Identifying Edgeworth Cycles using Unsupervised Machine Learning

by

Andrew Eckert[†], Douglas Silveira[*], and Yizhe Zhuang[‡]

July 21, 2026

Abstract

Retail gasoline prices often exhibit distinct dynamic patterns, yet reliable classification is difficult. Using Canadian retail gasoline data from 2017 to 2023, we construct indicators from the Edgeworth-cycle literature and apply unsupervised clustering to city–quarter observations. To interpret assignments, we train supervised models and compute permutation importance. Across algorithms, three coherent clusters emerge: cycling, rigidity, and wholesale-price following. Indicators based on daily first differences are most informative. We relate clusters to margins, market size, brand presence, and concentration. Tracking assignments over time, only about one third of cities initially classified as cycling remain so after the COVID-19 pandemic. Econometric evidence indicates this decline is not fully explained by observed market structure or the pandemic period, and its timing coincides with the entry of a pricing-algorithm provider. Our findings provide regulators with practical tools to assess pricing dynamics, including the influence of algorithmic pricing systems.

Keywords: Machine Learning, Gasoline Market, Edgeworth Cycle, Competition Policy
JEL Codes: C38, C55, L13, L41

† Department of Economics, University of Alberta, Edmonton, Alberta, Canada (aeckert@ualberta.ca).

∗ Federal University of Juiz de Fora, Juiz de Fora, Minas Gerais, Brazil (douglas.silveira@ufjf.br).

‡ Department of Economics, University of Alberta, Edmonton, Alberta, Canada (yizhe10@ualberta.ca).

Acknowledgements. Douglas Silveira thanks CNPQ (Grant 169328/2023-6) for financial support.

1 Introduction

An industry that frequently attracts public attention is gasoline retailing. A long-standing policy and academic concern has been the observation of different patterns of pricing behaviour. While in some markets station level retail prices appear rigid, remaining unchanged for weeks, in other markets prices have been shown to change on a daily or higher frequency, often following cyclic patterns in which they increase rapidly, and then fall back to initial levels more gradually (Castanias and Johnson 1993; Eckert 2003; Noel 2007; Zimmerman, Yun, and Taylor 2013). In other markets, prices may adjust from day to day in a smooth fashion, exhibiting neither rigidity nor cyclic behaviour (Doyle, Muehlegger, and Samphantharak 2010; Lewis and Noel 2011).

The literature examining these pricing patterns is substantial.[1] A common theoretical framework for thinking about these patterns is Maskin and Tirole (1988)'s alternating move duopoly model of price setting. A key feature of this model is the existence of different equilibria, including both focal pricing, matching each other at a price above marginal cost, and Edgeworth cycles, in which firms alternate between periods of undercutting and price restorations to high prices.[2] Empirical studies have considered whether Edgeworth cycles are associated with less market power than rigid prices, and with less competitive market structures (Noel 2007; Eckert 2003; Doyle, Muehlegger, and Samphantharak 2010).

A challenge in the empirical literature is distinguishing between markets exhibiting cyclic prices or price rigidity, as well as other patterns such as wholesale-price following behaviour. Often, past studies have relied on visual inspection or simple empirical rules of thumb to identify markets with price cycles (Castanias and Johnson 1993; Eckert 2002). Unfortunately, this creates the potential for miscategorization, which can be problematic if conclusions regarding market power are sensitive to the classification approach. In a recent analysis, Holt, Igami, and Scheidegger (2024) compare the accuracy of different empirical indices for identifying gasoline price cycles. Time series graphs of retail-wholesale margins for individual stations over calendar quarters in Australia and Germany were shown to research assistants, who classified them as exhibiting cycles or not. The classified data were then divided into training and testing samples. Different indices drawn from the literature were calibrated on the training data and employed to predict cycling in the testing data. In addition, new methods, including supervised machine learning approaches based on the constructed indices, were employed. The authors find that the best performance is

1. See Eckert (2013) for a review; more recent contributions will be discussed in Section 2.

2. Other studies suggest that gasoline price cycles represent a form of collusion between asymmetric firms, in which firms coordinate restorations, and lower cost firms are permitted greater profits on the downward portion of the cycle; see for example Clark and Houde (2013).

achieved using neural networks and machine learning algorithms.

One potential concern with this approach is that it relies on visual inspection to identify different pricing patterns. The purpose of the current paper is to take a different approach. While we follow Holt, Igami, and Scheidegger (2024) in using machine learning techniques to study gasoline pricing dynamics, we do not rely on visual inspection at the labelling stage to classify cities and time periods. Rather, we construct variables based on different methods for identifying Edgeworth cycles developed in the literature and use unsupervised machine learning methods to group city-quarter observations into clusters.[3] This approach also broadens the object of analysis relative to a simple cycle versus non-cycle classification, allowing the data to distinguish among multiple pricing regimes, including cycle pricing, rigidity, and wholesale-price following. We then use permutation importance to examine which variables the clustering algorithms appear to rely on most heavily.[4] Finally, we assess whether the resulting clusters are economically meaningful by asking whether they separate cities into distinct pricing regimes, whether they remain stable over time and robust to changes in variables or clustering methods, and whether they are associated with key market characteristics emphasized in the economic literature.

To anticipate conclusions, we find that unsupervised machine learning algorithms generally group observations into three coherent clusters, corresponding to cycle pricing, price rigidity, and wholesale-price following behaviour. Permutation importance analysis suggests that statistics based on daily first differences of retail prices, especially the number of zero first differences in a quarter and the average absolute change in price, are among the most informative indicators. The economic analysis indicates that the resulting clusters are not merely statistical artifacts. Rigid-cluster observations exhibit the highest margins and occur disproportionately in smaller cities, while wholesale-price-following observations are concentrated in the largest cities, especially those surrounding Toronto. The relationship between pricing regimes and standard concentration measures is more nuanced and appears to be economically modest. Finally, we document a substantial decline in price cycling over the sample period. This decline remains after conditioning on market structure, wholesale-price conditions, market size, and pandemic-related factors, and overlaps with the period following the entry through acquisition of a company providing algorithmic pricing services.

3. In supervised learning, models are trained to predict pre-existing labels, such as human-coded cycle/non-cycle classifications. In unsupervised learning, the algorithm forms groups directly from the observed features, without pre-assigned labels; see, for example, Hastie et al. (2009) and James et al. (2013).

4. Permutation importance measures how much predictive performance deteriorates when the values of a given variable are randomly shuffled, breaking its relationship with the target classification; see Altmann et al. (2010). We provide more detail on our use of permutation importance in Section 4.1.

The remainder of the paper proceeds as follows. Section 2 reviews the literature on retail gasoline pricing patterns. Section 3 describes the data and Canadian context. Section 4 presents the clustering methodology and indicator variables. Section 5 reports the clustering results, Section 6 examines their economic implications, and Section 7 concludes.

2 Literature Review

The empirical literature on asymmetric cycles in retail gasoline prices can be traced back at least to Castanias and Johnson (1993), who examined cycles in Los Angeles, in which retail prices rose over a single week, and declined back to original levels over several weeks; this behaviour was not observed in wholesale prices. Similar asymmetric price patterns have since been documented in many areas, including elsewhere in the United States, Canada, Australia, and parts of Europe.[5] In markets without price cycles, different forms of pricing have been observed, including rigid prices that remain constant for days or weeks, and prices that appear to adjust frequently in response to wholesale prices; see for example Eckert (2003) and Noel (2007) in the Canadian context.

A frequently cited framework to understand retail price cycles is the alternating-move price-setting duopoly model of Maskin and Tirole (1988).[6] In this setting, different classes of equilibria are identified. In one class, firms continuously match each other at a 'focal' price above marginal cost. In contrast, in an Edgeworth Cycle equilibrium, starting from a high price, firms undercut each other until a lower bound is reached, triggering a war of attrition before prices are increased and a new cycle is initiated. The model is extended in Noel (2008), who uses computational methods to consider the effects of fluctuating marginal costs, capacity constraints, additional firms, and product differentiation.

A key focus of the empirical literature is on which markets exhibit cycles versus other patterns, and whether price cycles are associated with higher or lower prices. Zimmerman, Yun, and Taylor (2013), in a study of price cycles in the U.S., find that price cycles are associated with lower average prices; in contrast, Doyle, Muehlegger, and Samphantharak (2010) find no difference in overall price levels across cycling and non-cycling markets. Cycles have been associated with markets with lower or intermediate levels of concentration; see Noel (2007), Eckert (2003), and Doyle, Muehlegger, and Samphantharak (2010). Doyle, Muehlegger, and Samphantharak (2010) also find that the presence of cycles is associated with independent stations with convenience stores.

5. See for example Zimmerman, Yun, and Taylor (2013) for the U.S., Eckert (2003) for Canada, and Holt, Igami, and Scheidegger (2024) for Australia and Germany.

6. In certain markets, price cycles have been associated with collusion between asymmetric retailers; see for example Clark and Houde (2013).

A challenge in studying markets that exhibit price cycles, and particularly in comparing markets with cycles to markets that exhibit other patterns, is in categorizing the markets. As discussed in Holt, Igami, and Scheidegger (2024), a range of different indices or rules of thumb have been used. These include comparing the length of positive and negative runs or the relative magnitudes of positive and negative first differences, or identifying a city as cycling if the median daily price change is negative or if the market exhibits frequent large price increases. Holt, Igami, and Scheidegger (2024) extend this list to include spectrum analysis and cubic splines, as well as neural networks and machine learning algorithms. The authors then test the effectiveness of these different methods by considering how well they can identify cities and quarters (in Australia and Germany) that have been identified through visual inspection by research assistants as exhibiting cycles.

A recent development in retail gasoline markets has been the introduction of algorithmic pricing. Algorithmic pricing software can monitor market conditions online, including potentially rival prices, and recommend price changes that respond quickly to these conditions. To the extent that such software allows for faster responses to rival price changes, it may have implications for the existence and characteristics of price cycles. Klein (2021) applies algorithmic pricing to the alternating price setting model of Maskin and Tirole (1988), supposing that firms choose prices using Q-learning algorithms; they find the emergence of fixed price behaviour at supra-competitive prices, as well as Edgeworth cycles, depending on parameter values. The authors conclude that collusion can occur through algorithmic pricing. In a recent empirical study of the impact of algorithmic pricing on gasoline prices in Germany, Assad et al. (2024) use structural breaks in pricing indicators to detect the adoption of algorithmic pricing; the authors find that adoption of algorithmic pricing increased mean prices, increasing margins by 15%.[7]

The Canadian gasoline market provides a useful setting in which to examine these issues because concerns have recently been raised about pricing software and data services in this industry. Kalibrate, an international company that collects gasoline-market data and provides pricing-related services to retailers, acquired Canada's Kent Marketing in February 2019 (Kalibrate 2019). In July 2024, the Canadian Competition Bureau announced an investigation into Kalibrate and whether its "data, pricing and consultation services" have "an adverse effect on competition between gas stations in Canada".[8]

7. Brazil's competition authority (CADE) is investigating allegations that an algorithmic pricing tool may have facilitated uniform retail gasoline pricing across cities (Athayde, Araújo Ferreira, and Diniz 2025); the case also involves a trade union alleged to have encouraged adoption. CADE's Department of Economic Studies reports statistically significant post-adoption price increases at adopting stations for most fuels (Technical Note 16/2024/DEE/CADE).

8. See "Competition Bureau advances an investigation into Kalibrate's gas pricing services" on the

Kalibrate's own public materials also suggest that its services are connected to retailers' pricing decisions and to the management of fuel-price volatility. In a 2018 blog post on fuel-price volatility, the company argues that undercutting rivals may trigger a "downward spiral" in prices and that, once such a volatility cycle begins, it can be difficult to stop (Kalibrate 2018). The same post states that large retailers with sufficient market presence may be able to "dampen volatility" and encourages firms to "use data science" to respond to fuel-price volatility. This background does not imply that Kalibrate caused the pricing-regime changes documented below. It does, however, motivate our descriptive analysis of whether retail gasoline pricing regimes changed during a period in which data-driven pricing tools became more salient in the Canadian market.

3 Data Sources and Description

The primary data set consists of city-level weekday gasoline prices for 66 Canadian cities, for the period January 2, 2017, to December 29, 2023.[9] Prices represent the average self-serve prices from a sample of stations in each city, collected each morning at 10:00 AM by telephone. Prices are not collected on weekends or statutory holidays.[10] Daily weekday wholesale prices were also obtained from the same source for 28 cities.[11]

Price data are available from Kalibrate for cities from all ten provinces and two territories. In our main analysis, we exclude the four Atlantic provinces: New Brunswick, Nova Scotia, Prince Edward Island, and Newfoundland and Labrador, as their prices are subject to regulation; all four provinces have price ceilings, while Prince Edward Island and Nova Scotia are also subject to price floors.[12]

While wholesale prices are available for some of the cities for which we have retail prices, for other cities in the retail price sample, we use the wholesale price at the geographically closest location. To account for transportation costs from the terminal location to the city, we obtained estimates of gasoline retail transportation costs for various retail markets for the province of Quebec[13] over the period 2017-2023, and used simple linear regressions to estimate the relationship between wholesale-retail distance and real transportation cost per litre over our sample period.[14] These results were then used to estimate real transportation

Competition Bureau of Canada website.

9. These data are collected by Kalibrate, and are publicly available at https://charting.kalibrate.com.

10. For more information on data collection methodology, see https://charting.kalibrate.com/KentReports/Methodology.

11. Tables B.1 and B.2 in Appendix B show summary statistics for before-tax (ex-tax) retail prices and wholesale prices for all cities in the sample, for January 2, 2017, to December 29, 2023.

12. We continue to include the province of Quebec, which has a price floor but no ceiling.

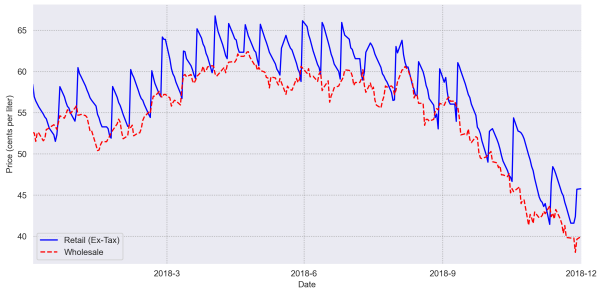
13. Data obtained from the Régie de l'énergie of the Government of Quebec.

14. A small number of cities, including in Yukon and Northwest Territories, were dropped because there

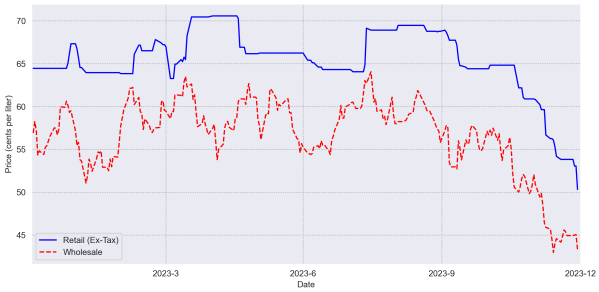
costs for each city in our sample not located at a wholesale location.

Figure 1: Examples of Retail and Wholesale Price Patterns

(a) Peterborough, ON, ex-tax retail and Toronto, ON, wholesale prices, 2018



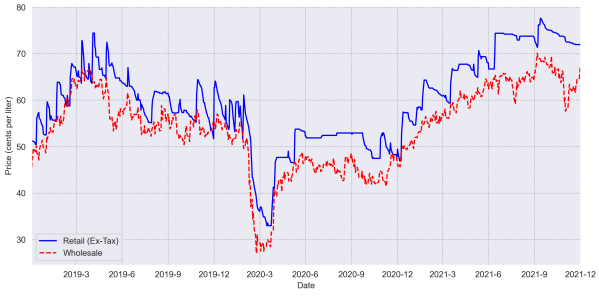
(b) Brandon, MB, ex-tax retail and Winnipeg, MB, wholesale prices, 2023



(c) Hamilton, ON, ex-tax retail and wholesale prices, 2018



(d) Kamloops, BC, ex-tax retail and wholesale prices, 2019–2021



In addition to pricing data, we compile city-level covariates that plausibly co-vary with pricing patterns. For market structure, we use Kalibrate's annual counts of gasoline stations by brand for each city from 2017–2023 to construct measures of retail concentration and brand presence. To proxy market size, we merge annual estimated populations by Census Metropolitan Area for the period 2017-2023.[15] To capture pandemic-related shocks, we incorporate province-level COVID-19 case counts.[16] All series are harmonized to our city–quarter panel.

To illustrate basic patterns observable in retail prices in our sample, Figure 1 shows time series plots of retail and wholesale prices for four cities. All prices are in 2002 dollars. Panel (a) shows ex-tax retail prices in Peterborough, ON, alongside Toronto wholesale prices in 2018. This city is an example of a market with clear asymmetric price cycles,

were no nearby wholesale locations in the data.

