Data Report

Luis M. B. Varona[1,2] Otoha Hanatani[3]

March 31, 2025

Introduction

In collaboration with the Union of Municipalities of New Brunswick and Dr. Craig Brett of Mount Allison University, we conduct a fixed-effects two-stage least squares (or FE-2SLS) regression analysis of average tax rates on police spending in New Brunswick municipalities, using median household income as an instrument variable to reduce simultaneity bias. We herein investigate whether police spending is a significant predictor of municipal tax rates and, if so, whether the specific policing provider plays into this correlation. Moreover, we leverage the fact that police expenditure (as per Provincial Police Service Agreement with the Royal Canadian Mounted Police) is largely an exogenous variable outside of municipal control to use this to approximate tax base elasticity with respect to tax rates. (In addition, we consider the relationship between population and this estimated elasticity.)

Background of the Problem

Price per unit of public goods—particularly police spending, in the context of this study—varies widely across municipalities in New Brunswick. We herein aim to regress regression municipal tax rates on the costs of several different public goods. We place particular emphasis on the significant variation in per capita cost of municipal bills under the Provincial Police Service Agreement (PPSA)—a contract between the Government of New Brunswick (GNB) and the Royal Canadian Mounted Police (RCMP) to provide smaller municipalities with policing services. As the RCMP provides the province with a single combined bill, the GNB charges different municipalities based on population, safety levels, and other factors, with this formula acting as an exogenous factor in the cost of policing services.

On the other hand, it is common for larger municipalities have their own direct contracts with the RCMP, further obscuring the relationship between municipal spending patterns and taxation. For instance, the Codiac Regional Policing Authority serves the municipalities of Dieppe, Moncton, and Riverview, none of which pay additional fees to the GNB under the PPSA. Others still maintain their own independent police forces like the Bathurst Police Force (although there remains an RCMP presence in Bathurst). (As shall be revealed in our section on Methodology, Amy Anderson of the Union of Municipalities of New Brunswick has provided us with data on municipal policing providers as of 2024 and a way to map this backwards to jurisdictions from previous years.)

[TODO: Explain why this setup is of interest not only insofar as showing how, and why, different level of exogeneity affect tax rates in different ways, but also in terms of … other stuff?]

Overview of Our Approach

[TODO: Introduce, at a high level, how our general FE-2SLS model works, how we are estimating stuff like elasticity, our general desired results, etc.]

[1]Department of Mathematics & Computer Science, Mount Allison University, Sackville, NB E4L 1E6
[2]Department of Politics & International Relations, Mount Allison University, Sackville, NB E4L 1A7
[3]Department of Economics, Mount Allison University, Sackville, NB E4L 1A7

Literature Review

[TODO: Elaborate on this now that we have a first draft, especially with regards to what methodology we took away]

We review two relevant articles on municipal taxation and spending (Brett and Pinkse 2003; Gadienne 2017). Brett and Pinkse (2003) investigate the "determinants of municipal tax rates in British Columbia," considering population, distance from major metropolitan centers (namely Vancouver), income, and several other factors as determinants of tax rates. They do not particularly emphasize spending patterns as a potential factor in municipal taxation, but the methodology presented in their paper will provide a useful framework on which to build for our project.

Gadienne (2017) provides a study in how extra tax revenue truly affects the quantity and quality of public services. Again, we are going the "other way" in that we are examining how public spnding costs affect taxation, but several of the case studies here are of interest. The relevant study also investigates municipalities specifically, so the methodology laid out (especially in Section B: Local Public Revenues) proves useful.

[TODO: Add lit. review for stuff like our elasticity estimates, median income as an instrument for tax base/capita, etc. specifically]

[TODO: Moreover, and this is important—add literature review on whether or not police expenditure is typically this exogenous (it should not be), and discuss further how the PPSA affects it. Maybe for the previous section too?]

(TODO: Add Brett and Tardif 2008; Dahlby 2024)

Methodology

In this section, we delineate our data collection process, data organization methods, and econometric models and analysis. We use Python (primarily the polars and linearmodels/statsmodels ecosystems) to parse and clean data from the Government of New Brunswick and Statistics Canada. Subsequently, we run several fixed-effects and correlated random-effects regression models on the resulting data in combination with an instrumental variable to account for simultaneity bias.

