model to derive the law of demand for potatoes in the United States.

Building off of these early models that were partially determined by available data, demand estimation has expanded into many goods and services. Moshary, et al. (2025) used stated-choice experiments to analyze elasticity of demand for firearms including substitution between different types of firearms. Capps, et al. (2024) constructed conditional demand for Greek yogurt using a probit model with scanner data finding that certain demographic characteristics lead to a higher likelihood of purchasing Greek yogurt. Meredith, Macy, and Meredith (2022) used survey data to estimate income elasticity of demand for tanning bed usage.

Examining demand and consumption during recessions is an important part of demand estimation literature, and it helps with understanding the impacts that recessions have on the overall economy, including how elasticities of demand can be time-varying. Field and Pagoulatos (1997) used instrumental variables to examine how price elasticity of demand in US food manufacturing varies over the business cycle. They found that price elasticities of demand were generally procyclical with heterogeneity across industries based on trade exposure, capital intensity, and product differentiation. Cho, Morley, and Singh (2024) modeled income and consumption surrounding the pre-GFC housing boom, finding that marginal propensity to consume increased following the end of the boom with a heterogeneous effect based on differing household balance sheet characteristics. Fleissig (2021) used the Fourier Flexible Form to estimate elasticities of demand for the "sin goods" of alcohol, tobacco, gambling, and lottery. He found that beer and lotteries exhibited higher own-price elasticity of demand during the GFC and that the level of substitutability or complementarity between the goods changed. Biswas, Chintagunta, and Dhar (2025) used scanner-level data in a logit model to examine how the wealth shocks of the GFC and COVID-19 pandemic impacted consumption patterns. They found that the GFC caused the share of consumer-packaged goods that were private label, as opposed to national brands, to decrease, and household income was found to be a factor in substitution patterns.

One popular demand estimation model is the Almost Ideal Demand System (AIDS) which was proposed by Deaton and Muellbauer (1980) and has been used to model demand in a variety of contexts, mainly in agricultural economics. Variations of the AIDS model have been used to estimate demand in many contexts including protein sources during the Global Financial Crisis (Yang, Raper, and Pruitt 2019); fresh-cut and artificial flowers (Girapunthong and Ward 2003); whale-related recreation trips and donations of time or money (Shaikh and Larson 2003); outpatient antibiotics (Filippini, Masiero, and Moschetti 2009); tobacco products (Zheng, et al. 2016); and gasoline, leisure, and related goods (West and Williams III 2004).

3 Data

Expenditure data are taken from the Consumer Expenditure Survey Diary Survey conducted by the Bureau of Labor Statistics, a panel survey of approximately 5,000 addresses conducted every calendar quarter that records two weeks of expenditures for each consumption unit (Bureau of Labor Statistics 2019). The data categories used in this analysis are "Cosmetics, Perfume, and Bath Preparations"; "Jewelry"; and "Women's Dresses" from Quarter 1 of 2002 through Quarter 3 of 2019. The category labeled "All Else" represented all expenditures that do not include these specific data categories. Price data are taken from the BLS Consumer Price Index for all Urban Consumers (CPI-U) not seasonally adjusted with a base of 1982-1984=100. The data categories used in this analysis are "Cosmetics, Perfume, Bath, Nail Preparations and Implements"; "Jewelry"; "Women's Dresses"; and "All Items" from January 2002 through December 2019. Price level for each quarter was denoted as the end-of-quarter value of CPI.

The data were categorized by quarter into three different time periods. The quarters Q1 of 2002 through Q4 of 2007 encompass the expansion before the Global Financial Crisis (or Pre- GFC), Q1 of 2008 through Q2 of 2009 represent the Global Financial Crisis (or GFC), and Q3 of 2009 through Q3 of 2019 encompass the expansion after the Global Financial Crisis up until the recession induced by the COVID-19 pandemic (or Post-GFC), as defined by NBER business cycle dates.

Table 3.1 displays summary statistics for the Price Level variable for the entire sample period, and it is notable that jewelry and dresses exhibited the larger variation in price than cosmetics, but overall CPI exhibited even more variability. The trend can also be seen in all other time periods except the Global Financial Crisis[1]. This could be due to overall CPI's inclusion of energy and food, both of which are generally more volatile in price than other goods, but core CPI only exhibits somewhat lower variability than overall CPI. For the purposes of this analysis, the inclusion of all other goods is meant to provide a level of total expenditure that can best approximate income for the purposes of estimating income elasticity of demand, so the greater price variability should not inhibit analysis. Estimation is repeated using core CPI in Section 7 to confirm.

[1] Additional summary statistics are included in the data appendix.

	Cosmetics	Jewelry	Dresses	All Items	Core
Min	165.9	125.6	96.8	178.8	189.8
Mean	180.5	157.9	116.9	220.3	224.9
Max	190.5	183.2	137.5	257.0	264.9
SD	7.5	18.6	10.3	22.7	22.1

Table 1: Summary Statistics of Quarter-End CPI for Different Goods Categories, 2002-2019

4 Model Specification

The equations used for this model are from Henningsen's "Demand Analysis with the 'Almost Ideal Demand System' in R: Package micEconAids." The AIDS model begins with an expenditure function that "defines the minimum expenditure necessary to attain a specific utility level at given prices" (Deaton and Muellbauer 1980). This utility level is unobserved and thus estimated through x_{it}, the consumed quantity of good i at time t, which depends on the overall price level and expenditure at time t: p_t and m_t, respectively. The resulting equation is:

$$x_{it}(p_t, m_t) = \frac{m_t}{p_{it}}\left[\alpha_i + \sum_j \gamma_{ij}\ln(p_{jt}) + \beta_i \ln\left(\frac{m_t}{p_{it}}\right)\right] \tag{1}$$

where p_{it} is the price of good i at time t, p_{jt} is the price of a corresponding good j, P_t is a translog price index at time t, and the estimated coefficients are α_i, γ_{ij}, β_i. The translog price index is an estimation of the unobserved price level p_i defined by:

$$\ln(P_t) = \alpha_0 + \sum_i \alpha_i \ln(p_{it}) + \frac{1}{2}\sum_i\sum_j \gamma_{ij}\ln(p_{it})\ln(p_{jt}) \tag{2}$$

The expenditure function is simplified by rewriting x_{it} using s_{it}, the unobserved utility-optimized proportion of the consumer's budget that would go to good i at time t:

$$s_{it} = x_{it}\frac{p_{it}}{m_t} \tag{3}$$

This yields the equivalent expenditure function:

$$s_{it}(p_t, m_t) = \alpha_i + \sum_j \gamma_{ij} \ln(p_{jt}) + \beta_i \ln\left(\frac{m_t}{P_t}\right) \tag{4}$$

For estimation, s_{it} is substituted with w_{it}, the observed budget share, under the assumption that consumers are optimizing their utility:

$$w_{it} = \alpha_i + \sum_j \gamma_{ij} \ln(p_{jt}) + \beta_i \ln\left(\frac{m_t}{P_t}\right) + u_{it} \tag{5}$$

To estimate the expenditure functions using only linear terms, the translog price index must be approximated by a linear price index, $\hat{P}_t$. Deaton and Muellbauer (1980) propose using the Stone price index:

$$\ln(\widehat{P}_t) = \sum_i w_{it} \ln(p_{it}) \tag{6}$$

Using this price index, the econometric estimation of the expenditure function is:

$$w_{it} = \hat{\alpha}_i + \sum_j \hat{\gamma}_{ij} \ln(p_{jt}) + \hat{\beta}_i \ln\left(\frac{x_t}{\widehat{P}_t}\right) + u_{it} \tag{7}$$

where $\hat{\alpha}_i$, $\hat{\gamma}_{ij}$, $\hat{\beta}_i$ are the estimated coefficients, x_t is the observed level of total expenditure at t, and u_{it} is a disturbance term such that $\sum_i u_{it} = 0$ for all values of t. Coefficients are estimated using the method of "Seemingly Unrelated Regression" first proposed by Zellner (1962).

The expenditure elasticity of demand for good i, η_i, is the percentage increase in the quantity demanded of good i associated with a one percent increase in total expenditure. In a standard AIDS model, expenditure elasticities are defined by:

$$\eta_i = 1 + \frac{\beta_i}{s_i} \tag{8}$$

which is estimated as:

$$\hat{\eta}_i = 1 + \frac{\hat{\beta}_i}{w_i} \tag{9}$$

However, as noted by Green and Alston (1991), "In the LA/AIDS model, expenditure elasticities also ought to account for the role of expenditure shares as variables in the Stone's price index." They account for this with the adjustment:

$$\hat{\eta}_i = 1 + \frac{\hat{\beta}_i}{w_i}\left[1 - \sum_j w_j \ln \widehat{P}_j \left(\hat{\eta}_j - 1\right)\right] \tag{10}$$

where η_j is the expenditure elasticity of related goods j, and the expenditure elasticities are calculated simultaneously. To ease the difficulty of simultaneous calculation, Buse (1994) finds an equivalent equation:

$$\hat{\eta}_i = 1 + \frac{\hat{\beta}_i}{w_i}\left[1 - \frac{\sum_k \hat{\beta}_k \ln p_k}{1 + \sum_k \hat{\beta}_k \ln p_k}\right] \tag{11}$$

Marshallian price elasticity of a good is the percentage change in the quantity demanded of a good i resulting from a 1% increase in the price of an associated good j, holding total expenditure constant. The Marshallian price elasticity in a standard AIDS model for goods i and j, θ_{ij}, can be calculated:

$$\hat{\theta}_{ij} = -\delta_{ij} + \frac{\hat{\gamma}_{ij}}{w_i} - \frac{\hat{\beta}_i}{w_i}\left(\frac{\partial \ln \widehat{P}}{\partial \ln p_j}\right) \tag{12}$$

where $\hat{P}$ is the estimated Stone Price Index for the entire sample and $\delta_i j$ is Kronecker delta, meaning that it is equal to 1 if $i = j$ and 0 otherwise. This accounts for own-price elasticities, the percentage increase in quantity demanded of a good associated with a 1% increase in its price. Buse (1994) has similarly derived an alternative calculation for the linear approximation:

$$\hat{\theta}_{ij} = -\delta_{ij} + \frac{\hat{\gamma}_{ij}}{w_i} - \frac{\hat{\beta}_i}{w_i}\left(\frac{w_j + \sum_k \hat{\gamma}_{kj} \ln p_k}{1 + \sum_k \hat{\beta}_k \ln p_k}\right) \tag{13}$$

Hicksian price elasticities measure the percentage change in the quantity demanded of a good i resulting from a 1% increase in the price of associated good j, holding total utility constant. The Hicksian price elasticity can be calculated by inserting the expenditure elasticity and Marshallian price elasticity into the Slutsky equation:

$$\theta^*_{ij} = \theta_{ij} + \eta_i s_j \tag{14}$$

estimated by:

$$\hat{\theta}^*_{ij} = \hat{\theta}_{ij} + \hat{\eta}_i\, w_i \tag{15}$$

5 Results

Expenditure elasticities of demand are calculated by inputting the estimated α and β coefficients from equation (7) for each time period and good into equation (11). These approximate income elasticity of demand, the percentage change in the quantity demanded of a good associated with a 1% increase in income. Not all consumers are "hand-to-mouth," consuming close to 100% of their income, and estimated proportions of the US population in this category vary widely. Kaplan, Violante, and Weidner (2014) estimate that, in 2010, 25-40% of households in the US were "hand-to-mouth." Nonetheless, total expenditures are very predictive of income (Paulin and Ferraro 1996), though income will be higher than expenditures. This may cause magnitudes of expenditure elasticities of demand to be smaller than the true income elasticities of demand, but sign of the estimates should remain consistent.