15. Data obtained from Statistics Canada.

16. Data obtained from the Canadian Institute for Health Information. *Canadian COVID-19 Intervention Timeline — Data Tables.* Ottawa, ON: CIHI; October 13, 2022.

where retail prices exhibit frequent and predictable fluctuations. Panel (b) shows ex-tax retail prices in Brandon, MB, alongside Winnipeg wholesale prices in 2023. Brandon's retail prices demonstrate substantial rigidity, with periods of extended price stability. The wholesale prices shown do not incorporate transportation costs to Brandon; since Brandon is approximately 214 km from Winnipeg, Winnipeg wholesale prices likely underestimate the true wholesale costs in this market. Panel (c) shows a different pattern. In Hamilton, ON, retail prices generally adjust daily and appear to follow movements in wholesale prices. Panel (d) illustrates that pricing behaviour can shift over time. In Kamloops, BC, retail pricing appears to transition from high-frequency cycles to either rigid pricing or cycles with much lower frequency during the 2019–2021 period.

4 Clustering Methodology

Our empirical analysis proceeds in several distinct stages. First we develop an initial set of indicators that we consider for clustering; these include variables that have been used previously in the literature to identify Edgeworth cycles, as well as new variables designed to distinguish between different forms of non-cycling behaviour, such as price rigidity and wholesale-price following behaviour. Different algorithms are used to cluster the city-quarter observations using the full set of variables. In the second stage of analysis, we use the results of the initial clustering stage, along with a consideration of correlations between variables and different sources of variation in the data to winnow our variables down to a final set of six; the clustering analysis is then repeated using these six indicators.

The remainder of the paper will proceed as follows. In this section, we first discuss different algorithms for clustering city/quarter observations, as well as our approaches to assessing consistency across algorithms, variable selection, and feature relevance. Subsection 4.2 then describes the initial set of indicators. Section 5 first describes the results using the full set of variables, and the process for reducing the variables to a final set of six, before presenting the final clustering analysis and results. Section 6 considers whether the resulting clusters are economically meaningful in terms of their persistence, location, and association with market characteristics.

4.1 Clustering Algorithms and Classifiers

In this subsection, we discuss the different algorithms we employ to cluster city-quarter observations, as well as our approach to assessing consistency across algorithms and the importance of different variables in the clustering exercise. We group cities by pricing behaviour using three standard unsupervised learning models: K-means (MacQueen 1967; Lloyd 1982; Gnanadesikan 2011), Agglomerative Hierarchical Clustering (Ward Jr 1963;

Murtagh 1983), and Gaussian Mixture Models (GMM) (Dempster, Laird, and Rubin 1977; McLachlan 1982; Bishop and Nasrabadi 2006; Izenman 2008). K-means and hierarchical clustering assign cities based on distance in the feature space, while GMM models the data as a mixture of Gaussian components and assigns membership probabilistically. We provide technical details in Appendix A.1.

We then examine how observed indicators explain the cluster assignments using supervised learning models. We estimate a multinomial logistic regression (Izenman 2008; Hosmer Jr, Lemeshow, and Sturdivant 2013) and Quadratic Discriminant Analysis (QDA) (Williams 1982). Logistic regression maps indicators to class probabilities, and QDA models class–conditional distributions with class–specific covariance matrices, yielding nonlinear decision boundaries. Using both models helps identify the indicators that most clearly separate cyclical, rigid, and wholesale-price following dynamics. Appendix A.2 provides estimation details for the classifiers.

To assess consistency across clustering methods, we calculate two widely used external validation metrics: the Adjusted Rand Index (ARI) and the Normalized Mutual Information (NMI). The ARI quantifies the similarity between two clusterings by comparing all pairs of observations and counting how many pairs are consistently grouped or separated across both partitions (Hubert and Arabie 1985). It adjusts for agreement expected by chance, with values ranging from -1 (complete disagreement) to 1 (perfect agreement), and 0 indicating random classification. NMI approaches the problem from an information-theoretic perspective (Vinh, Epps, and Bailey 2009). It measures how much knowing the cluster assignment from one method reduces uncertainty about the other. NMI values range from 0 (no mutual information) to 1 (perfect correspondence) and remain invariant to the absolute size or number of clusters. In addition, we also use pairwise confusion matrices to inspect agreement at the observation level (Landis and Koch 1977).

To examine the role of each indicator in shaping the observed cluster structures, we use Permutation Importance (PI) in a supervised classification setting, where the cluster labels from the unsupervised phase serve as the target variable (Silveira et al. 2023). PI measures how much the predictive performance of a model deteriorates when the values of a given feature are randomly shuffled, breaking its relationship with the target (Altmann et al. 2010). Features that cause larger performance drops are considered more important for distinguishing between clusters.

The procedure involves four steps: (i) Train a classification model on the full set of indicators to learn the mapping between variables and cluster labels; (ii) Evaluate the model's baseline performance on a hold-out test set; (iii) Shuffle the values of each indicator

one at a time and measure the resulting performance loss; (iv) Rank indicators based on the magnitude of the performance drop.

We calculate PI scores using both Logit and QDA classifiers to ensure that our variable importance rankings are not model-dependent. To complement PI, we examine the correlation matrix of the indicators. This helps identify and exclude highly collinear variables that may introduce multicollinearity and add noise to the clustering process. Variables that are both highly correlated and show low PI scores are strong candidates for exclusion. Prior work shows that targeted variable selection – guided by relevance measures and correlation checks – can sharpen cluster separation by filtering noisy or redundant features (Dy and Brodley 2004; Cai et al. 2018).

After selecting a reduced set of indicators based on PI rankings and the correlation analysis, we re-run the unsupervised clustering algorithms. We then reapply the same agreement metrics (ARI, NMI, and confusion matrices) to assess whether reducing the indicator set improves the level of agreement across clustering methods. This final diagnostic step allows us to evaluate whether focusing on a smaller, more informative set of variables enhances both the stability of the clustering results and the interpretability of the observed group structures.

4.2 Indicator Variables

This section describes our initial set of indicator variables used to characterize pricing behaviour before clustering. We separate the construction into two parts. First, we define five broad cycle-detection methods adapted from the previous Edgeworth-cycle literature: run-based asymmetry measures, average increase-decrease comparisons, median price-change measures, Many Big Price Increases (MBPI), and spectral analysis using the Lomb–Scargle periodogram. Second, we add variables designed to distinguish pricing regimes beyond cycling. These additional variables capture price rigidity, retail-wholesale margin behaviour, wholesale-price volatility, and the degree to which retail prices move with wholesale prices.

A summary of different variables that have been used in the prior literature to identify Edgeworth cycles can be found in Holt, Igami, and Scheidegger (2024); see Holt, Igami, and Scheidegger (2024) for citations to the previous literature. They classify these approaches into four methods; see Table 1 for definitions of all variables discussed.

Method 1: Positive Runs vs Negative Runs. Here, a run is a period of time over which price has been moving in the same direction; one indicator is the difference between the average length of time of positive runs and negative runs over the sample period.

Since cities experiencing Edgeworth cycles often exhibit price increases lasting one or two days, but prices that decrease over several days or a week, cities that exhibit much longer negative runs than positive runs are more likely to belong in the cycling cluster.

Method 2: Mean Increase vs Mean Decrease. This approach compares the average size of all positive first differences to all negative first differences over a period of time. A market is more likely to be cycling if the average (daily) price increase is large relative to the average daily price decrease. For this method, we compute for each city and each quarter the mean positive first difference minus the mean negative first difference in absolute value, and also the ratio of the mean positive and negative first differences.

Method 3: Negative Median Change. A city is considered to be more likely to be cycling when it has many daily price decreases relative to increases. Hence, this approach introduces the median daily first difference of price within the city/quarter as a variable.

Method 4: Many Big Price Increases. Since Edgeworth cycles involve large price increases within a short period, this approach counts the number of price increases within a city/quarter that exceed a specified threshold. In our analysis, we construct variables using thresholds of three, four and five cpl. To account for large wholesale price increase, we construct similar variables counting the number of large increases in the retail-wholesale margin; we also construct `largediff_diff`, which takes the difference between the number of large increases (more than three cpl) in the retail price and the number of large increases in the wholesale price.

In addition, Holt, Igami, and Scheidegger (2024) introduce methods to capture periodicity in the time series of prices of different cities or stations. From these, we select as our *Method 5* the *Lomb-Scargle Periodogram*, which is appropriate for unevenly sampled time series, and include the maximum power in the periodogram as a variable.[17]

Finally, to distinguish between non-cycling pricing patterns, we include variables to capture general price rigidity and volatility, and the relationship between retail and wholesale prices. These include the number of zero first differences, and the average length of runs without price change, the mean absolute value of the first difference and its standard deviation, the mean of the daily margin, the standard deviations of the retail and wholesale prices and the margin, and the correlation coefficient between retail and wholesale prices. Table 1 provides definitions of the 21 indicators used in the initial clustering analysis.

17. See Holt, Igami, and Scheidegger (2024) for details.

One issue in constructing these indicators is that retail price data are available only on weekdays. It is well established that retail gasoline pricing dynamics differ between weekdays and weekends. Studies of Edgeworth cycles find that price restorations are least likely on weekends but are frequent on Mondays, and that prices are less responsive and more likely to remain constant on weekends; see for example, Foros and Steen (2013), Foros, Nguyen-Ones, and Steen (2021), Byrne and De Roos (2019), Atkinson, Eckert, and West (2009), and Atkinson (2009). Because of nonlinear price dynamics and differences in behaviour between weekends and weekdays, interpolating weekend prices from weekday data may be unreliable, especially since our cycle indicators rely heavily on first differences. Dropping days for which the previous day's data are unavailable can also cause cycles to be missed, since price restorations may occur on Mondays.

Table 1: Full Set of Indicators Used in Initial Clustering Analyses

Category	Variable	Description
Cycle-detection indicators (Methods 1–5)		
Method 1: Positive Runs vs Negative Runs	`runs_diff`	Difference between the number of positive and negative price runs
Method 2: Mean Increase versus Mean Decrease	`size_diff`	Difference between mean size of increases and decreases
	`size_ratio`	Ratio of mean size of increases to mean size of decreases
Method 3: Negative Median Change	`med_diff`	Median daily retail price change
Method 4: Many Big Price Increases	`largeposdiff3`	Daily retail price increases above 3 cpl threshold
	`largeposdiff4`	Daily retail price increases above 4 cpl threshold
	`largeposdiff5`	Daily retail price increases above 5 cpl threshold
	`largeposdiff3_marg`	Margin-based daily price increases above 3 cpl threshold
	`largeposdiff4_marg`	Margin-based daily price increases above 4 cpl threshold
	`largeposdiff5_marg`	Margin-based daily price increases above 5 cpl threshold
	`largediff_diff`	Difference between large retail price increases and large wholesale price increases (3 cpl threshold)
Method 5: Lomb–Scargle Periodogram	`Lomb_Scargle`	Maximum power in Lomb–Scargle periodogram (first-differenced data)
Additional regime-classification indicators		
Retail price rigidity	`zero_count`	Days with zero retail price change
	`zero_length`	Length of zero-price-change runs (consecutive observed price changes)
Price-change dispersion	`absdiff`	Mean absolute daily retail price change
	`absdiff_sd`	Standard deviation of absolute daily retail price changes
Margin level and volatility	`margin`	Average retail-wholesale margin
	`margin_sd`	Standard deviation of retail-wholesale margin
Wholesale and retail volatility	`ex_tax_sd`	Standard deviation of ex-tax retail price
	`wholesale_sd`	Standard deviation of wholesale price
Retail-wholesale comovement	`rw_corr`	Retail-wholesale price correlation coefficient

Because of these concerns, our approach is to treat all observed prices as consecutive. For example, if the price is lower on Monday than on Friday, this is treated as a negative run

of one observed step. Since our algorithms group cities and quarters based on observable data, this approach avoids both omitting available observations and adding approximated data. Given that restorations and troughs are concentrated on specific weekdays and rarely initiated on weekends, and that relenting phases typically unfold over several days, treating Friday–Monday gaps as single-step changes in the business-day data should preserve the features that matter for distinguishing cycling, rigidity, and wholesale-price following.

Table 2: Summary Statistics of the Full Set of Indicators (2017–2023)

	Obs	Mean	Std. Dev.	Min	25%	50%	75%	Max	Skew	Kurtosis
Cycle indicators										
runs_diff	1176	0.78	1.30	-2.12	-0.06	0.39	1.28	10.80	1.71	4.94
size_diff	1176	0.71	0.94	-0.98	0.04	0.40	1.19	5.64	1.41	2.45
size_ratio	1176	2.23	1.78	0.18	1.06	1.63	2.94	18.53	2.83	14.23
med_diff	1176	-0.07	0.17	-1.31	-0.09	0.00	0.00	0.24	-2.68	10.36
largeposdiff3	1176	2.01	2.32	0	0	1.00	3	16	1.70	3.76
largeposdiff4	1176	1.25	1.78	0	0	1.00	2	11	2.17	5.69
largeposdiff5	1176	0.76	1.34	0	0	0.00	1	11	2.84	10.84
largeposdiff3_marg	1176	2.81	2.73	0	1	2.00	5	15	1.12	0.82
largeposdiff4_marg	1176	1.63	1.96	0	0	1.00	3	11	1.58	2.66
largeposdiff5_marg	1176	0.96	1.43	0	0	0.00	1	11	2.15	6.09
largediff_diff	1176	1.14	2.61	-7.00	0.00	1.00	2.00	14.00	0.95	3.69
Lomb_Scargle	1176	0.18	0.07	0.08	0.14	0.16	0.19	0.63	2.31	6.42
Regime indicators										
zero_count	1176	0.30	0.25	0.02	0.08	0.20	0.51	0.94	0.68	-0.85
zero_length	1176	2.45	2.27	1.00	1.08	1.50	2.87	20.67	3.13	13.60
absdiff	1176	0.66	0.40	0.01	0.37	0.60	0.84	2.59	1.17	2.22
absdiff_sd	1176	0.92	0.44	0.06	0.59	0.84	1.20	2.90	0.76	0.57
margin	1176	5.44	2.25	0.24	4.06	5.07	6.71	17.06	0.93	2.19
margin_sd	1176	2.22	1.16	0.57	1.44	2.03	2.75	8.66	1.34	2.78
ex_tax_sd	1176	3.55	2.12	0.12	1.99	2.83	5.04	13.43	0.94	0.50
wholesale_sd	1176	3.52	2.16	1.02	1.96	2.50	5.29	12.44	1.15	0.60
rw_corr	1176	0.68	0.32	-0.72	0.54	0.80	0.91	0.99	-1.61	2.63

Table 2 reports summary statistics for the 21 indicators used in the initial clustering analysis. The variables display substantial variation in scale, distribution, and dispersion. Several indicators exhibit substantial skewness or kurtosis, and standard deviations vary widely across indicators. These patterns reinforce the decision to standardize and normalize the variables prior to clustering, ensuring that differences in scale and distribution do not distort the clustering outcomes (Revelle 1979).[18]

18. The standardization step uses a Z-score transformation (mean-centered and divided by the standard deviation), followed by a Yeo–Johnson normalization. For details, see Python documentation for sklearn.preprocessing.StandardScaler and sklearn.preprocessing.PowerTransformer.

5 Clustering-Based Evidence

This section presents the main clustering evidence. In Section 5.1, we explain how the initial set of 21 indicators is reduced to the six indicators used in the baseline specification. We then report the clustering results obtained with the reduced set in Section 5.2, focusing on the number of clusters, agreement across algorithms, cluster composition, indicator profiles, and permutation-importance diagnostics. The objective is to show that the final classification is not only statistically coherent, but also interpretable as three economically meaningful pricing regimes: cycle pricing, rigidity, and wholesale-price following.

5.1 Reduction in Variable Set

We begin by discussing the selection of six variables from the larger set of 21 indicators. The transition to six variables is central to the empirical design: a compact set of indicators can improve interpretability by preserving the main pricing dimensions while limiting redundancy in the feature space.

We reduce the feature set using correlation diagnostics, consistency across algorithms, permutation importance rankings, and economic interpretation. The correlation diagnostics identify groups of variables that encode similar information. The permutation-importance exercise identifies indicators that help recover the cluster labels generated by the unsupervised methods. The cross-algorithm consistency checks indicate whether a feature set generates a classification that is stable across distinct clustering procedures rather than being an artifact of one algorithm's geometry or distributional assumptions. Economic interpretation then guides the final choice of indicators so that the reduced set continues to capture the three pricing regimes of interest. This interpretive step also involved checking representative city-quarter price series to verify that the selected variables separated retail-driven cycles from rigidity and wholesale-price following in an economically coherent way. We also considered alternative reduced feature sets to assess the sensitivity of the three-regime interpretation.[19] The detailed full-set diagnostics are reported in Appendix C.1; here, we summarize the main evidence motivating the reduction.

The correlation diagnostics reveal blocks of highly correlated variables; the full correlation matrix is reported in Figure C.1. Strong correlations frequently exist within variables

19. These alternatives included feature sets guided by principal component analysis (PCA), which summarizes common variation across correlated variables, as well as PI-oriented specifications and sets that placed greater weight on retail-wholesale comovement or median price-change measures (Jolliffe 2005). The alternatives generally preserved the broad distinction between cycling, rigidity, and wholesale-price following, with most changes occurring near the boundary between rigidity and wholesale-price following. We therefore view the final six-variable set as a parsimonious and interpretable reduction, rather than as the unique specification capable of generating the three-regime taxonomy.

associated with the different methods; for example, correlation coefficients exceed 0.8 for many variable pairs within the MBPI category, leading us to select just one of the seven MBPI variables for the final specification. High correlations are also observed among some additional regime-classification indicators that are intended to capture similar features.