Data Collection and Sources

We use an unbalanced panel of annual data from 2000–2018 on New Brunswick municipalities, received via personal correspondence with the GNB and Dr. Craig Brett of Mount Allison University; however, this data is also publicly available at ("2000–2018 Annual Reports of Municipal Statistics for New Brunswick" 2000–2018), albeit in a less structured format. (The year 2005 is excluded due to missing/improperly formatted tokens, but we may coordinate further with the GNB to obtain this data in the future.) Each set of annual data contains 95 to 103 municipalities, with a total of 104 unique municipalities across all years.

This is supplemented by 2024 data on municipal policing provider agreements (Anderson 2025). We map this data backwards to municipal jurisdictions and boundaries from previous years and integrate indicators into interaction terms in our panel as described below.

Finally, the instrument variable in the first stage of our 2SLS regression is median household income, given in census data from Statistics Canada. Data is only available from 2000 ("Table 95F0437XCB2001006" 2001), 2005 ("Table 97-563-XCB2006052" 2006), 2015 ("Table 98-400-X2016099" 2016), and 2020 ("Table 98-10-0061-01" 2021); hence, linear interpolation is applied for the intervening years. The resulting income data (typically correlated with tax base per capita but not with tax rate) is then used to reduce simultaneity bias in our fixed-effects model.

Data Cleaning and Organization

Primary Data

Primary data is cleaned in the `data_pipeline/` directory. The original Excel files extracted from `.zip` archives provided by the GNB and the UMNB are contained in the `data_raw/` subdirectory. These contain annual data from 2000–2022 on New Brunswick municipalities, as well as 2024 data on municipal policing providers. Given that some of these files are `.xls` and `.xlw` workbooks, we copy and convert them all to `.xlsx` format in the `data_xlsx/` subdirectory. The `helper_scripts/_raw_to_xlsx_.py` script is used for this purpose.

Files in this `data_xlsx/` subdirectory are cleaned and organized by `helper_scripts/_xlsx_to_clean_.py`. Finding that data from 2005 and 2019–2022 is unusable due to missing/improperly formatted tokens, our output (placed in the `data_clean/` subdirectory) excludes these time periods. No original data is discarded during this process (save for metadata and notes)—it is all simply reorganized into parseable form.

Addressing inconsistent municipality naming conventions across years/categories and concatenating all annual panels within each category (budget expenditures, budget revenues, comparative demographics, and tax bases), the `helper_scripts/_clean_to_final_.py` script then writes all four resulting worksheets—plus a fifth for provider data—to a single `data_fina;/data_master.xlsx` workbook. (The new municipal naming convention is also used to map provider data on newer, reformed 2024 municipalities and districts to past jurisdictions all the way back to 2000.)

All scripts are called and run by the main executable of the associated directory, `main.py`.

Instrumental Variable Data

Data on the instrumental income data is stored and processed in the `data_iv/` directory. There is one folder each for 2001, 2006, 2016, and 2021 (the years in which the census data were released) containing the original files downloaded from the Statistics Canada website. For 2016 and 2021, the downloads are straightforward, nicely formatted `.csv` files requiring no further processing. For 2001 and 2006, however, full data is only available in `.ivt` and `.xml` format; no schemas are available to parse the XML data, so we use the Government of Canada's Beyond 20/20 Browser to extract and download the data in `.csv` format. (Unfortunately, this process is not easily documentable, as the browser requires manual processing.)

With CSV files for all four years, the `main.py` executable script is finally used to clean and combine the relevant columns and rows into a single polars DataFrame. This is then saved as an `.xlsx` file in the `results/` subdirectory for immediate usage in the data analysis stage. (The aforementioned data interpolation—performed using Python's numpy library—is not applied until this stage and is thus not considered part of the data cleaning and organization pipeline.)

It is worth noting that although household income data from Canada censuses is publicly accessible for municipal-level geographic localities in 2000, 2005, 2015, and 2020, the only available source for 2010 is aggregated data from the 2011 National Household Survey. This survey refrained from providing disaggregated data at lower levels of geography, so we are unable to map it to most of the 104 municipalities in our dataset. Hence, linear interpolation is used to estimate the missing data for 2010, just as for all the other missing years. In the future, we may collaborate further with Statistics Canada to obtain the geographically disaggregated data, if it remains in their records.

