Procurement Dynamics: Analyzing Federal Contract Awards to Microsoft*

Trends of Microsoft's Future in Canadian Government from 2020

Deyi Kong

December 2, 2024

This paper analyzes the federal procurement contracts awarded to Microsoft, focusing on trends in award values, contract categories, and the relationships between buyers and Microsoft. The analysis reveals that Microsoft dominates high-value contracts and demonstrates consistent partnerships with key government departments. This paper utilizes a linear model and random forest to predict the future contract amount trend for Microsoft. By examining these procurement dynamics, this paper providing insights into the broader implications of tech-industry participation in government procurement.

Table of contents

---

*Code and data are available at: https://github.com/eeeee-cmd/Procurement/.

1 Introduction

Government procurement serves as a critical avenue for public sector innovation, economic stimulus, and the fulfillment of strategic objectives. Within this landscape, the relationship between major technology providers and governmental bodies has become increasingly prominent. Among these, Microsoft has emerged as a pivotal partner in driving IT modernization and infrastructure development. Despite the growing reliance on tech giants, limited research has comprehensively examined the dynamics of such partnerships, particularly in the context of federal procurement contracts.

This paper addresses this gap by conducting a detailed analysis of federal procurement contracts awarded to Microsoft between 2020 and September 2024. Using publicly available data from the Investigative Journalism Foundation, the study explores trends in contract values, contract types, and Microsoft's relationships with key government buyers. Emphasis is placed

on identifying patterns in contract allocation and the implications of Microsoft's dominance in high-value contracts.

The main estimand in this study is the amount of federal procurement contracts awarded to Microsoft. This metric is analyzed across multiple dimensions, including contract categories, duration, and partnerships, to quantify Microsoft's role in federal procurement comprehensively. The findings reveal that Microsoft plays a dominant role in IT infrastructure and modernization, securing the majority of high-value contracts and establishing consistent partnerships with departments such as Employment and Social Development Canada and National Defence. These insights highlight the broader implications of the private tech sector's influence on public sector procurement, raising questions about competitiveness, dependency, and strategic alignment.

This study contributes to understand the procurement landscape, providing valuable information for policymakers, industry stakeholders, and academics. It underscores the importance of transparency, strategic planning, and balancing innovation with equity in government contracts.

The structure of the paper is as follows: Section 2 outlines the data sources and variables considered, followed by the model setup in Section 3.1 and justification. The results in Section 4 presents the key findings of the analysis, with a discussion on the implications. Section 5 then discusses potential limitations and suggestions for future research. **?@sec-appx** provides additional detailed information about the data, model and methodology.

2 Data

2.1 Overview

The data used in this analysis comes from publicly available procurement data (Investigative Journalism Foundation 2024a). The analysis uses the statistical programming language R (R Core Team 2023) and several libraries, including `tidyverse` (Wickham et al. 2019), `janitor` (Firke 2023), `knitr` (Xie 2024), `dplyr` (Wickham et al. 2023), `arrow` (Richardson et al. 2024), `purrr` (Wickham and Henry 2023), and `here` (Müller 2020) for data manipulation. `ggplot2` (Wickham 2016) and `kableExtra` (Zhu 2024) for visualization. The dataset covers the information of chosen Federal Procurement Supplier - Microsoft - capturing the contracts from 2020 to 2024 (up to September).

2.2 Measurement

The dataset used in this analysis originates from Canadian governments of all levels post solicitations for companies to search for business opportunities, as documented in Investigative

Journalism Foundation (2024b), then collects data from each of these platforms and unifies into standard format.

The measurement process delegates contracts routed through Public Service and Procurement Canada (PSPC), and blanket disclosures organized by Treasury Board of Canada Secretariat (TBS) to understand the federal procurement. The methodologies of federal solicitations are either competitive or non-competitive. Competitive contracts accept bids from all or some subset of suppliers. Non-competitive contracts do not, citing certain exceptions, including a "pressing emergency in which delay would be injurious to the public interest," or when the estimated cost to the buyer is below certain thresholds defined for some goods and services.

More information about solicitation method can be found in **?@sec-appx**.

2.3 Data Cleaning

After imported the raw dataset using `read_csv` function, I first select specific columns to focus on relevant information towards this study by omitting redundant information, such as the contract link and same value of region. Then I rename key variables from raw data for clarity. The key variables of interest in this analysis include Contract, StartDate, AwardDate, EndDate, Buyer, and Amount.

For any missing data (NA) of StartDate, I infer the StartDate from the AwardDate to replace the missing value as the value of AwardDate. Then I standardize the date format of all date related datas for proper parsing by using `mdy()` function and change the dollar format of amount into numeric numbers for future calculation purposes. Due to the error with unknown reason that EndDate is before StartDate or AwardDate, those entries were excluded from this paper.

PreparatoryPhase, PhaseDays, and ContractDays variables were created for the purpose of analyzing preparatory phase and performance phase corresponding to each contract. PreparatoryPhase calculates the days from AwardDate to StartDate. ContractDays calculates the days from StartDate to EndDate. PhaseDays take the absolute value of PreparatoryPhase because of the minor effect of either prepare before contract being awarded or prepare after contract being awarded. Last but not least, I created a new variable called ContractType, which categorizes the contracts based on keywords.

This dataset contains only two entries for 2019 federal procurement contracts awarded to Microsoft. So, I filter out the contracts before 2020 to focus on the records between 2020 and September 2024 as desired. Then I check the character type variables, Supplier and Contract, using `table()` function for the purpose of manipulating input error. I standardize the supplier's name which combine the similar name due to capitalization or some notation differences, and format them with first letter of each word capitalized.

After completing the cleaning, we saved the final dataset in both Parquet and CSV formats for later analysis.

