1 Empirical Strategy

1.1 Data

I employ a novel administrative dataset from the Canadian Intellectual Property Office, the IP Horizons Patent Researcher Datasets. The dataset identifies patents in Canada from 1860 to 2023, with information on when the application for the patent was filed, granted, the parties involved in the application and other information. Parties can be identified to provinces based on their location, which can be based in Canada or other countries.

With these data, I compute quarterly and monthly patent application counts at the province level from January 2001 to June 2021, based on the application filing date. This period corresponds to the modern Canadian intellectual property institutional background, as reviewed on Section **??**. I assign patents to provinces based on where the majority of parties involved in a patent application report their location[1]. I only include 2021 as a partial year since the data present a downward trend for all provinces in late 2021, suggesting patent applications are yet to be updated for the most recent periods of the IP Horizons data. Further, I drop Newfoundland and Labrador, Prince Edward Island, Yukon and Nunavut due to missing observations on most explanatory variables.

The explained variable of interest is the count of patent applications. To allow for heterogeneity in treatment, I separate patents by their International Patent Classification (IPC) section, which defines a broad classification of the technology being patented. The IPC sections are divided into eight categories: A (Human Necessities), B (Performing Operations; Transporting), C (Chemistry; Metallurgy), D (Textiles; Paper), E (Fixed Constructions), F (Mechanical Engineering; Lighting; Heating; Weapons; Blasting), G (Physics) and H (Electricity), as defined by the Canadian Intellectual Property Office (2023). For robustness checks, I also consider the number of Canadian parties involved in a patent application as explained variables for robustness checks, separating by the different types of parties, as reviewed in Section **??**: all parties, inventors, owners and applicants[2].

1. Patent applications without information of party provinces or with an equal number of interested parties from two provinces are dropped from the sample.

2. I do not consider agents as a separate category due to them typically being hired legal professionals, which may not be informative about the innovative capacity of who files for the patent.

For my explanatory variables, I extract province-level data at the monthly frequency from Statistics Canada. These include data from the Labour Force Survey (LFS), such as labour force characteristics, employment wages, among others. Further, I also consider the consumer price index, international merchandise exports and imports, retail, wholesale and manufacturing trade sales, food services receipts, the new housing price index and electric power generation. I also include the number of business insolvencies as reported by Innovation, Science and Economic Development Canada and the number of foreign parties involved in patent applications, which I obtain from the IP Horizons data. I aggregate data at the quarterly level by summing all variables except the consumer and new housing indices, for which I take arithmetic averages. Table 1 presents a summary of the main variables used in the analysis for the province-quarter panel. Table **??** in Appendix **??** presents the same for the province-month panel.

Table 1: Descriptive statistics for the province-quarter sample

	Mean	SD	Min	Median	Max
Ln +1 Patent applications	4.261	1.405	1.099	4.107	6.691
Ln Full-time employment	8.026	1.034	6.726	7.831	9.814
Ln Median wage	2.949	0.192	2.523	2.956	3.395
CPI	119.145	12.668	95.400	119.400	148.900
Ln +1 Business insolvencies	4.403	1.396	0.693	4.197	6.957
Ln Intl. exports	15.810	1.139	13.694	15.848	17.804
Ln Intl. imports	15.646	1.198	13.715	15.369	18.372
Ln Retail sales	15.963	1.028	14.424	15.774	17.913
Ln Wholesale sales	15.910	1.292	13.907	15.892	18.490
Ln Manufacturing sales	16.027	1.179	14.398	15.729	18.213
Ln International travellers	12.470	1.779	4.344	12.387	15.929
Ln Arriving vehicles	11.944	3.562	0.000	12.516	15.801
Ln Electric power generation	16.213	0.997	14.344	16.219	17.990
Ln Average actual hours	3.545	0.050	3.311	3.550	3.676
New housing price index	88.064	16.987	42.900	94.250	129.500
Ln Food services receipts	13.737	1.108	12.255	13.575	15.857
Ln Average job tenure	4.636	0.088	4.399	4.653	4.830
Ln +1 Foreign patent parties	3.609	1.918	0.000	3.842	6.671

Notes: All statistics based on a balanced panel of $N = 656$ province-quarter observations from 2001Q1 to 2021Q2. The sample includes all Canadian provinces except Newfoundland and Labrador, Prince Edward Island, Yukon and Nunavut.

1.2 Empirical Strategy

I implement a two-way fixed effects (TWFE) difference-in-differences (DD) design, where I define treatment and control groups based on the first period of eligible expenditures for the AITC intervention, which was April 2016 (Alberta Economic Development and Trade 2017). The treatment group is Alberta, and the treatment period is composed of all periods after April 2016. The control group is all remaining Canadian provinces considered in my data. Thus, treated observations are those from Alberta after April 2016, where I believe the AITC affected Albertan patent applications. The DD design is implemented in a regression framework with both the quarter and month panels in order to better understand the dynamic effects of the intervention.

