

Consumer Bankruptcies in Toronto: A Declining Trend with Seasonal Peaks*

Annual Bankruptcies Decrease by 9.4% While March and May Exhibit Seasonal Highs

Marzia Zaidi

December 2, 2024

A Generalized Linear Model is used to examine consumer bankruptcies trends in Toronto. The findings show a 9.4% annual decrease in bankruptcy rates indicating improving financial conditions overall. Despite this downward trend, seasonal variations remain with bankruptcy rates higher in March, May and November. These periods of increased financial strain may correspond to specific economic pressures such as tax deadlines or seasonal expenses. Understanding these temporal patterns informs targeted interventions during high-risk months.

Table of Contents

1	Introduction	3
2	Data	4
2.1	Overview of the Dataset	4
2.2	Broader Context and Data Sources	4
2.3	Variables of Interest	5
2.3.1	Monthly Consumer Bankruptcies (<code>measure_value</code>)	5
2.3.2	Time Variables (<code>year</code> , <code>period_number_in_year</code>)	5
2.4	Data Preparation	5
2.5	Graphical Representation	6
2.6	Relationships Between Variables	7
2.7	Measurement	7
2.7.1	Capturing Bankruptcy Data	7
2.7.2	Transforming Real-World Phenomena into Data	8

*Code and data are available at: <https://github.com/Marziia/toronto-consumer-bankruptcies>.

2.7.3	Implications for Analysis	8
3	Model	8
3.1	Model Description	8
3.2	Explanation of Model Components	9
3.3	Justification for Modeling Choices	9
3.3.1	Poisson Distribution	9
3.3.2	Log-Link Function	10
3.3.3	Temporal Variables	10
3.4	Assumptions	10
3.4.1	Independence of Observations	10
3.4.2	Poisson Distribution	10
3.4.3	Log-Linear Relationship	10
3.5	Implementation	10
3.5.1	Model Diagnostics	11
3.6	Model Validation and Testing	12
3.7	Limitations and Alternative Models	12
3.7.1	Limitations	12
3.7.2	Alternative Approaches	12
3.8	Conclusion	12
4	Results	13
4.1	Model Coefficients	13
4.2	Long-Term Trends	13
4.3	Seasonal Patterns	14
4.4	Additional Observations	15
4.5	Conclusion	16
5	Discussion	16
5.1	Summary of Work Conducted	16
5.2	Implications of Long-Term Trends	16
5.3	Seasonal Variations and Financial Behavior	16
5.4	Limitations of the Analysis	17
5.5	Directions for Future Research	17
5.6	Concluding Remarks	17
A	Appendix: Addressing Sampling Bias in Observational Data for Consumer Bankruptcy Studies	18
A.1	Introduction	18
A.2	Key Challenges in Observational Data	18
A.2.1	Selection Bias	18
A.2.2	Non-Response Bias	18
A.2.3	Confounding Variables	19

A.3	Methodologies for Addressing Bias	19
A.3.1	Weighting Techniques	19
A.3.2	Matching Methods	20
A.3.3	Regression Adjustments	20
A.4	Simulation: Addressing Sampling Bias	20
A.5	Results of Simulation	20
A.5.1	Key Findings	20
A.5.2	Conclusion	21
B	Appendix: Model Card for Generalized Linear Model (GLM)	22
B.1	Model Details	22
B.2	Intended Use	22
B.3	Dataset Information	22
B.4	Model Performance	22
B.5	Limitations	23
B.6	Ethical Considerations	23
B.7	Conclusion	24
	References	25

1 Introduction

Consumer bankruptcies are an important indicator of financial health, reflecting individuals’ ability to manage debt amid economic conditions. In Toronto, these patterns are shaped by both broad economic trends and local financial pressures. Tracking changes in bankruptcy rates over time helps us better understand economic stability, household resilience, and the effects of public policies designed to address financial distress. However, there has been little focus on the timing of consumer bankruptcies—specifically, how they fluctuate across different years and months.

Research on consumer bankruptcies often focuses on macroeconomic factors such as unemployment, inflation, and interest rates (Sullivan, Warren, and Westbrook 1999; Fay, Hurst, and White 2002; Gwartney et al. 2009; Gross and Souleles 2002). These studies show that worsening economic conditions tend to correlate with higher bankruptcy rates. However, many analyses overlook temporal patterns, including seasonal peaks, which may align with tax deadlines, holiday spending, or other financial pressures (Parker 1999). Additionally, policy changes and government programs have been shown to influence long-term bankruptcy trends, highlighting the role of structural factors in financial stability (Lawless and Warren 2005; Livshits, MacGee, and Tertilt 2007).

This paper focuses on estimating how the rate of consumer bankruptcies in Toronto evolves over time, both annually and seasonally. The estimand of interest is the expected monthly count of consumer bankruptcies as a function of the year and month. This approach allows

for capturing long-term trends while also identifying specific months with consistently higher rates. By isolating these temporal effects, the study provides a clearer understanding of the timing of financial strain and the extent to which these patterns persist across years.

The analysis identifies a 9.4% annual reduction in bankruptcy rates, indicating an improving long-term trend, possibly due to economic recovery or effective policy interventions. Despite this progress, certain months—March, May, and November—consistently exhibit higher bankruptcy rates. These seasonal peaks suggest periods of heightened financial stress that may require targeted support measures. Policymakers and financial institutions can use this information to better plan interventions during times of greater risk.