Goods can be broken into two categories based on their income elasticities of demand: normal and inferior. Normal goods have a positive income elasticity of demand, meaning that the quantity demanded of such goods will increase as a consumer's income increases. Inferior goods have a negative income elasticity of demand, meaning that the quantity demanded will increase as a consumer's income decreases. Necessities have an income elasticity of demand small in magnitude, as consumers will not greatly change their consumption of essential goods. These can be thought of as having an income elasticity of demand in between 0 and 1, where a 1% change in income is associated with a less than 1% increase in consumption of said good. Luxuries, on the other hand, will have an income elasticity of demand greater than 1, as higher income will greatly increase one's propensity to consume such goods.

Table 2 presents expenditure elasticities of demand for the goods categories over the different time period specifications. Over the entire sample period of 2002-2019, all studied goods are normal with positive expenditure elasticities of demand. Cosmetics fall into the category of necessities while jewelry and dresses are luxuries. During the Global Financial Crisis, the expenditure elasticities of demand for the studied goods changed sharply with cosmetics and dresses becoming inferior goods and jewelry becoming much more of a luxury good.

	Cosmetics	Jewelry	Dresses
Pre-GFC	1.37	1.85	1.16
GFC	-2.77	8.22	-6.28
Post-GFC	0.97	1.83	2.12
2002-2019	0.61	1.94	1.38

Table 2: Expenditure Elasticities of Demand

Hicksian own-price elasticities measure the percentage change in the quantity demanded of a good associated with a 1% increase in its price, holding utility constant. Marshallian own- price elasticities measure this change, holding expenditure constant. According to the Law of Demand, consumers will demand less of a good the more it costs, so Hicksian and Marshallian own-price elasticities must be negative to be consistent with economic theory. Inelastic demand occurs when own-price elasticities are between -1 and 0, meaning that consumers do not change their consumption very much resulting from a change in price. Demand is more elastic when own- price elasticity is less than -1, meaning that a 1% increase in price is associated with a decrease in quantity demanded of more than 1%.

Hicksian cross-price elasticities measure the percentage change in the quantity demanded of a good associated with a 1% increase in the price of another good, holding utility constant. In indifference curve analysis, these elasticities represent the substitution effect where a consumer stays on the same indifference curve with a change in the slope of the budget constraint as opposed to the income effect, where the slope of the budget constraint is constant with a change in indifference curve. Marshallian cross-price elasticities measure the percentage change in the quantity demanded of a good associated with a 1% increase in the price of another good, holding utility constant. In indifference curve analysis, Marshallian cross-price elasticities are the total effect, combining the income and substitution effects.

Based on cross-price elasticities, goods can be complements, substitutes, or unrelated. When goods are complementary, they will have a negative cross-price elasticity; an increase in the price of one is associated with a decrease in the quantity demanded of the other. If an increase in the price of a good causes consumers to demand more of another, they have a positive cross-price elasticity of demand and are thus substitutes. Goods that are unrelated, neither substitutes nor complements, will have a cross-price elasticity near zero. As the magnitude of the cross-price elasticity of demand increases, the strength of complementarity or substitutability grows. Goods with a cross-price elasticity with a magnitude between 0.5 and 1 are weak substitutes, greater than 1 are strong substitutes. If two goods are strong substitutes, an increase in price of one good by 1% causes the quantity demanded of the other good to increase by more than 1%. Likewise, goods are weak complements if they have a cross-price elasticity between -1 and -0.5, and strong complements if their cross-price elasticity is less than -1.

Table 3 presents Marshallian own-price and cross-price elasticities of demand and Table 4 presents Hicksian own-price and cross-price elasticities for the expansion before the Global Financial Crisis. The estimate for Marshallian and Hicksian cross-price elasticities of demand are very similar for all the goods, meaning that the income effect is minimal. This pattern persists throughout all time specifications, so further analysis will look only at Marshallian cross-price elasticities[2]. The own-price elasticities, which are diagonal entries, are negative, consistent with the Law of Demand. All three goods exhibit similarly elastic demand. Since the matrix is not symmetric, dresses and cosmetics could be seen as either weak or strong substitutes since a 1% increase in the price level of cosmetics is associated with a 1.65% increase in the quantity demanded of dresses, but a 1% increase in the price level of dresses is only associated with a 0.65% increase in the quantity demanded of dresses. Other pairs of goods do not exhibit notable levels of complementarity or substitutability.

	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-2.13	0.08	0.65
$Q_{Jewelry}$	0.07	-2.19	-0.17
$Q_{Dresses}$	1.65	-0.49	-1.95

Table 3: Marshallian Price Elasticities, Pre-GFC

[2]All other Hicksian price elasticities are included in the data appendix.

	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-2.13	0.09	0.66
$Q_{Jewelry}$	0.08	-2.18	-0.17
$Q_{Dresses}$	1.66	-0.49	-1.95

Table 4: Hicksian Price Elasticities, Pre-GFC

Table 5 presents Marshallian own-price and cross-price elasticities of demand during the Global Financial Crisis. Demand for cosmetics became much more inelastic and dresses somewhat more elastic but not notably so. Demand for jewelry became significantly more elastic which is consistent with the findings of Browning and Crossley (2000) that "luxury goods tend also to have high intertemporal substitution elasticities." Dresses and cosmetics can be seen to change from being weak substitutes to strong complements, and jewelry became a strong complement to dresses during the Global Financial Crisis.

	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-0.67	0.30	-1.83
$Q_{Jewelry}$	0.06	-11.39	-3.21
$Q_{Dresses}$	-3.62	-8.65	-3.33

Table 5: Marshallian Price Elasticities, GFC

Table 6 presents Marshallian own-price and cross-price elasticities for the expansion after the Global Financial Crisis. Demand for cosmetics became more elastic, increasing to a magnitude closer to what it was before the Global Financial Crisis. Jewelry's own price elasticity of demand decreased in magnitude but remained elevated compared to before the Global Financial Crisis. Dresses exhibit much more inelastic demand than during or before the Global Financial Crisis. All cross-price elasticities became much weaker than during the Global Financial Crisis, and cosmetics changed to a complement for jewelry.

	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-1.61	-0.40	-0.84
$Q_{Jewelry}$	-0.56	-4.20	-0.57
$Q_{Dresses}$	-1.88	-0.94	-0.60

Table 6: Marshallian Price Elasticities, Post-GFC

6 Application to "Lipstick Effect"

An application of this demand estimation is to test the existence of the "Lipstick Effect." The "Lipstick Effect" or "Lipstick Index" is a hypothesis that women buy more of beauty products like lipstick during times of economic distress. It is credited to Leonard Lauder, former chairman of Estee Lauder: "When things get tough, women buy lipstick" (Finch, The Guardian 2001). This theory came after Estee Lauder saw strong net earnings and profits in the fiscal year 2001 and 2002 (SEC Form 10-K) despite both encompassing the economic contraction of 2001. Many assume that cosmetics, especially because there are a variety of luxury offerings, are a normal good, but the "Lipstick Effect" suggests that they are an inferior good.

This is not the first paper to test the "Lipstick Effect," as there have been both economics and psychology papers that test this hypothesis. Papers in the field of psychology tend to assume the existence of the "Lipstick Effect" and perform experiments to determine why the phenomenon happens. A notable example is by Netchaeva and Rees (2016) who found that women purchased more cosmetics during times of economic distress to improve perception in the workplace and to attract romantic partners which both have the potential to improve financial conditions. However, the assumption that the "Lipstick Effect" exists in these papers makes them of less use in performing the intended economic analysis of income elasticity of demand for cosmetics.

Macdonald and Dildar (2020) test the existence of the "Lipstick Effect" and some of the explanations put forth by psychology. They analyzed household consumption data from the Bureau of Labor Statistics Consumption Expenditure Survey and found that average expenditure on cosmetics did increase during the GFC among women aged 18-40 but not more so among unmarried than married women, going against the hypothesis that this increased expenditure was to attract a mate. Additionally, they found that employed women spend more on cosmetics in general, but the frequency and amount of spending did not increase notably more than unemployed women which goes against the hypothesis that women buy more cosmetics to improve their perception in the workplace. Given the lack of evidence for the two most popular psychological hypotheses, the authors proposed that the increased consumption of cosmetics was motivated by substitution away from more expensive luxury goods and found a substitution away from goods such as clothing which cost more on average than cosmetics.

Li, Zhen, and Dorfman (2020) model demand for cosmetics using a panel smooth transition regression (PSTR) using scanner data for the period 2006 to 2016 and found that income elasticity of demand for cosmetics notably decreased during the Global Financial Crisis before returning to pre-recession levels around Q1 of 2014, though remaining positive during the whole sample period. They also tested whether the ratio of job vacancies to unemployed job seekers had an impact on the income elasticity of demand and found no statistically significant relationship.

As can be seen in Table 2, cosmetics are a normal good over the entire sample period of 2002-2019 with a positive expenditure elasticity of demand, but this expenditure elasticity of demand actually became negative during the Global Financial Crisis meaning that cosmetics were an inferior good. This aligns quite closely with the "Lipstick Effect" because income decreased during the recession which would then cause increased consumption of inferior goods, exactly what the "Lipstick Effect" hypothesizes will occur for cosmetics during a recession. Even looking at the entire sample period of 2002-2019, cosmetics are estimated to have more income inelastic demand than other goods typically consumed by women such as jewelry or dresses which supports the hypothesis that cosmetics are less of a luxury good than jewelry. Cosmetics were found to be a moderate substitute for jewelry over all sample periods with no notable change in substitutability during the Global Financial Crisis which does not lend credibility to the hypothesis that cosmetics demand increases because women are substituting away from other luxury goods. Cosmetics and dresses went from being weak substitutes to strong complements during the Global Financial Crisis which lends itself to the hypothesis that women buy more cosmetics during recessions to improve workplace perceptions or for job interviews since dresses would be an item commonly purchased for that same purpose.