2.4 Outcome Variables

The main outcome variable of interest is the amount of federal procurement contracts awarded to Microsoft. This variable is need to represent the financial magnitude of Microsoft with the federal government and measurement of evaluating procurement dynamics. Figure 1 illustrates the annual total amounts of federal procurement contracts awarded to Microsoft from 2020 to 2024 (up to September). The histogram reveals a generally increasing trend of total value of each year's contract. This may reflect the growing of technology in government operations, including software, cloud solutions, and IT infrastructure. The amount in 2023 doubles the amount in 2021, which suggests a heightened investment in federal procurement funding and large-scale projects.
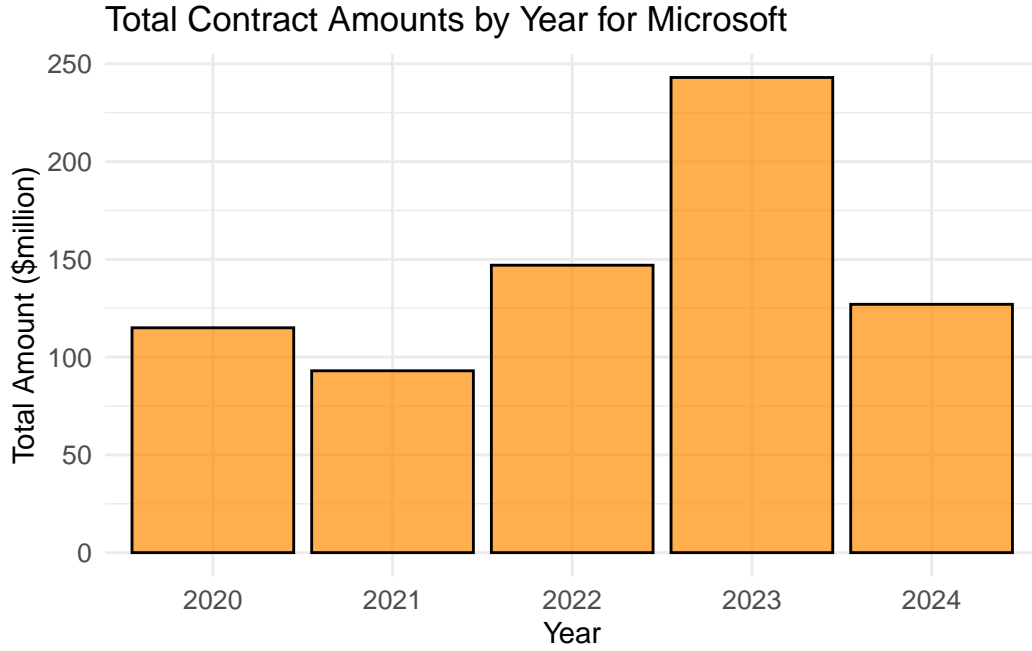
Figure 1: The histogram shows the distribution of the total amount($million) of each year's federal contract (since 2020).

2.5 Predictor Variables

In this analysis, several key predictors were identified to evaluate procurement data effectively. The first variable, **ContractType**, is a character-type variable that identifies each procurement contract awarded to Microsoft into Categories. Table 1

Table 1: The table summarizes the contract type which has over 25 contracts with the total dollar amount, and average dollar amount per contract.

Contract Type	Counts	Total Amount	Average
License/Maintenance Fees Related	495	$250,972,089	$507,014
Computer Services Related	118	$208,084,958	$1,763,432
Distributed Computing Environment Related	99	$25,125,005	$253,788
Consulting Related	83	$72,849,085	$877,700
Database Access Services Related	49	$82,342,191	$1,680,453
Software Related	45	$31,407,778	$697,951

Buyer is a character-type variable representing the federal government department or agency who initiate and award the contract. Table 2 shows the top 10 buyer's name who has over 50

contracts with Microsoft, which identifies key partners and understand the distribution of contracts across sectors. The table highlights that Employment and Social Development Canada (ESDC) and National Defence account over 50% of Microsoft's total awards. [CALCULATED 52.02%, BUT THINKING IF THIS SHOULD BE ADDED TO THE TABLE]

Table 2: The table summarizes top 10 buyer's name who has over 25 contracts with Microsoft in total with contract count, total dollar amount, and average dollar amount per contract.

Buyer	Contract	Total Amount	Average
Employment and Social Development Canada	116	$136,615,708	$1,177,722
Global Affairs Canada	71	$30,011,093	$422,691
National Defence	64	$81,913,883	$1,279,904
Natural Resources Canada	35	$12,856,204	$367,320
National Research Council Canada	32	$4,436,990	$138,656
Transport Canada	31	$16,297,432	$525,724
Indigenous Services Canada	28	$5,000,374	$178,585
Natural Sciences and Engineering Research Council of Canada	26	$4,626,898	$177,958
Shared Services Canada	26	$17,588,415	$676,478
Social Sciences and Humanities Research Council of Canada	26	$4,626,898	$177,958
Health Canada	25	$33,987,726	$1,359,509
Innovation, Science and Economic Development Canada	22	$10,136,919	$460,769
Correctional Service of Canada	20	$8,745,210	$437,260
Royal Canadian Mounted Police	20	$45,570,082	$2,278,504
Statistics Canada	20	$115,057,275	$5,752,864
Canada Revenue Agency	19	$27,559,347	$1,450,492
Immigration, Refugees and Citizenship Canada	18	$14,660,541	$814,474
Privy Council Office	18	$3,144,178	$174,677
Treasury Board of Canada Secretariat	18	$18,770,370	$1,042,798
Fisheries and Oceans Canada	16	$7,278,132	$454,883

This dataset contains a series of date-type variables capturing key milestones in the procurement process. **Award Date** is the date when the contract was officially granted to Microsoft. **Start Date** is the date when the contracted services or goods provision begins. **End Date** is the date when the contract is completed or terminated. These variables are essential for analyzing procurement timelines and trends over the study period.

PhaseDays is a numeric variable demonstrates the days between AwardDate and StartDate.