Data Analysis and Modelling

All data analysis is performed in the `data_analysis/` directory. Our included variables are:

- **Average Tax Rate**, or **AvgTaxRate** – unitless
- **Police Spending per Capita**, or **PolExpCapita** – 10^5 CAD / person
- **Non-Police Spending per Capita**, or **OtherExpCapita** – 10^5 CAD / person
- **Non-Warrant Revenue per Capita**, or **OtherRevCapita** – 10^5 CAD / person
- **Tax Base for Rate per Capita**, or **TaxBaseCapita** – 10^5 CAD / person

- **Policing Provider** – boolean, three categories:
  - *Provincial Police Service Agreement* (excluded control variable)
  - *Municipal Police Service Agreement*, or *Provider_MPSA* (included)
  - *Municipal Police*, or *Provider_MPSA* (included)
- **Median Household Income**, or **MedHouseInc** – 10^5 CAD / person

Our dependent variable is *AvgTaxRate*, which is calculated as a weighted average of the residential and non-residential tax rates in a municipal jurisdiction. (That is—as per government formulae, non-residential rates are multiplied by a factor of 1.5 before being integrated into the calculated average. Said averages are already available in the raw data ("2000–2018 Annual Reports of Municipal Statistics for New Brunswick" 2000–2018), not calculated by us; we take note of the process simply to clarify the layout of our data.) Our exogenous explanatory variables are *PolExpCapita*, *OtherExpCapita*, *OtherRevCapita*, *PolExpCapita*∗*Provider_MPSA*, and *PolExpCapita*∗*Provider_Muni*. Our sole endogenous explanatory variable is *TaxBaseCapita*, for which we control simultaneity bias using *MedHouseInc* as an instrumental variable.

Each of these variables is used throughout our FE-2SLS regression model, carried out by the `helper_scripts/_3_fe_2sls_analysis_.py` script. In addition, we have also included vanilla correlated random-effects (CRE) and fixed-effects (FE) models, run by `helper_scripts/_1_cre_analysis_.py` and `helper_scripts/_2_fe_analysis_.py`, to determine which variables are relevant and to demonstrate the need for an instrument variable. All helper scripts are called and run by the main executable of the associated directory, `main.py`.

Our final choice of FE in conjunction with 2SLS arose from [TODO: Elaborate, particularly on why *TaxBaseCapita* causes simultaneity bias]

It is worth noting that we chose not to use non-linear functional forms—with the most obvious candidate for a study in this particular real-world context being log transformation—as we found that the [TODO: Elaborate on relative lack of skewedness in the data]

Finally, we also estimate tax base elasticity by [TODO: Elaborate]

We now turn to describing our vanilla CRE and FE analyses, then proceed to more thoroughly delineate our final FE-2SLS regression model.

Correlated Random-Effects (CRE)

[TODO: Elaborate]

Fixed-Effects (FE)

After deeming the potential benefits of including the policing provider indicators directly (not in interaction terms) insufficient to warrant [TODO: Elaborate]

Fixed-Effects Two-Stage Least Squares (FE-2SLS)

Finally, we decided on [TODO: Elaborate]

Stage 1 We begin by estimating *MedHouseInc* data for the years missing from the Statistics Canada census data, which we do using simple linear interpolation. (As this project continues to develop, we may investigate more sophisticated approximation approaches, but this shall do for now.) After this is done, we perform an ordinary least squares regression of *TaxBaseCapita* on *MedHouseInc* to obtain

$$TaxBaseCapita_{it} = \alpha_0 + \alpha_1 MedHouseInc_{it} + v_{it}.$$

By performing this regression before proceeding to a fixed-effects model, we manage to reduce simultaneity bias, as *MedHouseInc* is correlated with *TaxBaseCapita* but not with *AvgTaxRate*. We use these predicted $\widehat{TaxBaseCapita}_{it} = TaxBaseCapita_{it} - v_{it}$ values in the second-stage regression, where we demean all variables over municipality.

Stage 2 Our primary fixed-effects regression model is now given by

$$\begin{aligned}\ddot{AvgTaxRate}_{it} = {} & \beta_1 \ddot{PolExpCapita}_{it} + \beta_2 \ddot{OtherExpCapita} + \beta_3 \ddot{OtherRevCapita} + \\ & \beta_4 \ddot{\hat{TaxBaseCapita}}_{it} + \beta_5 \ddot{PolExpCapita}_{it} * Provider\_MPSA_{it} + \\ & \beta_6 \ddot{PolExpCapita}_{it} * Provider\_Muni_{it} + \ddot{u}_{it},\end{aligned}$$

where we use the notation $\ddot{X}_{it} = X_{it} - \bar{X}_i$ to denote the difference between the value of X for municipality i in year t and the mean value of X for municipality i over all years. (Note that $\ddot{\hat{TaxBaseCapita}}_{it}$ is not the demeaning of $TaxBaseCapita_{it}$ itself but rather the demeaned prediction from our first-stage regression.)

