The Eastasouth Journal of Information System and Computer Science
Vol. 3, No. 01, August 2025, pp. 211-223
ISSN: 3025-566X, DOI: 10.58812/esiscs.v3i01

Business Analytics for United States Retail Channel Planning Under Structural Change

Md Jubayar Hossain[1], Semanto Saha Utsab[2], Sadia Afrin Jarir[3], Evha Rozario[4], Zerin Akter Tanni[5], Tushar Roy[6]

[1] St. Francis College, Brooklyn, New York, USA
[2] Montclair State University, Montclair, New Jersey, United States
[3] International University of America, Los Angeles, California 90010
[4] International University of America, Los Angeles, California 90010
[5] St. Francis College, Brooklyn, New York, USA
[6] St. Francis College, Brooklyn, New York, USA

Article Info

Article history:

Received August, 2025
Revised August, 2025
Accepted August, 2025

Keywords:

Business Analytics;
Channel Planning;
E-Commerce;
Forecast Bias;
Rolling-Origin Evaluation;
Structural Change;
United States Retail

ABSTRACT

Retail channel planning requires forecasts that remain useful when purchasing patterns change abruptly. This study examines the quarterly e-commerce share of United States retail sales using 99 observations from 2000 Q1 through 2024 Q3 in the data vintage available on December 31, 2024. Census Bureau estimates distributed through ALFRED provide the sole empirical data source. Six transparent specifications are compared at 59 successive one-quarter forecast origins using chronological estimation. Evaluation distinguishes 2010–2019, 2020–2021, and 2022–2024 Q3 and combines absolute error, squared error, directional bias, and asymmetric planning-loss sensitivity. The lowest mean absolute errors in these periods are 0.0755, 0.9375, and 0.1215 percentage points, respectively. Method rankings depend on the evaluation period and criterion. The findings support period-specific assessment, simple benchmarks, and explicit attention to planning consequences. Results use a fixed historical data vintage; they do not establish causal pandemic effects or realized firm-level savings.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Name: Tushar Roy
Institution: St. Francis College, Brooklyn, New York, USA
Email: troy3@sfc.edu

1. INTRODUCTION

A retail business can experience growing online sales while still making poor decisions about digital capacity. Purchasing patterns may change faster than investment plans, staffing arrangements, or fulfillment contracts. A forecast that understates online demand can leave a business short of picking capacity. A forecast that overstates it can encourage commitments that remain costly after an unusual surge has passed. Business analytics becomes useful when it makes these tensions visible and helps managers decide how much confidence to place in the next planning number.

This study addresses that problem through the national share of retail sales completed through e-commerce in the United States. Channel share is a useful starting point

because it describes how retail activity is distributed between ordering channels. It does not directly measure profitability, order volume, or delivery workload. Nevertheless, it provides a public and consistently defined indicator against which a firm can compare its own channel exposure. The empirical question is whether straightforward forecasting methods maintain their relative performance as the national channel mix moves through steady expansion, abrupt disruption, and subsequent adjustment.

The paper treats forecasting as a business analytics task rather than as an exercise in fitting the most elaborate model. [1] position business intelligence and analytics as a field concerned with the business implications of data. Within that broad field, prediction requires an evaluation design that measures performance beyond the observations used for estimation. [2], [3] distinguish predictive questions from explanatory ones. Here, the objective is to assess the behavior of forecasts, not to identify why households adopted online shopping or to estimate the causal effect of public health restrictions.

The setting is especially instructive because historical growth and temporary disruption can be mistaken for each other. A model using a long sample may adapt too slowly to a higher channel share. A model using only recent observations may interpret a sharp, temporary movement as a continuing trend. Both problems can arise without computational complexity. They are concerned about which part of the past is allowed to shape the future and how the resulting error is judged.

Three research questions guide the analysis. First, how do forecast rankings differ across the pre-pandemic, disruption, and later evaluation periods? Second, how sensitive are local trend forecasts to the length of their estimation window? Third, does a greater penalty for underallocation change the practical interpretation of model performance? These questions are examined using the same U.S. series, common forecast dates, and directly reproducible calculations.

The contribution is a focused empirical comparison that links estimation memory, changing error direction, and planning-loss sensitivity. The models themselves are established baselines. The study does not introduce a new algorithm, claims a general forecasting superiority result, or present national aggregates as evidence about individual retailers. Its value lies in showing, with an auditable U.S. application, how a model that appears attractive in one period can become unreliable in another. The reported methods, tables, and figures make the comparison open to inspection.

2. RELATED RESEARCH AND ANALYTICAL PERSPECTIVE

2.1 Forecasting Within Business Analytics

Predictive analytics should be judged against the decision it is intended to support. A small average error can conceal repeated underprediction, while a model with slightly larger typical errors may avoid a particularly costly direction of mistake. This distinction is central to the present analysis. The empirical work measures both the magnitude and direction of forecast errors, then adds a transparent sensitivity analysis for unequal planning penalties.

[3] explain how predictive evaluation contributes to information systems research. Their argument motivates the separation between a plausible description of historical retail change and evidence that a forecasting rule works on later observations. A line drawn through all available data can be visually persuasive without being an appropriate prediction rule. Conversely, a simple rule can provide a useful benchmark even when it offers little explanation of the underlying economic process.