In Figure 2, there is a high concentration around 0 indicates rapid procurement processes or emergency software services. Figure 2 also shows a few outlying contracts with a mobilization period of over one year. These long mobilization periods reflects issues such as delays in the procurement process, a long negotiation period, or administrative approvals.
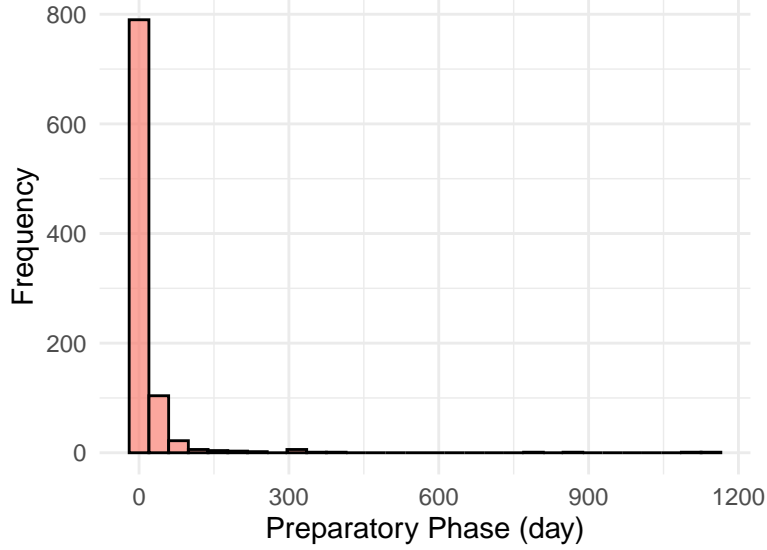

Figure 2: The histogram shows the distribution of preparatory period.

ContractDays is a numeric variable which is the days between StartDate and EndDate. Figure 3 shows the distribution of performance period. The right skewed distribution in Figure 3 indicates that the majority of contract duration is within a year. There is an over 5 years performance period in Figure 3 implies a massive project that requires extensive preparation.

3 Model

3.1 Model Set-Up

This paper utilizes a Bayesian model to examine the factors influencing the dollar amount of federal procurement contracts awarded to Microsoft. Define $Amount_i$ as the dollar amount of the federal procurement contract awarded to Microsoft and μ_i is the expected value of the contract amount modeled as a linear combination of the predictors. The model for predicting
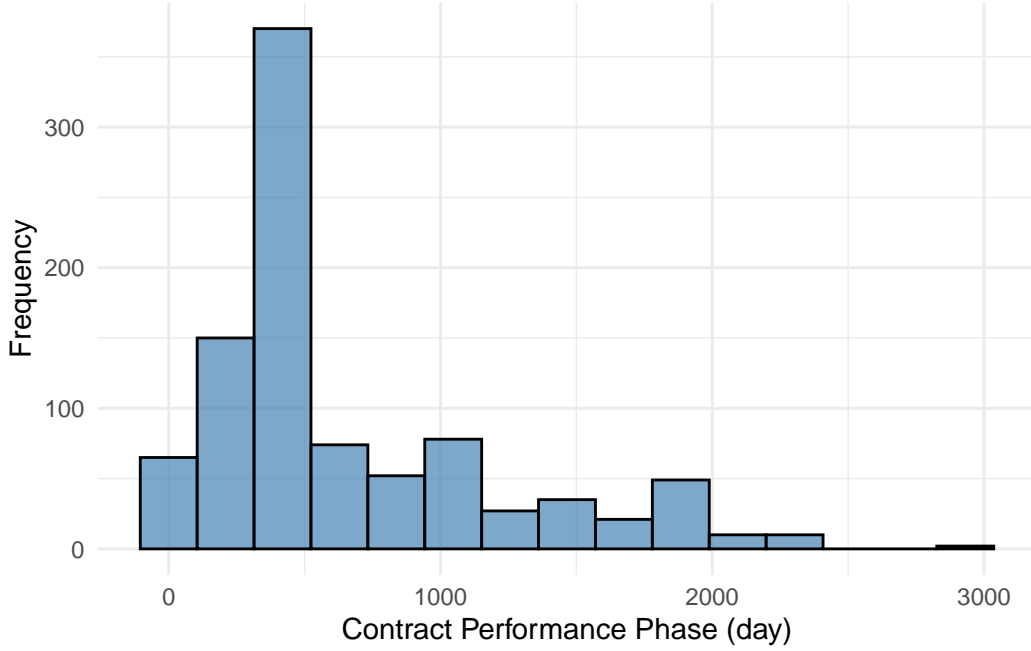
Figure 3: The histogram shows the distribution of performance period, which demonstrates the days between start and end date.

$Amount_i$ is specified as follows:

$$Amount_i \mid \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta_1 \cdot \text{ContractType}_i + \beta_2 \cdot \text{Buyer}_i + \beta_3 \cdot \text{ContractDays}_i + \beta_4 \cdot \text{PhaseDays}_i$$
$$\alpha \sim \text{Normal}(0, 2)$$
$$\beta_1, \beta_2, \beta_3, \beta_4 \sim \text{Normal}(0, 2)$$
$$\sigma \sim \text{Exponential}(1)$$

Where:
- α is the intercept term.
- $\beta_1, \beta_2, \beta_3, \beta_4$ are the coefficients associated with the predictors:
- $ContractType_i$: A categorical variable representing the contract type.
- $Buyer_i$: A categorical variable representing the buyer of the contract.
- $ContractDays_i$: A numeric variable representing the duration of the contract is active.
- $PhaseDays_i$: A numeric variable representing the days between AwardDate and StartDate.