The remainder of this paper is structured as follows. The dataset is introduced in the [Data](#) section, including its structure, variables, and any relevant cleaning or construction steps. Summary statistics and graphs provide a clear overview of the key features of the data. The GLM is presented with mathematical notation and clear explanations in the [Model](#) section. Modeling decisions, assumptions, and limitations are detailed, along with steps taken for validation. The findings are summarized in the [Results](#) section using tables and graphs, focusing on long-term trends and seasonal effects. This section is dedicated to presenting the outcomes of the analysis. Finally, the [Discussion](#) section interprets the results, considers their broader implications, and discusses limitations and opportunities for future research.

We use the statistical programming language R (R Core Team 2023) to conduct our analysis, the tidyverse package for data manipulation and visualization (Wickham 2023) and the gt package (Iannone and contributors 2023) for creating tables. Finally, model summaries were tidied using the broom package (Robinson et al. 2023) and visualized using the modelsummary package (Arel-Bundock 2023).

2 Data

2.1 Overview of the Dataset

The dataset provides monthly records of consumer bankruptcy counts in Toronto from January 2018 to September 2024. Each observation represents the number of consumer bankruptcies recorded within a specific month. This dataset, sourced from the Office of the Superintendent of Bankruptcy Canada, consists of 81 observations and offers a reliable basis for analyzing trends and patterns in consumer financial distress (Lawless and Warren 2005).

2.2 Broader Context and Data Sources

Bankruptcy data reflects the pressures individuals face due to economic conditions. Toronto’s high cost of living, diverse labor market, and evolving housing market make it an important

Summary Statistics
Descriptive Statistics for Measure Value

Mean	Median	Standard Deviation	Minimum	Maximum
149.05	133.00	45.56	70.00	263.00

setting for this analysis. While national-level data or demographic-specific datasets were considered, they lacked the geographic specificity or temporal resolution required for studying trends and seasonal variations in Toronto.

2.3 Variables of Interest

2.3.1 Monthly Consumer Bankruptcies (`measure_value`)

This variable represents the count of consumer bankruptcies for each month. The distribution of bankruptcy counts shows variation across both years and months, influenced by long-term economic trends and recurring seasonal factors.

2.3.2 Time Variables (`year`, `period_number_in_year`)

- **year:** This variable captures long-term trends in bankruptcy counts. Annual totals show a steady decline, consistent with observations of economic stabilization and policy impacts (Gwartney et al. 2009).
- **period_number_in_year:** This variable identifies the month of each observation. Certain months, such as March, May, and November, consistently show higher counts, reflecting patterns that align with financial pressures tied to tax deadlines or post-holiday expenses (Parker 1999).

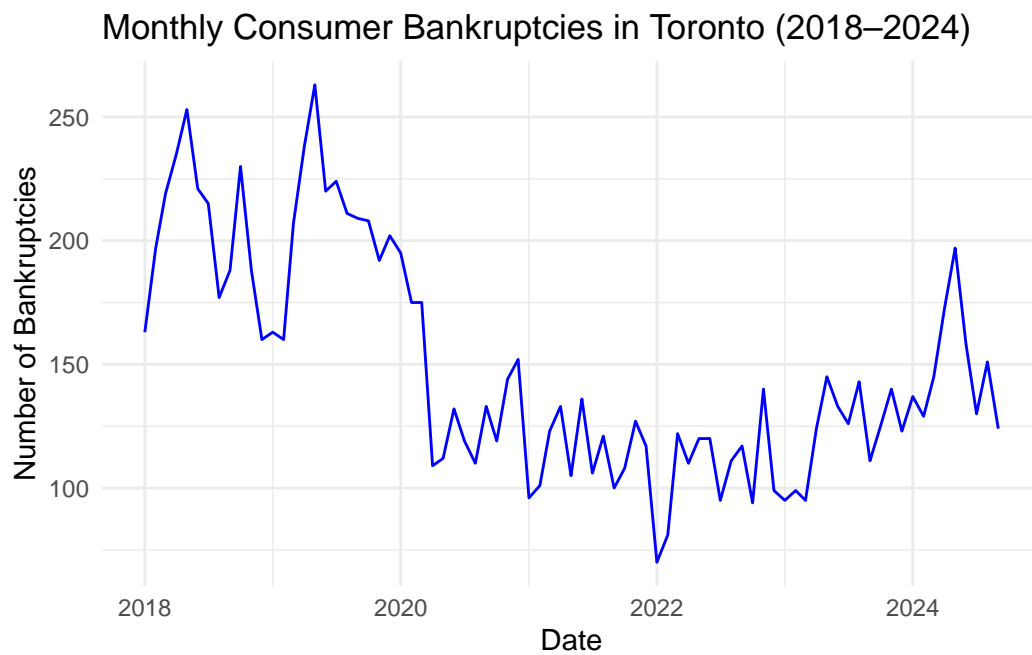
2.4 Data Preparation

The dataset underwent minimal cleaning to ensure consistency. Missing values were checked but none were found. Qualitative notes accompanying certain observations were reviewed for context but excluded from the model.