7 Robustness Checks

The use of survey data in this analysis raises concerns regarding the quality and accuracy of the responses provided. One concern regarding survey data is known as rotation group bias and arises from surveys where respondents will be prompted with additional questions if they respond affirmatively and will respond with less affirmative answers the next time they are interviewed or surveyed. In the case of expenditure data, the concern is that an individual reporting that they purchased an item of a certain category will then have to respond with how much they spent. Bach and Eckman (2020) analyzed the Consumer Expenditure Surveys which were the survey data used in this analysis and found that there was evidence of rotation group bias but that there was actually improved data accuracy in terms of rounded and missing amounts spent in subsequent waves of questioning. This is important to keep in mind during data analysis but will not meaningfully change results. Another concern regarding Consumption Expenditure Survey data is how representative the sample chosen is. Bee et al. (2012) found underrepresentation of the top of the income distribution and that "diary respondents are much more likely to report zero spending for a consumption category." They note that this may lead to biased and misleading results when using diary data to research inequality trends, which is not of particular relevance to the results of this analysis.

There are potential issues regarding the estimation of demand using the AIDS model that can be tested for. One of these arises from the use of the Stone Price Index, which includes w_{it},meaning that w_{it} is on both sides of the estimated expenditure function. This can cause simultaneity bias where the estimated coefficients are correlated with the error term. Eales and Unnevehr (1988) address this by using a lag of the Stone Price Index in their estimation. This replaces the price index in equation (6) with:

$$\ln(\widehat{P}_t) = \sum_i (w_{it} - 1)\ln(p_{it}) \tag{16}$$

in the estimation of the expenditure function. Repeating prior analysis using this lagged Stone Price Index does not produce dramatically different results, except for some greater variation during the Global Financial Crisis that could be from the smaller sample size[3]

To check whether the additional price variation from the inclusion of food and

[3] Results calculated using the lagged Stone Price Index are included in the data appendix.

energy in CPI created inconsistency in the estimation of elasticities of demand, the model was run again using quarter-end Core CPI as the price level variable. Expenditure elasticities, displayed in Table 7 were approximately the same, with the same exception of the Global Financial Crisis where dresses became a luxury rather than inferior, and price elasticities were generally similar[4]. The result of interest, that cosmetics became an inferior good during the Global Financial Crisis, remains both when Core CPI and the lagged Stone Price Index are used.

	Cosmetics	Jewelry	Dresses
Pre-GFC	1.19	1.57	1.39
GFC	-1.62	9.65	5.87
Post-GFC	1.02	1.90	2.09
2002-2019	0.74	1.86	1.32

Table 7: Expenditure Elasticities of Demand using Core CPI

8 Conclusions

The severity and length of the Global Financial Crisis may have impacted consumer demand beyond what would be expected from decreased income alone. This leaves room to study how demand for a variety of consumer goods changed throughout the Global Financial Crisis, in this case cosmetics were of interest. Using an Almost Ideal Demand System (AIDS), cosmetics were found to be a normal good with relatively inelastic price and expenditure elasticity of demand compared to the related goods of jewelry and dresses. This relative inelasticity alone could explain anecdotal experiences of the "Lipstick Effect" since consumers may not cut back on their consumption of cosmetics as much as would be expected if they were treated the same as luxuries like jewelry. However, expenditure elasticity of demand for cosmetics became negative during the Global Financial Crisis which is exactly what would happen if there truly were a "lipstick effect." Further research could be done using retail scanner data to see if specific cosmetics products, such as lipstick, are driving this effect or

[4]Marshallian and Hicksian price elasticities calculated using Core CPI are included in the data appendix.

if it is cosmetics overall. Additionally, retail scanner data could be used to examine whether there is substitution in terms of cosmetics brands being purchased away from brands more broadly considered to be "luxury" toward "drugstore" makeup that is generally more cost-effective.

In addition to seeing the sign on expenditure elasticity of demand for cosmetics switch during the Global Financial Crisis, there were other notable changes to demand patterns during that time. Dresses also became inferior goods, and the magnitude of expenditure elasticity of demand for jewelry increased more than fourfold. Cosmetics became a strong complement to dresses during the Global Financial Crisis, regardless of the previous relationship. Additionally, rates of substitutability or complementarity between cosmetics and jewelry or dresses became much more inelastic after the Global Financial Crisis. These large changes in the relationship between these related goods during the Global Financial Crisis warrant further examination, potentially using a model that can incorporate additional determinants of demand.

References

[1] Attanasio, Orazio, Kieran Larkin, Morten O. Ravn, and Mario Padula. 2022. "(S)Cars and the Great Recession." *Econometrica* 90 (5): 2319–2356.

[2] Bach, Ruben L., and Stephanie Eckman. 2020. "Rotation group bias in reporting of household purchases in the U.S. Consumer Expenditure Survey." *Economics Letters* 187.

[3] Bee, Adam, Bruce D. Meyer, and James X. Sullivan. 2012. "The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?" *NBER Working Paper Series*.

[4] Biswas, Shirsho, Pradeep Chintagunta, and Sanjay Dhar. 2025. "How do U.S. households change their expenditure patterns in response to income or wealth shocks? Insights from NielsenIQ Data." *Quantitative Marketing and Economics* 23: 419–445.

[5] Browning, Martin, and Thomas F. Crossley. 2000. "Luxuries Are Easier to Postpone: A Proof." *Journal of Political Economy* 108 (5).

[6] Bureau of Labor Statistics. 2019. *Consumer Expenditure Surveys PUMD Microdata Diary Survey*. Washington, DC.

[7] Bureau of Labor Statistics. n.d. *Consumer Price Index for All Urban Consumers (CPI-U)*. Washington, DC. Accessed November 8, 2025. https://data.bls.gov/dataViewer/view/timeseries/CUUR0000SA0.

[8] Buse, Adolf. 1994. "Evaluating the Linearized Almost Ideal Demand System." *American Journal of Agricultural Economics* 76 (4): 781–793.

[9] Capps, Oral Jr., Ruixin Jia, Vikas Mishra, and Macson Ogieriakhi. 2024. "A Micro-perspective Analysis of the Demand for Greek and Non-Greek Yogurt in the United States Over Calendar Years 2018 to 2020." *Journal of Agricultural and Applied Economics* 56 (2): 329–351.

[10] Chakrabarti, Rajashri, Donghoon Lee, Wilbert van der Klaauw, and Basit Zafar. 2010. "Household Debt and Saving during the 2007 Recession." *Federal Reserve Bank of New York*, October. Accessed November 10, 2025. https://www.newyorkfed.org/medialibrary/media/research/economists/chakrabarti/Households_during_the_Great_Recession_282010.pdf.

[11] Cho, Yunho, James Morley, and Aarti Singh. 2024. "Did marginal propensities to consume change with the housing boom and bust?" *Journal of Applied Econometrics* 39 (1): 174–199.

[12] Deaton, Angus, and John Muellbauer. 1980. "An Almost Ideal Demand System." *The American Economic Review* 70 (3): 312–326.

[13] Eales, James S., and Laurian J. Unnevehr. 1988. "Demand for Beef and Chicken Products: Separability and Structural Change." *American Journal of Agricultural Economics* 70 (3): 521–532.

[14] Field, Martha K., and Emilio Pagoulatos. 1997. "The Cyclical Behavior of Price Elasticity of Demand." *Southern Economic Journal* 64 (1): 118–129.

[15] Filippini, M., G. Masiero, and K. Moschetti. 2009. "Regional Consumption of Antibiotics: A Demand System Approach." *Economic Modeling* 26 (6): 1389–1397.

[16] Finch, Julia. 2001. "Lauder's Large on the Lippie Index." *The Guardian*. November 30. Accessed October 6, 2025. https://www.theguardian.com/business/2001/dec/01/6.

[17] Fleissig, Adrian R. 2021. "Estimating Elasticities of Substitution for Sin Goods." *Applied Economics* 53 (30): 3549–3561.

[18] Girapunthong, Napaporn, and Ronald W. Ward. 2003. "Demand Drivers for Fresh-Cut Flowers and Their Substitutes: An Application of Household Expenditure Allocation Models." *AgEcon Search*. July. Accessed December 11, 2025. https://ageconsearch.umn.edu/record/22178/files/sp03wa01.pdf.

[19] Green, Richard, and Julian M. Alston. 1991. "Elasticities in AIDS Models: A Clarification and Extension." *American Journal of Agricultural Economics* 73 (3): 874–875.

[20] Henningsen, Arne. n.d. "Demand Analysis with the 'Almost Ideal Demand System' in R: Package micEconAids." *The Comprehensive R Archive Network*. Accessed October 12, 2025. https://cran.r-project.org/web/packages/micEconAids/vignettes/micEconAids_vignette.pdf.

[21] Kaplan, Greg, Giovanni L. Violante, and Justin Weidner. 2014. "The Wealthy Hand-to-Mouth." *Brookings Papers on Economic Activity*: 77–153.

[22] Lehfeldt, R. A. 1914. "The Elasticity of Demand for Wheat." *The Economic Journal* 24 (94): 212–217.

[23] Li, Wenying, Chen Zhen, and Jeffrey H. Dorfman. 2020. "Modelling with Flexibility through the Business Cycle: Using a Panel Smooth Transition Model to Test for the Lipstick Effect." *Applied Economics* 52: 2697–2704.

[24] Meredith, Neil R., Anne Macy, and Amy Meredith. 2022. "Income Elasticity of Demand for Tanning Bed Usage: Evidence." *Journal of Applied Economics* 25 (1): 1156–1181.

[25] Moore, Henry Ludwell. 1922. "Elasticity of Demand and Flexibility of Prices." *Journal of the American Statistical Association* 18 (137): 8–19.

[26] Moshary, Sarah, Bradley T. Shapiro, and Sara Drango. 2025. "Preferences for Firearms." *AER: Insights* 7 (3): 340–356.

[27] National Bureau of Economic Research. 2023. "US Business Cycle Expansions and Contractions." March 14. Accessed October 12, 2025. https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions.

[28] Pashardes, Panos. 1993. "Bias in Estimating the Almost Ideal Demand System with the Stone Index Approximation." *The Economic Journal* 103 (419): 908–915.

[29] Paulin, Geoffrey D., and David L. Ferraro. 1996. "Do Expenditures Explain Income? A Study of Variables for Income Imputation." *Journal of Economic and Social Measurement* 22: 103–128.