We then carried out the clustering exercise using the full set of 21 indicators; results are reported in Appendix C.1. The full-set clustering results show uneven cross-algorithm agreement. K-means and GMM-Spherical align closely, with an overall agreement of 90.39%, ARI of 0.724, and NMI of 0.710. Hierarchical clustering is less aligned with K-means, with agreement of 51.62%, ARI of 0.331, and NMI of 0.392. GMM-Full also shows relatively weak agreement with K-means, with agreement of 50.51%, ARI of 0.257, and NMI of 0.284. [20] The permutation-importance results suggest removing `Lomb_Scargle`, `margin`, and `runs_diff`, which have the three lowest average rankings across models, and retaining `zero_count` and `size_diff`, which have the strongest average rankings. At the same time, permutation importance does not imply a mechanical ranking rule; because several indicators are correlated or capture related aspects of pricing behaviour, a variable with a moderate standalone PI score may still be useful when combined with other indicators. The final set therefore does not simply correspond to the six highest average PI rankings.

Table 3: Reduced Indicator Set Used in the Main Clustering Analysis

Indicator	Main pricing dimension	Role in the reduced set
`zero_count`	Retail price rigidity	Captures days with no retail price change
`absdiff`	Short-run retail price variability	Measures average absolute daily retail price changes
`margin_sd`	Retail-wholesale margin volatility	Captures variation in margins over the quarter
`largediff_diff`	Retail-wholesale adjustment asymmetry	Summarizes MBPI-type asymmetry net of wholesale movements
`size_diff`	Asymmetric price changes	Measures difference between average price increases and decreases
`size_ratio`	Asymmetric price changes	Measures ratio of average price increases to decreases

Economic interpretation also guided the reduction by requiring the retained indicators to cover the three pricing regimes the analysis seeks to distinguish, rather than simply reproduce the highest average PI rankings. Table 3 summarizes the resulting six-indicator specification. We retain `zero_count` as a direct measure of rigidity, since frequent unchanged prices are central to distinguishing rigid markets from both cycling and wholesale-price-following markets. We retain `absdiff` to measure the intensity of day-to-day retail price adjustment and `margin_sd` to capture variation in the retail-wholesale margin. The remaining three variables capture asymmetric price-adjustment dynamics. `size_diff` measures the difference between average price increases and decreases in levels,

20. We also tested the GMM-Tied and GMM-Diagonal covariance specifications. Their results fall between those of GMM-Spherical and GMM-Full.

while `size_ratio` measures the same asymmetry in scale-free terms. Although these two variables are highly correlated, they are not interchangeable; in alternative reduced sets, dropping either variable lowered cross-algorithm agreement, and both remain informative in the reduced-set permutation-importance profiles reported below. We therefore retain both as complementary measures of the restoration-relenting asymmetry associated with Edgeworth-cycle behaviour. Finally, `largediff_diff` represents the MBPI method by capturing large retail restorations net of large wholesale price increases. This compact set is useful because permutation importance evaluates variables one at a time, whereas the clustering exercise requires indicators that jointly separate cycle pricing, rigidity, and wholesale-price following in a stable and economically interpretable way.

5.2 Clustering Results: Reduced Set of Six Variables

We now turn to the clustering results using the reduced set of six variables. A first consideration is the number of clusters. Table 4 reports the internal cluster-number diagnostics for the six-indicator specification, comparing the three- and four-cluster solutions.

Table 4: Cluster-Number Diagnostics Using the Reduced Six-Indicator Set

Algorithm	**Criterion**	$k = 3$	$k = 4$	**Preferred** k
K-means	Silhouette Score	0.367	0.345	3
K-means	Davies–Bouldin Index	1.036	1.226	3
Hierarchical	Silhouette Score	0.340	0.327	3
Hierarchical	Davies–Bouldin Index	1.024	1.362	3
GMM-Spherical	AIC	16502.335	15624.835	4
GMM-Spherical	BIC	16618.942	15782.002	4
GMM-Tied	AIC	13698.328	13466.958	4
GMM-Tied	BIC	13906.192	13710.312	4
GMM-Diagonal	AIC	16049.765	15370.155	4
GMM-Diagonal	BIC	16242.421	15628.718	4
GMM-Full	AIC	11725.668	10906.125	4
GMM-Full	BIC	12146.467	11468.881	4

Notes: For the Silhouette Score, higher values indicate better separation. For the Davies–Bouldin Index, AIC, and BIC, lower values indicate better fit. The distance-based criteria favor three clusters for both K-means and hierarchical clustering. The GMM information criteria favor four clusters, while the main non-Atlantic analysis uses three clusters to preserve a parsimonious taxonomy of cycle pricing, rigidity, and wholesale-price following. The four-cluster specification is analyzed when the Atlantic provinces are included.

We also examined two- and five-cluster solutions. These specifications were less attractive either according to the internal diagnostics or because they produced less interpretable partitions: the two-cluster solution collapses economically distinct pricing regimes, while

the five-cluster solution tends to subdivide existing regimes without producing a clearer economic taxonomy. We therefore report the comparison between three and four clusters, the two candidate specifications that are both empirically relevant and economically interpretable. For both K-means and hierarchical clustering, the Silhouette Score and the Davies–Bouldin Index favor three clusters. Moving from three to four clusters lowers the Silhouette Score and increases the Davies–Bouldin Index for both distance-based algorithms, indicating weaker average separation and compactness. For GMM, the AIC and BIC favor four clusters across covariance structures. We use three clusters as the main non-Atlantic specification because this choice preserves a parsimonious and economically interpretable taxonomy: cycle pricing, rigidity, and wholesale-price following. The four-cluster case becomes economically relevant when regulated Atlantic markets are added, and we analyze that case separately in Appendix C.2.

Using the reduced set of six indicators, cross-algorithm agreement improves substantially relative to the full 21-indicator specification for two of the three main model comparisons. Agreement between K-means and hierarchical clustering rises from 51.62% with the full set to 90.48% under the reduced set, with ARI increasing from 0.331 to 0.737 and NMI from 0.392 to 0.716.[21] Agreement between K-means and GMM-Spherical also improves, from 90.39% to 92.77%, with ARI rising from 0.724 to 0.791 and NMI from 0.710 to 0.757. By contrast, K-means and GMM-Full show only a small increase in matching labels, from 50.51% to 54.42%, alongside lower ARI and NMI relative to the full set. Taken together, these results show that the reduced set provides a more stable and interpretable basis for distinguishing rigidity, wholesale-price following, and cycling tendencies across cities.[22]

Given the high agreement between K-means, hierarchical clustering, and GMM-Spherical under the reduced set, we use K-means as the benchmark classification for the descriptive and econometric analysis that follows. K-means provides a transparent partition of the normalized indicator space, and its labels are closely aligned with the two alternative algorithms that show the strongest agreement in the reduced specification. The resulting clusters also have clear indicator profiles, allowing us to interpret them as Cycle, Rigidity, and Wholesale-price Following regimes. The remaining algorithms are used to assess robustness and to identify areas where the classification is more sensitive to modelling choices. The mapping between the three K-means clusters and the labels Cycle, Rigidity,

21. Hierarchical clustering is sensitive to noisy or correlated features because it relies on distance calculations. Reducing the indicator set can remove distortions in the feature space, yielding clearer and more stable groups.

22. We also tested alternative subsets that included other high-ranking variables, but none matched the agreement or interpretive clarity of the final six-variable selection.

and Wholesale-price Following is based on the cluster-level indicator profiles. The Cycle cluster is characterized by stronger asymmetric price adjustments, the Rigidity cluster by a higher incidence of unchanged prices, and the Wholesale-price Following cluster by smoother adjustment patterns more closely aligned with wholesale-price movements. The representative city examples in Figure 1 illustrate these profiles.

Table 5: Cluster Composition by Model Using the Reduced Indicator Set

Model	Cycle	Rigidity	Wholesale-price Following
***K*-Means**	286	504	386
GMM-Spherical	268	574	334
Hierarchical	209	605	362
GMM-Full	283	372	521

Table 5 presents the distribution of observations across clusters for each algorithm. Relative to K-means, hierarchical clustering assigns more observations to rigidity and fewer observations to the cycle group. GMM-Full shifts toward the wholesale-price following group, while GMM-Spherical assigns more observations to rigidity. These composition shifts help explain the lower alignment of GMM-Full and the remaining areas of disagreement for GMM-Spherical.

At the class level, the reduced indicator set further highlights the advantage of hierarchical clustering and GMM-Spherical relative to GMM-Full. Using K-means as a benchmark, GMM-Spherical replicates the rigidity class perfectly and performs well for both cycle and wholesale-price following labels. Hierarchical clustering also matches K-means exactly for rigidity and closely for wholesale-price following, with somewhat lower alignment for the cycle class. By contrast, GMM-Full shows weaker alignment across all three regimes, consistent with its lower overall agreement measures.

Figure 2 presents the temporal and cross-city distribution of pricing regimes identified by K-means and hierarchical clustering using the reduced six-indicator set. Both models reveal a broadly consistent structure, with strong visual alignment in the classification of cycle, rigidity, and wholesale-price following dynamics across cities and over time. Several cities display a high degree of temporal consistency in their dominant pricing dynamics, with extended stretches of the same regime for both algorithms. While regime transitions occur, they tend to persist for multiple quarters, suggesting that pricing behaviour is spatially differentiated and temporally stable within cities. These visual patterns reinforce the statistical agreement observed in ARI, NMI, and per-class metrics.

Figure 3 complements the visual labelling based on the example cities by showing how the six selected indicators vary across the three K-means pricing regimes. The rigidity

cluster stands out most clearly in `zero_count`, which captures days with no retail price change. The cycle cluster displays higher values of `absdiff`, `largediff_diff`, `size_diff`, and `size_ratio`. These indicators capture larger absolute daily retail price changes and stronger asymmetries between price increases and decreases, consistent with the rapid-restoration and gradual-decline pattern associated with Edgeworth cycles. The wholesale-price following cluster generally shows lower values for the asymmetry indicators and a lower incidence of zero retail price changes, consistent with smoother adjustment rather than prolonged rigidity or pronounced cyclical dynamics.

Figure 2: Temporal and Cross-City Distribution of Pricing Dynamics

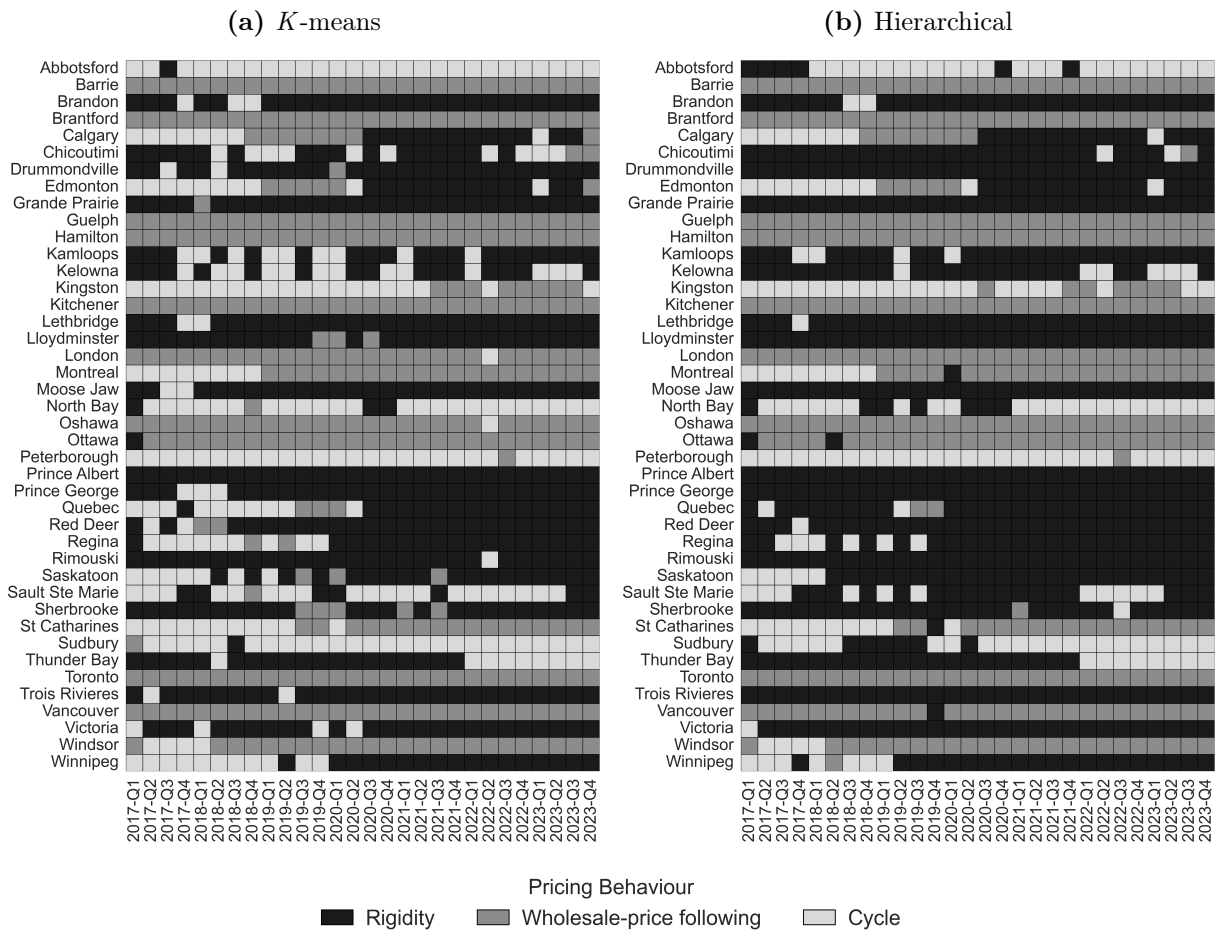
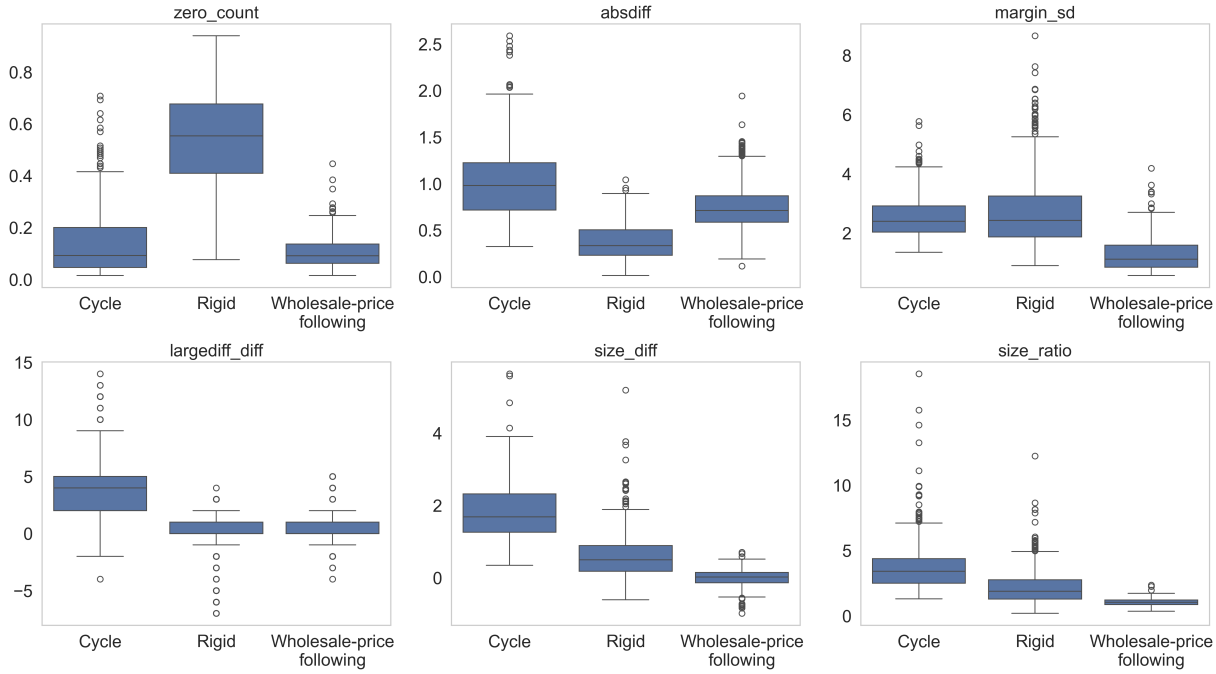


Figure 4 reports the permutation-importance profiles for K-means and hierarchical clustering combined with Logit in the three-cluster specification. The permutation-importance exercise evaluates how strongly each indicator contributes to recovering the cluster assignments in a supervised classification setting. Both profiles rely on the same economic dimensions: price stickiness through `zero_count`, dispersion and short-run variability through

`absdiff` and `margin_sd`, scale-related asymmetries through `size_diff` and `size_ratio`, and retail-wholesale restoration asymmetry through `largediff_diff`. The emphasis shifts by method: K-means+Logit draws relatively more signal from the size measures, while Hierarchical+Logit places more weight on dispersion and stickiness. The importance of both `size_diff` and `size_ratio` supports retaining them jointly despite their correlation, since the classifiers use both level-based and proportional asymmetry information.[23]

Figure 3: Distribution of Selected Indicators Across Pricing Regimes: K–means Classification

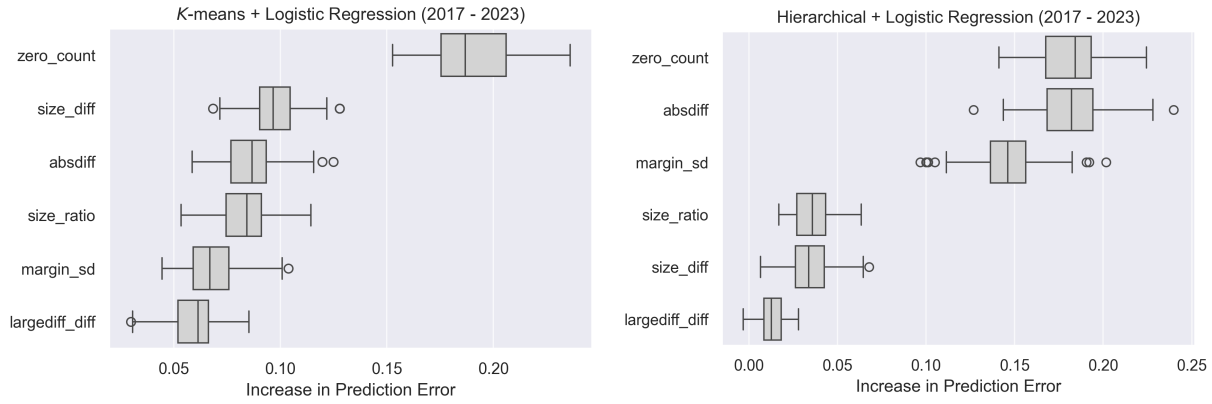


The more modest permutation-importance ranking of `largediff_diff` should be interpreted in this context. Its role is not to dominate classification on its own, but to add complementary information on the relationship between large retail restorations and wholesale-price movements. This helps distinguish retail-driven cycle dynamics from wholesale-price following. Consistent with this interpretation, `largediff_diff` remains informative in the reduced-set analysis, performs reasonably well among the full 21 indicators in the GMM-based permutation-importance profiles in Appendix C.1, and continues to contribute in the nine-indicator four-cluster specification with Atlantic provinces in Appendix C.2.