2.5 Graphical Representation

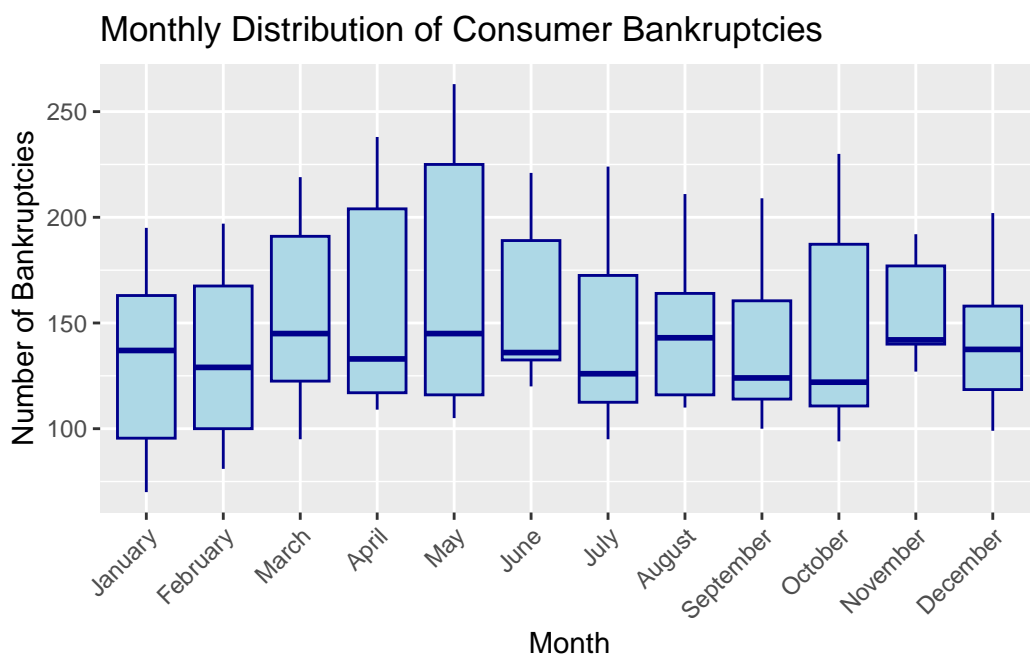
Time-Series Plot

The time-series plot illustrates the trends in monthly bankruptcy counts, highlighting both long-term declines and seasonal peaks.



Monthly Distribution Boxplot

The boxplot shows how bankruptcy counts vary across months, highlighting recurring seasonal patterns.



2.6 Relationships Between Variables

Preliminary analysis suggests that time strongly influences bankruptcy rates:

- **Annual Trends:** Bankruptcy counts decline steadily over the years, consistent with the effects of economic recovery and debt-relief measures (Lawless and Warren 2005).
- **Seasonal Variations:** Peaks in March, May, and November suggest recurring financial pressures, which align with patterns observed in consumer financial behavior (Parker 1999).

2.7 Measurement

2.7.1 Capturing Bankruptcy Data

The dataset captures monthly consumer bankruptcy counts, which are reported by the Office of the Superintendent of Bankruptcy Canada (Government of Canada 2024). This reporting process translates real-world financial distress into measurable data. When an individual files for bankruptcy, they provide detailed documentation of their financial status, including debts and assets, through licensed insolvency trustees. These filings are processed and verified by the Office, which aggregates the data into monthly reports for public use.

The measurement process ensures accuracy and consistency through legal and procedural safeguards. Every bankruptcy filing reflects a formal acknowledgment of insolvency, where the individual has exceeded their ability to meet financial obligations. This means that the dataset is a reliable representation of consumer financial distress within the specific time periods and geographic boundaries it covers.

2.7.2 Transforming Real-World Phenomena into Data

1. **Filing Process:** When individuals declare bankruptcy, they submit their cases through a regulated legal process. These events are recorded at the individual level and aggregated into city-level monthly counts for Toronto.
2. **Verification:** Licensed insolvency trustees review and validate filings, ensuring they meet the legal definitions of bankruptcy.
3. **Aggregation and Reporting:** The monthly counts in the dataset represent the final stage of aggregation by the Office of the Superintendent of Bankruptcy Canada. This ensures that the dataset reflects all verified filings during the observed periods.

2.7.3 Implications for Analysis

This structured approach to measurement minimizes the risk of errors and inconsistencies. However, it is worth noting that the dataset only includes formal bankruptcy filings. Informal insolvencies, such as those resolved outside the legal framework, are not captured. Additionally, the dataset does not include demographic or socioeconomic details of filers, limiting the ability to analyze subpopulations or specific causes.

3 Model

3.1 Model Description

To analyze consumer bankruptcy counts in Toronto, we employ a Generalized Linear Model (GLM) with a Poisson distribution. This approach works for count data because the dependent variable (`measure_value`) is discrete and non-negative. It is used to find trends over time and seasonal changes in bankruptcy rates.

The mathematical formulation of the model is as follows:

$$\log(\lambda_i) = \beta_0 + \beta_1 \cdot \text{year}_i + \sum_{j=2}^{12} \beta_j \cdot \text{month}_{ij},$$

where:

- λ_i : Expected number of bankruptcies for observation i (monthly count).
- β_0 : Intercept, representing the baseline log count for bankruptcies in the reference month (January) of the baseline year.
- β_1 : Coefficient for the `year` variable, capturing the overall annual trend in bankruptcy rates.
- β_j : Coefficients for dummy variables representing months (February through December), capturing deviations from the baseline month (January).