[30] Shaikh, Sabina L., and Douglas M. Larson. 2003. "A Two-Constraint Almost Ideal Demand Model of Recreation and Donations." *The Review of Economics and Statistics* 85 (4): 953–961.

[31] The Estee Lauder Companies Inc. 2002. *Form 10-K 2002*. New York, NY: The Estee Lauder Companies Inc.

[32] West, Sarah E., and Roberton C. Williams III. 2004. "Estimates from a Consumer Demand System: Implications for the Incidence of Environmental Taxes." *Journal of Environmental Economics and Management* 47 (3): 535–558.

[33] Yang, Ruoye, Kellie Curry Raper, and J. Ross Pruitt. 2019. "The Influence of Recession and Income Strata on Consumer Demand for Protein Sources." *Applied Economics* 51: 4615–4628.

[34] Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57 (298): 348–368.

[35] Zheng, Yuqing, Chen Zhen, Daniel Dench, and James M. Nonnemaker. 2016. "U.S. Demand for Tobacco Products in a System Framework." *Health Economics* 26 (8): 1067–1086.

A Data Appendix

Table A1: Summary Statistics of Quarter-End CPI for Different Goods Categories, Pre-GFC

	Cosmetics	Jewelry	Dresses	All Items	Core
Min	165.9	125.6	96.8	178.8	189.8
Mean	170.9	134.3	107.5	192.5	199.4
Max	177.7	145.6	121.6	208.5	211.6
SD	3.7	5.1	6.4	9.6	7.0

Table A2: Summary Statistics of Quarter-End CPI for Different Goods Categories, GFC

	Cosmetics	Jewelry	Dresses	All Items	Core
Min	176.4	139.7	105.9	210.0	212.4
Mean	179.8	152.2	115.0	214.3	216.2
Max	183.5	157.1	122.6	218.8	219.3
SD	2.9	6.1	5.5	3.7	2.3

Table A3: Summary Statistics of Quarter-End CPI for Different Goods Categories, Post-GFC

	Cosmetics	Jewelry	Dresses	All Items	Core
Min	181.7	154.0	106.4	216.0	220.1
Mean	186.1	172.1	122.8	235.9	239.8
Max	190.5	183.2	137.5	256.8	264.5
SD	2.4	7.8	8.4	11.5	13.6

Table A4: Hicksian Price Elasticities, GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-0.68	0.28	-1.84
$Q_{Jewelry}$	0.06	-11.34	-3.19
$Q_{Dresses}$	-3.62	-8.65	-3.34

Table A5: Hicksian Price Elasticities, Post-GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-1.60	-0.40	-0.84
$Q_{Jewelry}$	-0.55	-4.20	-0.57
$Q_{Dresses}$	-1.87	-0.93	-0.60

Table A6: Expenditure Elasticities of Demand using Lagged Stone Price Index			
	Cosmetics	Jewelry	Dresses
Pre-GFC	1.29	2.00	1.08
GFC	-1.73	5.95	-10.41
Post-GFC	0.99	1.97	2.16
2002-2019	0.60	1.97	1.40

Table A7: Marshallian Price Elasticities using Lagged Stone Price Index, Pre-GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-1.86	0.05	0.66
$Q_{Jewelry}$	0.04	-1.18	0.01
$Q_{Dresses}$	1.67	0.03	-2.43

Table A8: Hicksian Price Elasticities using Lagged Stone Price Index, Pre-GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-1.85	0.05	0.66
$Q_{Jewelry}$	0.05	-1.17	0.01
$Q_{Dresses}$	1.67	0.04	-2.43

Table A9: Marshallian Price Elasticities using Lagged Stone Price Index, GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-1.50	6.20	-2.88
$Q_{Jewelry}$	4.16	-6.88	-7.92
$Q_{Dresses}$	-5.70	-21.93	0.33

Table A10: Hicksian Price Elasticities using Lagged Stone Price Index, GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-1.51	6.19	-2.88
$Q_{Jewelry}$	4.19	-6.84	-7.91
$Q_{Dresses}$	-5.74	-21.99	0.31

Table A11: Marshallian Price Elasticities using Lagged Stone Price Index, Post-GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-1.53	-0.48	-0.80
$Q_{Jewelry}$	-0.67	-4.30	-0.61
$Q_{Dresses}$	-1.79	-0.99	-0.54

Table A12: Marshallian Price Elasticities using Core CPI, Pre-GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-1.40	-0.32	0.55
$Q_{Jewelry}$	-0.28	-2.79	-0.45
$Q_{Dresses}$	1.38	-1.29	-1.73

Table A13: Hicksian Price Elasticities using Core CPI, Pre-GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-1.40	-0.31	0.55
$Q_{Jewelry}$	-0.28	-2.78	-0.45
$Q_{Dresses}$	1.38	-1.28	-1.73

Table A14: Marshallian Price Elasticities using Core CPI, GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-7.68	-2.06	-1.60
$Q_{Jewelry}$	-1.63	-13.61	-3.67
$Q_{Dresses}$	-3.23	-10.58	3.69

Table A15: Hicksian Price Elasticities using Core CPI, GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-7.68	-2.07	-1.60
$Q_{Jewelry}$	-1.59	-13.56	-3.65
$Q_{Dresses}$	-3.21	-10.55	3.70

Table A16: Marshallian Price Elasticities using Core CPI, Post-GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-1.57	-0.28	-0.77
$Q_{Jewelry}$	-0.38	-4.10	-0.41
$Q_{Dresses}$	-1.72	-0.68	-0.54

Table A17: Hicksian Price Elasticities using Core CPI, Post-GFC			
	$P_{Cosmetics}$	$P_{Jewelry}$	$P_{Dresses}$
$Q_{Cosmetics}$	-1.57	-0.27	-0.77
$Q_{Jewelry}$	-0.38	-4.09	-0.41
$Q_{Dresses}$	-1.71	-0.67	-0.54

Estimating Labor Impacts of Environmental Policy: How Has California's Cap-and-Trade Scheme Affected Industry-Specific Employment?

M. Lucy Pfeiffer
George Washington University

Abstract

This paper estimates the unemployment rate effect of California's Cap-and-Trade Program (CATP) using a Difference-in-Differences (DiD) approach applied to microdata from the Current Population Survey (CPS) from 2006 to 2019. Building on Yip (2018), the analysis disaggregates treatment effects by industry group and implementation timing, using both benchmark and robustness specifications. While the 2013 treatment group demonstrates no significant effect of the CATP on employment, the 2015 expansion to fuel distributors leads to a statistically significant decline in the average probability of employment for individuals working in those industries (-2.6%). Robustness checks using contiguous-county samples confirm these patterns. Additional industry-level regressions reveal heterogeneity in treatment effects, consistent with compliance cost burdens. These results suggest that labor market effects from environmental policy depend largely on sector exposure and policy design over time. **JEL Codes:** Q58, J23, C21

1 Introduction

In the 21st century, climate policy resistance in the United States has largely been motivated by perceived trade-offs between environmental goals and economic outcomes, particularly around employment effects. Cap-and-Trade Programs (CATPs) are one popular example of such environmental policy. CATPs are aimed at reducing emissions by setting a general "cap" on the maximum total emissions-level before allowing regulated entities to "trade" emissions permits through open buying and selling. These programs attempt to arrive at the socially optimal level of emissions by forcing firms to internalize the negative externalities of pollution.

California's Cap-and-Trade Program (CATP), launched in 2013 and expanded in 2015, remains one of the most ambitious environmental programs enacted at the state level. As more states seek to emulate California's approach, it is increasingly important to have a robust understanding of the ripple effects of such policies. While most studies of the CATP have focused on emissions reductions and GDP effects, little has been done to establish an impact on labor market outcomes, despite employment being a politically salient concern in debates around climate legislation.

This paper fills that gap by offering a publicly replicable Difference-in-Differences (DiD) analysis using CPS microdata to estimate how the CATP has impacted employment outcomes. By separating industries into early and late treatment groups, we are able to extrapolate differential treatment effects across groups and identify industry-specific heterogeneity. A set of robustness checks using contiguous-county controls and disaggregated treatment effects ensures results are not artifacts of sampling or state-level confounders.

The results suggest that while early-treated industries (ie. electricity generators and large industrial facilities) demonstrate no statistically significant effect, late-treated industries (ie. distributors of transportation, natural gas, and other fuels) experienced significant job losses. Specifically, the generalized regression identifies a 2.6% decrease in the probability of being employed after treatment for individuals working in late-treated industries in CA compared to other states. These findings support the view that carbon pricing can impose labor market frictions–but that such effects are uneven and depend heavily on sectoral exposure, timing, and design features like permit prices and revenue reinvestment.

2 Literature Review

Current literature on the California CATP is mostly focused on emissions and GDP effects; for example, see Mascia and Onalib (2023). Metcalf and Stock (2023) explore the labor market impacts of carbon pricing in Europe; Yip (2018) does the same with data from British Columbia. Gray, Linn, and Morgenstern (2016) use confidential data from Californian firms to find that there are industry specific job losses related to the CATP, but that study has not been replicated and it does not include net employment changes nor trends across states.

Mascia and Onalib (2023) attempt to ascertain the GDP and emissions effects of the California CATP. They use a DiD model with county-level data from the BEA–made available in 2019–to find that there is no discernible effect on GDP of the policy nor is there a short run reduction in emissions resulting from the policy. Mascia and Onalib (2023) explain this by hypothesizing that the emissions cap chosen may not have been low enough relative to the level of business-as-usual emissions. While findings show that some companies rebalanced their portfolios to shift regulated business outside the state, they do not find evidence of corresponding emissions leakage to facilities in other states. Notably, Mascia and Onalib (2023) do not decompose the treatment effects between 2013, when the policy was initially implemented and 2015, when it was expanded to include a greater number of industries.

Metcalf and Stock (2023) examine the macroeconomic impact of carbon taxes in European countries across the last 30 years. The paper uses OLS to estimate a sequence of panel data regressions which include a large number of controls, including fixed country and year effects; long run effects on GDP and unemployment are generally within one standard deviation of zero. The authors identify the possibility of a double dividend effect from reinvesting tax revenue. The paper also offers commentary on how results can be extrapolated to the United States, hypothesizing a smaller negative impact because of relatively lower marginal abatement costs for broad American environmental policies, as opposed to European policies generally applied to narrow, high cost sectors.