Retail forecasting research also emphasizes the connection between predictions and operational decisions. [4] discuss forecasting at different levels of retail aggregation. That distinction

matters here: a national channel-share series is far removed from a store-item demand series. The current study uses the national measure to examine broad planning exposure and model behavior. It does not evaluate replenishment, promotions, customer response, or inventory optimization.

[5] connect consumer analytics with sales forecasting and supply chain planning. Their discussion supports using richer information when it serves a defined business decision. In this study, national channel share remains an external indicator; it cannot substitute for a retailer's own order, product, and customer information.

[6] organize demand-forecasting applications of big data analytics and machine learning. This broader methodological literature identifies possibilities for future extensions, but it does not establish that a complex method will improve this study's short quarterly series. [7] examine interpretable regression for retail performance indicators, providing related motivation for transparent models. Their store-level targets differ from the national share evaluated here.

2.2 *Benchmarks and Chronological Evaluation*

Simple forecasts provide a reference against which added complexity can be assessed. [8] describe last-observation and drift methods as standard forecasting tools. The last-observation method assumes that the next value will equal the most recently observed value. Drift adds an average historical change. These rules make different commitments about whether movement should continue, making them useful comparators in a channel-share series with an upward long-run path.

[9] discusses the importance of evaluation across successive forecast origins. Following that principle, this paper repeatedly estimates each specification using earlier observations and predicts the next quarter. The resulting comparison avoids random mixing of past and future observations at the model-fitting stage. It also allows the same method to be observed under several different economic conditions.

The limits of validation procedures still require care. [10] consider conditions relevant to cross-validation in autoregressive settings. This study uses a chronological design because the operational question concerns predicting later quarters from earlier ones. However, the observations available today include revisions. Chronological fitting with a revised data vintage is not equivalent to reconstructing the exact information available to managers on historical release dates.

Window selection under changing parameters has a direct econometric foundation. [11] show that estimation-window choice involves a trade-off: retaining earlier observations can reduce variance while introducing bias after a break. The present fixed-window comparison illustrates this issue descriptively. It neither implements their selection procedures nor establishes that discarding all earlier data is optimal.

2.3 *Error Measures and Changing Environments*

[12] examine problems in common forecast accuracy measures. Because this study evaluates one consistently measured percentage-share series, mean absolute error in percentage points is easy to interpret. Root mean squared error gives greater influence to large misses. Mean error identifies persistent directional mistakes. Reporting these measures together prevents a single summary statistic from carrying more meaning than it can support.

The M4 competition illustrates a broad empirical approach to comparing forecasting methods across many series [13]. The present study is narrower: it examines one U.S. national series and a

small set of transparent specifications. It therefore cannot establish which method should dominate across retail applications. Its results instead demonstrate why performance claims should name the forecast target, horizon, evaluation period, and criterion.

Pandemic-era retail behavior raises a further difficulty. [14] provide a post-script discussing forecast research in the context of the pandemic and machine learning. The analysis below uses the pandemic years as a clearly labeled evaluation block, without estimating a causal treatment effect. The period labels organize the evidence. They are not a formal statistical test that identifies the date or mechanism of a structural break.

The M5 accuracy study evaluates forecasting on a large hierarchy of Walmart sales series [15]. It provides a relevant retail benchmark for the scale of modern forecast comparisons, while differing substantially from a single quarterly national share. Its results cannot be transferred directly to this target. The companion uncertainty study evaluates predictive quantiles [16], highlighting a separate task that the present point-forecast comparison does not perform.

[17], [18] review adaptation when data-generating relationships evolve. This literature motivates monitoring forecast errors and revisiting estimation assumptions. The current period-by-period results are consistent with the need for such monitoring, but they do not constitute a concept-drift detection test or validate an adaptive learning algorithm.

3. DATA AND RESEARCH DESIGN

3.1 United States Data Source and Coverage

The empirical input is E-Commerce Retail Sales as a Percent of Total Sales, series ECOMPCTSA. The U.S. Census Bureau produces the underlying estimates. The series is quarterly, seasonally adjusted, and expressed in percent. The analysis uses the ALFRED vintage dated December 31, 2024 [19].

The sample contains 99 observations from 2000 Q1 through 2024 Q3. This is the latest reference quarter available by December 31, 2024. The November 19, 2024 Census release reports the third-quarter estimate [20]. Fourth-quarter estimates are excluded because they were not available by the information cutoff. All observed empirical values refer to the United States.

The Census release describes survey-based estimates subject to sampling and nonsampling error. Seasonally adjusted sales are not adjusted for price changes [20]. Sales shares therefore measure nominal expenditure composition. Relative prices and changes in product mix can affect the share even when physical quantities follow a different path.

The observed series is rounded to one decimal place. Differences of a few hundredths of a percentage point in average forecast accuracy should consequently be interpreted cautiously. The source contains no missing values within the retained interval, and no interpolation, outlier deletion, winsorization, or additional seasonal adjustment is applied. The exceptional movements in 2020 remain part of the evaluation because they are central to the business problem.