23. The corresponding GMM–QDA diagnostics point to the same six variables but assign different rankings. In both GMM-Spherical and GMM-Full, `size_diff`, `size_ratio`, `absdiff`, and `zero_count` are among the leading indicators. `margin_sd` and `largediff_diff` remain informative but rank lower. Thus, the GMM diagnostics place more weight on adjustment-size and asymmetry measures than the K–means and hierarchical diagnostics, but they support the same reduced six-indicator specification.

As a robustness check, we reintroduce the Atlantic provinces and estimate a four-cluster specification. This allows regulated pricing environments to form their own group rather than being forced into the three-regime non-Atlantic benchmark. The additional cluster predominantly captures Atlantic markets, while the allocation of non-Atlantic city-quarters to the cycle, rigidity, and wholesale-price following regimes remains broadly consistent with the baseline three-cluster specification. Appendix C.2 reports the full results.

Figure 4: PI Using the Reduced Six-Indicator Set (Three-Cluster Specification)



6 Economic Analysis

In the previous section, we showed that the clustering algorithms divide observations into groups corresponding to cycle pricing, rigidity, and wholesale-price following behaviour. In this section, we examine whether those clusters are economically meaningful in terms of their persistence, location, and association with market characteristics. We also document the reduction in cycle pricing over the sample period and assess whether that shift remains after conditioning on observable factors, including market structure, wholesale-price conditions, market size, and pandemic-related variables. Subsection 6.1 presents the descriptive analysis of the clusters, while Subsection 6.2 reports the multinomial logit results.[24] Throughout, we use the K-means classification as the benchmark and use the alternative algorithms as robustness checks for the main descriptive patterns.

6.1 Descriptive Analysis

As noted above, over our sample period, K-means classified 386 city-quarter pairs as being in the wholesale-price following cluster, 286 in the cycle cluster, and 504 in the

24. We report an additional econometric check in Appendix D, where we examine whether the nonlinear HHI profile is sensitive to excluding observations in the right tail of the HHI distribution.

rigid cluster. Clusters exhibit a high degree of persistence, with a low frequency of cities switching from one cluster to another from quarter-to-quarter. A city in the wholesale-price following cluster in period t remains in the wholesale-price following cluster in $t+1$ in 93% of cases, switching to cycles in only 4% of cases and to rigid pricing in only 3% of cases. The rigid cluster shows a similar degree of persistence, with a city in the rigid cluster remaining rigid next quarter in 87% of cases, and switching to the cycle cluster in 10% of cases. Somewhat less persistence is exhibited by the cycle cluster; a cycling city remains in the cycle cluster next quarter in 75% of cases, switching to rigid pricing in 19% and to wholesale-price following pricing in 6%.

The persistence of the wholesale-price following cluster is highlighted by the fact that seven cities are in the cluster in all 28 quarters, while another three are in the wholesale-price following cluster in 27 out of 28 quarters; together, these ten cities account for 71% of all city-quarter pairs in the cluster. Nine of these cities are in Ontario (the exception being Vancouver), with eight consisting of Toronto and cities in the surrounding area. Notably, Toronto and Vancouver are Canada's first and third largest cities, respectively; the second largest city, Montreal, falls in the wholesale-price following cluster in 20 quarters.[25]

In contrast, rigid pricing is often observed in the smaller cities in our sample. The median (mean) population of city-quarter pairs in the rigid cluster is 100,558 (192,464), compared to 143,163 (277,061) for cycle city-quarters, and 266,170 (648,409) for city-quarters in the wholesale-price following cluster.[26] Cities in the rigid cluster also exhibit higher margins (retail price - wholesale price and estimated transportation costs) than the other two clusters; the median (mean) margin for city/quarters in the rigid cluster is 6.24cpl (6.42cpl), compared to 4.58cpl (4.40cpl) for the cycle cluster and 4.64cpl (4.92cpl) for the wholesale-price following cluster; note that there is little appreciable difference between median and mean margins in the latter two clusters.[27]

To evaluate the association between cluster and concentration, we obtained for each city annual counts of gasoline stations by brand from Kalibrate. This allowed us to compute annual HHI indices for each city, based on the percentage of total stations in the city accounted for by each brand.[28] Perhaps unintuitively, we find that the wholesale-price

25. In addition, there is one city (Prince Albert) that is in the rigid cluster for the entire sample period.

26. Populations are for Census Metropolitan Areas in the 2021 Canadian Census.

27. Recall that the level of the margin was not used as an input variable in the clustering algorithm with the reduced set of six variables.

28. Unfortunately, we were unable to obtain brand-level volume data; to the extent that per-station volume varies systematically by brand or company type, this may not be an accurate representation of market concentration. Also note that certain companies operate multiple brands within the same market; in computing concentration, we combined such brands into one firm. Finally, because we do not have contract data for individual stations, we assume that all stations of the same brand are of the same firm,

following cluster is associated with higher concentration, with a median (mean) value of 1576 (1572) compared to 1458 (1489) for cycle cities/quarters, and 1458 (1450) for the rigid cluster. Dividing brands into national refinery brands, regional refinery brands, and brands not affiliated with a refinery, we find that the highest presence of national refinery brands occurs in the wholesale-price following cluster (mean market share of 58% compared to 53% for the cycle cluster and 51% for the rigid cluster). Regional refinery brands represent a higher proportion of stations in rigid-cluster cities than in the other clusters (mean market share of 20%, compared to 9% for wholesale-price following cluster cities and 12% for cycle cities).

Figure 5: Number of Cities in Each Cluster, 2017 - 2023

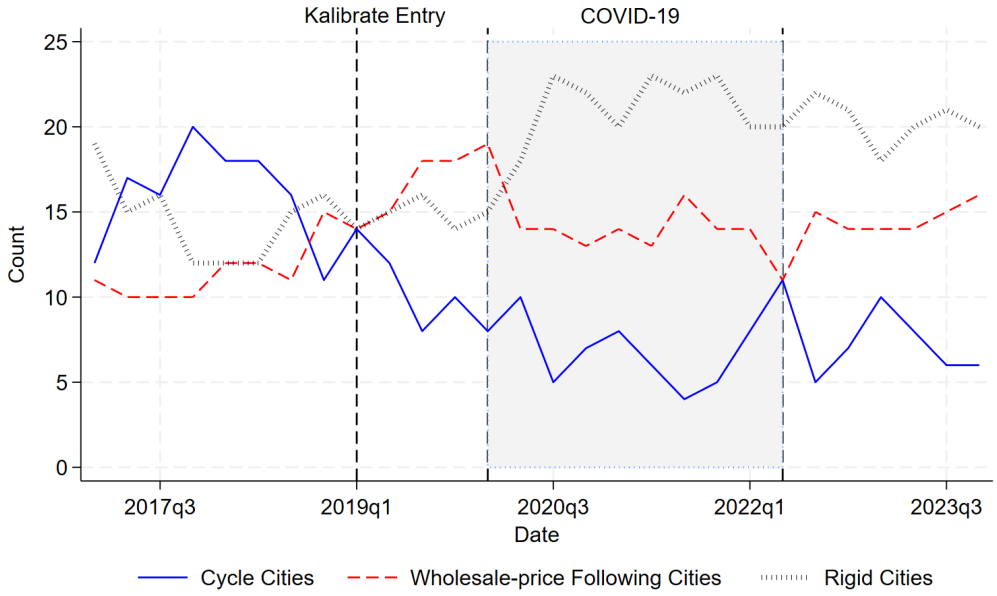


Figure 5 plots the number of cities in each cluster over our sample period. The leftmost vertical line corresponds to the acquisition of Kent Marketing by Kalibrate, while the shaded area represents the COVID-19 period, taken to be 2020 Q1 to 2022 Q2. Figure 5 shows a reduction in the number of cities exhibiting cycles over the sample period, from 22 in the fourth quarter of 2017 to six cities at the end of the sample period. In contrast, the number of cities exhibiting rigid prices increases, particularly in the first year of the COVID-19 pandemic. The timing of the decline in cycling overlaps with broader industry changes, including the period following Kalibrate's acquisition of Kent Marketing, but this evidence is descriptive and should not be interpreted as identifying a causal effect of Kalibrate or algorithmic pricing.

regardless of the contractual relationship between the station and the brand's parent company.

To describe further the changes in patterns exhibited by different cities over the sample period, Table 6 shows a transition matrix of cities exhibiting different patterns over the 2017-2018 period, and the patterns they exhibited post-COVID-19 (i.e. after the second quarter of 2022). Specifically, we identify the pattern exhibited in the most quarters over these two time periods; for example, if a city was in the cycle cluster for six of the eight quarters of 2017-2018, it would be identified as a cycle city over this time period. Over 2017-2018, out of the 42 cities in our sample, 10 were in the wholesale-price following cluster for the majority of quarters, 15 were primarily in the cycle cluster, and 17 were in the rigid cluster. In the period after Q2 2022, 14 cities were in the wholesale-price following cluster in the majority of periods, 7 were in the cycle cluster, and 21 were in the rigid cluster. Notably, all 10 cities that were classified as wholesale-price following over the first two years of the sample were also classified as wholesale-price following in the post-COVID-19 period. Similarly, of the 17 rigid cities over 2017-2018, all but two were still in the rigid cluster for the majority of the post-COVID period, with two switching to the cycle cluster. The main transition across these two periods is from the cycle cluster, which lost six cities to the rigid cluster and four to the wholesale-price following cluster.

Table 6: Transition Matrix: 2017-2018 to post-COVID-19

		post-COVID-19		
2017-2018	**Wholesale-price following**	**Cycle**	**Rigid**	**Total**
Wholesale-price following	10	0	0	**10**
Cycle	4	5	6	**15**
Rigid	0	2	15	**17**
Total	**14**	**7**	**21**	**42**

Notes: The table reports the benchmark K-means classification. The same pre/post transition matrix is obtained under hierarchical clustering. Under GMM-Spherical, the number of cities primarily classified as Cycle falls from 14 in 2017–2018 to five after Q2 2022. GMM-Spherical assigns relatively more observations to Rigidity and implies more switching between Wholesale-price Following and Rigidity, indicating that the decline in Cycle pricing is robust while the boundary between the two non-cycle regimes is more algorithm-sensitive.

6.2 Econometric Analysis

The descriptive evidence above points to a reallocation of pricing regimes over time. In this subsection, we examine whether the decline in cycle pricing remains after conditioning on observed changes in market structure, wholesale-price conditions, market size, and pandemic-related factors. This exercise is descriptive and hypothesis-generating rather than causal. For this purpose, we estimate a multinomial logit model that takes the K-means classification as the dependent variable. The three outcomes are *Cycle*, *Rigid*, and *Wholesale-price following*, with *Cycle* as the base outcome. The specification includes city

fixed effects, year fixed effects, market-structure controls, wholesale-price controls, market size, and a pandemic-related control. Since the relationship between concentration and pricing regimes may be nonlinear, the baseline specification includes both HHI_{1000} and its square, where HHI_{1000} is the Herfindahl-Hirschman Index divided by 1,000.[29]

The estimation sample excludes cities with no within-city variation in the K-means regime over the sample period. This removes the seven cities that are always classified as *Wholesale-price following*, and Prince Albert, which is always classified as *Rigid*. The resulting sample contains 952 city-quarter observations: 286 classified as *Cycle*, 476 as *Rigid*, and 190 as *Wholesale-price following*.

Table 7: Summary Statistics for Regressors Used in the Baseline Multinomial Logit

Variable	Unit/coding	N	Mean	S.D.	Min	P25	Median	P75	Max
HHI_{1000}	HHI / 1000	952	1.475	0.275	0.968	1.262	1.478	1.611	2.510
HHI^2_{1000}	squared HHI_{1000}	952	2.250	0.896	0.938	1.593	2.184	2.595	6.302
Independent share	share (0–1)	952	0.324	0.124	0.061	0.238	0.313	0.412	0.609
Major-brand share	share (0–1)	952	0.531	0.094	0.304	0.474	0.537	0.587	0.778
Log population	log(population)	952	11.995	1.045	9.894	11.303	11.856	12.387	14.427
Mean wholesale price	real cpl	952	80.476	20.755	40.785	65.620	77.151	94.584	148.563
Wholesale-price s.d.	real cpl	952	5.790	3.913	1.327	2.931	4.006	8.037	22.513
Log COVID incidence	$\log(\text{incidence} + 10^{-8})$	952	-12.905	7.892	-18.421	-18.421	-18.421	-3.034	1.947

Notes: Summary statistics are computed on the baseline estimation sample. City and year fixed effects are omitted because they are indicator variables. Share variables are coded on the unit interval. HHI_{1000} is the Herfindahl-Hirschman Index divided by 1,000. Wholesale prices are expressed in real cents per litre (cpl). Log COVID incidence is defined as $\log(\text{new patient normalized} + 10^{-8})$.

Table 7 reports summary statistics for the regressors used in the baseline specification. *Independent share* and *Major-brand share* measure outlet-based market shares of independent stations and major refinery brands, with the share of stations affiliated with regional refinery brands being the omitted category. *Log population* is the natural logarithm of annual city population. *Mean wholesale price* and *Wholesale-price s.d.* are measured in real cents per litre. The pandemic control is $\log(\text{new patient normalized} + 10^{-8})$, where *new patient normalized* is the provincial new-case measure normalized by population.[30]

Because the multinomial logit includes many city fixed effects in a relatively short panel, the estimates may be sensitive to incidental-parameter bias. To assess this sensitivity, we implement a split-panel jackknife correction.[31] In practice, we randomly split the

29. As a check, we also estimated a specification with a linear HHI_{1000} term. The main time-pattern conclusion is unchanged: cycle pricing remains substantially less common from 2019 onward after conditioning on market structure, wholesale-price conditions, market size, and pandemic-related controls.

30. We add 10^{-8} before taking logs to keep zero-case observations in the logarithmic transformation. We also examined alternative treatments of pandemic controls in robustness checks. These alternatives do not qualitatively change the main time-pattern results.

31. The split-panel jackknife follows the general bias-correction logic for nonlinear fixed-effects panels

panel into two halves, estimate the model on the full sample and on each half, construct the corresponding jackknife-corrected estimator, and repeat this procedure 100 times using different random splits. In the baseline AME exercise, 81 of the 100 random splits produced valid corrected estimates for all three outcomes. The HHI-profile calculations are conducted separately and yield 80 successful replications. We report each set of results as averages across the successful replications for the corresponding exercise.