Yip (2018) provides the regression that this paper adopts. It uses individual-level data to estimate labor market effects associated with a carbon tax in British Columbia (BC). A DiD approach to estimate the local labor market impact of the revenue neutral carbon tax, using the Canadian Labor Force Survey (CLFS) data, which includes educational attainment, to differentiate effects across levels of skilled labor. Despite revenue neutrality, Yip (2018) finds the tax increases the unemployment rates for middle and low educated males by 1.4 and 2.4 percent respectively.

Gray, Linn, and Morgenstern (2016) look at how changes in energy prices, resulting from the California cap-and-trade program, affect output, employment, and value added, indicating leakages from the policy. The paper forecasts the effect of an assumed increase in compliance costs associated with the cap and trade scheme after using a regression to model linkages, finding that short-run effects are somewhat significant but become minimal in the long-run, defined as 5 years.

3 Theoretical Framework

Cap and trade programs (CATPs) establish a limit on the total amount of emissions allowed and allocate those emissions across firms. In the case of California's CATP, permits that allow for a certain amount of emissions are auctioned off to firms for a given year; revenue is then diverted into climate-related programs and investments.

During the initial phase of implementation of the policy, beginning on January 1, 2013, the CATP applied only to electricity generators and large industrial facilities emitting 25,000 $MTCO_2e$ or more annually. Firms which fall into this category are thus part of the early treatment group. On January 1st, 2015, the emissions cap was extended to include distributors of transportation, natural gas, and other fuels, making that group the late treatment group.

As explored by Mascia and Onalib (2023), the price of allowances may have initially been set too low to see significant effects in emissions reductions or GDP growth. However, as the prices of allowances rise and the total cap decreases, employment effects would be likely to increase as a result of the increasing burden of compliance costs borne by firms.

The price of allowances was in the range US $10-14 in the period from 2013 to 2015, after which it increased continuously, reaching $27 in 2021. Thus, the regression undertaken in this paper, which will separate the employment effects for early and late treatment groups, may also capture the effect of a better-designed program. In other words, by 2015, the program had been adjusted such that the emissions cap was below business-as-usual emissions, making it more effective at reducing total emissions and therefore potentially exhibiting a more significant effect on employment than would be seen after initial treatment. Specifically, the emissions cap has been continuously adjusted downward at a rate of around 3-4% per year each year since the cap was expanded in 2015 because of the addition of the fuel distribution industry.

3.1 Justification of Controls and Fixed Effects

This analysis includes both individual-level control variables–age, educational attainment, sex, citizenship, and marital status–and fixed effects. Controlling for age helps distinguish policy effects from demographic shifts; likewise, including educational attainment ensures the reported effect is not confounded by changes in the composition of the workforce. Controlling for sex makes sure that gender-based labor market dynamics are not influencing the output; this is important because gender differences can influence employment outcomes, particularly in industries differentially exposed to cap and trade.

State fixed effects (STATEFIP) controls for time-invariant differences across states, such as industrial structure, baseline employment levels, and other institutional factors. Including STATEFIP accounts for persistent differences between California and control states that might otherwise indicate biased estimates. The analysis also includes year fixed effects (YEAR). These absorb national economic shocks and time trends that affect all states equally, such as recessions or changes in federal policy. By holding year constant, the analysis separates national labor market changes from those induced by California's policy. IND controls for industry fixed effects, ensuring that we control for systematic differences in employment levels across industries that are not due to the policy treatment.

Taken together, these controls and fixed effects help reduce omitted variable bias to create an unbiased estimate of the true causal relationship between employment and the CATP. The remaining variation which is then used to identify the treatment effect thus comes from within-state changes over time, relative to trends in untreated states, embodying the core logic of the Difference-in-Differences model framework employed by this paper.

3.2 Model Outline

Generalized Treatment Effects (pooled by treatment timing)

$$\begin{aligned} Y_{ist} = \beta_0 + \beta_1(earlyIND_i \cdot Post2013_t \cdot CA_s) \\ + \beta_2(lateIND_i \cdot Post2015_t \cdot CA_s) + X'_{ist}\delta \\ + \gamma_{ind} + \delta_{year} + \theta_{state} + \varepsilon_{ist} \end{aligned} \tag{1}$$

Industry-specific Late Treatment Effects

$$\begin{aligned} Y_{ist} = \beta_0 + \sum_{j \in LateTreat} \beta_{2k}\left(IND_{ik} \cdot Post2015_t \cdot CA_s\right) \\ + X'_{ist}\delta + \gamma_{ind} + \delta_{year} + \theta_{state} + \varepsilon_{ist} \end{aligned} \tag{2}$$

Industry-specific Early Treatment Effects

$$\begin{aligned} Y_{ist} = \beta_0 + \sum_{k \in EarlyTreat} \beta_{1j}\left(IND_{ij} \cdot Post2013_t \cdot CA_s\right) \\ + X'_{ist}\delta + \gamma_{ind} + \delta_{year} + \theta_{state} + \varepsilon_{ist} \end{aligned} \tag{3}$$

Robustness Check (County-Level)

$$\begin{aligned} Y_{isct} = \beta_0 + \beta_1\left(earlyIND_i \cdot Post2013_t \cdot CA_{sc}\right) \\ + \beta_2\left(lateIND_i \cdot Post2015_t \cdot CA_{sc}\right) \\ + X'_{isct}\delta + \gamma_{ind} + \delta_{year} + \theta_{state} + \varepsilon_{icst} \end{aligned} \tag{4}$$

Y_{ist}	Dummy variable equal to 1 if the survey respondent is employed, 0 otherwise (industry *i* in state *s* and year *t*)
$earlyIND_i$	Equals 1 if industry i is in the early-treated group, 0 otherwise (IND codes = 380, 470, 2070, 2170, 2180, 2190, 2270, 2280, 2290, 2370, 2390, 2870, 3080)
$lateIND_i$	Equals 1 if industry *i* is in the late-treated group, 0 otherwise (IND codes = 370, 4490, 5680, 587, 580, 590)
IND_{ij}	Equals 1 if the survey response is associated with a given early-treated industry *j*, 0 otherwise
IND_{ik}	Equals 1 if the survey response is associated with a given late-treated industry *k*, 0 otherwise
$Post2013_t$	Dummy variable, equals 1 if year $t \geq 2013$, 0 otherwise
$Post2015_t$	Dummy variable, equals 1 if year $t \geq 2015$, 0 otherwise
CA_s	Dummy variable, equals 1 if the survey respondent works in California (state FIPS = 6), 0 otherwise
X_{ist}	Vector of individual-level control variables for the survey respondent, which specifically includes: age of the respondent, educational attainment (categorical), sex (binary), marital status (categorical), and citizenship status (categorical)
γ_{ind}	Industry fixed effects controlling for time-invariant differences between industries
δ_{year}	Year fixed effects capturing nationwide trends affecting all individuals in a given year
θ_{state}	State fixed effects controlling for time-invariant differences between states
ε_{ist}	Error term

Table 1: Variable Descriptions

In order to isolate the employment effects of California's CATP, we first use a difference-in-differences framework that breaks down treatment effects by late and early treatment groups. The dependent variable is a binary indicator for whether the individual survey respondent associated with industry i in state s and year t is employed. The relevant independent variables are the interactions between industry indicators for early (and late) periods and a set of treatment dummy variables equal to one for observations taken in California during years 2013 or later (2015 or later) and equal to zero otherwise.

Because CPS is a repeated cross section and the same individuals are not followed across time, the coefficients on the interaction terms represent group-level probability change rather than individual-level change. In other words, the coefficients will represent the change in the probability that a person in the population who belongs to a certain group (e.g., industry $\times$ state $\times$ year) is employed, on average. Assuming representative sampling, that change should be an estimate of the population change in employment within a certain industry.

To estimate heterogeneous treatment effects by industry, we run two subsequent regressions: one for late-treated industries (2) and one for early-treated industries (3). In each case, we restrict the sample to include only those industries exposed to the relevant treatment and never-treated industries. This ensures that treatment effects are identified relative to a clean control group unaffected by a previous or subsequent phase of the CATP.

Allowing for variable employment effects across industries allows users to interpret unique industry factors that may contribute to particularly large effects given that numerous variables such as compliance costs and exposure to emissions regulations differ widely across sectors. This is why, for example, we might anticipate seeing a negative employment effect in manufacturing–an emissions intensive industry–while a simultaneous positive employment effect is identified in a sector where revenue from the CATP is diverted, indicating green job creation.

Each model also includes individual-level controls for age, education level, sex, marital status, and citizenship in an attempt to account for standard determinants of employment. We also include state, year, and industry fixed effects to control for unobserved heterogeneity across states, national, and industry-based trends, respectively. As a robustness measure, the regression is weighted using CPS sampling weights (WTFINL) to ensure representative estimates. Standard errors are also clustered at the state level to account for within-state serial correlation.

3.3 Parallel Trends Test

This model identifies the causal effect of the CATP on employment under the assumption of parallel trends–that in the absence of the policy, employment outcomes in California would have evolved similarly to those in the control states. Mirroring the analysis carried out by Mascia and Onalib (2023), we verify parallel trends in GDP growth and emissions across California, Oregon, Arizona, and Nevada prior to the treatment period. Preliminary results from an event study point to no significant pretreatment deviations when trends are specified at the industry level. However, due to restrictions which are further discussed in the caveat section, this is an important point for further research to ensure robustness of results when interpreting industry-level coefficients.