3.2 Target and Scope Of Inference

The target is next quarter's national e-commerce share. A percentage-point error is the difference between two percentage shares. It is not a relative percentage error. For example, a forecast of 15.0% and an observed share of 15.5% imply an underprediction of 0.5 percentage points. The distinction is important for both reporting accuracy and translating a channel-share forecast into an initial capacity discussion.

The target is bounded between zero and 100, but all estimated one-quarter forecasts remain comfortably inside those limits. No clipping is applied. The narrow comparison does not attempt to forecast the total retail market, convert national share into firm revenue, or estimate demand elasticity. Those tasks require additional variables and a different research design.

3.3 Forecast Specifications

Six specifications are evaluated. The naive forecast uses the latest observed share. Drift uses that share plus the average quarterly change from the first retained observation to the current origin. The expanding global trend estimates an ordinary least squares line on every observation available at the origin and extrapolates it by one quarter. Three local trend specifications estimate the same type of line using only the latest eight, twelve, or twenty quarters.

In the forecast definitions, an origin with t available observations uses values indexed from zero through t minus one. Drift adds the difference between the last and first observed values divided by t minus one. The global line is fitted to integer time indices from zero through t minus one and evaluated at t. A local line with window w uses indices from zero through w minus one and is evaluated at w. These definitions are sufficient to reproduce the point forecasts without ambiguity.

The window lengths represent two, three, and five years of quarterly history. They provide a sensitivity comparison rather than a claim that one window was prospectively optimized. No grid search chooses a winning window from the reported test periods. Nevertheless, the design is exploratory and was not preregistered. Selecting a specification for deployment after inspecting these results would require a further, genuinely untouched evaluation period.

3.4 Rolling-Origin Procedure

The first 40 observations, covering 2000–2009, establish the initial estimation history. The first forecast is for 2010 Q1. Each target observation then becomes available to the next estimation origin. The process continues through 2024 Q3, producing 59 forecasts per method and 354 method-quarter predictions. Every method predicts the same dates.

The 59 targets are divided into 2010–2019 with 40 forecasts, 2020–2021 with eight, and 2022–2024 Q3 with eleven. These are evaluation blocks rather than independent datasets. Expanding methods retain their earlier history, and local methods update their fixed-length windows. No model receives the future period label as a predictor.

The forecast horizon is one quarter ahead in observation time. Because official statistics are released after their reference quarter, this is not claim that the latest reference quarter value was available before the next quarter began. The exercise evaluates a standard retrospective sequence on revised quarterly data. An operational implementation must explicitly incorporate release dates, publication lags, and whatever earlier indicators a business can access.

3.5 Accuracy and Planning-Loss Measures

Forecast error is defined as observed share minus forecast share. A positive error indicates underprediction. Mean absolute error, or MAE, averages absolute errors. Root mean squared error, or RMSE, is the square root of average squared errors. Bias is average signed error. The underprediction rate is the fraction of quarters with strictly positive errors; exact matches count as neither underprediction nor overprediction.

The planning sensitivity measure applies a relative penalty r to positive errors and a penalty of one to negative errors. For each quarter, loss

equals the larger of r times error and negative errors. This produces zero loss for an exact forecast, r times the error for underprediction, and the absolute error for overprediction. The main table uses r equal to three. Figure 4 examines one, two, three, and five.

These weights are analyst-defined sensitivity settings, not estimated U.S. business costs. The measure assumes that forecast share acts as a planning target and that deviations have linear consequences in either direction. It does not solve a capacity optimization problem. A business facing asymmetric costs would generally need a predictive distribution and a decision rule tailored to its constraints, rather than simply adopting a point forecast unchanged.

Point forecasts and predictive distributions require different evaluation tools. [21] develop proper scoring rules for probabilistic predictions. Their framework would be relevant to an extension that estimates complete forecast distributions. The asymmetric score used here instead compares existing point forecasts under assumed penalties; it should not be interpreted as a probabilistic calibration measure.

3.6 Reproducibility and Statistical Restraint

The analysis uses a fixed ALFRED historical vintage dated December 31, 2024. All methods use the published percentage shares without interpolation, and calculations retain full precision until display. The common sample, estimation windows, forecast horizon, and error definitions specify the comparison. Historical vintages at every individual forecast origin are not reconstructed.

The study reports descriptive comparisons without significance claims. The disruption block has eight quarters and the later block eleven. Forecast errors are temporally related, and specifications are compared on the same targets. Lower observed errors are not proof of stable population superiority. Prediction interval coverage and formal break inference remain outside the analysis.

4. RESULTS

4.1 United States Channel-Share Trajectory

Figure 1 shows the long-run rise in e-commerce share and the abrupt movement around 2020. A sudden increase followed by partial reversal creates a difficult forecasting environment: continuity with earlier growth may fail, while extrapolating the immediate increase can also fail.

The final observed share is 16.2% in 2024 Q3. The November 2024 Census release also reports a third-quarter share of 16.2%. The sample ends there to respect the December information cutoff.

Quarterly shares describe the composition of retail sales in individual quarters. An arithmetic average of quarterly percentages is not an annual e-commerce sales total divided by an annual total retail sales figure. No full-year estimate is reported for the incomplete final year.