The general specification for the DD model is:

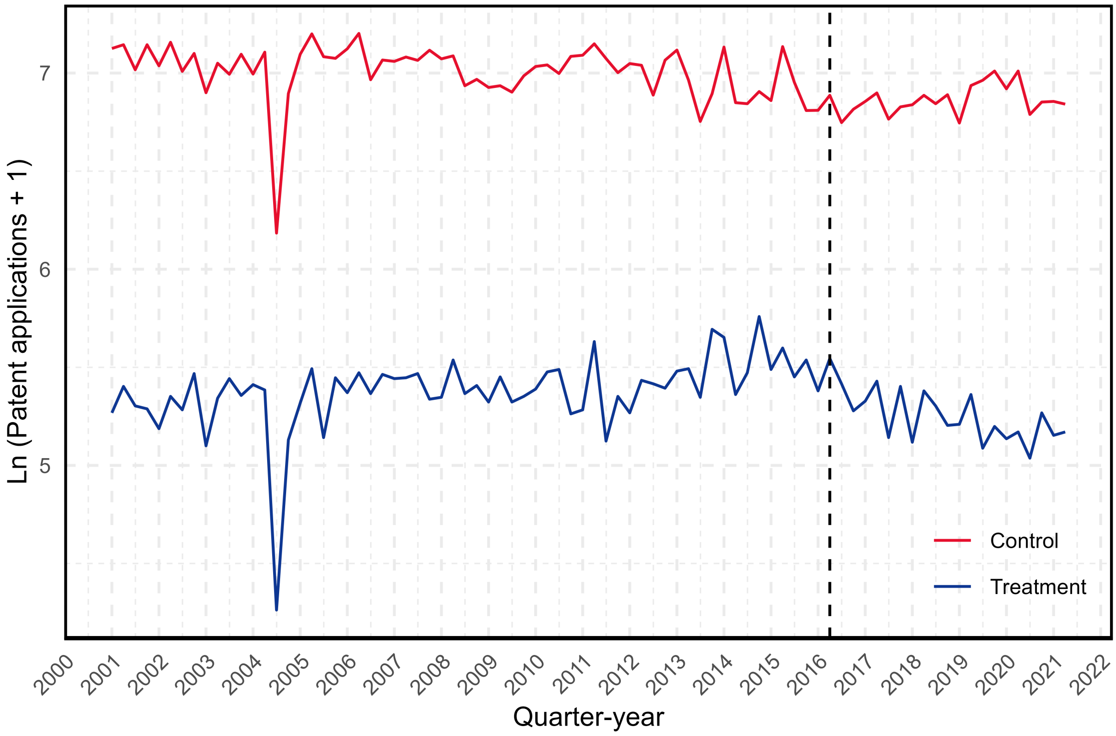
$$\ln(P_{it} + 1) = \theta_i + \theta_t + \beta + T_{it} + \mathbf{x}_{it}^{'}\gamma + u_{it} \tag{1}$$

where P_{it} is the explained variable; in most specifications, P_{it} is the number of patents filed in a province i and period t. θ_i and τ_t are sets of province and period fixed effects. I use a natural logarithm transformation along with the addition of one to correct for provinces with small amounts of patent applications on some periods. T_{it} is a binary variable equal to unity for observations for treated observations and zero otherwise. Hence, the estimated parameter $\hat{\beta}$ is the coefficient of interest, which is my estimate for the average treatment effect of the AITC on the explained variable. $\mathbf{x}_{it}$ is a vector of time and province-varying controls, as described in the previous subsection, and γ is the associated vector of parameters. u_{it} is a stochastic error term which varies between provinces and periods. For my results, I cluster standard errors at the province and period level.

Tables **??** and **??** in Appendix **??** present the difference in means between treated and control provinces for the province-quarter and province-month samples for all considered explained variables. This presents the simplest version of the DD model, where I compare the average number of patent applications between Alberta and the control provinces before and after the AITC intervention. This simple comparison suggests a small or null effect; the regression analysis described above provides a more robust DD estimate.

The key identifying assumption of the DD framework is that, absent of treatment, the trend of the explained variable in Alberta would follow a similar pattern to control provinces. Figure 1 shows the quarterly time series of patent applications between Alberta and control provinces from 2001Q1 to 2021Q2. This visual representation of the trends shows that Alberta's patent applications follow a similar pattern to control provinces before the AITC intervention, however, some deviations are present in the leading months before the intervention.

Figure 1: Quarterly time series of patent applications between Alberta and control provinces



Notes: The figure shows the quarterly time series of patent applications between Alberta and control provinces from 2001Q1 to 2021Q2. The vertical line represents the start of the AITC intervention (first expense eligibility date) in April 2016.

To allay the concern of unobservable factors impacting patent application trends across provinces, I estimate event study regressions following Equation 2 below and provide supporting evidence for causal identification of $\hat{\beta}$.

$$\ln(P_{it} + 1) = \theta_i + \tau_t + \beta_{\mathbf{t}}(t \cdot A_t) + \mathbf{x}_{it}^{\prime}\gamma + u_{it} \tag{2}$$

$\theta_i, \tau_t, \mathbf{x}_{it}, \gamma$ and u_{it} represent the same as in Equation 1. t is a set of binary variables for each of the periods for which there is data available, with the reference level set to one period before AITC eligibility (March 2016). A_t is a binary variable equal to unity if the observation is mapped to Alberta and zero otherwise. $t \cdot A_t$ is the interaction term between these two variables, and β_t is the associated vector of coefficients, which will show the difference between the treatment and control groups in the explained variable for all t. For these regressions, I show the values of the interaction terms in event study plots, along with their 95% confidence intervals. I cluster standard errors at the province and period level.

Evidence in favour of the identifying assumption will be observed if the interaction terms before April 2016 are not statistically significant. This supports the idea that Alberta had no significant differences in the trend of patent applications to other provinces before the intervention. Thus, I use the event study regressions to provide evidence of the causal identification of the average treatment effect of the AITC on patent applications. Further, I examine the effectiveness of the AITC by looking at post-treatment interaction terms, which should show statistically significant differences if the AITC affected Albertan patent applications.