The model is fit using the `rstanarm` package of Goodrich et al. (2022) in R (R Core Team 2023), with the specified priors guiding the model while allowing it to adapt to the data.
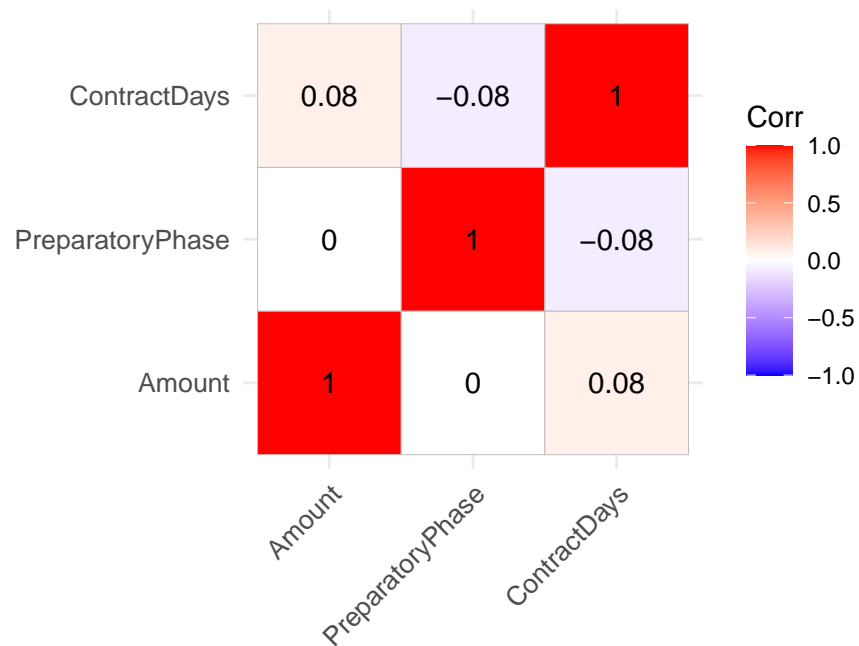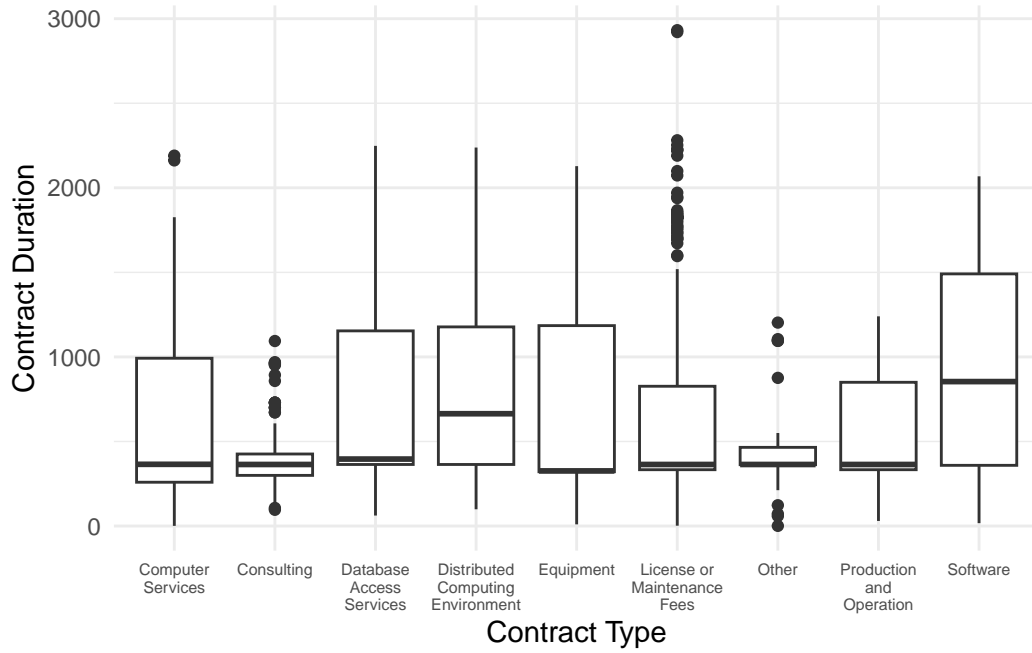
Figure 4: This correlation matrix visualizes the relationships between key numerical predictor variables and the outcome variable, Amount. The figure shows correlations include a weak negative correlations between PreparatoryPhase and ContractDays; and weak positive correlations between predictors and Amount.

(a) Contract Duration in each Contract Type
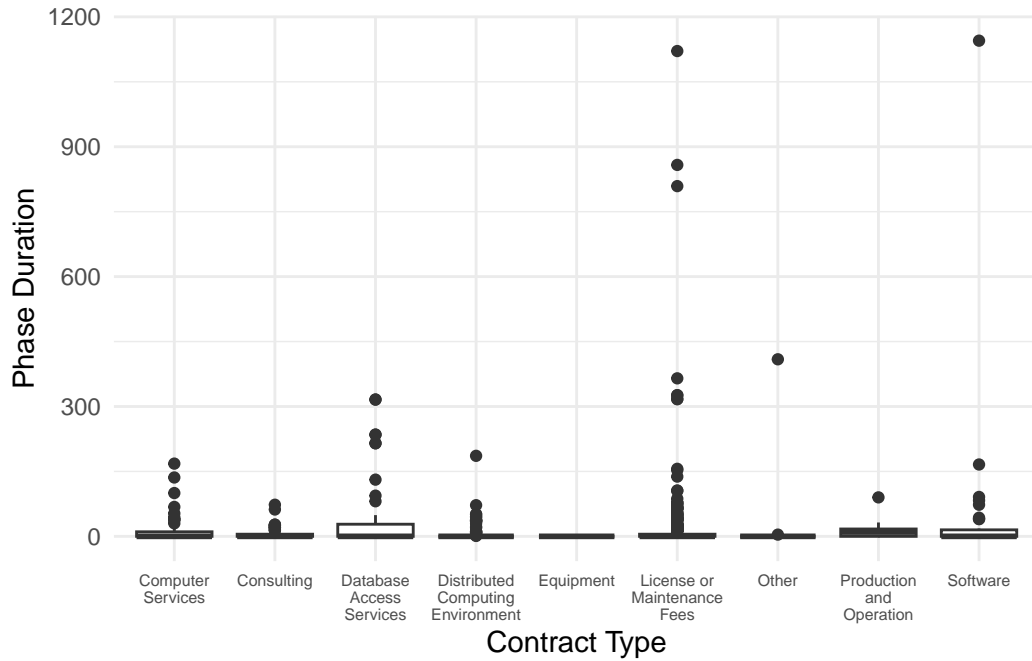
(b) Phase Duration in each Contract Type

Figure 5: The Boxplots shows the relationship of ContractDays and PhaseDays with Contract-Type.