Our covariance estimator in this model is clustered by municipality, as [TODO: Elaborate]

Tax Base Elasticity Estimates

Given these results, we now approximate tax base elasticity with respect to tax rates by multiplying our obtained coefficient on **PolExpCapita**—one of the most exogenous expenditure categories, as previously discussed—by *TaxBaseCapita*. [TODO: Elaborate]

[TODO: Show the calculus for this, i.e., the $\frac{1}{1+\eta}$ formula]

Results

[TODO: Elaborate]

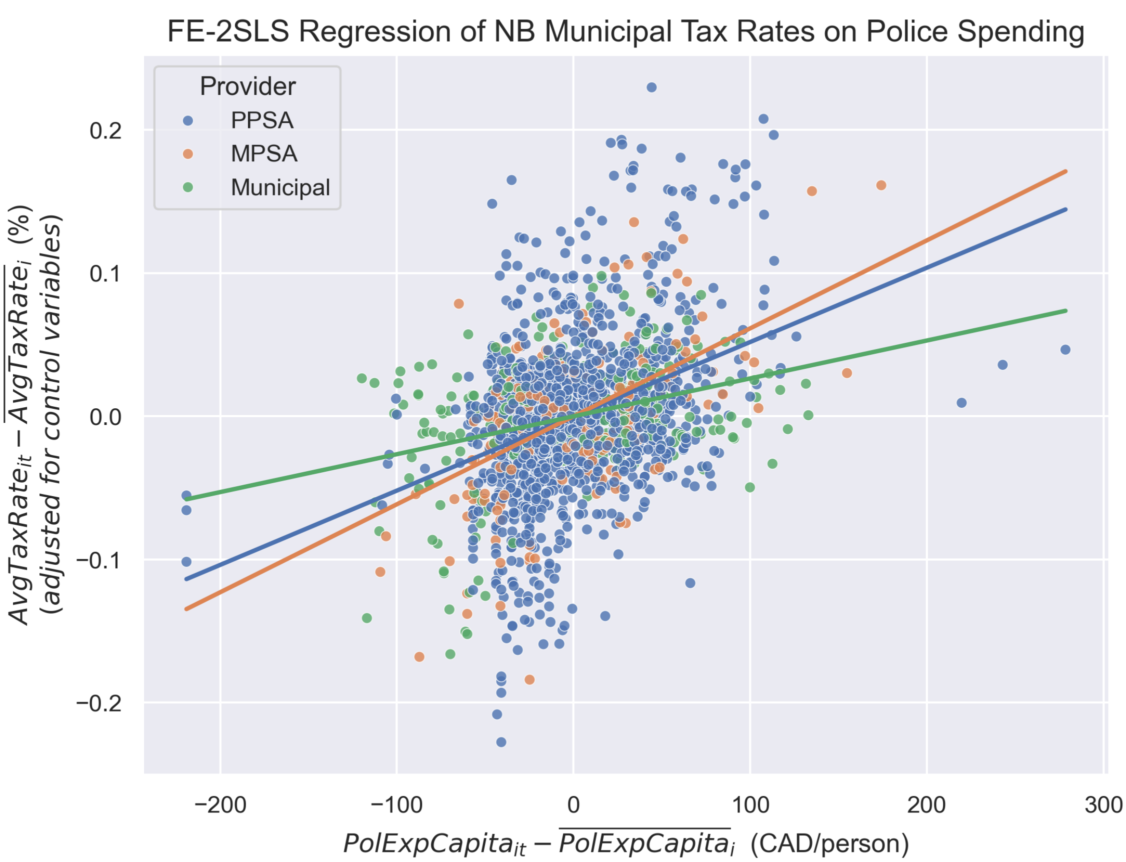

Figure 1: TODO

[TODO: Add explanation of the above figure]