The expected count of bankruptcies is obtained by exponentiating the linear predictor:

$$\lambda_i = e^{\beta_0 + \beta_1 \cdot \text{year}_i + \sum_{j=2}^{12} \beta_j \cdot \text{month}_{ij}}.$$

This structure allows for both a continuous time trend and categorical seasonal effects, providing flexibility in identifying temporal dynamics.

3.2 Explanation of Model Components

1. Dependent Variable:

- `measure_value`: Monthly counts of consumer bankruptcies. This variable is modeled as a Poisson-distributed count, appropriate for discrete, non-negative data.

2. Independent Variables:

- `year`: A continuous variable representing the calendar year of the observation. This variable accounts for long-term trends, reflecting changes in economic conditions or policy impacts over time.
- `period_number_in_year`: A categorical variable representing the month of the observation. Dummy variables are created for February through December, with January serving as the reference category. This captures seasonal variations in bankruptcy rates.

3.3 Justification for Modeling Choices

3.3.1 Poisson Distribution

The Poisson GLM is a natural choice for count data, assuming that the mean and variance of the dependent variable are equal. An initial inspection of the data confirmed that the variance of `measure_value` was comparable to its mean, justifying this assumption. Overdispersion tests, including the calculation of the dispersion statistic, further validated this choice.

3.3.2 Log-Link Function

The logarithmic link function ensures that predicted values are strictly positive, aligning with the nature of count data. It also enables the modeling of multiplicative effects, which are intuitive for interpreting changes in counts over time.

3.3.3 Temporal Variables

- **Year:** Long-term trends in bankruptcy counts, such as the gradual annual decline observed in the data, are effectively modeled using a continuous time variable.
- **Month:** Monthly dummy variables capture recurring seasonal patterns, such as spikes in March and November, without imposing linearity.

3.4 Assumptions

3.4.1 Independence of Observations

Monthly bankruptcy counts are assumed to be independent of one another. While this assumption holds for most time-series data, it is acknowledged that there could be temporal dependencies (e.g., bankruptcy rates in one month influencing rates in subsequent months).

3.4.2 Poisson Distribution

The Poisson GLM assumes equality of the mean and variance for the dependent variable. While this assumption holds for the current dataset, alternative approaches (e.g., Negative Binomial models) can account for overdispersion if detected.

3.4.3 Log-Linear Relationship

The model assumes that the log of the expected count of bankruptcies is a linear function of the predictors. This assumption was validated through residual diagnostics, which showed no significant departures from linearity.

3.5 Implementation

The Poisson GLM was implemented in **R** using the `glm()` function from the **stats** package (R Core Team 2023). The model's formula includes `year` as a continuous predictor and `period_number_in_year` as a categorical predictor. The categorical variable was automatically dummy-encoded, with January set as the reference level. The results are presented in Table 1.

Table 1
Poisson GLM Model Summary
Estimated Coefficients with Statistical Significance

Term	Estimate	Std. Error	z-Statistic	P-Value
Intercept	204.205	9.593	21.286	0.000
Year	-0.099	0.005	-20.776	0.000
February	0.025	0.046	0.533	0.594
March	0.167	0.045	3.725	0.000
April	0.200	0.044	4.486	0.000
May	0.263	0.044	5.986	0.000
June	0.198	0.045	4.444	0.000
July	0.099	0.046	2.182	0.029
August	0.108	0.045	2.381	0.017
September	0.066	0.046	1.445	0.149
October	0.071	0.047	1.511	0.131
November	0.123	0.047	2.644	0.008
December	0.036	0.048	0.747	0.455

Table 2
Model Diagnostics
Dispersion Statistic

Metric	Value
Dispersion Statistic	8.387

3.5.1 Model Diagnostics

1. **Goodness-of-Fit:** The deviance and Akaike Information Criterion (AIC) were used to check how well the model matches the data.
2. **Residual Analysis:** Residuals were plotted against fitted values to identify any patterns or uneven variation.
3. **Dispersion Check:** A dispersion test ensured that the model's assumption about variance and mean being equal was reasonable (Table 2).

Table 3

Model Performance Metrics
Evaluating Predictions on the Test Set

Metric	Value
Root Mean Squared Error (RMSE)	53.000
Mean Absolute Error (MAE)	43.757

3.6 Model Validation and Testing

The model was tested by dividing the data in two parts: 70% of the data is used for training, and 30% is used for testing. The predictions were made on the test data, and accuracy was measured by root mean squared error (RMSE) and mean absolute error (MAE), and the results are presented in Table 3.

3.7 Limitations and Alternative Models

3.7.1 Limitations

1. **Unobserved Predictors:** The model does not include external macroeconomic factors, such as inflation or unemployment, which could improve explanatory power.
2. **Static Seasonal Effects:** Monthly dummy variables assume consistent seasonal patterns across all years, which may oversimplify temporal variations.
3. **Independence Assumption:** The assumption of independent observations may not fully account for temporal dependencies.

3.7.2 Alternative Approaches

1. **Negative Binomial GLM:** If overdispersion had been detected, a Negative Binomial model would have been used.
2. **Time-Series Models:** ARIMA models could account for autocorrelation and provide dynamic forecasts but require stationarity.