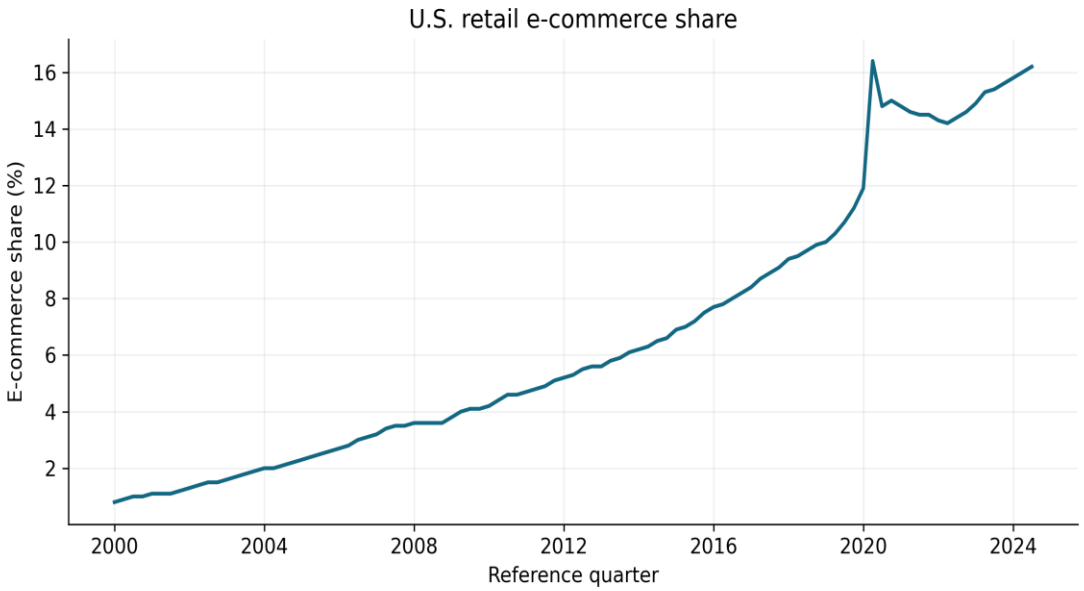


Figure 1. United States Quarterly Retail E-Commerce Share from 2000 Q1 through 2024 Q3

4.2 Pre-Pandemic Forecast Performance

Table 1 and Figure 2 compare accuracy. In 2010–2019, the lowest MAE is achieved by Local trend 8, at 0.0755 percentage points. Naive and drift record 0.1775 and 0.0870, respectively. These values refer to the December 2024 vintage.

The expanding global trend has an MAE of 0.6028 and mean signed error of 0.6028. A single straight line fitted from 2000 onward does not adequately capture the later path. This does not establish that long historical records are intrinsically unhelpful.

The naive method underpredicts 95.0% of pre-disruption quarters. The smallest RMSE is 0.1083, achieved by Local trend 8. The smallest three-to-one planning loss is 0.1845, achieved by Local trend 8.

Table 1. Forecast Accuracy by Evaluation Block

Period	Method	MAE	RMSE
2010–2019	Naive	0.1775	0.2031
2010–2019	Drift	0.0870	0.1192
2010–2019	Global trend	0.6028	0.7540
2010–2019	Local trend 8	0.0755	0.1083
2010–2019	Local trend 12	0.0822	0.1162
2010–2019	Local trend 20	0.1368	0.1669
2020–2021	Naive	0.9375	1.7114
2020–2021	Drift	0.9889	1.6928
2020–2021	Global trend	3.7784	3.9885
2020–2021	Local trend 8	1.7031	2.1137
2020–2021	Local trend 12	1.6600	2.1055
2020–2021	Local trend 20	1.5231	2.0757
2022–2024 Q3	Naive	0.2091	0.2236
2022–2024 Q3	Drift	0.1215	0.1625
2022–2024 Q3	Global trend	1.8474	1.8512
2022–2024 Q3	Local trend 8	0.3409	0.4128
2022–2024 Q3	Local trend 12	0.8646	1.0627
2022–2024 Q3	Local trend 20	1.0735	1.2302

Note. MAE and RMSE are in percentage points. Forecast counts per method are 40, 8, and 11 in chronological block order. All calculations use the December 31, 2024 data vintage.

4.3 Disruption-Period Performance

In 2020–2021, Naive has the lowest MAE, at 0.9375. Naive and drift record MAEs of 0.9375 and 0.9889. The eight-quarter trend records 1.7031. Responsiveness alone does not prevent errors when a sharp movement is followed by reversal.

The global trend has an MAE of 3.7784 during the disruption block. Figure 3 places selected one-quarter forecasts beside observations. Each prediction is fitted at its own chronological origin; the curves are not multiquarter forecasts fitted using future observations.