Table 8: Average Marginal Effects from the Baseline Multinomial Logit

	Cycle		**Rigid**		**Wholesale-price following**	
Variable	**dy/dx**	**SE**	**dy/dx**	**SE**	**dy/dx**	**SE**
2018	0.0064	0.0107	-0.0237***	0.0063	0.0174	0.0145
2019	-0.0845***	0.0115	0.0067	0.0071	0.0778***	0.0159
2020	-0.2292***	0.0129	0.1163***	0.0080	0.1129***	0.0169
2021	-0.2777***	0.0132	0.1513***	0.0086	0.1264***	0.0167
2022	-0.2678***	0.0152	0.1982***	0.0098	0.0696***	0.0181
2023	-0.2120***	0.0151	0.1027***	0.0107	0.1094***	0.0200
HHI_{1000}	-0.1050***	0.0287	0.1153***	0.0202	-0.0103	0.0314
Independent share	0.3047***	0.0960	-0.2169***	0.0612	-0.0878	0.0815
Major-brand share	-0.1956	0.1222	0.2322***	0.0706	-0.0367	0.1026
Log population	0.4477***	0.1025	0.1496*	0.0878	-0.5973***	0.0818
Mean wholesale price	0.0005***	0.0001	-0.0006***	0.0001	0.0001	0.0001
Wholesale-price s.d.	0.0069***	0.0004	-0.0084***	0.0004	0.0016***	0.0003
Log COVID incidence	0.0055***	0.0003	-0.0012***	0.0003	-0.0043***	0.0003

Notes: Entries are average marginal effects and standard errors are reported in the adjacent columns. The model includes city and year fixed effects, with 2017 as the omitted year. The baseline specification includes both HHI_{1000} and HHI^2_{1000}; the HHI_{1000} row reports the average marginal effect of HHI accounting for both terms. Entries are based on the average of the split-panel jackknife-corrected estimates across 81 successful random splits. HHI_{1000} = HHI/1000, so a 100-point increase in the raw HHI corresponds to 0.1 unit in HHI_{1000}. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

We focus on AMEs because they summarize the multinomial logit results in terms of changes in predicted probabilities, which makes the magnitude of the associations across regimes easier to interpret. For continuous regressors, an AME gives the average change in the predicted probability of belonging to a given cluster associated with a one-unit increase in the corresponding regressor. For indicator variables, such as the year dummies, it gives the average discrete change in predicted probability when the indicator switches from 0 to 1 relative to the omitted category. The year effects should therefore be read as residual time

developed by Hahn and Newey (2004), Dhaene and Jochmans (2015), and Fernández-Val and Weidner (2016). We also draw on the recent discussion of fixed-effects multinomial logit in Chen et al. (2025). Following guidance on reporting failures and non-convergence in repeated computational procedures (Pawel et al. 2026) and on summarizing repeated simulation estimates (White 2010), we report the number of successful SPJ splits used to compute the AMEs.

shifts relative to 2017, conditional on the included covariates. Because several regressors are shares coded on the unit interval, it is often more natural to interpret them in 0.10 increments, that is, as 10-percentage-point changes.

Table 8 reports the baseline AMEs from the split-panel jackknife multinomial logit. The strongest result concerns the time pattern. Relative to 2017, there is no statistically meaningful decline in the probability of belonging to the cycle cluster in 2018. The decline becomes clear from 2019 onward: the estimated probability of cycling is about 8.5 percentage points lower in 2019, 22.9 points lower in 2020, 27.8 points lower in 2021, 26.8 points lower in 2022, and 21.2 points lower in 2023. This conditional time pattern is consistent with the descriptive evidence: after accounting for observed changes in market structure, wholesale-price conditions, market size, and pandemic-related conditions, cycle pricing remains substantially less common in the later part of the sample.

The probability mass associated with the decline in cycling is split between the other two regimes. In 2019, the decline in cycling is almost entirely mirrored by a higher probability of wholesale-price following. During 2020–2022, both rigidity and wholesale-price following increase, with rigidity accounting for a larger share of the offsetting change in 2022. In 2023, both non-cycle regimes remain more likely than in 2017.

Wholesale-price conditions also help characterize regime membership. A one-unit increase in the within-quarter standard deviation of wholesale prices, measured in real cents per litre, is associated with an increase of about 0.69 percentage points in the probability of the cycle regime, a reduction of about 0.84 points in the probability of rigidity, and an increase of about 0.16 points in the probability of wholesale-price following. The sign pattern suggests that more volatile upstream conditions are less compatible with prolonged rigidity and more compatible with regimes in which retail prices adjust more frequently. Higher mean wholesale prices are also associated with a slightly higher probability of cycling and lower probability of rigidity, although the magnitudes are small.

The market-structure controls also have informative sign patterns. A 10-percentage-point increase in the independent-station share is associated with a roughly 3.0-percentage-point increase in the probability of cycling and a 2.2-point reduction in the probability of rigidity. A 10-percentage-point increase in the major-brand share is associated with a roughly 2.3-point increase in the probability of rigidity. These associations should be read descriptively because station-type shares vary substantially across cities and may proxy for persistent differences in local retail structure that are not fully summarized by market shares alone.

The concentration result requires separate discussion because the baseline specification

is quadratic in HHI and because HHI varies largely across cities rather than within cities. The HHI_{1000} row in Table 8 is an average marginal effect that accounts for both the linear and squared HHI terms. It should not be read as a single linear concentration effect. Table 9 therefore reports selected predicted probabilities and local marginal effects from the quadratic specification. Predicted probabilities are computed by setting HHI_{1000} to the indicated value and averaging over the observed distribution of the remaining covariates. The local marginal effects are derivatives with respect to HHI_{1000} at the indicated value, again accounting for the quadratic term.

Table 9: Predicted Cycle and Rigid Probabilities at Selected HHI Values

HHI_{1000}	Location in support	Pr(Cycle)	Pr(Rigid)	AME on Pr(Cycle)
1.0	low HHI	0.506	0.250	-0.694***
1.3	near P25	0.327	0.363	-0.393***
1.5	near median	0.280	0.395	-0.091***
1.6	near P75	0.277	0.401	0.034*
2.0	right tail	0.386	0.350	0.467***

Notes: Predicted probabilities and local marginal effects are computed from the baseline quadratic HHI specification. The model sets HHI_{1000} to the indicated value and averages predictions over the observed covariate distribution. The AME column reports the local marginal effect of HHI_{1000} on the probability of the cycle regime at each HHI value. A 100-point increase in the raw HHI corresponds to one-tenth of the reported AME. The values 1.3, 1.5, and 1.6 are close to the 25th percentile, median, and 75th percentile of HHI_{1000} in the estimation sample. Entries are based on 80 successful SPJ replications. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 9 shows that the relationship between concentration and cycling is non-monotonic. Over the lower and middle parts of the HHI distribution, the predicted probability of cycling falls as HHI rises. The predicted probability of cycling is about 0.506 at $HHI_{1000} = 1.0$, falls to about 0.327 near the first quartile, and reaches about 0.280 near the median. The local marginal effect is negative at these values and remains negative near the median. Around the upper quartile, the marginal effect is close to zero and only marginally significant. The pattern changes in the right tail: at $HHI_{1000} = 2.0$, the predicted probability of cycling rises to about 0.386 and the local marginal effect becomes positive.

This differs from a simple monotonic concentration story and from the cleaner relationship one might expect from standard theory or from the comparison with Doyle, Muehlegger, and Samphantharak (2010). Much of the variation in HHI is cross-sectional across cities, and several cities exhibit limited within-city HHI variation during the sample period. With city fixed effects, persistent city-level pricing tendencies are absorbed by the fixed effects, so the HHI association is identified from a narrower source of variation. The right tail of the HHI distribution is also relatively sparse. For these reasons, we treat the

concentration estimates as descriptive evidence of a nonlinear association rather than as structural effects of concentration. Appendix D shows that the nonlinear pattern remains visible when the top 10 percent of HHI observations are excluded, although the increase in cycling in the upper tail becomes more muted.

The remaining controls help characterize regime membership but are less central to the main argument. Log population enters significantly for all three outcomes at the 10% level, increasing the predicted probabilities of cycling and rigidity and reducing the probability of wholesale-price following. Since the regressor is in logs, the reported AME corresponds to a one-unit increase in log population; a 1 percent increase in population corresponds approximately to 0.01 log points and therefore to about one hundredth of the reported AME. Note that the finding that the probability of wholesale-price following decreases in population is at odds with the descriptive evidence that the average population was greater in wholesale-price following cities than other cities. However, in the multinomial logit we have removed the largest cities in the sample because they exhibited wholesale-price following behaviour in all quarters; as well, city fixed effects mean that the effect estimated in Table 8 is identified off of limited within-city population changes. The pandemic control is positively associated with cycling and negatively associated with wholesale-price following.

The AME results show that the decline in cycle pricing from 2019 onward remains after conditioning on observed market structure, wholesale-price conditions, market size, and pandemic-related conditions. The estimated year effects capture residual shifts in regime membership relative to 2017. This pattern is consistent with broader changes in the industry after 2018, including the period following Kalibrate's acquisition of Kent Marketing, but the model does not identify a causal effect of that entry. The econometric exercise therefore reinforces the descriptive evidence: the decline in cycling remains the central pattern, while the covariates help describe the market conditions associated with each pricing regime.

7 Conclusion

Retail gasoline prices exhibit different dynamic patterns in different jurisdictions, including rigidity, wholesale-price following, and Edgeworth cycles. To date, the empirical literature studying these cycles has largely relied on visual inspection or simple rules of thumb and basic indicators to classify markets as following different patterns. In this paper, using Canadian city-quarter data from 2017 to 2023, we construct indicators grounded in the Edgeworth-cycle literature and apply unsupervised clustering methods to classify pricing dynamics without ex ante visual labelling. We then use supervised classifiers and

permutation importance to interpret the resulting assignments. The approach recovers three economically meaningful regimes: cycle pricing, rigid pricing, and wholesale-price following.

Our analysis points to certain indicators playing a more important role in generating stable clusters that correspond to the types of pricing patterns recognized in the literature. In particular, variables measuring the frequency of first differences equal to zero, the average size of the first difference in absolute value, and the size of price increases relative to decreases, appear to be strongly relied on by the clustering algorithms. In contrast, measures of the retail-wholesale margin, the relative length of negative to positive runs, and our measure of periodicity, were not found to be important in the clustering exercise, and were excluded from the final feature set. Our results therefore suggest focusing on first differences of the retail price when trying to distinguish gasoline markets exhibiting different pricing patterns.

Regarding how our clusters relate to market structure and other economic variables, rigid pricing is more common in smaller cities and is associated with higher margins, while wholesale-price following is concentrated in the largest urban markets, especially in and around Toronto. Both the descriptive evidence and the econometric analysis point to a substantial shift away from cycles after 2018. This pattern remains after conditioning on observed market structure, wholesale-price conditions, market size, and pandemic-related variables. Its timing overlaps with the period following Kalibrate's acquisition of Kent Marketing and is consistent with the possibility that pricing software and related data services may have affected retail pricing dynamics.

Our analysis suggests several possible directions for future research. First, using station-level pricing data, across a country or other jurisdiction, would allow the methodology to be extended to not only identify markets exhibiting different patterns, but also the different roles played by stations in those patterns; this could involve, for example, distinguishing price leaders from price followers. A richer data set may also permit the identification of different regimes within broader categories; this is suggested by the idea in the literature that different forms of price cycles might represent relatively competitive markets, or possibly collusive behaviour. More generally, unsupervised clustering algorithms may be applied to pricing or bid data to identify groups of firms behaving differently from others, possibly as a way of identifying coordinated behaviour.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used OpenAI's ChatGPT to assist with language editing, consistency checks, development, refinement, and documentation of replication code based on author-designed procedures and existing scripts, and preparation of the replication-package documentation. The authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Altmann, André, Laura Tološi, Oliver Sander, and Thomas Lengauer. 2010. "Permutation Importance: A Corrected Feature Importance Measure." *Bioinformatics* 26 (10): 1340–1347.

Assad, Stephanie, Robert Clark, Daniel Ershov, and Lei Xu. 2024. "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market." *Journal of Political Economy* 132 (3): 723–771.

Athayde, Amanda, Carolina Araújo Ferreira, and Maria Diniz. 2025. *2.0 "What About Bob Test": Pricing Algorithms and a Proposal for New Additional Parameters that Can Inform Antitrust Investigations Around the World.* SSRN Working Paper. Available at SSRN. SSRN, October. https://doi.org/10.2139/ssrn.5554238. https://ssrn.com/abstract=5554238.

Atkinson, Benjamin. 2009. "Retail Gasoline Price Cycles: Evidence from Guelph, Ontario Using Bi-Hourly, Station-Specific Retail Price Data." *The Energy Journal* 30 (1): 85–109.

Atkinson, Benjamin, Andrew Eckert, and Douglas West. 2009. "Price Matching and the Domino Effect in a Retail Gasoline Market." *Economic Inquiry* 47 (3): 568–588.

Bishop, Christopher M, and Nasser M Nasrabadi. 2006. *Pattern Recognition and Machine Learning.* Vol. 4. 4. Springer.

Byrne, David P, and Nicolas De Roos. 2019. "Learning to Coordinate: A Study in Retail Gasoline." *American Economic Review* 109 (2): 591–619.

Cai, Jie, Jiawei Luo, Shulin Wang, and Sheng Yang. 2018. "Feature Selection in Machine Learning: A New Perspective." *Neurocomputing* 300:70–79.

Castanias, Rick, and Herb Johnson. 1993. "Gas Wars: Retail Gasoline Price Fluctuations." *The Review of Economics and Statistics,* 171–174.

Chen, Mingli, Ian Meeker, Marc Rysman, and Sarah Wang. 2025. "An MM Algorithm for Fixed Effects Multinomial Logit Models." Working Paper, Boston University. https://sites.bu.edu/mrysman/files/2025/08/MM_Paper.pdf.

Clark, Robert, and Jean-Francois Houde. 2013. "Collusion with Asymmetric Retailers: Evidence from a Gasoline Price-Fixing Case." *Americal Economic Journal: Microeconomics* 5 (3): 97–123.

Davies, David L, and Donald W Bouldin. 1979. "A Cluster Separation Measure." *IEEE Transactions on Pattern Analysis and Machine Intelligence,* no. 2, 224–227.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1): 1–38.

Dhaene, Geert, and Koen Jochmans. 2015. "Split-Panel Jackknife Estimation of Fixed-Effect Models." *The Review of Economic Studies* 82 (3): 991–1030.

Doyle, J., E. Muehlegger, and K. Samphantharak. 2010. "Edgeworth Cycles Revisited." *Energy Economics* 32:651–660.

Dy, Jennifer G, and Carla E Brodley. 2004. "Feature Selection for Unsupervised Learning." *Journal of machine learning research* 5 (Aug): 845–889.

Eckert, Andrew. 2002. "Retail Price Cycles and Response Asymmetry." *Canadian Journal of Economics/Revue Canadienne d'Économique* 35 (1): 52–77.

———. 2003. "Retail Price Cycles and the Presence of Small Firms." *International Journal of Industrial Organization* 21:151–170.

———. 2013. "Empirical Studies of Gasoline Retailing: A Guide to the Literature." *Journal of Economic Surveys* 27 (1): 140–166.

Fernández-Val, Ivan, and Martin Weidner. 2016. "Individual and Time Effects in Nonlinear Panel Models with Large N, T." *Journal of Econometrics* 192 (1): 291–312.

Foros, Øystein, Mai Nguyen-Ones, and Frode Steen. 2021. "The Effects of a Day Off from Retail Price Competition: Evidence on Consumer Behavior and Firm Performance in Gasoline Retailing." *International Journal of the Economics of Business* 28 (1): 49–87.

Foros, Øystein, and Frode Steen. 2013. "Vertical Control and Price Cycles in Gasoline Retailing." *Scandanavian Journal of Economics* 115 (3): 640–661.

Gnanadesikan, Ram. 2011. *Methods for Statistical Data Analysis of Multivariate Observations.* John Wiley & Sons.

Hahn, Jinyong, and Whitney Newey. 2004. "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models." *Econometrica* 72 (4): 1295–1319.

Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Vol. 2. Springer.

Holt, Timothy, Mitsuru Igami, and Simon Scheidegger. 2024. "Detecting Edgeworth Cycles." *The Journal of Law and Economics* 67 (1): 67–102.

Hosmer Jr, David W, Stanley Lemeshow, and Rodney X Sturdivant. 2013. *Applied Logistic Regression.* Vol. 398. John Wiley & Sons.

Hubert, Lawrence, and Phipps Arabie. 1985. "Comparing Partitions." *Journal of classification* 2:193–218.

Izenman, Alan J. 2008. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning.* Springer Texts in Statistics. New York, NY: Springer. ISBN: 978-0-387-78188-4. https://doi.org/10.1007/978-0-387-78189-1.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning.* Vol. 112. Springer.

Jolliffe, Ian. 2005. "Principal Component Analysis." *Encyclopedia of statistics in behavioral science.*

Kalibrate. 2018. "How to Tell If You're Contributing to Fuel Price Volatility." Kalibrate. Archived version available through the Internet Archive's Wayback Machine, September 5, 2018. Accessed June 1, 2026. https://web.archive.org/web/20210619002311/https://kalibrate.com/insights/blog/fuel-pricing/fuel-price-volatility/.

Kalibrate. 2019. "Kalibrate Acquires Kent Group Ltd." Kalibrate. Archived version available through the Internet Archive's Wayback Machine, February 1, 2019. Accessed June 1, 2026. https://web.archive.org/web/20251114030649/https://kalibrate.com/insights/press-release/uncategorized/kalibrate-kent-group/.

Klein, T. 2021. "Autonomous Algorithmic Collusion: Q-learning under Sequential Pricing." *RAND Journal of Economics* 52:538–599.

Landis, J Richard, and Gary G Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *biometrics,* 159–174.

Lewis, Matthew, and Michael Noel. 2011. "The Speed of Gasoline Price Response in Markets with and without Edgeworth Cycles." *Review of Economics and Statistics* 93 (2): 672–682.

Lloyd, Stuart. 1982. "Least Squares Quantization in PCM." *IEEE Transactions on Information Theory* 28 (2): 129–137.