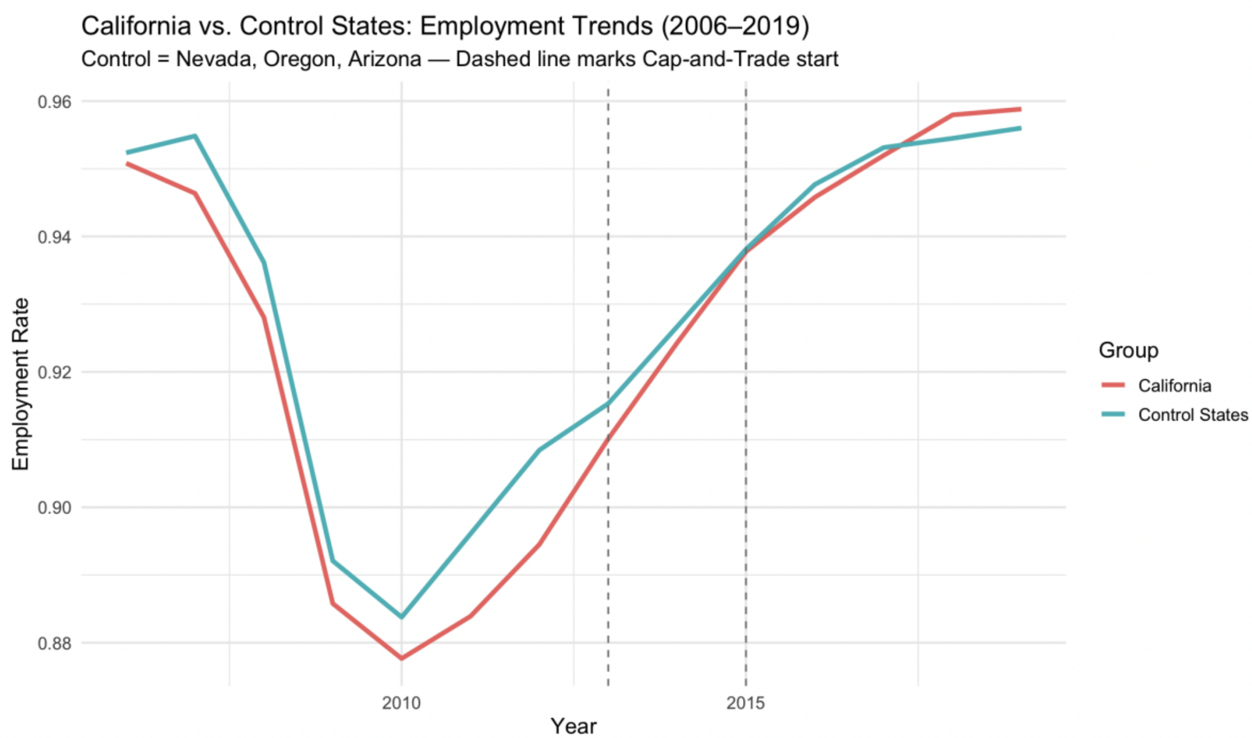

Figure 1: Parallel Trends Test (California vs. Aggregated Control States)

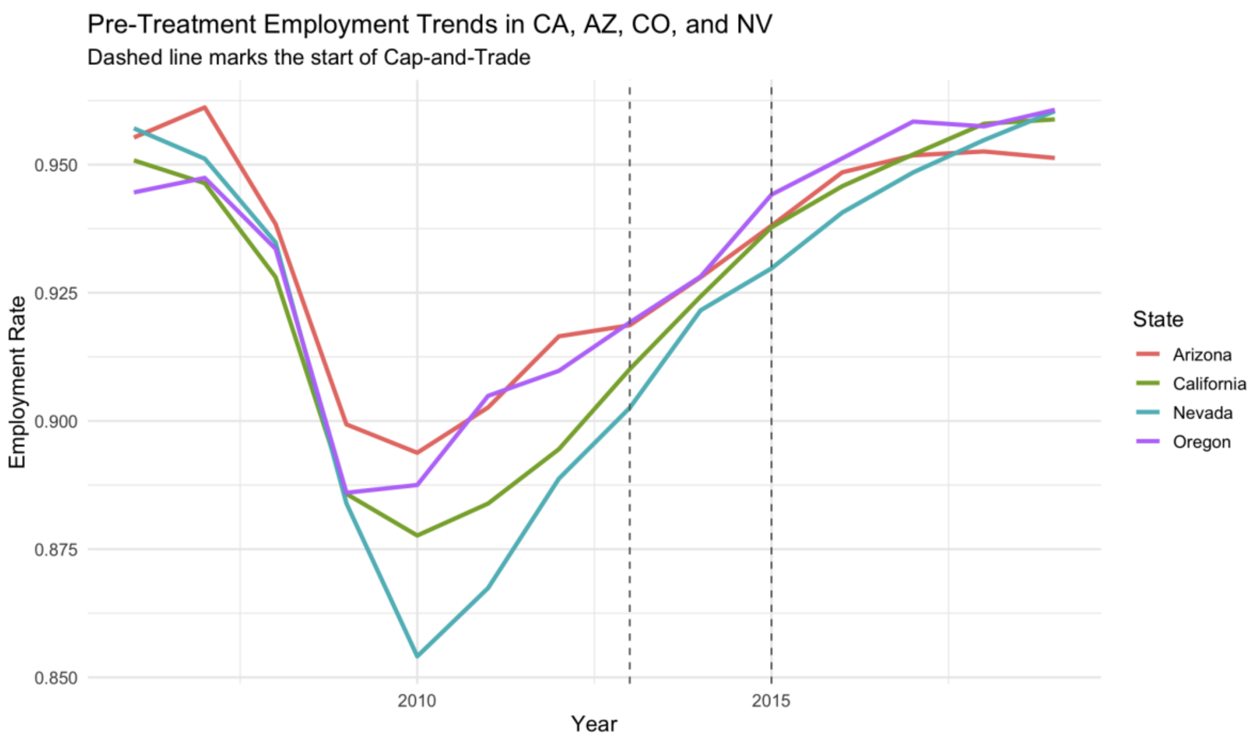

Figure 2: Parallel Trends Test (California vs. Individual Control States)

4 Data Descriptions

The model uses a dataset that consists of individual-level observations from the Current Population Survey (CPS) between the years 2006 to 2019. This data is collected as a monthly repeated cross-section, meaning each observation corresponds to a unique individual surveyed during a particular month, but individuals are not followed over time. Unlike a panel dataset, where the same respondents are tracked longitudinally, a repeated cross-section enables analysis of population-level trends across time periods without individual-level continuity. The CPS interviews a new sample of individuals each month. To facilitate policy evaluation, annual indicators are constructed by aggregating across these monthly samples. Although finer monthly variation is available in the data, we rely on annual aggregation to maintain consistency with the timing of California's Cap-and-Trade implementation and to avoid overfitting to short-run fluctuations.

Because the treatment began in January of 2013, we chose the period to be a symmetrical 7 year period before and after the treatment. This conveniently ends the analysis in 2019, allowing us to forego adjustments that would be necessary if we incorporated data overlapping with the COVID-19 pandemic period. Likewise, as identified by Mascia and Onalib (2023), there was a spike in auction prices for allocating emissions permits that occurred in 2021; consequently, we decide to limit the data to a period of 7 years after treatment which will allow this paper to make limited commentary on long-run effects, supplementing the analysis in Gray, Linn, and Morgenstern (2016) which indicates that long run impacts of the policy are likely to dissipate by 5 years.

We also filter the initial data set to include only those observations in California or contiguous states (Nevada, Arizona, Oregon). This approach is modeled from the analysis carried out in Mascia and Onalib (2023), where they test several levels of analysis in order to discern where the common trends assumption is fulfilled for an analysis on emissions level and GDP growth. Their analysis concludes that analyses undertaken with both contiguous counties and contiguous states yield similar results because of their geographic proximity; on the other hand, an analysis including all national observations as controls against observations in California is significantly different, yielding the conclusion that contiguous states or contiguous counties are the preferred control needed to satisfy the common trends assumption for a DiD analysis in this case. Although Mascia and Onalib (2023) looks at GDP growth and emissions as opposed to employment, replicating their common trends analysis yields similar results, as demonstrated by Figures 1 and 2.

Note also that this geographic analysis uses locations based on place of work state instead of place of residence state. This method is preferred, because it is a labor market analysis and regulation will affect individuals based on the location of their job and not that of their residence. Similarly, we opt to filter out any observations from an industry with less than 100 total observations. This is because the causal effects identified in industries with fewer than 100 total observations are not reliable with small changes in total number indicating large (and unrepresentative) swings in percent change in employment.

The IND variable is taken from a specific subset of questions on the Current Population Survey–namely questions 42 b through d. The responses to those questions are then coded as follow: for employed individuals, IND reflects the respondent's main job during the reference week; for unemployed individuals (and those not in the labor force but who have held a job within the past five years), IND refers to the most recent job they held; for individuals who have never worked, or whose last job was more than five years ago, IND is typically coded as missing or zero, (such individuals are excluded from this analysis). The coding of the IND variable with past industry association for unemployed individuals is key to capturing variation in the left-hand-side variable for this analysis.

After filtering the dataset explicitly to include only those individuals in the labor force, the final sample includes approximately 1.4 million weighted observations. The key dependent variable is a binary indicator of employment status, which takes a value of 1 if the respondent is employed and 0 otherwise. In the sample, the overall employment rate is approximately 90%. The average respondent is 41.6 years old, with ages ranging from 15 to 85.

Educational attainment is captured using a recoded categorical variable which ranges from 2 to 125, with a mean of 86.3. Values equal to or greater than 70 correspond to individuals who have achieved at least a 12th grade education, with values greater than or equal to 110 corresponding to at least 4 years of post-secondary education. Observations where educational attainment is equal to 2 are used for the reference group in order to avoid the dummy variable trap; a code of 2 corresponds to individuals with no education or those with only a preschool or kindergarten level. Each of the 16 categorical education variables used is also defined explicitly alongside the corresponding coefficient in the data appendix where results are presented.

The sample is roughly balanced by sex (mean = 1.5, where 1= male and 2=female). Observations where sex is coded as male are used as the reference group for the purpose of avoiding the dummy variable trap. Marital status is also included as a categorical variable and is coded from 1-9 where 1 = married, spouse present; 2 = married, spouse absent; 3 = separated; 4 = divorced; 5 = widowed; 6 = never married, single; 7 = widowed or divorced; and 9 = NIU. Observations where marital status is equal to 7 and 9 are dropped. Observations where marital status is equal to 1 are used as the reference group for the purpose of avoiding the dummy variable trap.

Similarly, a categorical citizenship status control is coded from 1-9 where 1 = born in US; 2 = born in US, outlying; 3 = born abroad of American parents; 4 = naturalized citizen; 5 = not a citizen; and 9 = NIU. Observations where citizenship status is equal to 9 are dropped. Observations where citizenship status is equal to 1 are used as the reference group for the purpose of avoiding the dummy variable trap.

The CPS-provided sampling weight (WTFINL) is used to improve external validity, with an average weight of approximately 3,029. Industry codes are drawn from the IPUMS CPS variable IND, which remains largely consistent across the 2003-2019 period. Minor differences in classification–such as dropped or reworded labels for 'not specified' categories–do not impact the core analysis.

