

# 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. All parties involved in a patent application can be identified along with their registered province of origin if the party comes from Canada and the state if the party is recorded in the United States.

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 context, which I reviewed on Section ???. I assign patents to provinces based on where the majority of parties involved in a patent application report their location. 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. For robustness checks, I also consider the number of parties involved in a patent application as explained variables for robustness checks, separating by party type (inventors, owners and applicants).

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. I aggregate data at the quarterly level by summing all variables

except the consumer and new housing indices, for which I take arithmetic averages.

## 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 had an effect in patent applications from Albertan provinces. 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:

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

where  $P_{it}$  is the explained variable; in most specifications,  $P_{it}$  is the number of patents filed in a province  $i$  and period  $t$ .  $\theta_i$  and  $\tau_t$  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  $\gamma$  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.

The key identifying assumption is that, absent of treatment, the trend of the explained variable in Alberta would follow a similar pattern to control provinces. I estimate models with explanatory variables to control for factors which may affect the comparability of the treatment and control groups. However, to allay the concern of unobservable factors impact-

ing patent application trends across provinces, I estimate event study regressions following Equation 2 below and provide supporting evidence for causal identification of  $\hat{\beta}$ .

$$\ln(P_{it} + 1) = \theta_i + \theta_t + \beta_t(t \cdot A_t) + \mathbf{x}_{it}'\gamma + u_{it} \quad (2)$$

$\theta_i$ ,  $\tau_t$ ,  $\mathbf{x}_{it}$ ,  $\gamma$  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  $\beta_t$  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 a statistically significant difference if the AITC had an effect on patent applications in Alberta.