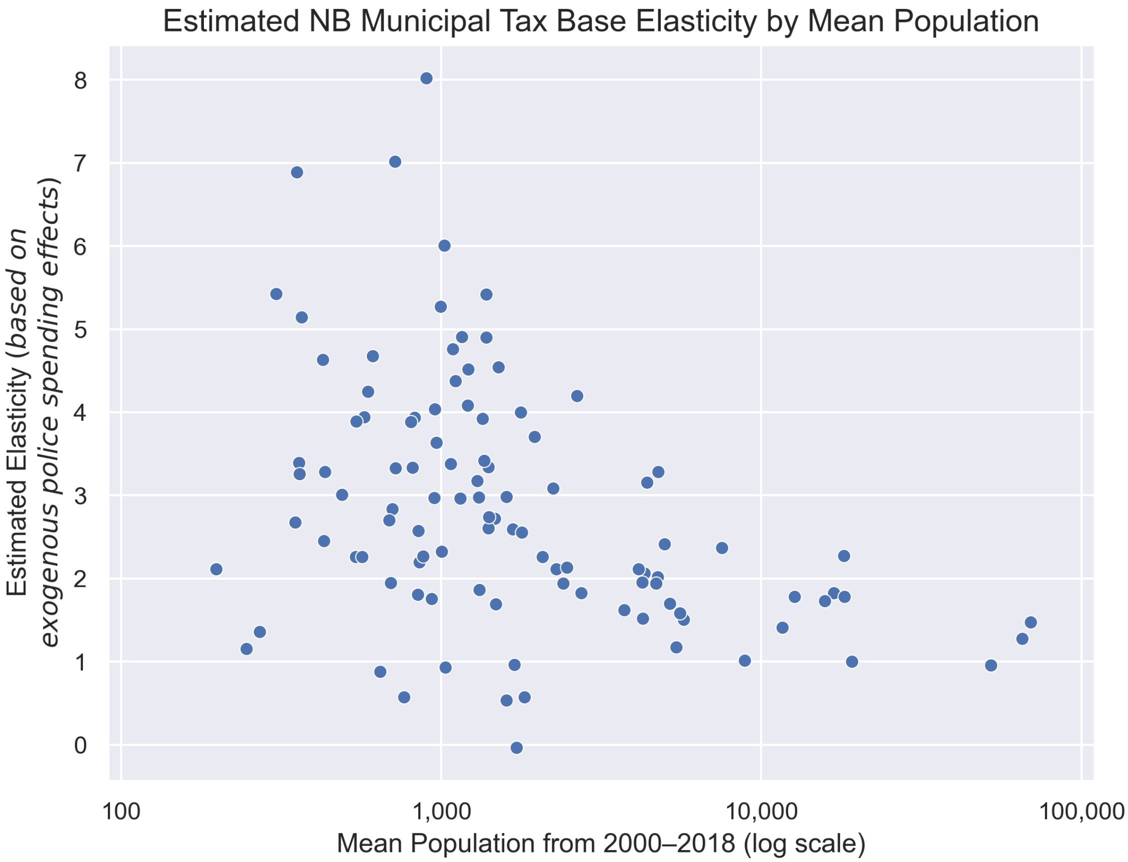

Figure 2: TODO

[TODO: Add explanation of the above figure]

Discussion

[TODO: Elaborate]

Conclusion

[TODO: Elaborate]

Appendix

The raw fixed-effects regression model used in our analysis is given below:

Dep. Variable:	AvgTaxRate	**R-squared:**	0.7216
Estimator:	PanelOLS	**R-squared (Between):**	0.1325
No. Observations:	1818	**R-squared (Within):**	0.7216
Date:	Fri, Apr 04 2025	**R-squared (Overall):**	0.1356
Time:	18:11:50	**Log-likelihood**	1.196e+04
Cov. Estimator:	Clustered		
		F-statistic:	738.02
Entities:	104	**P-value**	0.0000
Avg Obs:	17.481	**Distribution:**	F(6,1708)
Min Obs:	6.0000		
Max Obs:	18.000	**F-statistic (robust):**	66.473
		P-value	0.0000
Time periods:	18	**Distribution:**	F(6,1708)
Avg Obs:	101.00		
Min Obs:	95.000		
Max Obs:	103.00		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
PolExpCapita	1.3092	0.0990	13.222	0.0000	1.1150	1.5034
OtherExpCapita	0.9665	0.0914	10.575	0.0000	0.7872	1.1458
OtherRevCapita	-0.8964	0.0991	-9.0486	0.0000	-1.0907	-0.7021
TaxBaseCapita	-0.0122	0.0012	-10.440	0.0000	-0.0145	-0.0099
PolExpCapita:Provider__MPSA	-0.5551	0.1951	-2.8459	0.0045	-0.9377	-0.1725
PolExpCapita:Provider__Muni	-0.6083	0.1377	-4.4162	0.0000	-0.8784	-0.3381

```
F-test for Poolability: 51.098
P-value: 0.0000
Distribution: F(103,1708)

Included effects: Entity
```