3.1.1 Model Justification

This paper expect that the factors - such as ContractType, Buyer, ContractDays and Phase-Days - will influence the outcome of contract amount. In particular, …

In this paper, I compared three different modeling approaches to examine the factors influencing the federal procurement contract amount: linear regression, random forest, and Bayesian regression. Each of these models offers different advantages and trade-offs. More detailed justification about model selection will be discussed in Section 5. The Bayesian model is selected due to better performance and flexibility.

Bayesian approach can explicitly model uncertainty by incorporating prior knowledge and provides a full posterior distribution of the model parameters. This is useful when the data is sparse and prior knowledge about the contract types, buyers, or duration is available. The linear regression structure was selected due to its simplicity and interpretability. The numeric dependent variable, $Amount_i$, gives a chance of linear regression for modeling the relationship between the contract amount and its predictors.

By including weakly informative priors for the coefficients and residual standard deviation, the model remains flexible and interpretable, while still preventing extreme parameter estimates. The use of an exponential prior for σ allows the model to adapt a scale of residuals in the data, providing estimates of the variance. The normal prior for the coefficients $\beta_1, \beta_2, \beta_3, \beta_4$ with mean 0 and standard deviation 2 shows the expectation that the coefficients should not be too larg but flexible enough to adapt the data.

This Bayesian linear regression model balances complexity with interpretability, and its flexibility allows us to draw meaningful observations from the data while quantifying uncertainty around the estimates, especially when the sample size is not large or when the predictors have a complex relationship with the outcome. [TOO GENERAL, CHANGE]

3.2 Assumptions and Limitations

The model takes several assumptions, including linear relationship between predictors, normality of the residuals, independence of observations, homoscedasticity, and reasonable chosen priors.

While the model presents some valuable observations of Microsoft's federal procurement contracts, there are some model limitations including prior function selection, data quality, potential for endogeneity, generalizability, and unobserved variables. These limitations will be discussed more detailed in Section 5.

4 Results

4.1 Data
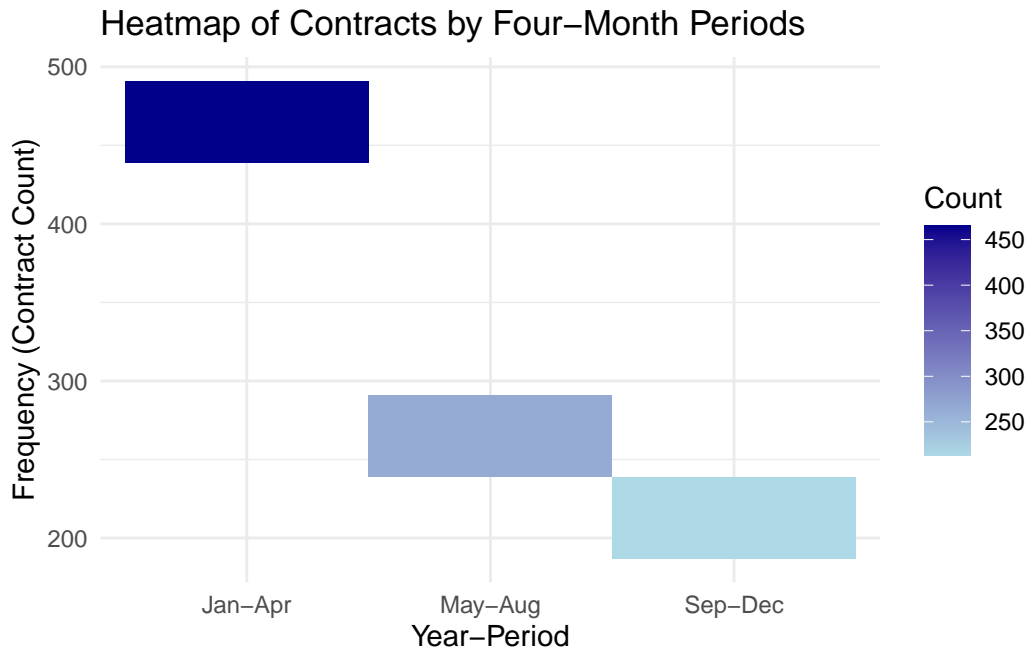

Figure 6: The heatmap ….

4.2 Model

R-squared, adjusted R-squared

tests (e.g. VIF, AIC, BIC).
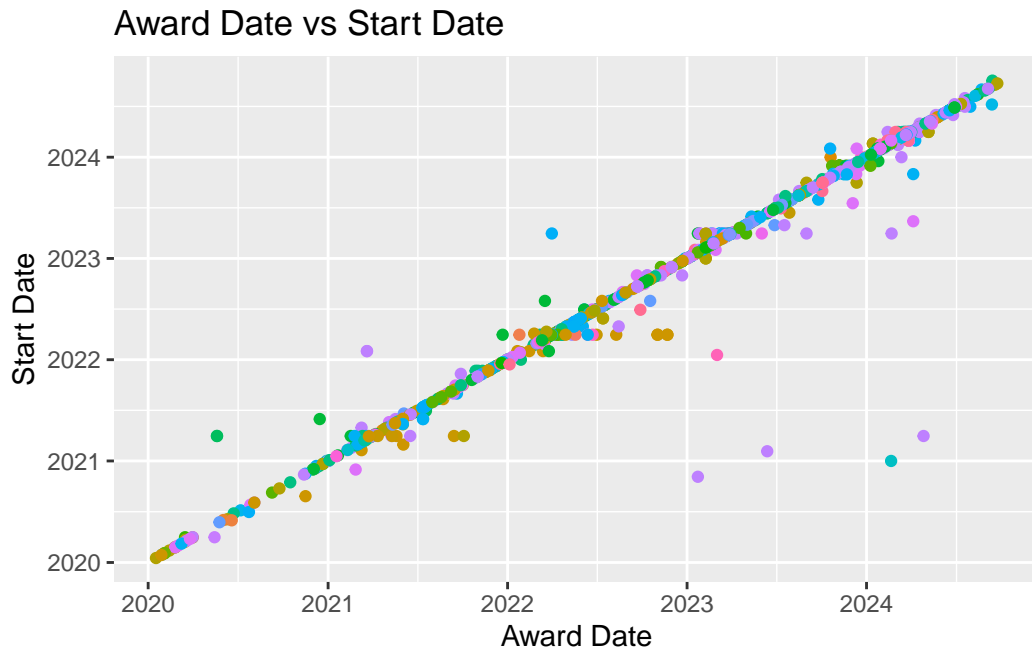
Our results are summarized in Table 3.