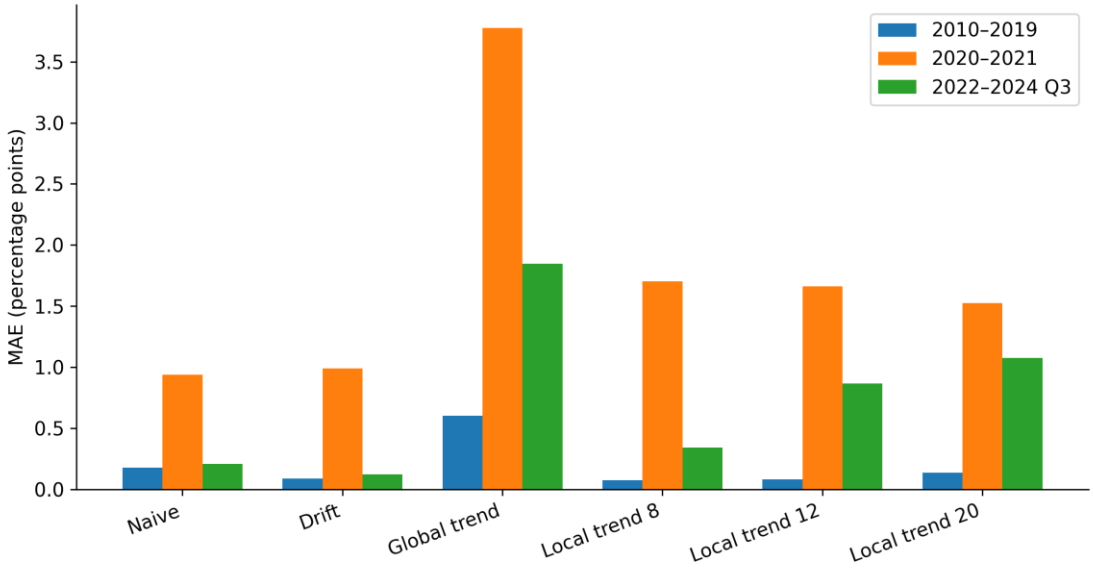


Figure 2. Mean Absolute Forecast Error by Evaluation Block

The eight-quarter trend has mean error of -0.3174, compared with 3.7784 for the global trend and 0.9963 for the twenty-quarter trend. Different estimation memories produce different directional errors during the same episode.

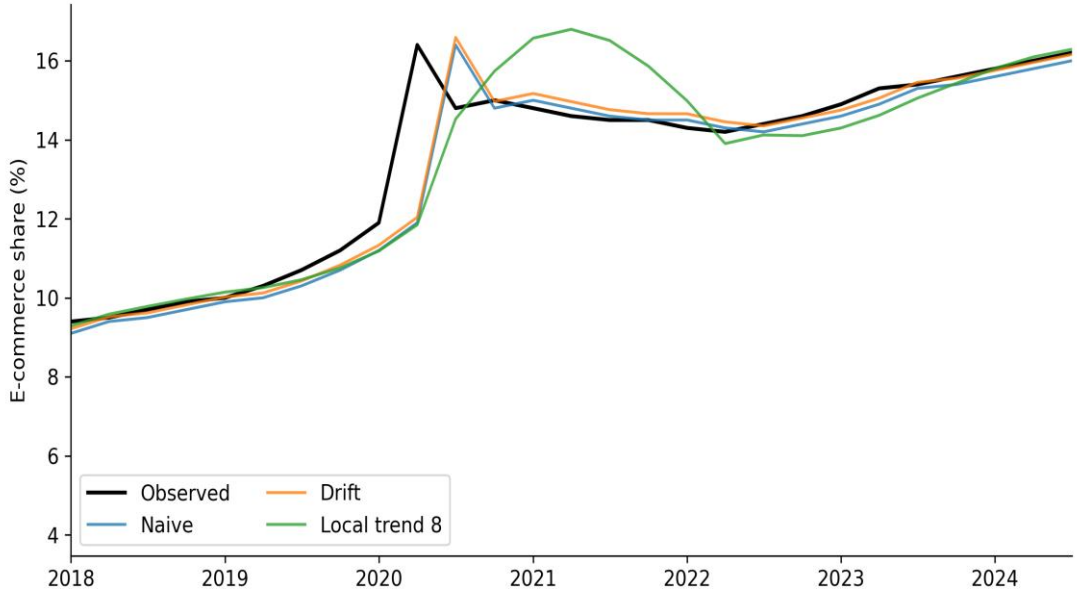


Figure 3. Observed Shares and Selected One-Quarter Forecasts

4.4 Later-Period Performance

In 2022–2024 Q3, Drift has the lowest MAE, at 0.1215. Naive and drift record 0.2091 and 0.1215. The smallest RMSE is 0.1625, achieved by Drift.

The eight-, twelve-, and twenty-quarter trends have later-period MAEs of 0.3409, 0.8646, and 1.0735, respectively. The twenty-quarter method has mean signed error of -1.0735 and underpredicts 0.0% of quarters. Its estimation window continues to include disruption-era observations.

The global trend has later-period MAE of 1.8474. These findings apply only to the six specifications and historical vintage examined. Exponential smoothing, autoregressive methods, machine learning, and firm-specific forecasts are not evaluated.

4.5 Planning-Loss Sensitivity

Under the three-to-one planning penalty, disruption-period naive and drift losses are 2.2875 and 2.2286. Comparing these with MAE shows why error direction matters. The weights represent sensitivity assumptions rather than estimated business costs.