[TODO: Elaborate]

[TODO: Explain first stage regression]

Dep. Variable:	TaxBaseCapita	**R-squared:**	0.011
Model:	OLS	**Adj. R-squared:**	0.011
Method:	Least Squares	**F-statistic:**	20.99
Date:	Fri, 04 Apr 2025	**Prob (F-statistic):**	4.93e-06
Time:	18:11:51	**Log-Likelihood:**	-363.29
No. Observations:	1818	**AIC:**	730.6
Df Residuals:	1816	**BIC:**	741.6
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.6668	0.020	33.737	0.000	0.628	0.706
MedHouseInc	-11.2497	2.455	-4.582	0.000	-16.066	-6.434

Omnibus:	920.376	**Durbin-Watson:**	1.434
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	8236.943
Skew:	2.196	**Prob(JB):**	0.00
Kurtosis:	12.458	**Cond. No.**	354.

```
Notes:
 [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

The final FE-2SLS regression model used in our analysis is given below:

Dep. Variable:	AvgTaxRate	**R-squared:**	0.4290
Estimator:	PanelOLS	**R-squared (Between):**	0.9794
No. Observations:	1818	**R-squared (Within):**	0.4290
Date:	Fri, Apr 04 2025	**R-squared (Overall):**	0.9783
Time:	18:11:51	**Log-likelihood**	1.131e+04
Cov. Estimator:	Clustered		
		F-statistic:	213.84
Entities:	104	**P-value**	0.0000
Avg Obs:	17.481	**Distribution:**	F(6,1708)
Min Obs:	6.0000		
Max Obs:	18.000	**F-statistic (robust):**	27.629
		P-value	0.0000
Time periods:	18	**Distribution:**	F(6,1708)
Avg Obs:	101.00		
Min Obs:	95.000		
Max Obs:	103.00		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
PolExpCapita	0.5188	0.1371	3.7845	0.0002	0.2499	0.7877
OtherExpCapita	0.0301	0.0255	1.1801	0.2381	-0.0199	0.0802
OtherRevCapita	0.1025	0.0644	1.5912	0.1117	-0.0238	0.2288
TaxBaseCapita	0.0225	0.0068	3.2987	0.0010	0.0091	0.0359
PolExpCapita:Provider__MPSA	0.0955	0.1941	0.4923	0.6226	-0.2851	0.4762
PolExpCapita:Provider__Muni	-0.2542	0.1821	-1.3962	0.1628	-0.6113	0.1029

```
F-test for Poolability: 115.50
P-value: 0.0000
Distribution: F(103,1708)

Included effects: Entity
```

[TODO: Explain results further; potentially show in the **Results** section instead, and show the other preliminary models here?]

References

"2000–2018 Annual Reports of Municipal Statistics for New Brunswick." 2000–2018. Fredericton, NB: Government of New Brunswick.

Anderson, Amy. 2025. "Personal Correspondence with Amy Anderson."

Brett, Craig, and Joris Pinkse. 2003. "The Determinants of Municipal Tax Rates in British Columbia." *Canadian Journal of Economics* 33 (3): 695–714. https://doi.org/10.1111/0008-4085.00037.

Brett, Craig, and Christina Tardif. 2008. "How Municipal Governments Changed Taxes in Response to Provincial Support in New Brunswick, 1983-2003." *Canadian Public Policy* 34 (4): 441–56. https://doi.org/10.3138/cpp.34.4.441.

Dahlby, Bev. 2024. "The High Cost of Raising Provincial Tax Revenues Has Gotten Even Higher." 17-13. Calgary, AB.

Gadienne, Lucie. 2017. "Sources of Public Finance and Government Accountability." *American Economic Journal: Applied Economics* 9 (1): 274–314. https://doi.org/10.1257/app.2015050.

"Table 95F0437XCB2001006." 2001. Statistics Canada; https://www12.statcan.gc.ca/english/census01/products/standard/themes/Download.cfm?PID=55710.

"Table 97-563-XCB2006052." 2006. Statistics Canada; https://www12.statcan.gc.ca/census-recensement/2006/dp-pd/tbt/Download.cfm?PID=94594.

"Table 98-10-0061-01." 2021. Statistics Canada; https://doi.org/10.25318/9810006101-eng.

"Table 98-400-X2016099." 2016. Statistics Canada; https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dt-td/CompDataDownload.cfm?LANG=E&PID=110192&OFT=CSV.