Variable	Unique Values	Missing (%)	Mean	SD	Min	Median	Max
Employed	2	0	0.9	0.3	0.0	1.0	1.0
Age	67	0	41.6	14.0	15.0	41.0	85.0
Educational Attainment (Recode)	16	0	86.3	25.7	2.0	81.0	125.0
Sex	2	0	1.5	0.5	1.0	1.0	2.0
Marital Status	6	0	3.0	2.3	1.0	1.0	6.0
Citizenship Status	5	0	2.0	1.6	1.0	1.0	5.0
Final Basic Weight (WTFINL)	753680	0	3029.4	945.8	0.0	3129.2	12507.8

Sample period: monthly (2006 - 2019)
Total observations: 1,405,806

Table 2: Basic Summary Statistics

Finally, this paper separates industries into three groups: early treatment, late treatment, and never treated. The early treatment group begins in 2013 and includes industries which are electricity generators or consistently include large industrial facilities emitting 25,000 $MTCO_2e$ or more annually. Specifically, these are: coal mining (380); nonmetallic mineral mining and quarrying (470); petroleum refining (2070); resin, synthetic rubber and fibers, and filaments manufacturing (2170); agricultural chemical manufacturing (2180); pharmaceutical and medicine manufacturing (2190); paint, coating, and adhesives manufacturing (2270); soap, cleaning compound, and cosmetic manufacturing (2280); industrial and miscellaneous chemicals (2290); plastics product manufacturing (2370); rubber products, except tires, manufacturing (2390); structural metals and tank and shipping container manufacturing (2870); construction mining and oil field machinery manufacturing (3080). The late treatment group begins in 2015 and includes industries which are distributors of transportation, natural gas, and other fuels. Specifically, these are: oil and gas extraction (370); petroleum and petroleum product wholesalers (4490); fuel dealers (5680); electric power generation, transmission and distribution (587); natural gas distribution (580); electric and gas, and other combinations (590). The numbers in parentheses after the industry specify the industry code used as identification in the data set. Industry-specific effects reported in the data appendix are identified using these codes.

5 Results

5.1 Analysis

Results are presented at length in Section VII. Data Appendix. We begin by estimating a DiD model that measures the average employment effect of California's CATP on early and late treated industries. The specification includes individual-level controls (age, education, sex, marital status, citizenship) and fixed effects for industry, year, and state. The results of that regression are shown in Column (1) of *Table 3*. The coefficient on the early treatment interaction term (β_1) is positive but not statistically significant (0.009, SE = 0.004), suggesting no meaningful change in employment for early-treated industries after the 2013 rollout. By contrast, the coefficient on the late treatment interaction term (β_2) is negative and highly significant (-0.026, SE = 0.001, ***), indicating that late-treated industries experience a statistically significant reduction in employment following the 2015 expansion of the program. While the adjusted R^2 for the benchmark regression is modest, this is consistent with prior research using individual-level employment data, where outcomes are influenced by many unobserved factors.

These findings are partially consistent with the theory that compliance costs imposed by the CATP may lead to a contraction in labor demand, especially among downstream firms. The lack of an effect for early-treated industries is easily explained by the imposition of an initial emissions cap at a level that was too far above business-as-usual emissions to have a significant effect on emissions, yielding zero compliance costs. However, the effect could also reflect a stronger capacity to absorb regulatory costs, longer adjustment periods, or stronger pre-treatment anticipatory effects among firms in the early treatment group.

5.2 Robustness Checks

Next, we restrict the sample to bordering counties within California and its neighboring states (Arizona, Nevada, and Oregon) in order to conduct a robustness check to test whether our benchmark results are driven by broader regional or geographic differences. While this does not eliminate state-level heterogeneity, it provides a more geographically concentrated comparison group and helps control for local labor market dynamics and spatial spillovers. The idea is to compare employment outcomes in regions that are physically and economically adjacent, thereby reducing concerns about differential trends across more distant areas within each state.

Column (4) of *Table 3* presents the results of that robustness check. The early treatment coefficient remains positive and statistically insignificant (0.013, SE = 0.007), consistent with the benchmark. The late treatment coefficient is also still negative, statistically significant, and similar in magnitude to the benchmark estimate (-0.037, SE = 0.002, **). By narrowing the sample to include only those counties which are geographically adjacent to California (on both sides of the state border), this robustness check limits the comparison to regions that are more likely to share labor market characteristics, reducing concerns about broader economic or policy differences driving results. The persistence of the treatment effect among late-treated industries in this more concentrated sample ultimately supports the validity of the benchmark findings.

5.3 Industry-Specific Effects

Finally, columns (2) and (3) of *Table 3* report the industry-specific treatment effects for late and early treated industries, respectively. These regressions replace the average treatment coefficients with parameters on interaction terms between an individually specified industry and treatment timing, allowing for heterogeneity in the policy's impact across sectors.

For the industries exposed to late treatment in 2015 during the expansion of the CATP, several coefficients are large and statistically significant in column (2). For instance, petroleum and petroleum product wholesalers (4490) shows a negative and highly significant employment effect (-0.027, SE = 0.002, ***), consistent with expectations given its direct exposure to compliance costs. In contrast, natural gas distribution (580) and electric and gas, and other combinations (590) exhibit positive and significant coefficients (0.014, SE = 0.001, **) and (0.018, SE = 0.002, **) respectively. This could suggest potential short-run adjustments in staffing or changes in local fuel distribution dynamics; alternatively, it could be a symptom of resilience to increasing compliance costs among consumer-oriented utilities. Other industries, like fuel dealers (5680), also show significant declines, consistent with cost pass-throughs or demand effects (-0.035, SE = 0.002, ***).

Column (3) presents the analogous estimates for early treated industries, subject to regulation during the initial policy implementation phase in 2013. The results here are more mixed, as we expect based on the statistically insignificant results on TreatEarly in the benchmark regression. Some industries do show positive and significant employment effects (380, 2180) while others exhibit large, negative effects (2270, 2170). These patterns could suggest substantial heterogeneity within the early-treatment group, possibly reflecting differences in capital intensity, ability to abate emissions, or exposure to trade competition.

Importantly, the inclusion of industry fixed effects and individual-level controls ensures that each of these interaction coefficients reflects the difference-in-differences estimate for a specific treated industry, comparing the change in employment in that industry in California post-treatment to the corresponding change in the same industry in control states and relative to never-treated industries. Also note that for these models, a consistent reference group was used, composed of all industries in the never-treated group.

5.4 Caveat

There are six key considerations related to limitations and caveats for the analysis carried out in this paper: leakage effects, anticipatory effects, confounding effects from macroeconomic shocks, heterogeneous treatment timing and intensity, possible endogeneity issues, and lack of revenue neutrality.

First, consider the leakage concern common to spatially defined DiD studies: where the observed employment effect includes both the decrease in California employment and the increase in external employment, the estimate can serve only as a lower bound for the true impact of the policy. To address this concern, we mimic the strategy used by Mascia and Onalib (2023), narrowing the analysis to contiguous counties where labor demand is most likely to be displaced rather than destroyed. The results remain qualitatively consistent with the benchmark, suggesting that the main findings are not driven by job displacement to nearby labor markets. Still, without firm-level relocation data, this leakage effect cannot be directly tested and may bias the estimates.

Second, if there is a significant anticipatory effect, the estimates would represent a lower bound for the true effect. This analysis assumes no significant anticipatory effects in response to the policy; however, Masica and Onalib (2023) suggest that carbon markets can induce firm-level adjustments even before initial implementation, either through expectations, capital planning, or advance price signaling. Preliminary evidence from event study plots does not clearly reject the possibility of such anticipatory behavior, but also does not suggest that it is significant if present. Potential robustness checks for future research might include re-estimating the model with treatment set to begin at the time of policy announcement, 2012, to account for potential pre-policy reactions. For the sake of the analysis carried out in this paper, it is assumed that these effects were not large enough to significantly distort estimates, not least because the high level of the initial emissions cap stoked expectations of small or zero abatement costs for firms.

Third, the study window includes the period following the Great Recession (2008-2009), and although fixed effects for year and state help account for national and regional shocks, some residual influence of macroeconomic volatility could nonetheless bias the estimates.

Fourth, there is the issue of heterogeneous treatment timing and intensity across all industries. Although the benchmark regression attempts to build on the Mascia and Onalib (2023) analysis by decomposing the effect into early and late treatment groups, the reality of industry treatment is more complex. Industries within the treatment group still vary in their exposure to emissions pricing depending on emissions volume, compliance requirements, and abatement capacity. These differences account for the variation in industry-specific treatment effects reported in columns (2) and (3) of *Table 3*. For example, while petroleum refineries faced heavy upfront costs at the beginning of policy implementation, smaller fuel distributors were phased in later with comparatively lighter regulation. The industry-specific regressions are aimed at addressing some of this variation, but it may violate the "sharp treatment" assumption made by standard DiD models. Future research could incorporate continuous measures of policy intensity, such as emissions per worker or facility-level permit coverage, in order to estimate heterogeneous treatment effects more precisely.

Fifth, there may be endogeneity issues associated with treatment timing if industry selection correlated with unobservable variables. For instance, perhaps regulators anticipated that early-treated industries would be better able to absorb compliance burdens or tolerate employment frictions, while later-treated industries were more politically sensitive or less flexible with regards to cost-pass-through ability. In reality, it appears that the selection of industries into the early and late treatment categories is a relatively exogenous process given that direct emissions volumes and the state's ability to monitor output were primary determinants. While the phased rollout may raise concerns about differential selection, treatment timing was determined by the emissions source type and infrastructure readiness, rather than labor market expectations. Nevertheless, we conduct industry-specific regressions and robustness checks to account for potential differences in exposure intensity and regulatory burden.

Finally, there is the issue of revenue neutrality. As opposed to the policy analyzed by Yip (2018), the California CATP is not revenue neutral. Instead, revenues from permit auctions are reinvested in climate-positive sectors. The observed net effect on employment may therefore understate or obscure distributional consequences across sectors and regions. A more nuanced analysis would require disaggregated data on where and how revenues were spent, or a triple differences framework comparing employment in green vs. non-green industries across treated and control regions. For the purpose of this paper, that approach was precluded by the lack of a distinct industry wherein revenue investment was concentrated.

Despite the need to thoroughly consider these limitations, the weaknesses in the model are not fatal to the central argument. Indeed, they reflect the unavoidable complexity of evaluating real-world environmental policy using observational data.

6 Conclusion

This paper's attempt to estimate the labor market effects of environmental policy in California is motivated by the consistent reliance on arguments related to employment consequences in climate policy debates. The California CATP is a leading model for state-level carbon pricing, yet its impacts remain understudied, with labor market impacts virtually untested. This paper fills that gap using a DiD framework leveraging public CPS microdata from IPUMS.

The benchmark regression shows no significant effect on employment for industries treated during the initial phase of policy rollout, but a statistically significant, negative effect (-2.6robustness specification that restricts the sample to counties geographically contiguous with the California state border.

Regressions (2) and (3) allow for identification of heterogeneous changes in by industry. The findings suggest that labor market effects from carbon pricing are not uniformly distributed across the economy. Instead, they depend on the compliance cost burden borne by different sectors, the stringency of the emissions cap, and perhaps most importantly, the timing and evolution of the policy itself. These findings offer practical insights for states planning to replicate the CATP in the future, including considerations around policy implementation timing and best use of revenues. Ultimately, we conclude that policymakers should weigh distributional labor market impacts when designing emissions regulations.