For 2022–2024 Q3, naive and drift losses are 0.5727 and 0.2432. The eight-quarter, twenty-quarter, and global trends record 0.8630, 1.0735, and 5.5422, respectively.

Figure 4 compares later-period average losses at underprediction penalties of one, two, three, and five. These ratios are assumptions. A retailer facing nonlinear overtime costs, irreversible facility investments, or strict service requirements may need a different loss function.

Table 2. Directional Errors and Illustrative Planning Loss

Period	Method	Bias	Under %	Loss 3 to 1
2010–2019	Naive	0.1775	95.0	0.5325
2010–2019	Drift	0.0752	85.0	0.2491
2010–2019	Global trend	0.6028	100.0	1.8084
2010–2019	Local trend 8	0.0334	57.5	0.1845
2010–2019	Local trend 12	0.0605	77.5	0.2250
2010–2019	Local trend 20	0.1274	92.5	0.4010
2020–2021	Naive	0.4125	37.5	2.2875
2020–2021	Drift	0.2509	37.5	2.2286
2020–2021	Global trend	3.7784	100.0	11.3352
2020–2021	Local trend 8	-0.3174	37.5	3.0888
2020–2021	Local trend 12	0.1718	50.0	3.4919
2020–2021	Local trend 20	0.9963	62.5	4.0424
2022–2024 Q3	Naive	0.1545	81.8	0.5727
2022–2024 Q3	Drift	0.0002	72.7	0.2432
2022–2024 Q3	Global trend	1.8474	100.0	5.5422
2022–2024 Q3	Local trend 8	0.1812	63.6	0.8630
2022–2024 Q3	Local trend 12	-0.1679	63.6	1.5613
2022–2024 Q3	Local trend 20	-1.0735	0.0	1.0735

Note. Bias equals actual minus forecast. Under % is the percentage of strictly positive errors. Loss assigns three times the penalty to underprediction. Weights are assumptions.

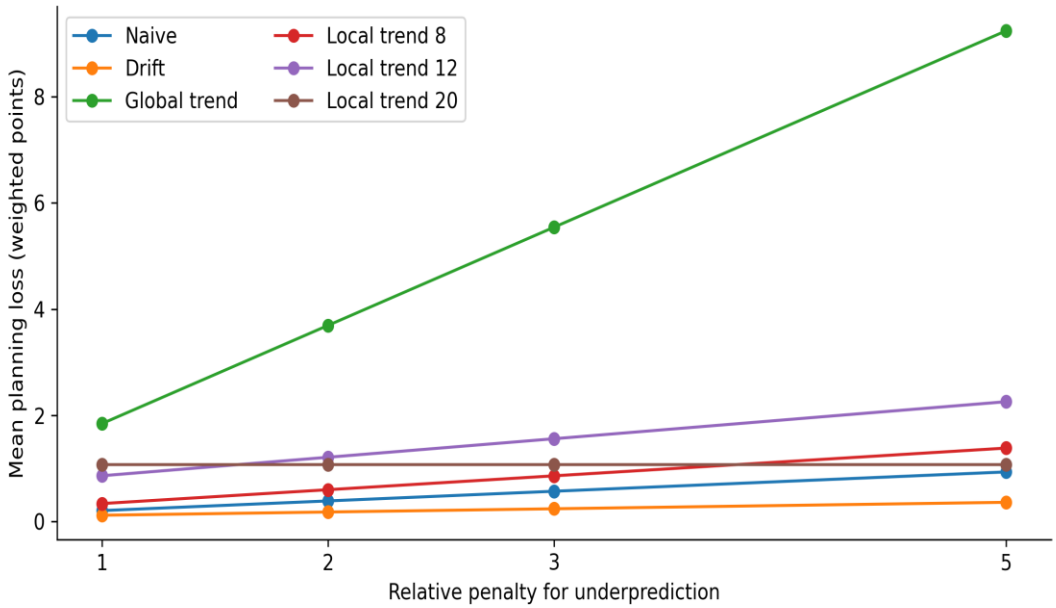


Figure 4. Later-period planning-loss sensitivity from 2022 through 2024 Q3

5. DISCUSSION

5.1 *Historical Memory is A Modeling Decision*

The evidence demonstrates a practical tension in business analytics. A larger estimation sample can stabilize a model while also preserving patterns that no longer describe the next planning interval. A smaller sample can adapt quickly while giving unusual observations too much influence. The expanding line and the local lines expose both sides of that tension with very little machinery.

Before 2020, the eight-quarter trend benefits from recent movement. During the disruption, it extrapolates from observations that are a poor guide to the subsequent reversal. Later, longer local windows continue to reflect the unusual period and often overpredict. These results suggest that model review should examine which observations are driving a forecast rather than treating window length as a purely technical setting.

The comparison also separates data retention from estimation choice. A business should preserve historical records for auditing, comparison, and understanding rare events. It does not follow that every historical observation must receive equal influence in every forecasting specification. Retaining data and deciding how to use them are distinct analytical responsibilities.