MacQueen, J. 1967. "Classification and Analysis of Multivariate Observations." In *In 5th Berkeley Symposium on Mathematical Statistics and Probability,* 5:281–297. 1.

Manning, Christopher D. 2009. *An Introduction to Information Retrieval.* New York, NY, USA: Cambridge University Press.

Maskin, Eric, and Jean Tirole. 1988. "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles." *Econometrica* 56 (3): 571–599.

McLachlan, G.J. 1982. "The Classification and Mixture Maximum Likelihood Approaches to Cluster Analysis." In *Classification Pattern Recognition and Reduction of Dimensionality,* 2:199–208. Handbook of Statistics. Elsevier.

Murtagh, Fionn. 1983. "A Survey of Recent Advances in Hierarchical Clustering algorithms." *The Computer Journal* 26 (4): 354–359.

Noel, Michael. 2007. "Edgeworth Price Cycles, Cost-Based Pricing, and Sticky Pricing in Retail Gasoline Markets." *The Review of Economics and Statistics* 89 (2): 324–334.

———. 2008. "Edgeworth Price Cycles and Focal Prices: Computational Dynamic Markov Equilibria." *Journal of Economics and Management Strategy* 17:345–377.

Pawel, Samuel, František Bartoš, Björn S Siepe, and Anna Lohmann. 2026. "Handling Missingness, Failures, and Non-convergence in Simulation Studies: A Review of Current Practices and Recommendations." *The American Statistician* 80 (1): 31–48.

Real, Raimundo, and Juan M. Vargas. 1996. "The Probabilistic Basis of Jaccard's Index of Similarity." *Systematic Biology* 45 (3): 380–385.

Revelle, William. 1979. "Hierarchical Cluster Analysis and the Internal Structure of Tests." *Multivariate Behavioral Research* 14 (1): 57–74.

Silveira, Douglas, Lucas B de Moraes, Eduardo PS Fiuza, and Daniel O Cajueiro. 2023. "Who Are You? Cartel Detection Using Unlabeled Data." *International Journal of Industrial Organization* 88:102931.

Vinh, Nguyen Xuan, Julien Epps, and James Bailey. 2009. "Information Theoretic Measures for Clusterings Comparison: Is a Correction for Chance Necessary?" In *Proceedings of the 26th annual international conference on machine learning,* 1073–1080.

Ward Jr, Joe H. 1963. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American statistical association* 58 (301): 236–244.

White, Ian R. 2010. "simsum: Analyses of Simulation Studies Including Monte Carlo Error." *The Stata Journal* 10 (3): 369–385.

Williams, Byron K. 1982. "A Simple Demonstration of the Relationship between Classification and Canonical Variates Analysis." *The American Statistician* 36 (4): 363–365.

Zimmerman, P., J. Yun, and T. Taylor. 2013. "Edgeworth Price Cycles in Gasoline: Evidence from the United States." *Review of Industrial Organization* 42:297–320.

Appendix A Machine Learning Methods

In Appendix A, we provide details on the different algorithms employed in this study. Subsection A.1 describes the unsupervised learning methods used to cluster the data, while Appendix A.2 describes the supervised machine learning techniques used in the permutation importance analysis.

Subsection A.1 Unsupervised Learning

A.1.1 *K*-means

K-means clustering partitions n observations into $K\,(\leq n)$ clusters, where each observation belongs to the cluster with the nearest mean (MacQueen 1967; Lloyd 1982; Gnanadesikan 2011). The algorithm refines clusters iteratively to minimize the sum of squared distances between observations and their cluster centroids. Figure A.1 outlines the steps. Although K-means is not a fully probabilistic model, its distance-based objective implicitly favors clusters that are compact, similarly scaled, and approximately spherical in the feature space. To find the optimal number of clusters K, we use the Silhouette score (Izenman 2008). This metric measures clustering quality by balancing intra-cluster cohesion and inter-cluster separation. For a clustering C_K with K clusters, the silhouette score $s_{i\,K}$ for the ith item is:

$$s_{i\,K} = \frac{b_i - a_i}{\max\{a_i,\, b_i\}},$$

where a_i is the mean intra-cluster distance, and b_i is the mean distance to the nearest cluster. The score ranges from -1 to 1, with values close to 1 showing well-clustered points and values near -1 indicating poor clustering.

The Davies-Bouldin Index (DBI) offers complementary insights by evaluating clustering quality through intra-cluster dispersion and inter-cluster separation (Davies and Bouldin 1979).

$$DBI = \frac{1}{K}\sum_{i=1}^{K} \max_{j \neq i} \left(\frac{s_i + s_j}{d_{ij}} \right)$$

where s_i and s_j are the average intra-cluster distances for clusters C_i and C_j, and d_{ij} is the distance between their centroids. A lower DBI indicates better clustering quality.

Figure A.1: The Lloyd's algorithm for K-means clustering.

procedure LLOYDS
  Choose K points as initial cluster centroids.
  while Clusters C change **do**
    for all $x \in X$ **do**
      Find cluster C_k with centroid c_k closest to x (based on distance).
      Assign x to C_k.
    end for
    for all Clusters k **do**
      Recalculate c_k as the mean of all points in C_k.
    end for
  end while
end procedure

A.1.2 Hierarchical Clustering

Hierarchical clustering organizes data into a tree-like structure called a dendrogram, which represents nested groupings of observations (Ward Jr 1963; Murtagh 1983). The algorithm builds clusters by either starting with individual observations and merging them iteratively (agglomerative clustering) or starting with all observations in one cluster and splitting them (divisive clustering). We focus on agglomerative clustering.

Figure A.2: The agglomerative clustering algorithm for hierarchical clustering.

procedure AGGLOMERATIVE CLUSTERING
  Compute the pairwise distance matrix for all observations.
  while More than one cluster exists **do**
    Identify the two closest clusters based on a linkage criterion (e.g., single, complete, average).
    Merge the two closest clusters.
    Update the distance matrix to reflect the new cluster.
  end while
end procedure

Figure A.2 outlines the steps. This method does not require specifying the number of clusters in advance. Instead, the number of clusters can be determined by cutting the dendrogram at a chosen height, which represents the similarity threshold for cluster separation.[32] Alternatively, the Silhouette score and Davies-Bouldin Index – commonly

32. A dendrogram is a tree-like diagram that illustrates how data points are merged into clusters at

used for K-means – can also assess clustering quality in hierarchical methods.

A.1.3 Gaussian Mixture Model (GMM)

GMM clusters data by modeling it as a mixture of Gaussian distributions (Dempster, Laird, and Rubin 1977; McLachlan 1982; Bishop and Nasrabadi 2006; Izenman 2008). Each cluster represents a Gaussian component, defined by its mean, covariance, and weight. This probabilistic approach offers a flexible alternative to methods like K-means. GMM handles complex data structures effectively by adapting to the distribution of the data.

Suppose we have a dataset $X = x_1, \cdots, x_N$ consisting of N observations of a D-dimensional variable. GMM assumes that each observation x_n is generated by a mixture of K Gaussian components. The density function for an observation x_n is:

$$p(x_n) = \sum_{k=1}^{K} \pi_k p(x_n|\mu_k, \Sigma_k), \tag{A.1}$$

where π_k represents the mixing probability of cluster k, $p(x_n|\mu_k, \Sigma_k)$ is the Gaussian likelihood, μ_k is the mean vector, and Σ_k is the covariance matrix for cluster k.

The Gaussian likelihood is given by:

$$p(x_n|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k)\right), \tag{A.2}$$

where $|\Sigma_k|$ is the determinant of Σ_k, and T denotes the transpose.

We estimate the parameters of GMM using the Expectation-Maximization (EM) algorithm. The EM algorithm iteratively computes the posterior probabilities of cluster membership (E-step) and updates the parameters π_k, μ_k, and Σ_k to maximize the log-likelihood of the observed data (M-step). The log-likelihood function for the dataset is:

$$\ell(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \log\left(\sum_{k=1}^{K} \pi_k p(x_n|\mu_k, \Sigma_k)\right). \tag{A.3}$$

The Bayesian Information Criterion (BIC) is commonly used to determine the optimal number of clusters K in GMM, with lower BIC values indicating better model fit.[33] The

different similarity levels, helping to visualize hierarchical relationships in the clustering process (Manning 2009).

33. However, combining BIC with other clustering validity metrics, such as those used for K-means (i.e., the Silhouette score and Davies-Bouldin Index), can provide a more comprehensive assessment and lead to a more robust decision regarding the optimal number of clusters.

specification of the covariance matrix determines the shape and orientation of clusters, significantly influencing the model's flexibility and complexity. The most common specifications are as follows. In the Full specification, each cluster is assigned its own covariance matrix, allowing for unique ellipsoidal shapes with distinct orientations and volumes. This provides maximum flexibility but increases the number of parameters to estimate. The Tied specification constrains all clusters to share a single covariance matrix, resulting in identical shapes and orientations, where differences between clusters are primarily driven by their mean vectors. This approach balances model complexity and interpretability. The Diagonal specification further simplifies the model by restricting covariance matrices to be diagonal, meaning clusters align with coordinate axes. This reduces the number of parameters while still allowing clusters to vary in size. The Spherical specification assumes covariance matrices are proportional to the identity matrix, producing clusters with equal variance in all directions. This is the simplest specification, enforcing isotropic clusters that are uniform in shape and orientation.[34]

Subsection A.2 Supervised Learning

A.2.1 Multinomial Logit

Let $y_i \in \{1, \ldots, K\}$ denote the pricing regime for observation i, and let x_i be the input (indicator) vector (including an intercept). We set class K as the baseline and estimate coefficient vectors $\beta_1, \ldots, \beta_{K-1}$. For $k = 1, \ldots, K-1$,

$$\Pr(y_i = k \mid x_i) = \frac{\exp(x_i^\top \beta_k)}{1 + \sum_{j=1}^{K-1} \exp(x_i^\top \beta_j)}, \qquad \Pr(y_i = K \mid x_i) = \frac{1}{1 + \sum_{j=1}^{K-1} \exp(x_i^\top \beta_j)}.$$

We fit the model by maximizing the multinomial log-likelihood (Izenman 2008; Hosmer Jr, Lemeshow, and Sturdivant 2013):

$$\mathcal{L}(\beta_1, \ldots, \beta_{K-1}) = \sum_{i=1}^{n} \left[\sum_{k=1}^{K-1} \mathbb{1}\{y_i = k\}\, x_i^\top \beta_k - \log\left(1 + \sum_{j=1}^{K-1} \exp(x_i^\top \beta_j)\right) \right].$$

We use the fitted model as a classifier and measure feature relevance with permutation importance.

34. Isotropic clusters have the same spread in all directions, meaning their shape is unaffected by rotation and remains perfectly spherical in multidimensional space.

A.2.2 Quadratic Discriminant Analysis (QDA)

Let $\{(x_i, y_i)\}_{i=1}^n$ denote the data, with $y_i \in \{1, \ldots, K\}$ and $x_i \in \mathbb{R}^D$. QDA classifies x_i by choosing the class that maximizes its posterior under class-specific Gaussian models (Williams 1982):

$$\hat{y}_i \;=\; \arg\max_{k \in \{1,\ldots,K\}} \; \pi_k \, \phi_D\big(x_i; \mu_k, \Sigma_k\big),$$

where π_k is the class prior, and ϕ_D is the D-variate Gaussian density

$$\phi_D(x; \mu, \Sigma) = (2\pi)^{-D/2} \, |\Sigma|^{-1/2} \exp\!\big(-\tfrac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\big)\,.$$

Equivalently, QDA uses the discriminant score

$$g_k(x) \;=\; -\tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}(x-\mu_k)^\top \Sigma_k^{-1}(x-\mu_k) + \log \pi_k,$$

and predicts the class with the largest $g_k(x)$. Allowing Σ_k to vary across classes yields quadratic decision boundaries. We use QDA as a classifier and rank indicators with permutation importance, following the same procedure used for the multinomial logit.

Appendix B Additional Tables

Tables B.1 and B.2 report summary statistics for the real price series used in the analysis over January 2, 2017 to December 29, 2023. Retail prices are measured net of taxes, and both retail and wholesale prices are expressed in real cents per litre.

Table B.1: Ex-tax Retail Price Summary Statistics, January 2 2017 to December 29 2023

City	Obs	Mean	Std. dev.	Min	Max	City	Obs	Mean	Std. dev.	Min	Max
Abbotsford	1,825	68.7	12.6	28.0	111.3	Barrie	1,825	56.4	10.1	20.4	87.6
Bathurst	1,825	57.8	10.0	26.3	88.0	Brandon	1,825	60.2	10.0	31.2	89.9
Brantford	1,825	56.5	9.8	21.0	86.7	Calgary	1,825	55.9	10.3	22.1	86.5
Campbellton	1,825	59.0	9.8	27.7	88.9	Charlottetown	1,825	58.5	10.0	28.5	87.8
Chicoutimi	1,825	59.2	10.6	26.6	94.6	City of Toronto	1,825	57.2	10.1	21.0	88.2
Corner Brook	1,825	61.0	10.4	31.1	93.2	Drummondville	1,825	59.0	11.0	29.7	92.4
Edmonton	1,825	54.4	10.2	21.1	84.3	Edmunston	1,825	58.3	10.1	27.1	88.6
Fort St. John	1,825	60.8	10.1	28.7	90.6	Fredericton	1,825	58.4	10.3	24.4	90.2
Gander	1,825	61.0	10.6	30.9	93.2	Gatineau	1,825	59.2	11.1	28.9	93.8
Grand Falls	1,825	64.2	10.5	32.6	96.1	Grande Prairie	1,825	56.0	9.9	23.7	85.1
Guelph	1,825	57.0	10.1	21.1	88.1	Halifax	1,825	58.6	11.0	25.4	94.4
Hamilton	1,825	56.6	9.9	20.9	87.1	Kamloops	1,825	67.2	12.2	33.0	105.8
Kelowna	1,825	67.4	12.1	33.4	105.8	Kentville	1,825	58.8	11.0	25.4	94.4
Kingston	1,825	55.3	10.2	20.8	87.4	Kitchener	1,825	57.1	10.1	21.2	88.2
Lethbridge	1,825	56.1	9.8	25.0	85.4	Lloydminster	1,825	55.5	10.0	24.5	84.8
London	1,825	56.5	10.0	21.0	87.3	Miramichi	1,825	58.6	10.0	26.9	88.3
Moncton	1,825	57.8	9.9	26.4	87.8	Montréal	1,825	58.9	10.6	28.1	92.1
Moose Jaw	1,825	61.5	10.3	31.2	93.3	New Glasgow	1,825	58.8	10.8	23.9	92.1
North Bay	1,825	58.8	10.5	26.6	92.5	Oshawa	1,825	57.0	9.9	20.9	88.0
Ottawa	1,825	57.0	10.8	19.7	91.1	Peterborough	1,825	55.2	10.1	21.9	87.3
Prince Albert	1,825	61.0	10.3	31.2	93.3	Prince George	1,825	61.8	10.2	29.7	92.3
Québec	1,825	59.7	11.4	30.0	94.5	Red Deer	1,825	54.7	10.2	21.0	84.3
Regina	1,825	61.0	10.3	31.2	93.3	Rimouski	1,825	60.5	10.8	30.2	93.2
Saint John	1,825	58.4	10.0	27.1	88.2	St. Johns	1,825	61.0	10.6	30.9	93.2
Saskatoon	1,825	61.0	10.3	31.2	93.3	Sault Ste. Marie	1,825	58.5	10.1	28.2	90.7
Sherbrooke	1,825	59.0	11.1	29.0	93.4	St. Catharines	1,825	56.4	10.2	21.6	87.7
Sudbury	1,825	59.7	10.6	29.7	93.7	Sussex	1,825	58.6	10.0	27.1	88.4
Sydney	1,825	58.9	11.0	25.5	94.4	Thunder Bay	1,825	60.8	10.7	30.3	95.0
Timmins	1,825	59.2	10.5	26.7	92.5	Trois-Rivières	1,825	59.2	11.2	29.4	93.3
Truro	1,825	58.7	11.0	25.4	94.4	Val-d'Or	1,825	60.9	10.8	27.0	95.7
Vancouver	1,825	69.4	12.0	29.0	108.9	Victoria	1,825	70.9	11.9	37.6	107.5
Windsor	1,825	55.8	10.2	21.6	87.5	Winnipeg	1,825	59.1	10.8	26.3	91.0
Woodstock	1,825	57.2	10.8	22.1	89.9	Yarmouth	1,825	58.7	10.8	23.8	92.0

Retail prices are observed for all cities in our sample, so Table B.1 summarizes city-specific retail distributions directly. By contrast, wholesale prices are not available for every city. To construct a balanced wholesale series for descriptive purposes, we assign to each city the wholesale price series from the geographically nearest city in our sample for which wholesale data are observed (based on a distance criterion). As a result, multiple cities may share the same underlying wholesale series, which is why some rows in Table B.2 report identical wholesale summary statistics.