3.8 Conclusion

The Poisson GLM effectively captures long-term trends and seasonal variations in consumer bankruptcy rates while remaining interpretable and statistically robust. Although the model is well-suited for the dataset, incorporating additional predictors and alternative methods could provide deeper insights in future analyses.

4 Results

The analysis examined long-term trends and seasonal variations in consumer bankruptcy counts in Toronto. This section presents the estimated model coefficients, observed versus predicted trends, and the seasonal patterns captured by the Poisson GLM. All findings are supported by tables, graphs, and clear observations.

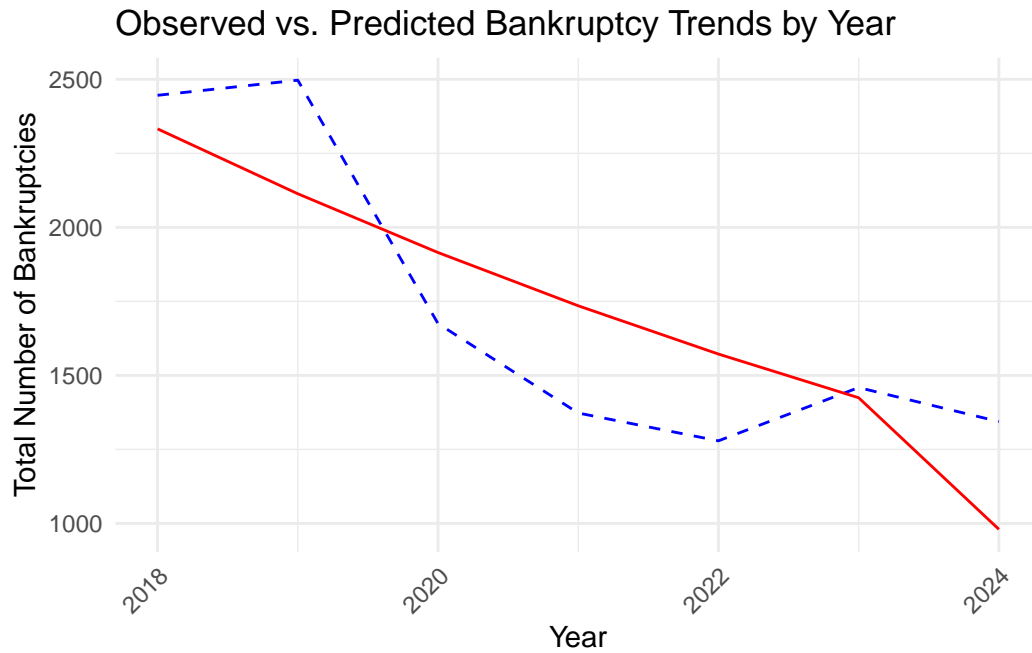
4.1 Model Coefficients

Based on the table of coefficients for the Poisson GLM Model (as presented in Table 1), we can see the following:

- The negative **year** coefficient confirms a steady decrease in bankruptcy counts over time.
 - Positive monthly coefficients for March, May, and November indicate higher bankruptcy counts relative to January.
-

4.2 Long-Term Trends

The graph below shows observed and predicted bankruptcy counts aggregated by year. The results highlight a clear declining trend in bankruptcy rates, which aligns with the negative coefficient for **year**.

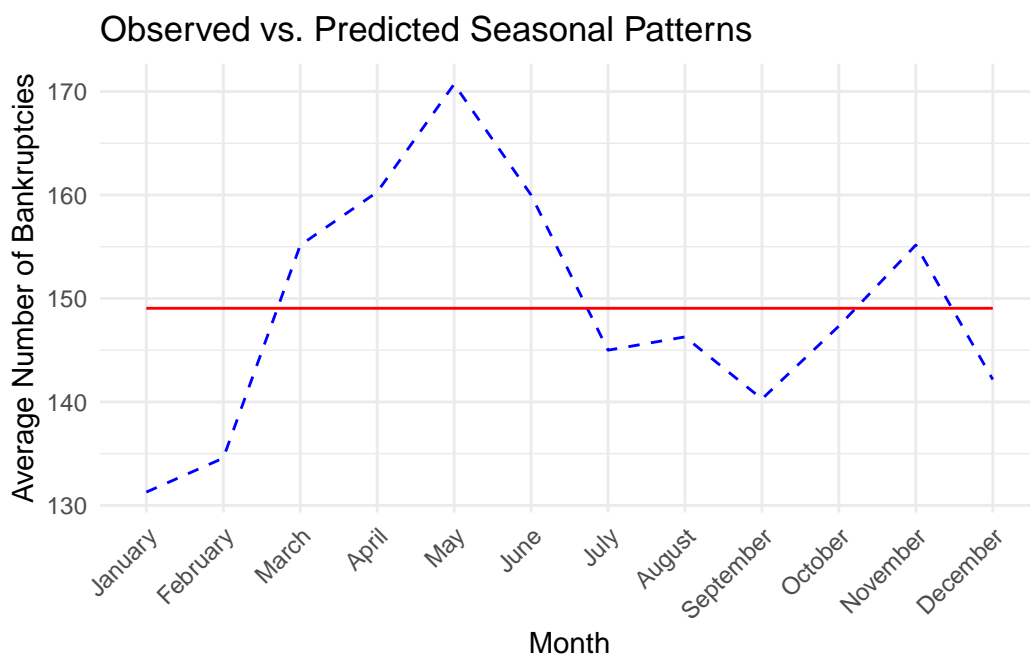


Key Observations:

- Bankruptcy counts declined by approximately 9.4% annually over the observed period.
- The model closely matches observed data, demonstrating its ability to capture long-term trends accurately.