5.2 *Forecast Rankings Require Context*

The method with the lowest MAE changes across periods: Local trend 8 leads in 2010–2019, Naive in 2020–2021, and Drift in 2022–2024 Q3. A single aggregate ranking would conceal this period dependence.

For management reporting, this means that a claim such as "the forecasting model improved accuracy" is incomplete without its target, horizon, dates, and criterion. An apparently successful model may owe its ranking to a period that differs from current conditions. Conversely, a model with a poor aggregate score may remain a useful fallback under particular circumstances. The present results support reporting separate evaluation blocks while avoiding an automatic switch based on knowledge of which block later occurred.

The paper does not implement a real-time regime detector or an adaptive model-selection policy. Designing such a policy after observing these rankings would introduce another layer of selection. A proper follow-up would

specify the switching rule in advance, account for data release timing, and evaluate the entire rule on new observations. Choosing the best method retrospectively in each block would overstate what a manager could have achieved.

The literature therefore supports two distinct follow-up questions. The first is whether an estimation-window policy, specified before evaluation, can improve adaptation [11]. The second is whether a monitoring mechanism can distinguish persistent change from temporary fluctuation [17]; Lu et al., 2019). Neither question is answered by selecting the lowest-error model after observing each evaluation block.

5.3 Translating A National Indicator Into Business Decisions

A retailer can use the national e-commerce share as an external reference for strategic discussions. If its own online share moves differently, that difference may warrant investigation into product mix, customer geography, channel strategy, or measurement practices. The national series alone cannot determine whether divergence is desirable. A specialist retailer may rationally operate far above or below the national average.

A practical planning process would combine the external reference with the firm's orders, margins, returns, delivery promises, and capacity constraints. Before converting a share forecast into a staffing or investment commitment, managers would need a forecast of total business volume and an estimate of workload per order. Online revenue and fulfillment burden need not move proportionately. Expensive products, small baskets, and changes in return behavior can alter that relationship.

The asymmetric-loss exercise is useful because it forces a conversation about the consequences of error. It is deliberately limited to relative penalties. No dollar savings, return on investment, productivity gain, or service improvement is estimated. Establishing such outcomes would require observed business decisions and costs, ideally evaluated against a credible operational comparison.

5.4 Implications for Analytics Governance

The most immediate governance implication is to keep an auditable benchmark beside any production forecast. A dashboard should show the latest prediction, recent errors, signed bias, and performance against simple alternatives. If a sophisticated system persistently loses to a basic rule, the business needs to know before the forecast shapes a major commitment. Complexity can still be justified, but it should earn that justification through relevant evidence.

A second implication concerns overrides. Managers may possess current information absent from a published national series, including contracts, platform changes, or impending promotions. Judgmental adjustments can be useful when they incorporate such information [8]. Recording the reason, direction, timing, and subsequent outcome of an override allows the organization to assess whether human intervention adds value.

A third implication is to preserve the information state behind each forecast. Revised official data are suitable for retrospective analysis but can create a more favorable picture than historical releases would support. A production system should retain source vintages, release timestamps, code versions, and the forecast issued at the time. This record makes later performance assessment credible and helps distinguish model failure from data revision.

6. LIMITATIONS AND FUTURE RESEARCH

First, the study contains one national series with 99 quarterly observations, rather

than independent observations on 99 firms. Serial dependence reduces the information available for comparison. Findings cannot be transferred directly to states, industries, stores, or products.

Second, the fixed December 31, 2024 vintage respects the information cutoff but does not reproduce each historical release. Seasonal adjustment and revisions available by that date may incorporate information unavailable at earlier forecast origins. The exercise is a retrospective comparison, rather than a real-time performance reconstruction.

Third, evaluation blocks are selected for substantive interpretation rather than estimated break dates. The disruption block contains eight forecasts and the later block eleven. Individual quarters can strongly influence either block, so statistical superiority and causal pandemic claims are unwarranted.

Fourth, the model set is intentionally restricted. It contains no exponential smoothing, autoregressive integrated moving average model, nonlinear trend, ensemble, or machine learning method. These omissions leave open the possibility that another specification would perform better. The current evidence supports the importance of benchmarks and evaluation design, not the conclusion that additional modeling effort has no value.

Fifth, uncertainty is only partially addressed. RMSE highlights large errors, and the sensitivity analysis examines unequal penalties, but neither produces a predictive distribution. Future work should evaluate prediction intervals, calibration, and decision rules based on relevant quantiles. Such work should also consider the bound nature of a share and changing measurement precision.

A probabilistic extension should report both calibration and distributional accuracy, using scores appropriate to the forecast object [21]. The M5 uncertainty study provides a retail example of quantile evaluation [16]. Such an extension would also allow a retailer to choose a planning quantile from its own undercapacity and overcapacity costs rather than treating every point forecast as the same operational decision.

Finally, the planning-loss measure is illustrative. Penalties are not inferred from U.S. company accounts, and the study observes no actual capacity decisions. Business value must be tested using real operational constraints and costs. A useful extension would connect a prespecified forecasting pipeline to documented U.S. retailer decisions while preserving customer confidentiality and avoiding selection of only favorable cases.