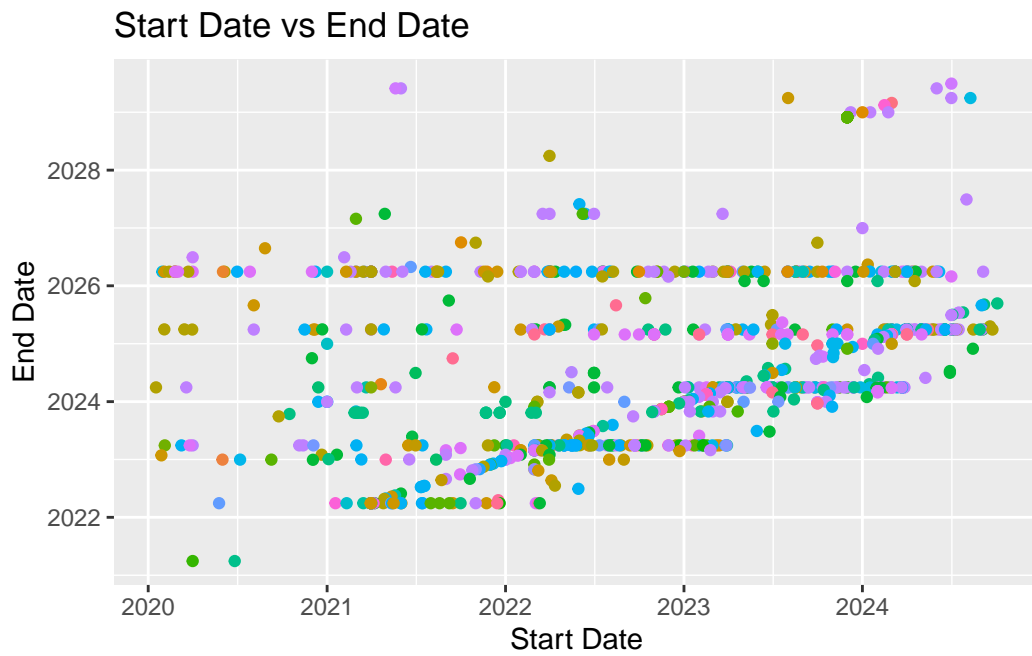Figure 7: The scatter plot shows ...

Figure 8: The scatter plot shows ....

Table 3: Explanatory models of contract amount based on ContractType, Buyer, Contract-Days, PhaseDays and Month

	Bayesian Model_Full	Bayesian Model
ContractDays	836.275	587.589
PhaseDays	−572.037	−703.976
Num.Obs.	943	943
R2	0.289	0.082
R2 Adj.	0.015	0.027
Log.Lik.	−15 351.795	−15 398.022
ELPD	−15 519.3	−15 444.8
ELPD s.e.	146.3	179.7
LOOIC	31 038.7	30 889.7
LOOIC s.e.	292.5	359.4
WAIC	30 992.3	30 894.8
RMSE	2 714 467.04	2 971 135.35

5 Discussion

5.1 Data

The completeness and reliability of the data. For the missing or erroneous records, despite the data cleaning process, introduce bias and distort the findings.

Since the data of the procurement contracts awarded to Microsoft in Canada is between 2020 and 2024 (up to September), the analysis provides a snapshot of recent procurement trends during this period. However, the data do not capture longer-term patterns or emerging shifts in procurement policies.

5.2 Limitations

One limitation of the model is that it may fail to account for certain external factors or unobserved variables such as political factors, budgetary constraints, and departmental priorities. These external factors could lead to omitted variable bias, which also influence procurement decision.

The model assumes a linear relationship between the independent variables and the dependent variable. However, procurement dynamics may be more complex, with non-linear interactions between variables. The model categorizes contract types and buyers into broad groups, which may oversimplify the true diversity of contracts awarded.

There may be issues of endogeneity, where causality runs in both directions between some predictors and the outcome variable. The amount of procurement from a buyer may influence the allocation of future budgets to that department, which in turn affects future contract amounts awarded to Microsoft. This reciprocal relationship could introduce bias in the coefficient estimates if not properly addressed.

5.3 Model Selection

5.4 Weaknesses and Next Steps

This paper provides a foundation for a deeper understanding of federal contract awards to Microsoft, but it also suggests several directions for future exploration. First, the study can expand to include comparisons with other major suppliers. By examining whether Microsoft's procurement trends are unique or reflective of broader patterns across industries. This gives a better understanding on assess whether certain dynamics are systemic within federal contracting practices.

Data quality and consistency issues are also of concern. The anomalies identified in this study, such as incomplete or inconsistent contract dates, highlight the need for improved data reporting standards. These discrepancies can be addressed in collaboration with government agencies to improve the reliability of procurement datasets and ensure that future analyses are more reliable and actionable.

Appendix

A Additional data details

B Model details

B.1 Bayesian Model

B.1.1 Posterior predictive check

In Figure 9a we implement a posterior predictive check. This shows...

In Figure 9b we compare the posterior with the prior. This shows...