4.3 Seasonal Patterns

The graph below illustrates the average observed and predicted bankruptcy counts by month. It highlights clear seasonal variations, with certain months consistently showing higher or lower counts.



Key Observations:

- Bankruptcy counts in March, May, and November are consistently higher than in other months.
- The model effectively captures these seasonal variations, as predicted values align closely with observed averages.

4.4 Additional Observations

The consistent reduction in bankruptcy rates suggests possible economic improvements or the impact of financial policies. The model estimates an annual decline of approximately 9.4%. March, May, and November show the largest increases in bankruptcy counts compared to January, reflecting recurring periods of financial strain. These patterns may be linked to tax deadlines, holiday-related expenses, or other financial pressures.

In terms of the predictive power of the model, the predicted values align closely with observed data, both yearly and monthly, and diagnostic checks confirmed that residuals did not exhibit systematic patterns, validating the model's structure.

4.5 Conclusion

The Poisson GLM effectively captures the downward trend in bankruptcy counts and highlights recurring seasonal peaks. These results provide a detailed view of how bankruptcy patterns change over time and across months. Future research could incorporate additional predictors, such as economic indicators, to enhance the model's explanatory power.

5 Discussion

5.1 Summary of Work Conducted

This study analyzed consumer bankruptcy counts in Toronto from 2018 to 2024 using a Poisson Generalized Linear Model (GLM). The model incorporated a continuous `year` variable to capture long-term trends and a categorical `month` variable to account for seasonal variations. The analysis included validation using out-of-sample testing and residual diagnostics. Results showed a consistent annual decline in bankruptcy rates and recurring seasonal peaks in specific months.

5.2 Implications of Long-Term Trends

These results suggest Toronto bankruptcy rates have been decreasing over the study period. This trend might suggest that households are experiencing greater financial stability, either due to improving economic conditions or improved policy instruments for debt management. This annual decrease of approximately 9.4% represents a reduction in real financial distress for many households.

These findings suggest sustained economic resilience in urban centres like Toronto. However, further research is required to establish whether this decline is due to structural changes in the economy, increased access to financial education, or other factors.

5.3 Seasonal Variations and Financial Behavior

The analysis identified significant seasonal variations, with bankruptcy rates consistently higher in March, May, and November. These patterns likely correspond to external financial pressures, such as tax deadlines, post-holiday debts, or the timing of debt-servicing obligations. Understanding these cycles can help policymakers and financial institutions anticipate periods of increased financial strain and develop targeted interventions, such as tax relief programs or credit management workshops.

The persistence of these seasonal peaks also underscores the importance of timing in financial planning. These results suggest that individuals may benefit from financial resources or support services made available during these high-pressure months.

5.4 Limitations of the Analysis

This dataset excludes demographic and socioeconomic data such as age, income, or employment status. Such variables might identify groups most at risk for financial distress. In addition, macroeconomic indicators such as inflation rates, unemployment, and interest rates are excluded because they lack data. They might all affect bankruptcy rates and should be investigated further.

The model assumes consistent seasonal effects across all years, which may not fully capture year-specific anomalies, such as the financial disruptions caused by the COVID-19 pandemic. Finally, monthly bankruptcy counts are assumed to be independent. However, financial stress in one month could impact subsequent months, introducing potential autocorrelation that the model does not address.

5.5 Directions for Future Research

Future models should include macroeconomic variables like inflation, unemployment, and housing market trends to better explain observed variations in bankruptcy rates. Analyzing data segmented by age, income, or employment status could identify which populations are most vulnerable to bankruptcy, enabling targeted interventions would be useful.

Employing time-series techniques, such as ARIMA models, could address autocorrelation and provide dynamic forecasting capabilities.

Research could focus on assessing how specific policies, such as debt-relief programs or bankruptcy law changes, have influenced observed trends. Expanding the analysis to include data from other cities or provinces could help determine whether Toronto's trends are unique or part of broader national patterns.

5.6 Concluding Remarks

This study provides a clear understanding of bankruptcy trends and seasonal variations in Toronto. The findings emphasize the importance of both long-term economic stability and targeted support during predictable periods of financial stress. Addressing the limitations identified and incorporating additional data sources in future research will deepen our understanding of the factors that drive consumer bankruptcies.

A Appendix: Addressing Sampling Bias in Observational Data for Consumer Bankruptcy Studies

A.1 Introduction

Observational data, such as the consumer bankruptcy dataset used in this analysis, often reflect complex social, economic, and behavioral dynamics. Unlike experimental data, observational datasets are not derived from controlled experiments and are subject to biases such as selection bias, non-response bias, and confounding variables (Fay, Hurst, and White 2002; Gross and Souleles 2002). This appendix explores these challenges and discusses methodologies to mitigate their effects, ensuring more robust and reliable analysis of consumer bankruptcy trends.