7. CONCLUSION

This study uses U.S. public information available through December 2024 to examine forecasting under changing retail channel patterns. The lowest-MAE methods are Local trend 8 before 2020, Naive during 2020–2021, and Drift during 2022–2024 Q3. The comparison highlights the importance of period-specific evaluation.

For business analytics, the implication is that estimation history, error direction, and decision criteria deserve explicit attention. A forecast should be assessed against simple alternatives on later observations and reviewed when recent behavior differs from the historical pattern it relies on. The analysis provides reproducible evidence for that practice while leaving causal effects, real-time historical performance, and realized business benefits for further research.

REFERENCES

[1] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Q.*, pp. 1165–1188, 2012, doi: https://doi.org/10.2307/41703503.

[2] G. Shmueli, "To Explain or to Predict?," *Stat. Sci.*, vol. 25, no. 3, pp. 289–310, 2010, doi: https://doi.org/10.1214/10-STS330.

[3] G. Shmueli and O. R. Koppius, "Predictive Analytics in Information Systems Research," *MIS Q.*, vol. 35, no. 3, pp. 553–572, 2011, doi: https://aisel.aisnet.org/misq/vol35/iss3/5/.

[4] R. Fildes, S. Ma, and S. Kolassa, "Retail forecasting: Research and practice," *Int. J. Forecast.*, vol. 38, no. 4, pp. 1283–

1318, 2022, doi: https://doi.org/10.1016/j.ijforecast.2019.06.004.

[5] T. Boone, R. Ganeshan, A. Jain, and N. R. Sanders, "Forecasting sales in the supply chain: Consumer analytics in the big data era," *Int. J. Forecast.*, vol. 35, no. 1, pp. 170–180, 2019, doi: https://doi.org/10.1016/j.ijforecast.2018.09.003.

[6] M. Seyedan and F. Mafakheri, "Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities," *J. Big Data*, vol. 7, no. 1, p. 53, 2020, doi: https://doi.org/10.1186/s40537-020-00329-2.

[7] B. Panay *et al.*, "Forecasting Key Retail Performance Indicators Using Interpretable Regression," *Sensors*, vol. 21, no. 5, p. 1874, 2021, doi: https://doi.org/10.3390/s21051874.

[8] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and practice*, 3rd ed. OTexts, 2021.

[9] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: an analysis and review," *Int. J. Forecast.*, vol. 16, no. 4, pp. 437–450, 2000, doi: https://doi.org/10.1016/S0169-2070(00)00065-0.

[10] C. Bergmeir, R. J. Hyndman, and B. Koo, "A note on the validity of cross-validation for evaluating autoregressive time series prediction," *Comput. Stat. Data Anal.*, vol. 120, pp. 70–83, 2018, doi: https://doi.org/10.1016/j.csda.2017.11.003.

[11] M. H. Pesaran and A. Timmermann, "Selection of estimation window in the presence of breaks," *J. Econom.*, vol. 137, no. 1, pp. 134–161, 2007, doi: https://doi.org/10.1016/j.jeconom.2006.03.010.

[12] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, 2006, doi: https://doi.org/10.1016/j.ijforecast.2006.03.001.

[13] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 Competition: 100,000 time series and 61 forecasting methods," *Int. J. Forecast.*, vol. 36, no. 1, pp. 54–74, 2020, doi: https://doi.org/10.1016/j.ijforecast.2019.04.014.

[14] R. Fildes, S. Kolassa, and S. Ma, "Post-script—Retail forecasting: Research and practice," *Int. J. Forecast.*, vol. 38, no. 4, pp. 1319–1324, 2022, doi: https://doi.org/10.1016/j.ijforecast.2021.09.012.

[15] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "M5 accuracy competition: Results, findings, and conclusions," *Int. J. Forecast.*, vol. 38, no. 4, pp. 1346–1364, 2022, doi: https://doi.org/10.1016/j.ijforecast.2021.11.013.

[16] S. Makridakis *et al.*, "The M5 uncertainty competition: Results, findings and conclusions," *Int. J. Forecast.*, vol. 38, no. 4, pp. 1365–1385, 2022, doi: https://doi.org/10.1016/j.ijforecast.2021.10.009.

[17] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Comput. Surv.*, vol. 46, no. 4, pp. 1–37, 2014, doi: https://doi.org/10.1145/2523813.

[18] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under Concept Drift: A Review," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 12, pp. 2346–2363, 2019, doi: https://doi.org/10.1109/TKDE.2018.2876857.

[19] U.S. Census Bureau, "E-Commerce Retail Sales as a Percent of Total Sales [ECOMPCTSA]." ALFRED, Federal Reserve Bank of St. Louis, 2024. [Online]. Available: https://alfred.stlouisfed.org/series?seid=ECOMPCTSA

[20] U.S. Census Bureau, "Quarterly Retail E-Commerce Sales: 3rd Quarter 2024," U.S. Department of Commerce, 2024. [Online]. Available: https://www2.census.gov/retail/releases/historical/ecomm/24q3.pdf

[21] T. Gneiting and A. E. Raftery, "Strictly Proper Scoring Rules, Prediction, and Estimation," *J. Am. Stat. Assoc.*, vol. 102, no. 477, pp. 359–378, 2007, doi: https://doi.org/10.1198/016214506000001437.