(a) Posterior prediction check

(b) Comparing the posterior with the prior
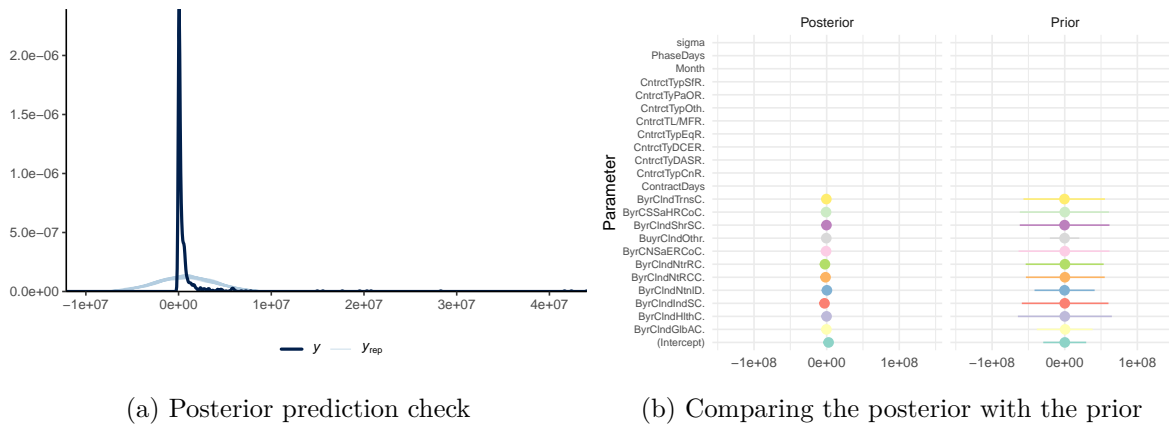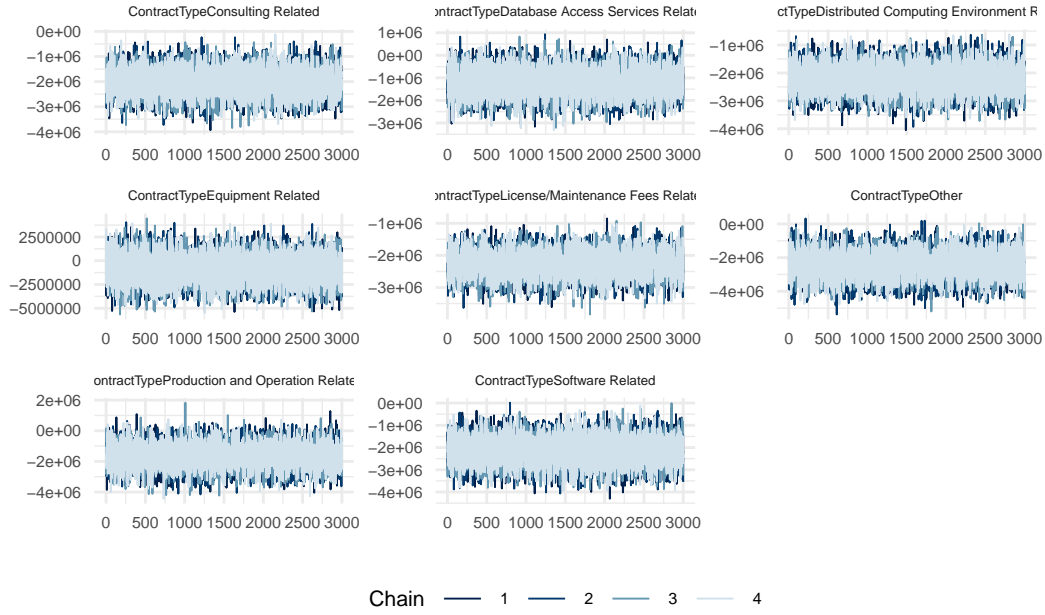
Figure 9: Examining how the model fits, and is affected by, the data
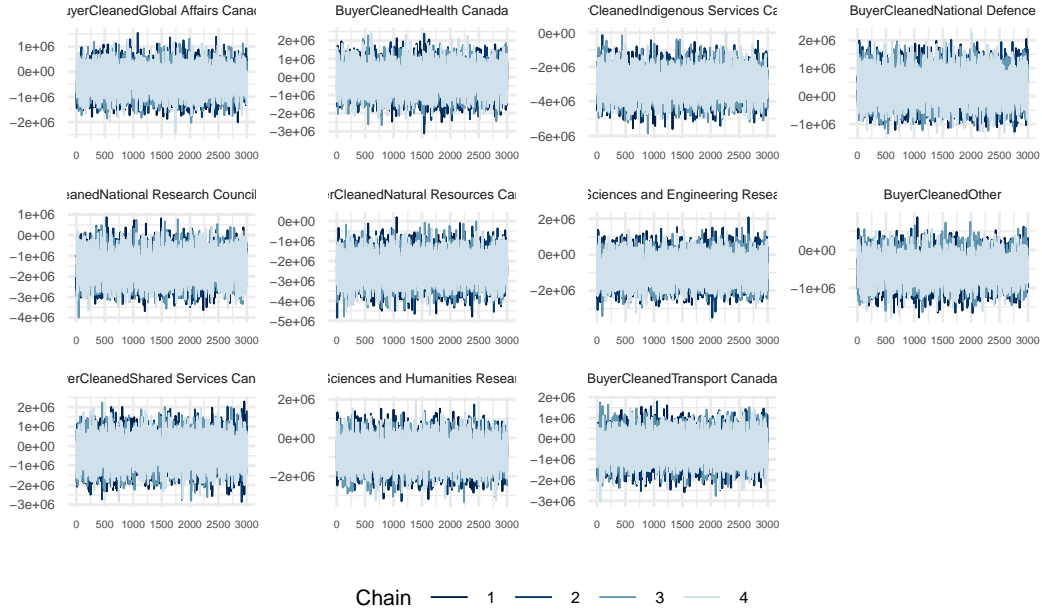
B.1.2 Diagnostics

Figure 10a is a trace plot. It shows... This suggests...

Figure 10b is a trace plot. It shows... This suggests...

Figure 11 is a Rhat plot. It shows... This suggests...