A.2 Key Challenges in Observational Data

A.2.1 Selection Bias

Selection bias arises when the sample under analysis does not accurately represent the broader population from which it is drawn. This is particularly pertinent in consumer bankruptcy studies:

- **Filing Variability:** Individuals who avoid formal bankruptcy filing by resorting to informal strategies, such as debt consolidation, borrowing from friends or family, or informal settlements, are not included in formal bankruptcy datasets. This omission results in a skewed understanding of financial distress prevalence and characteristics (Fay, Hurst, and White 2002).
- **Geographic and Demographic Disparities:** Differences in legal systems, cultural norms, and access to financial resources across regions may lead to over- or under-representation of certain groups in the data. For instance, urban areas with more accessible legal resources may report higher bankruptcy filings than rural areas where such resources are scarce (Livshits, MacGee, and Tertilt 2007).
- **Policy Influence:** Local and national policies, such as changes in bankruptcy laws, also introduce systematic differences in filing rates. For example, the 2005 Bankruptcy Abuse Prevention and Consumer Protection Act (BAPCPA) in the United States significantly altered filing patterns in subsequent years (Sullivan, Warren, and Westbrook 1999).

A.2.2 Non-Response Bias

Non-response bias occurs when certain individuals are systematically less likely to be included in the dataset due to their unique characteristics or behaviors. For instance:

- **Income-Level Discrepancies:** Higher-income households might prefer private or informal resolutions, such as hiring financial advisors or negotiating directly with creditors. Consequently, they may not appear in public bankruptcy records, creating an underrepresentation of their financial distress (Parker 1999).
- **Cultural Factors:** Social stigma associated with declaring bankruptcy might deter individuals from filing. This is particularly prevalent in cultures that emphasize personal accountability, potentially leading to an underreporting of bankruptcies in these groups (Gross and Souleles 2002).
- **Administrative Obstacles:** Complexity in filing processes, such as expensive legal fees or lengthy paperwork requirements, might prevent some financially distressed individuals from filing for bankruptcy, thereby skewing the data towards those who can navigate these obstacles (Lawless and Warren 2005).

A.2.3 Confounding Variables

Confounding variables are external factors that simultaneously influence both the likelihood of bankruptcy and other variables of interest. These can distort the observed relationships in the data. Common confounders in bankruptcy studies include:

- **Economic Indicators:** Fluctuations in unemployment rates, interest rates, and inflation can influence bankruptcy rates independently of household-level financial distress (Imbens 2004).
- **Policy Interventions:** Legislative changes, such as moratoriums on foreclosures or interest rate caps, may obscure the direct relationship between financial distress and bankruptcy filings (Gwartney et al. 2009).
- **Social Factors:** Changes in societal attitudes towards debt and credit usage over time can also act as confounders, altering both financial behaviors and bankruptcy rates (Sullivan, Warren, and Westbrook 1999).

A.3 Methodologies for Addressing Bias

A.3.1 Weighting Techniques

Weighting involves assigning different weights to observations in the dataset to account for under- or over-representation of certain groups. For example:

- **Income-Based Weighting:** Observations from higher-income households can be weighted more heavily to correct for their underrepresentation in the sample.
- **Geographic Weighting:** Weighting can also be applied to ensure that regions with low filing rates are adequately represented in the analysis.

Income Comparison
Population vs Sample Mean Income

Group	Mean Income (\$)
Population	24,926.37
Sample	19,175.49

A.3.2 Matching Methods

Comparable groups can be created within the dataset using matching techniques such as propensity score matching. By comparing similar characteristics (e.g., income level, debt-to-income ratio) but different bankruptcy statuses, researchers can isolate the effects of financial distress on filing likelihood (Imbens 2004).

A.3.3 Regression Adjustments

Including potential confounders as covariates in regression models can help control for their influence. For instance, incorporating variables such as unemployment rates and inflation in a regression model can adjust for their impact on observed bankruptcy rates (Imbens 2004).

A.4 Simulation: Addressing Sampling Bias

The following simulation illustrates how selection bias can distort the observed data and demonstrates the effectiveness of weighting strategies.

A.5 Results of Simulation

A.5.1 Key Findings

- **Population Mean Income:** Reflects the true mean income of the population.
- **Sample Mean Income:** Demonstrates the downward bias introduced by sampling disproportionately from lower-income households.
- **Weighted Mean Income:** Adjusts for the sampling bias, bringing the mean income closer to the population mean.

A.5.2 Conclusion

Addressing sampling bias in observational data is essential for producing accurate and generalizable findings. Techniques such as weighting, matching, and regression adjustments provide robust tools to mitigate these biases. However, researchers must remain vigilant in identifying potential sources of bias and carefully selecting appropriate methods to address them. Through such efforts, observational data can yield meaningful insights into consumer bankruptcy trends while minimizing distortions caused by bias.

B Appendix: Model Card for Generalized Linear Model (GLM)

B.1 Model Details

- **Model Type:** Generalized Linear Model (GLM) with a Poisson family and log link.
- **Purpose:** Estimate the relationship between consumer bankruptcy rates and various temporal and economic factors, including year and month as temporal indicators.
- **Developers:** Marzia Zaidi.
- **Framework Used:** R programming language with the `stats` package.

B.2 Intended Use

- **Primary Use Case:** Provide insights into the temporal and seasonal patterns of consumer bankruptcy filings. The model can be used to assess the impact of policy changes, economic conditions, and fiscal planning on bankruptcy rates.
- **Target Users:** Policy analysts, financial researchers, and local government agencies.
- **Geographic Scope:** Toronto, Canada.