Table B.2: Wholesale Price Summary Statistics, January 2 2017 to December 29 2023

City	Obs	Mean	Std. dev.	Min	Max	City	Obs	Mean	Std. dev.	Min	Max
Abbotsford	1,825	62.82	12.26	22.72	103.88	Barrie	1,825	51.34	10.04	16.73	83.78
Bathurst	1,825	52.57	10.16	17.37	84.85	Brandon	1,825	52.97	9.54	18.92	84.45
Brantford	1,825	51.35	10.04	17.16	83.84	Calgary	1,825	49.97	9.22	17.75	81.18
Campbellton	1,825	52.57	10.16	17.37	84.85	Charlottetown	1,825	50.95	10.43	17.37	84.48
Chicoutimi	1,825	54.21	10.44	21.45	88.96	City of Toronto	1,825	51.34	10.04	16.73	83.78
Corner Brook	1,825	50.72	10.20	17.94	83.73	Drummondville	1,825	54.21	10.44	21.45	88.96
Edmonton	1,825	48.40	9.31	16.07	79.89	Edmunston	1,825	52.57	10.16	17.37	84.85
Fort St. John	1,825	56.85	12.20	22.46	98.03	Fredericton	1,825	51.29	10.36	16.84	83.78
Gander	1,825	50.72	10.20	17.94	83.73	Gatineau	1,825	54.21	10.44	21.45	88.96
Grand Falls	1,825	50.72	10.20	17.94	83.73	Grande Prairie	1,825	49.90	9.24	17.48	81.26
Guelph	1,825	51.34	10.04	16.73	83.78	Halifax	1,825	51.82	10.47	17.47	86.11
Hamilton	1,825	51.34	10.04	16.73	83.78	Kamloops	1,825	62.82	12.26	22.72	103.88
Kelowna	1,825	62.82	12.26	22.72	103.88	Kentville	1,825	51.82	10.47	17.47	86.11
Kingston	1,825	51.22	10.06	16.95	83.91	Kitchener	1,825	51.34	10.04	16.73	83.78
Labrador City	1,825	50.72	10.20	17.94	83.73	Lethbridge	1,825	49.66	9.14	17.28	80.52
Lloydminster	1,825	49.90	9.24	17.48	81.26	London	1,825	50.99	10.01	16.37	83.21
Miramichi	1,825	52.57	10.16	17.37	84.85	Moncton	1,825	52.57	10.16	17.37	84.85
Montréal	1,825	54.21	10.44	21.45	88.96	Moose Jaw	1,825	53.68	10.02	18.72	87.49
Nanaimo	1,825	62.82	12.26	22.72	103.88	New Glasgow	1,825	51.82	10.47	17.47	86.11
North Bay	1,825	53.40	9.96	19.24	85.49	Oshawa	1,825	51.34	10.04	16.73	83.78
Ottawa	1,825	51.22	10.06	16.95	83.91	Peterborough	1,825	51.22	10.06	16.95	83.91
Prince Albert	1,825	53.96	10.01	19.02	87.72	Prince George	1,825	56.85	12.20	22.46	98.03
Québec	1,825	54.21	10.44	21.45	88.96	Red Deer	1,825	49.90	9.24	17.48	81.26
Regina	1,825	52.97	9.54	18.92	84.45	Rimouski	1,825	54.21	10.44	21.45	88.96
Saint John	1,825	52.57	10.16	17.37	84.85	Saskatoon	1,825	52.97	9.54	18.92	84.45
Sault Ste. Marie	1,825	53.40	9.96	19.24	85.49	Sherbrooke	1,825	54.21	10.44	21.45	88.96
St. Catharines	1,825	51.35	10.04	17.16	83.84	St. Johns	1,825	50.72	10.20	17.94	83.73
Sudbury	1,825	53.40	9.96	19.24	85.49	Sussex	1,825	51.29	10.36	16.84	83.78
Sydney	1,825	51.82	10.47	17.47	86.11	Thunder Bay	1,825	54.97	9.98	20.82	88.86
Timmins	1,825	53.40	9.96	19.24	85.49	Trois Rivieres	1,825	54.21	10.44	21.45	88.96
Truro	1,825	51.82	10.47	17.47	86.11	Val-Dor	1,825	54.21	10.44	21.45	88.96
Vancouver	1,825	62.82	12.26	22.72	103.88	Victoria	1,825	62.82	12.26	22.72	103.88
Windsor	1,825	50.99	10.01	16.37	83.21	Winnipeg	1,825	52.97	9.54	18.92	84.45
Woodstock	1,825	50.99	10.01	16.37	83.21	Yarmouth	1,825	51.82	10.47	17.47	86.11

Appendix C Additional Results

Appendix C presents two different sets of additional clustering results. Subsection C.1 describes the results of clustering using the full set of 21 variables, and provides additional detail on the process reducing the feature set to six variables. Subsection C.2 presents the results of clustering when the Atlantic provinces are included, and four clusters are generated.

Subsection C.1 Full Indicator Set and Variable Reduction

This appendix documents the analysis underlying the reduction from the full set of 21 candidate indicators to the six-indicator specification used in the main clustering results. The full set combines cycle-oriented indicators, margin-based indicators, rigidity measures, wholesale-volatility measures, and retail-wholesale adjustment measures. This broad candidate set is useful because it allows us to evaluate indicators associated with

the Edgeworth-cycle detection literature, including MBPI variables and the Lomb–Scargle periodogram, while also considering variables designed to capture rigidity and wholesale-price following. The purpose of the reduced set is to avoid overweighting highly similar indicators in the clustering step. Several variables in the 21-indicator set are variants of the same underlying screen, especially within the MBPI family and its margin-based counterparts. We therefore combine three pieces of evidence: internal cluster-number diagnostics using the full set, correlation patterns among indicators, and permutation-importance rankings from supervised classifiers trained to recover the unsupervised cluster labels.

Table C.1: Cluster-Number Diagnostics Using the Full 21-Indicator Set

Algorithm	**Criterion**	$k = 2$	$k = 3$	$k = 4$	**Preferred** k
K-means	Silhouette Score	0.274	0.229	0.228	2
K-means	Davies–Bouldin Index	1.465	1.535	1.575	2
Hierarchical	Silhouette Score	0.233	0.217	0.209	2
Hierarchical	Davies–Bouldin Index	1.663	1.748	1.605	4
GMM-Spherical	AIC	62548.663	58612.650	56004.808	4
GMM-Spherical	BIC	62776.807	58957.402	56466.167	4
GMM-Tied	AIC	37560.946	36918.451	37770.110	3
GMM-Tied	BIC	38950.091	38419.134	39382.330	3
GMM-Diagonal	AIC	52377.931	44156.169	41694.444	4
GMM-Diagonal	BIC	52808.870	44805.113	42561.392	4
GMM-Full	AIC	27833.163	23970.827	21751.996	4
GMM-Full	BIC	30393.450	27813.791	26877.639	4

Notes: For the Silhouette Score, higher values indicate better separation. For the Davies–Bouldin Index, AIC, and BIC, lower values indicate better fit. The table reports the two-, three-, and four-cluster solutions because these are the relevant alternatives for the three-regime taxonomy used in the main analysis.

Cluster-number diagnostics using the full 21-indicator set

Table C.1 reports the internal cluster-number diagnostics using the full 21-indicator set. The distance-based criteria do not provide a clear three-cluster recommendation. For K-means, both the Silhouette Score and the Davies–Bouldin Index favor two clusters. For hierarchical clustering, the Silhouette Score favors two clusters, while the Davies–Bouldin Index favors four clusters. GMM provides mixed evidence: the spherical, diagonal, and full-covariance specifications favor four clusters by AIC and BIC, whereas the tied-covariance specification favors three clusters. These results indicate that the full 21-indicator feature space contains useful information but does not yield a clean data-driven recommendation for the three-regime taxonomy used in the main text. This ambiguity is consistent with the correlation patterns discussed below: when several indicators capture

closely related dimensions of price adjustment, internal fit measures may favor partitions that reflect redundant variation rather than economically distinct pricing regimes.

Table C.2 reports per-class agreement relative to the K-means labels. GMM-Spherical shows high agreement across all three classes. GMM-Full shows moderate agreement for rigidity and cycle, but weak agreement for wholesale-price following. Hierarchical clustering shows the opposite pattern: it performs well for wholesale-price following, but shows weaker agreement for rigidity and cycle. This reinforces the central point that the full feature set contains meaningful signal, but also generates algorithm-specific differences in how the feature space is partitioned.

Table C.2: Per-Class Agreement with K-Means Labels Using Full Set of 21 Indicators

Model	Rigidity (%)	Cycle (%)	Wholesale-price following (%)
GMM-Spherical	**86.0**	**100.0**	**83.0**
GMM-Full	62.0	63.0	17.0
Hierarchical	17.0	56.0	90.0

Correlation and redundancy among indicators

The correlation diagnostics clarify why the full set can overweight some dimensions of pricing behaviour. Figure C.1 reports the correlation heatmap for the full 21-indicator set before transformation. The heatmap shows clear blocks of related variables, especially among the MBPI threshold indicators, their margin-based counterparts, and several measures of price variability and asymmetry. These blocks indicate where multiple indicators are capturing similar underlying pricing dynamics.

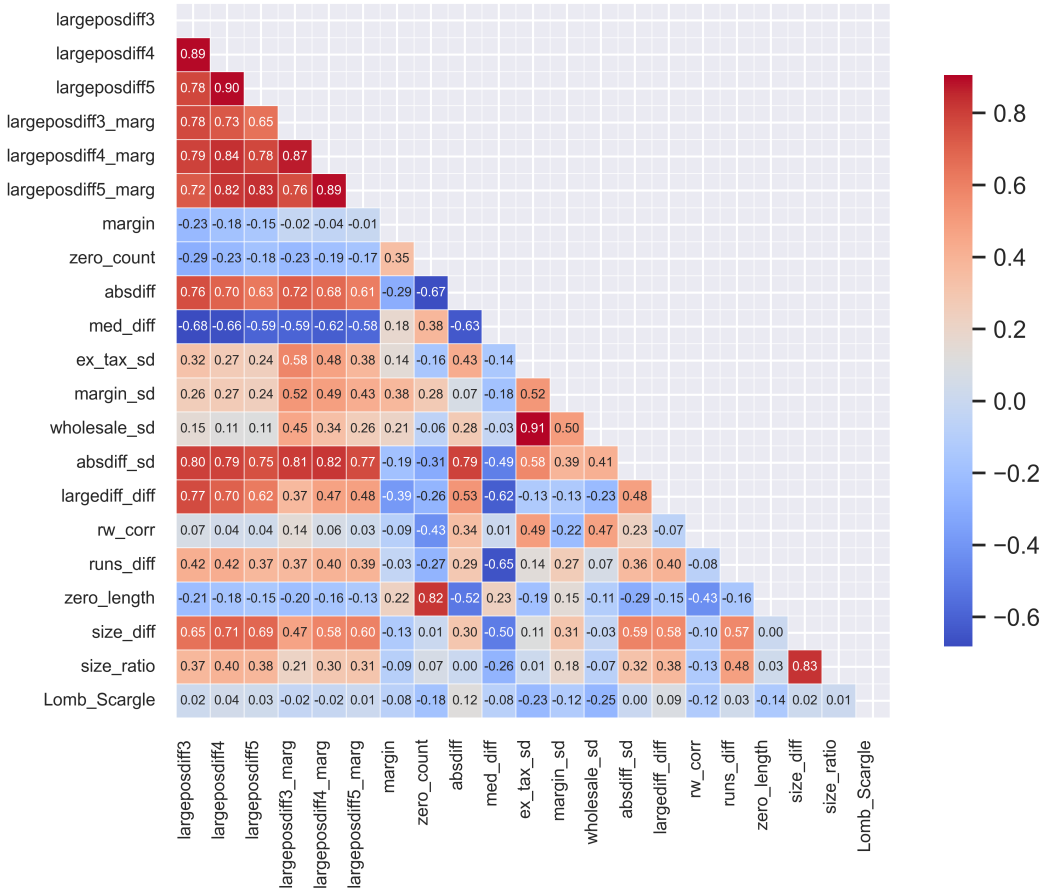
The strongest redundancy appears within the MBPI family. The retail MBPI threshold variables are highly correlated with one another, and the same pattern appears among the margin-based MBPI variables. The heatmap also shows positive correlations between the retail and margin MBPI measures, indicating that many of these indicators capture closely related restoration behaviour. Retaining all MBPI variants would therefore give excessive weight to the same cycle-restoration dimension. This motivates the use of `largediff_diff` as a compact measure of MBPI-type asymmetry that also incorporates wholesale-price movements.

The heatmap also highlights overlap among rigidity and volatility measures. `zero_count` and `zero_length` move closely together, so `zero_count` provides a compact representation of retail price stickiness. At the same time, `zero_count` is negatively related to `absdiff`, consistent with the idea that markets with more unchanged prices tend to display less short-run retail price movement. We retain both variables because they capture

complementary aspects of pricing behaviour: stickiness and day-to-day variability.

A similar logic applies to volatility. `ex_tax_sd` and `wholesale_sd` are strongly correlated, indicating that both variables largely reflect a common broad volatility dimension. By contrast, `margin_sd` remains related to these variables while capturing variability in the retail-wholesale margin more directly. We therefore retain `margin_sd` in the reduced set. Among the asymmetry measures, `size_diff` and `size_ratio` are also strongly correlated, but both are kept because they summarize asymmetric price adjustment in complementary ways: one in levels and the other as a scale-free ratio.

Figure C.1: Correlation Heatmap for the Full 21-Indicator Set



The Lomb–Scargle periodogram plays a different role. It generally has low correlations with the other indicators, so it provides a distinct periodicity signal rather than a redundant one, as Figure C.1 shows. We therefore keep it in the full-set analysis as part of the broader cycle-detection evidence. The final reduced specification, however, is designed to separate three pricing regimes – cycle pricing, rigidity, and wholesale-price following – rather than to maximize periodicity detection alone. This is why the main specification

retains the six indicators that provide the clearest and most parsimonious three-regime separation.

Permutation-importance evidence

We next use permutation importance to assess which variables help recover the cluster assignments generated by the unsupervised models. Specifically, we apply Quadratic Discriminant Analysis (QDA) to classify the labels generated by GMM, using both Full and Spherical covariance specifications, and a multinomial Logit model to classify the labels produced by K-means and hierarchical clustering. This supervised step does not replace clustering. It provides a diagnostic summary of which indicators carry the most information about the cluster assignments (Silveira et al. 2023).

Figure C.2: Permutation Importance Analysis Using the Full Set of 21 Indicators

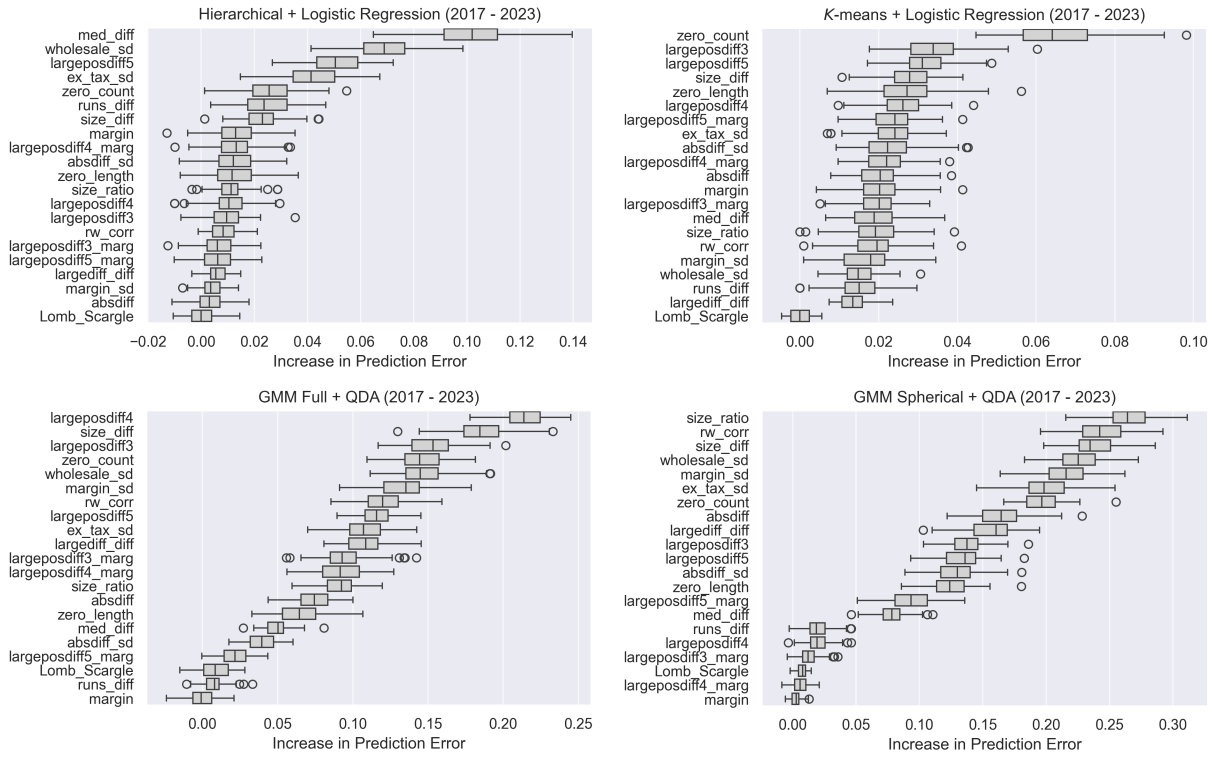


Figure C.2 reports the permutation-importance profiles for the full 21-indicator set. The profiles show that the most informative variables vary across clustering-supervised learning combinations. Some variables rank highly in specific models, but the full rankings are not stable across all methods. This is why the reduction does not rely on permutation importance alone. Instead, we combine permutation-importance rankings with correlation evidence and economic interpretation.