(a) Trace plot - Contract Type

(b) Trace plot - Buyers

Figure 10: Checking the convergence of the MCMC algorithm

(a) Rhat plot - Full model   (b) Rhat plot - Reduced model
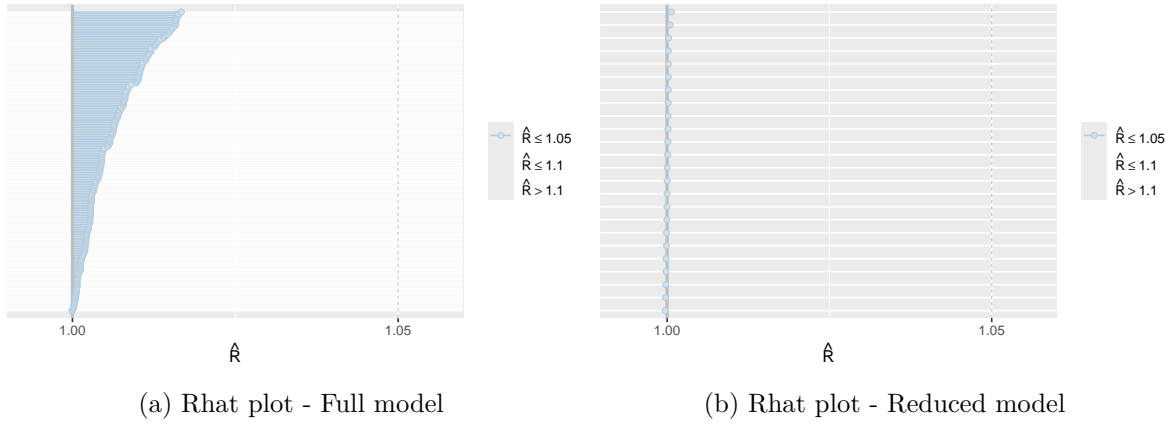
Figure 11: Rhat plot for bayesian model

Table 4: Summary of key model estimates for Microsoft, including coefficients for predictors like xxxx, with standard errors for each estimate. Model performance statistics, such as sample size, R^2, and adjusted R^2, are also displayed.

	Linear Model_Full	Linear Model
ContractDays	839.999	589.045
	(232.671)	(193.849)
PhaseDays	−596.421	−711.726
	(1707.643)	(1343.646)
Num.Obs.	943	943
R2	0.220	0.064
R2 Adj.	0.051	0.042
AIC	30 952.8	30 833.9
BIC	31 772.2	30 950.3
Log.Lik.	−15 307.375	−15 392.958
F		2.870
RMSE	2 713 355.91	2 971 129.56

Table 5: Summary of key model estimates for Microsoft, including coefficients for predictors like xxxx, with standard errors for each estimate. Model performance statistics, such as sample size, R², and adjusted R², are also displayed.

```
                Length Class  Mode
call                4  -none- call
type                1  -none- character
predicted         943  -none- numeric
mse               500  -none- numeric
rsq               500  -none- numeric
oob.times         943  -none- numeric
importance          5  -none- numeric
importanceSD        0  -none- NULL
localImportance     0  -none- NULL
proximity           0  -none- NULL
ntree               1  -none- numeric
mtry                1  -none- numeric
forest             11  -none- list
coefs               0  -none- NULL
y                 943  -none- numeric
test                0  -none- NULL
inbag               0  -none- NULL
terms               3  terms  call
```

B.2 Linear Regression Model

B.3 Random Forest Model

C Methodology

The methodology employed in this study involves analyzing federal procurement contracts awarded to Microsoft, leveraging data from two primary sources: Public Services and Procurement Canada (PSPC) and the Treasury Board of Canada Secretariat (TBS). These two datasets provide information on government procurement, but they come with distinct limitations and reporting structures (Investigative Journalism Foundation 2024c).

C.1 Data Sources and Structure

Federal procurement data is categorized into two main streams:

- PSPC serves as the central procurement body for most federal government departments and is responsible for managing a significant portion of federal contracts. It maintains a public platform called CanadaBuys, where procurement information is published. However, due to migration from the older BuyAndSell platform and the incomplete nature of pre-2022 data, some variables, such as contract types and trade agreement information, are missing or inconsistently reported. The PSPC data is particularly useful for understanding contracts where the PSPC is the managing body, but only contracts above a certain dollar threshold are publicly available.

- TBS consolidates procurement data from all government departments and requires proactive publication under the Access to Information Act. The TBS dataset is more comprehensive in terms of the number of contracts, but it suffers from inconsistencies and gaps in the reported fields. TBS publishes data quarterly, and records before 2017 should not be considered complete. Unlike PSPC, TBS data includes the full details of contract awards but does not incorporate tender or solicitation data.

C.2 Data Integration and Challenges:

There is an overlap between TBS and PSPC data, such that many contracts reported by PSPC are also included in the TBS dataset. However, the two datasets do not always align perfectly. Contracts reported by TBS may contain different details from the PSPC records.

This mismatch is exacerbated by the lack of a consistent unique identifier that would allow for easy linking between the two datasets. While about 20% of PSPC awards have matching TBS records, many contracts do not, making it impossible to perfectly merge the two datasets.

To address this, both datasets were kept separate in (Investigative Journalism Foundation 2024c). PSPC data was used primarily for its comprehensive tender information, while TBS data was relied upon for its broader coverage of contract awards, spanning more years and a greater volume of contracts.

C.3 Solicitation Methods and Contract Types:

Federal procurement processes are governed by various solicitation methods, which determine how contracts are awarded. These methods are categorized into competitive and non-competitive processes.

Four competitive options:

- Open bidding: Any supplier can bid in response to an online solicitation.
- Selective tendering: Only some prequalified suppliers can bid.
- Limited tendering: Deviation from requirements of the relevant trade agreements (see Trade Agreements), still allowing bids.
- Traditional: Many suppliers can bid but not in response to a public, online solicitation; for example, responding to an email sent to suppliers by the government.

Two non-competitive options:

- Advanced Contract Award Notice: Notice to suppliers that the buyer intends to award a contract to a pre-identified supplier, while accepting a Statement of Capabilities, not a bid, from challenging suppliers.
- Non-competitive: Blanket category for all other non-competitive processes.

References

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Investigative Journalism Foundation. 2024a. "Procurement – Microsoft." Toronto, Canada. https://theijf.org/procurement/supplier/microsoft?region=Federal.

———. 2024b. "Procurement Handbook Introduction." Toronto, Canada. https://theijf.org/procurement-introduction.

———. 2024c. "Procurement Sources." Toronto, Canada. https://theijf.org/procurement-sources#solicitation-method.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'.* https://CRAN.R-project.org/package=arrow.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Lionel Henry. 2023. *Purrr: Functional Programming Tools.* https://CRAN.R-project.org/package=purrr.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.