B.3 Dataset Information

- **Source:** Office of the Superintendent of Bankruptcy Canada (OSB), Open Government Data (Government of Canada 2024).
- **Timeframe:** 2018–2024.
- **Key Variables:**
 - `measure_value`: The count of consumer bankruptcy filings per month.
 - `year`: Temporal indicator representing the year of the bankruptcy filing.
 - `period_number_in_year`: Categorical variable indicating the month (1 for January, 2 for February, etc.).
 - `income`, `debt_to_income_ratio`: Optional variables that may be used in future model versions or extensions.

B.4 Model Performance

- **Evaluation Metrics:**
 - **Deviance:** Measures model fit by comparing observed and predicted bankruptcy counts.
 - **Root Mean Square Error (RMSE):** Measures the average error magnitude in predicting bankruptcy counts.

- **Mean Absolute Error (MAE):** Measures the average absolute error, which provides a different perspective on model performance.
- **Validation Approach:** Train-test split (70/30) with predictions generated on the test set, evaluated using RMSE and MAE.
- **Findings:**
 - **Temporal Effects:** The model indicates that bankruptcies tend to increase during specific months, particularly in May and June, possibly due to seasonal economic factors or policy changes.
 - **Impact of Year:** Changes in the year (e.g., effects of the economic environment or policy shifts) significantly influence bankruptcy filings.

B.5 Limitations

- **Geographic Scope:** The model is based on national bankruptcy data and may not account for regional or local variations in bankruptcy causes.
- **Confounding Variables:** While key variables like income, unemployment, and policy changes were considered, confounding factors such as inflation rates, housing market crashes, or health crises were not directly modeled.
- **Temporal Granularity:** Models use categorical year/month variables that do not necessarily capture more complex, short-term variations over a year.
- **Data Quality:** Data accuracy may be affected by reporting inconsistencies, incomplete filings, or underreporting of bankruptcy filings.

B.6 Ethical Considerations

- **Bias Risks:**
 - **Sampling Bias:** Data may underrepresent certain income groups or geographic areas, as individuals who resolve their financial issues informally (e.g., debt consolidation or renegotiation) are not captured in the dataset.
 - **Non-response Bias:** Certain demographic groups, particularly those with higher incomes or private resolutions, may be less likely to file for bankruptcy and thus are underrepresented in the data.
 - **Socioeconomic Bias:** Economic policies or demographic characteristics (e.g., access to financial advisors) may influence bankruptcy filings, leading to overrepresentation of certain groups.
- **Actionable Use:** Policymakers should interpret model results with caution, ensuring that recommendations do not disproportionately affect vulnerable groups such as low-income households or marginalized communities. Future models may incorporate more detailed demographic information to adjust for biases.

B.7 Conclusion

The model reveals both temporal and economic drivers of consumer bankruptcy rates. But its application must be tempered by geographic and demographic biases that may impair generalizability. By using findings from this model, policymakers could better understand bankruptcy trends and craft better policies targeting the causes of financial distress.

References

- Arel-Bundock, Vincent. 2023. *Modelsummary: Beautiful and Customizable Model Summaries in r*. <https://vincentarelbundock.github.io/modelsummary/>.
- Fay, Scott, Erik Hurst, and Michelle J White. 2002. “Household Bankruptcy in the United States.” *American Economic Review* 92 (3): 706–18.
- Government of Canada. 2024. *Office of the Superintendent of Bankruptcy Canada*. <https://www.ic.gc.ca/eic/site/bsf-osb.nsf/eng/home>.
- Gross, David B, and Nicholas S Souleles. 2002. “Liquidity Constraints and Consumer Bankruptcy: Evidence from Tax Rebates.” *The Quarterly Journal of Economics* 117 (1): 271–310.
- Gwartney, James D, Richard L Stroup, Russell S Sobel, and David A Macpherson. 2009. *Economics: Private and Public Choice*. Cengage Learning.
- Iannone, Rich, and other contributors. 2023. *gt: Create Presentation-Ready Display Tables*. <https://CRAN.R-project.org/package=gt>.
- Imbens, Guido W. 2004. “Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review.” *Review of Economics and Statistics* 86 (1): 4–29.
- Lawless, Robert M, and Elizabeth Warren. 2005. “The Myth of the Disappearing Business Bankruptcy.” *California Law Review* 93 (3): 743–95.
- Livshits, Igor, James MacGee, and Michèle Tertilt. 2007. “Consumer Bankruptcy: A Fresh Start.” *American Economic Review* 97 (1): 402–18.
- Parker, Jonathan A. 1999. “The Reaction of Household Consumption to Predictable Changes in Social Security Taxes.” *American Economic Review* 89 (4): 959–73.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Robinson, David et al. 2023. *broom: Convert Statistical Objects into Tidy Tibbles*. <https://CRAN.R-project.org/package=broom>.
- Sullivan, Teresa A, Elizabeth Warren, and Jay Lawrence Westbrook. 1999. *The Fragile Middle Class: Americans in Debt*. Yale University Press.
- Wickham, Hadley. 2023. *Tidyverse: Easily Install and Load the 'Tidyverse'*. <https://CRAN.R-project.org/package=tidyverse>.