Table C.3 summarizes the cross-model rankings for the retained variables and for selected variables not retained in the final six-indicator set. The retained variables combine predictive relevance, economic interpretation, and complementary signal content. Other variables rank highly in particular models but are either less stable across specifications or more redundant with other indicators. For example, `largeposdiff4` ranks first in the GMM-Full specification and `largeposdiff3` ranks second in the K-means specification, but both belong to the highly correlated MBPI family. Similarly, `med_diff` ranks first under hierarchical clustering, but its importance is more model-specific.

The full-set diagnostics, correlation evidence, and permutation-importance results support the reduced six-indicator specification used in the main text. The final set preserves the dimensions needed to classify the three pricing regimes: rigidity through `zero_count`; short-run price and margin variability through `absdiff` and `margin_sd`; and asymmetric price-adjustment dynamics through `largediff_diff`, `size_diff`, and `size_ratio`. The remaining indicators are not treated as irrelevant. Rather, they are documented in the full-set analysis and excluded from the main clustering specification to avoid redundant weighting and preserve a parsimonious, interpretable feature space.

Table C.3: Summary of Permutation-Importance Rankings Across Models

Feature	GMM-Spherical	GMM-Full	K-means	Hierarchical	Notes
`size_diff`	3rd	2nd	4th	7th	Strong in GMM; asymmetric size effect
`absdiff`	8th	14th	11th	20th	Retail price variability
`margin_sd`	5th	6th	17th	19th	Margin volatility signal
`zero_count`	7th	4th	**1st**	5th	Stickiness proxy; robust across models
`size_ratio`	**1st**	13th	15th	12th	Complements `size_diff`
`largediff_diff`	9th	10th	20th	18th	MBPI-type asymmetry net of wholesale movements
`rw_corr`	2nd	7th	16th	15th	Strong in GMM-Spherical; less stable elsewhere
`largeposdiff4`	17th	**1st**	6th	13th	High in one model but redundant with MBPI group
`largeposdiff3`	10th	3rd	2nd	14th	Correlated with other MBPI indicators
`med_diff`	15th	16th	14th	**1st**	Model-dependent; high in hierarchical clustering

Subsection C.2 Clustering including Atlantic provinces

In this subsection, we present clustering results after returning the regulated Atlantic provinces to the data set. When the Atlantic provinces are included, the key question is whether the algorithms identify a distinct "Regulated" regime in addition to the cycle, rigidity, and wholesale-price following regimes documented in the main analysis. To better separate regulated from non-regulated pricing patterns, the four-cluster robustness checks use a slightly richer but still targeted indicator set chosen from the full set in Table 1. The added indicators are designed to capture dimensions that are especially informative for regulated markets. `zero_length` and `med_diff` refine the description of retail price stickiness by distinguishing long runs of unchanged prices and very small day-to-day ad-

justments that are common under administered pricing rules. `rw_corr` and `wholesale_sd` track the interaction between wholesale volatility and retail pass-through, allowing the algorithms to distinguish markets where large wholesale shocks are absorbed by the regulatory regime from markets where they translate more directly into retail movements. At the same time, the original core variables, `size_diff`, `absdiff`, `margin_sd`, `zero_count`, and `largediff_diff`, continue to anchor the structure documented in the three-cluster analysis.

Table C.4: Cluster-Number Diagnostics Using the Reduced Nine-Indicator Set and Including the Atlantic Provinces

Algorithm	**Criterion**	$k = 2$	$k = 3$	$k = 4$	**Preferred** k
K–means	Silhouette Score	0.388	0.255	0.246	2
K–means	Davies–Bouldin Index	1.188	1.213	1.554	2
Hierarchical	Silhouette Score	0.338	0.246	0.206	2
Hierarchical	Davies–Bouldin Index	1.305	1.171	1.727	3
GMM-Spherical	AIC	39695.494	35630.710	34394.047	4
GMM-Spherical	BIC	39810.140	35805.409	34628.798	4
GMM-Tied	AIC	36158.601	34870.415	34407.625	4
GMM-Tied	BIC	36507.999	35274.406	34866.210	4
GMM-Diagonal	AIC	23903.080	21277.662	20407.390	4
GMM-Diagonal	BIC	24105.076	21583.385	20816.841	4
GMM-Full	AIC	18336.515	17066.069	16427.778	4
GMM-Full	BIC	18931.583	17961.401	17623.374	4

Notes: For the Silhouette Score, higher values indicate better separation. For the Davies–Bouldin Index, AIC, and BIC, lower values indicate better fit. The preferred k is reported criterion by criterion.

Table C.4 summarizes the internal fit diagnostics for the all-provinces specification. The criteria do not give a single mechanical recommendation across methods. For K–means, both the Silhouette Score and the Davies–Bouldin Index favor two clusters. For hierarchical clustering, the Silhouette Score favors two clusters, while the Davies–Bouldin Index favors three clusters. By contrast, for GMM, both AIC and BIC decline monotonically as we move from two to three to four clusters across all covariance structures. This pattern indicates that an additional latent group helps capture systematic variation in the joint distribution of the nine indicators. We therefore treat the four-cluster specification as the preferred model for the all-provinces robustness analysis. This choice is not based solely on internal fit criteria. It is motivated by the strong *a priori* distinction between regulated and non-regulated markets, by the GMM information criteria, and by the purpose of this robustness check: to verify whether introducing a Regulated regime preserves the core non-Atlantic cycle/rigidity/wholesale-price following structure.

Table C.5: Cluster Composition by Model Including Atlantic Provinces

Model	Wholesale-price Following	Cycle	Rigidity	Regulated
GMM-Spherical	347	212	395	782
GMM-Full	267	290	383	796
***K*-means**	448	174	425	689
Hierarchical	328	271	460	677

Notes: Entries are counts of city-quarter observations in the four-cluster all-provinces specification. Rows sum to 1,736 observations.

Table C.5 reports the distribution of observations across the four labelled clusters. As in the baseline three-cluster analysis, labels are assigned ex post using the cluster-level indicator profiles. The Regulated label is assigned to the cluster with the largest concentration of Atlantic-province observations and with pricing patterns consistent with administered pricing, especially stronger price stickiness and muted day-to-day retail adjustment. The remaining clusters are labelled Cycle, Rigidity, and Wholesale-price Following using the same economic interpretation as in the non-Atlantic baseline.

Table C.6: Four-Cluster Robustness: Regulated Classification and Cycle Stability

	Regulated cluster			Cycle stability	
Algorithm	**Atlantic → Regulated**	**Regulated in Atlantic**	**Non-Atlantic Regulated from K3 Rigidity**	**Jaccard**	**K3 Cycle → K4 Cycle**
K–means	78.2	63.6	95.6	0.581	58.7
Hierarchical	87.0	71.9	94.7	0.711	94.3
GMM-Spherical	92.5	66.2	97.7	0.744	75.7
GMM-Full	89.1	62.7	22.2	0.479	65.0

Notes: Entries are percentages, except for the Jaccard index, which is reported on a 0–1 scale. The first three percentage columns use different denominators: Atlantic observations; K4 Regulated observations; and non-Atlantic K4 Regulated observations, respectively. The final percentage column is computed relative to K3 Cycle observations in the non-Atlantic sample. K3 refers to the three-cluster non-Atlantic specification, and K4 refers to the four-cluster all-provinces specification.

Table C.6 reports diagnostics for the Regulated cluster and for the stability of the Cycle cluster on non-Atlantic markets. The first two columns describe how strongly the four-cluster solution isolates Atlantic observations into the Regulated regime. The third column describes the source of the non-Atlantic observations that are also assigned to Regulated. The last two columns measure the overlap between Cycle assignments in the baseline three-cluster non-Atlantic specification and the four-cluster all-provinces specification. Two patterns emerge from the four-cluster robustness check. First, the Regulated cluster is dominated by Atlantic markets: depending on the algorithm, 78.2–92.5% of Atlantic city-quarters are assigned to Regulated, and 62.7–71.9% of all Regulated observations are located in the Atlantic provinces. For the relatively small set of non-Atlantic observations that are nevertheless labelled Regulated, most come from the Rigidity class

in the baseline three-cluster runs: 95.6%, 94.7%, and 97.7% under K–means, hierarchical clustering, and the spherical GMM, respectively. The full-covariance GMM is an exception, with only 22.2% of its non-Atlantic Regulated assignments coming from Rigidity. Second, the cyclical regime on non-Atlantic markets remains broadly stable when moving from three to four clusters. The Cycle Jaccard index (Real and Vargas 1996) ranges from 0.58 to 0.74 for K–means, hierarchical clustering, and the spherical GMM, indicating substantial overlap in Cycle assignments across specifications, while the full-covariance GMM again shows weaker agreement at 0.48.[35] Taken together, these diagnostics suggest that introducing a Regulated cluster primarily isolates regulated Atlantic markets and reclassifies a subset of previously rigid non-Atlantic observations, without materially altering the cycle/rigidity/wholesale-price following structure identified in the baseline three-cluster analysis.

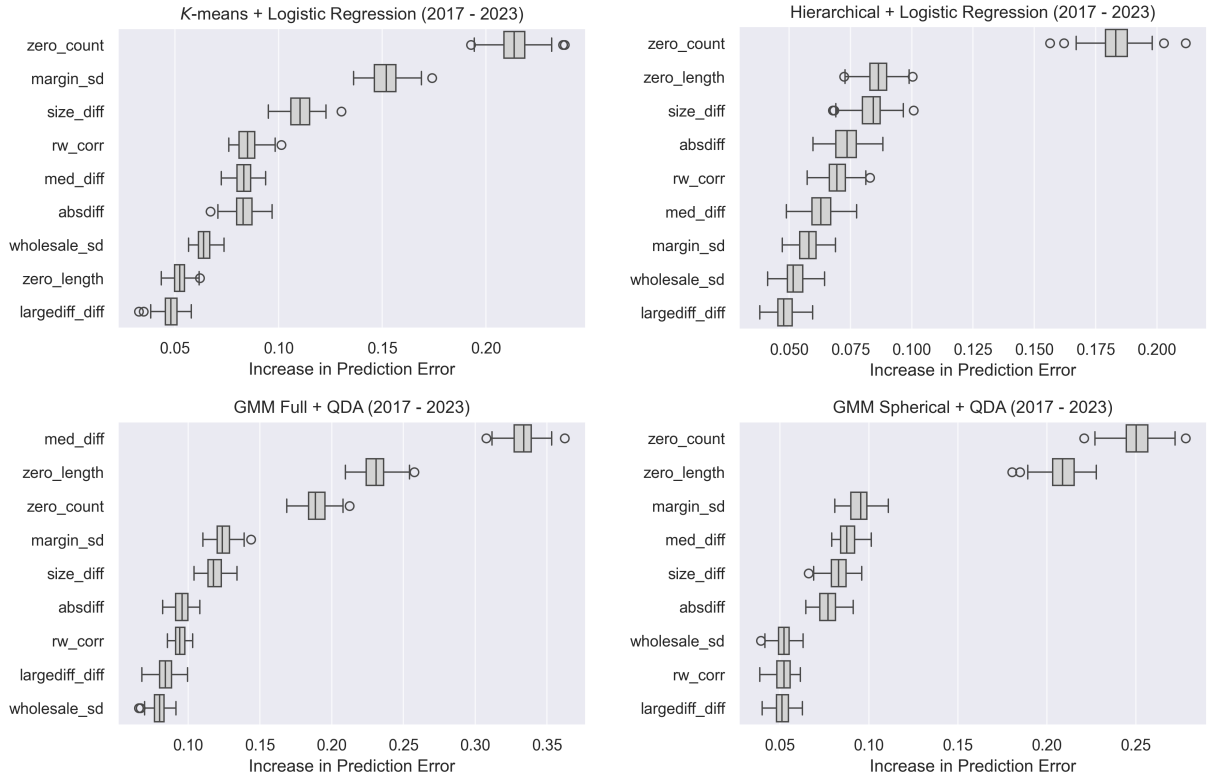
Figure C.3 shows that, even after adding the Regulated regime and expanding the indicator set to nine variables, the basic permutation-importance structure from the three-cluster case in Figure 4 is preserved. Across both K–means+Logit and hierarchical+Logit, `zero_count` remains the dominant indicator, with permutation-importance values between 0.18 and 0.21, and `margin_sd` and `size_diff` remain among the most informative variables. The additional indicators mainly layer information on top of this core structure rather than displacing it. In the four-cluster specification, `zero_length`, `med_diff`, `rw_corr`, and `wholesale_sd` all have non-trivial importance, but their roles differ by algorithm: under K–means+Logit they enter with moderate importance, whereas under hierarchical+Logit `zero_length` rises to the second position. Taken together, these results indicate that the same core dimensions of stickiness, dispersion, and size asymmetries continue to organize the cycle/rigidity/wholesale-price following structure, while the added indicators, especially run length and wholesale volatility, primarily help refine the separation between regulated and non-regulated regimes rather than redefining the original three pricing regimes.

This conclusion is reinforced in the GMM–QDA specifications. Permutation importance in the four-cluster GMM–QDA models yields profiles that are closely aligned with the K–means and hierarchical patterns, but place even more weight on the refined measures of stickiness. For both the spherical and full-covariance GMMs, the following indicators emerge as dominant: `zero_count`, `zero_length`, and `med_diff`. This underscores the central role of days without price changes and very small day-to-day adjust-

35. The Jaccard index is bounded between 0, indicating no shared observations, and 1, indicating identical sets.

ments in separating the four regimes, particularly in distinguishing regulated markets from other pricing dynamics. Dispersion and volatility measures such as `margin_sd` and `size_diff` form the next tier of importance, while `absdiff`, `rw_corr`, `largediff_diff`, and `wholesale_sd` provide smaller but non-negligible incremental signal. Overall, the permutation-importance evidence across all four model combinations confirms that the richer nine-indicator set sharpens regime separation mainly by adding information on stickiness and, to a lesser extent, volatility, while preserving the core dimensions identified in the six-indicator, three-cluster analysis.

Figure C.3: Permutation Importance Using the Reduced Nine-Indicator Set in the Four-Cluster Specification



Appendix D Econometric robustness checks: additional HHI profile

Section 6.2 reports the quadratic HHI specification as the main econometric specification. This appendix provides an additional check on that nonlinear profile by restricting the estimation sample to observations at or below the 90th percentile of HHI. This exercise is useful because the main-text profile shows that the positive association between HHI and cycling appears in the right tail of the HHI distribution. The 90th percentile cutoff

is $\text{HHI}_{1000} = 1.758$, leaving 860 observations in the restricted sample. The split-panel jackknife produces 52 successful replications out of 100, compared with 80 successful replications in the unrestricted quadratic specification. Hence, the p90-restricted quadratic specification should also be read as a diagnostic check. Trimming the right tail of HHI reduces the support over which the nonlinear profile is identified and lowers the number of successful SPJ replications. The exercise is therefore useful for assessing whether the upward turn in cycling depends entirely on the most concentrated observations, but it should not be interpreted as a competing baseline estimate.

Table D.1: HHI Profile in the Quadratic Specification Excluding the Top 10 Percent of HHI

HHI_{1000}	Location in support	Pr(Cycle)	Pr(Rigid)	AME on Pr(Cycle)	AME on Pr(Rigid)
1.000	low HHI	0.470	0.211	-0.483***	0.575***
1.262	Q1/Q2 cutoff	0.341	0.361	-0.389***	0.431***
1.478	Q2/Q3 cutoff	0.297	0.407	-0.027	0.016
1.611	Q3/Q4 cutoff	0.307	0.393	0.190***	-0.244***
1.758	p90 cutoff	0.352	0.332	0.403***	-0.586***

Notes: Predicted probabilities and local marginal effects are computed from the quadratic HHI specification restricted to observations with HHI_{1000} at or below the 90th percentile. The model sets HHI_{1000} to the indicated value and averages predictions over the observed covariate distribution. The AME columns report local marginal effects of HHI_{1000} at each HHI value. A 100-point increase in the raw HHI corresponds to one-tenth of the reported AME. Estimates are based on 52 successful SPJ replications. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table D.1 reports predicted probabilities and local marginal effects from the restricted quadratic specification. The main pattern remains nonlinear, but the upper-tail increase in cycling becomes more muted once the top 10 percent of HHI observations are excluded. The predicted probability of cycling falls from about 0.470 at $\text{HHI}_{1000} = 1.000$ to about 0.297 near the median. It then rises to about 0.307 near the third quartile and to about 0.352 at the 90th percentile cutoff. The rigid outcome follows the opposite pattern, increasing through the middle of the HHI distribution and then declining near the upper end of the restricted support.

This restricted-sample check supports two points from the main text. First, the nonlinear HHI pattern does not come only from the extreme right tail: the predicted probability of cycling still declines through the lower and middle parts of the HHI distribution and then begins to rise around the upper-middle part of the support. Second, the stronger increase in cycling at high HHI values in the unrestricted specification depends partly on observations in the right tail. This reinforces our cautious interpretation of the HHI result. The concentration estimates describe a nonlinear association between HHI and regime membership, but they should not be read as structural effects of concentration.