Future research could benefit from the inclusion of post-2019 data in order to conduct a robust, long-run analysis. More granular geographic controls or firm-level employment data could also help to better disentangle within-industry heterogeneity. Additionally, future work could incorporate measures of green job creation, permit price variation, or continuous treatment measures. Overall, the evidence presented herein serves to establish a statistically significant short run decline in the average probability of employment for individuals working in treated industries when CATP is implemented with a sufficiently low emissions cap.

References

[1] California Air Resources Board. 2015. *Cap-and-Trade Program Overview*. Sacramento, CA: California Environmental Protection Agency. https://ww2.arb.ca.gov/sites/default/files/cap-and-trade/guidance/cap_trade_overview.pdf.

[2] Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler, and Michael Westberry. 2024. *IPUMS CPS: Version 12.0 [dataset]*. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D030.V12.0.

[3] Gray, Wayne B., Joshua Linn, and Richard D. Morgenstern. 2016. *Employment and Output Leakage Under California's Cap-And-Trade Program*. Resources for the Future Discussion Paper.

[4] International Carbon Action Partnership (ICAP). 2024. *USA - California Cap-and-Trade Program*. Last modified January 15, 2024. https://icapcarbonaction.com/en/ets/usa-california-cap-and-trade-program.

[5] Mascia, D. V., and E. Onali. 2023. "Keep Calm and Carry On Emitting: Cap-and-Trade Rules, Local Emissions and Growth." *Regional Studies* 58 (1): 220–237. https://doi.org/10.1080/00343404.2023.2194315.

[6] Metcalf, Gilbert E., and James H. Stock. 2023. *The Macroeconomic Impact of Europe's Carbon Taxes*. MIT Center for Energy and Environmental Policy Research Working Paper Series.

[7] Yip, Chi Man. 2018. "On the Labor Market Consequences of Environmental Taxes." *Journal of Environmental Economics and Management* 89 (May): 136–152. https://doi.org/10.1016/j.jeem.2018.03.004.

A Data Appendix

Note: The following pages comprise the entirety of Table 3

(1) $Y_{ist} = \beta_0 + \beta_1 \cdot (earlyIND_i \cdot Post2013_t \cdot CA_s) + \beta_2 \cdot (lateIND_i \cdot Post2015_t \cdot CA_s) + X_{ist}'\delta + \gamma_{ind} + \delta_{year} + \theta_{state} + \varepsilon_{ist}$

(2) $Y_{ist} = \beta_0 + \sum_{j\epsilon LateTreat} \beta_{2k}(IND_{ik} \cdot Post2015_t \cdot CA_s) + X_{ist}'\delta + \gamma_{ind} + \delta_{year} + \theta_{state} + \varepsilon_{ist}$

(3) $Y_{ist} = \beta_0 + \sum_{j\epsilon EarlyTreat} \beta_{1j}(IND_{ij} \cdot Post2013_t \cdot CA_s) + X_{ist}'\delta + \gamma_{ind} + \delta_{year} + \theta_{state} + \varepsilon_{ist}$

(4) $Y_{isct} = \beta_0 + \beta_1 \cdot (earlyIND_i \cdot Post2013_t \cdot CA_{sc}) + \beta_2 \cdot (lateIND_i \cdot Post2015_t \cdot CA_{sc}) + X_{isct}'\delta + \gamma_{ind} + \delta_{year} + \theta_{state} + \varepsilon_{icst}$

Model Term	(1) Staggered DiD	(2) Late Industries vs Untreated	(3) Early Industries vs Untreated	(4) Contiguous County Sample
β_1 early-IND x Post-2013 x CA	0.009 (0.004)	—	—	0.013 (0.007)
β_2	-0.026***	—	—	-0.037**

late-IND x Post-2015 x CA	(0.001)			(0.002)
β_{2k} **k = IND 370** (oil and gas extraction)	—	-0.002 (0.001)	—	—
β_{2k} **k = IND 4490** (petroleum and petroleum product wholesalers)	—	-0.027*** (0.002)	—	—
β_{2k} **k = IND 570** (fuel dealers)	—	-0.003 (0.002)	—	—
β_{2k} **k = IND 580** (natural gas distribution)	—	0.014** (0.001)	—	—
β_{2k} **k = IND 590** (electric and gas)	—	0.018* (0.002)	—	—
β_{2k} **k = IND 5680** (fuel dealers)	—	-0.035*** (0.002)	—	—
β_{1j} **j = IND 380** (coal mining)	—	—	0.124*** (0.001)	—
β_{1j} **j = IND 470** (nonmetallic mineral mining and quarrying)	—	—	0.050*** (0.001)	—
β_{1j} **j = IND 2070** (petroleum refining)	—	—	0.020*** (0.001)	—
β_{1j} **j = IND 2170** (resin, synthetic rubber and fibers, and filaments manufacturing)	—	—	-0.061*** (0.001)	—
β_{1j} **j = IND 2190**	—	—	0.002 (0.001)	—

(pharmaceutical and medicine manufacturing)				
β_{1j} **j = IND 2270** (paint, coating, and adhesives manufacturing)	—	—	-0.112*** (0.001)	—
β_{1j} **j = IND 2280** (soap, cleaning compound, and cosmetic manufacturing)	—	—	0.033*** (0.001)	—
β_{1j} **j = IND 2290** (industrial and miscellaneous chemicals)	—	—	-0.006* (0.001)	—
β_{1j} **j = IND 2180** (agricultural chemical manufacturing)	—	—	0.040*** (0.002)	—
β_{1j} **j = IND 2370** (plastics product manufacturing)	—	—	0.036*** (0.001)	—
β_{1j} **j = IND 2390** (rubber products, except tires, manufacturing)	—	—	0.055*** (0.001)	—
β_{1j} **j = IND 2870** (structural metals and tank and shipping container manufacturing)	—	—	0.012** (0.001)	—
β_{1j} **j = IND 3080** (construction mining and oil field machinery manufacturing)	—	—	-0.023*** (0.001)	—
AGE	0.000** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001* (0.000)

EDUC = 10 (grades 1, 2, 3, or 4)	0.013 (0.007)	0.017 (0.007)	0.016 (0.007)	-0.023 (0.032)
EDUC = 20 (grades 5 or 6)	0.023* (0.006)	0.030* (0.007)	0.030* (0.007)	0.011 (0.027)
EDUC = 30 (grades 7 or 8)	0.011 (0.009)	0.017 (0.012)	0.017 (0.012)	0.006 (0.029)
EDUC = 40 (grade 9)	0.015 (0.009)	0.012 (0.012)	0.012 (0.012)	-0.029 (0.036)
EDUC = 50 (grade 10)	-0.008 (0.011)	-0.037+ (0.013)	-0.036+ (0.013)	-0.022 (0.024)
EDUC = 60 (grade 11)	-0.021* (0.006)	-0.039** (0.005)	-0.038** (0.005)	-0.056 (0.024)
EDUC = 71 (grade 12, no diploma)	0.001 (0.007)	0.004 (0.007)	0.005 (0.008)	-0.034 (0.040)
EDUC = 73 (high school diploma or equivalent)	0.024* (0.007)	0.043* (0.008)	0.043* (0.008)	0.014 (0.029)
EDUC = 81 (some college, no degree)	0.035* (0.006)	0.062** (0.007)	0.062** (0.008)	0.028 (0.031)
EDUC = 91 (Associate's degree, occupational/vocational program)	0.038* (0.007)	0.068** (0.008)	0.068** (0.008)	0.030 (0.031)
EDUC = 92 (Associate's degree, academic program)	0.044** (0.006)	0.075** (0.007)	0.075** (0.008)	0.046 (0.035)
EDUC = 111 (Bachelor's degree)	0.053** (0.007)	0.084** (0.008)	0.084** (0.008)	0.053 (0.031)
EDUC = 123 (Master's degree)	0.055** (0.009)	0.087** (0.009)	0.087** (0.009)	0.060 (0.031)
EDUC = 124 (Professional school degree)	0.059** (0.007)	0.100** (0.009)	0.100** (0.009)	0.057 (0.028)
EDUC = 125 (Doctoral degree)	0.057** (0.007)	0.093** (0.009)	0.094** (0.009)	0.064 (0.028)

SEX = 2 (Female)	-0.004+ (0.002)	0.002 (0.001)	0.002 (0.001)	-0.005 (0.002)
MARST = 2 (married, spouse absent)	-0.019*** (0.001)	-0.019** (0.002)	-0.020** (0.002)	-0.020+ (0.005)
MARST = 3 (seperated)	-0.037*** (0.003)	-0.038** (0.003)	-0.038** (0.003)	-0.042* (0.009)
MARST = 4 (divorced)	-0.025*** (0.002)	-0.027*** (0.002)	-0.026** (0.002)	-0.031* (0.005)
MARST = 5 (widowed)	-0.023** (0.002)	-0.029** (0.003)	-0.029** (0.003)	-0.056+ (0.013)
MARST = 6 (never married/ single)	-0.031*** (0.001)	-0.041*** (0.001)	-0.041*** (0.001)	-0.035*** (0.001)
CITIZEN = 2 (born in US, outlying)	-0.018 (0.008)	-0.013 (0.009)	-0.013 (0.009)	0.004 (0.004)
CITIZEN = 3 (born abroad of American parents)	0.005* (0.001)	0.007* (0.001)	0.006* (0.002)	-0.005+ (0.002)
CITIZEN = 4 (naturalized citizen)	0.008** (0.001)	0.011*** (0.001)	0.011*** (0.001)	0.019** (0.001)
CITIZEN = 5 (not a citizen)	0.012*** (0.000)	0.012*** (0.001)	0.012*** (0.001)	0.021+ (0.007)
Number of Observations	1,398,383	1,383,194	1,390,412	88,255
R^2	0.134	0.036	0.036	0.161
Adjusted R^2	0.134	0.036	0.036	0.158
R^2 Within	0.010	0.025	0.025	0.016
R^2 Within Adjusted	0.010	0.025	0.025	0.015
AIC	9365.2	155107.1	158127.5	24892.4
BIC	13168.5	155702.0	158807.7	27793.3
RMSE	0.24	0.26	0.26	0.28

Std.Errors	by: STATEFIP	by: STATEFIP	by: STATEFIP	by: STATEFIP
FE: STATEFIP	X	X	X	X
FE: YEAR	X	X	X	X
FE: IND	X			X

p < 0.1, * p < 0.05, ** p < 0.01, ***

Table 3: Regression Results