# Estimating Income-Heterogeneous Inter-Provincial Mobility Margins of Income Tax: Assessing the Prevalence of Tax Avoidance Practices

ECO2460 Research Paper

Malek Hassouneh - 1005724806

Ismail Landolsi - 1009470010

April 12, 2023

**Abstract**

Using machine learning techniques, we study the determinants of inter-provincial relocation decisions of labor in Canada, spanning over a segmented 15-year window, and across the 10 Canadian provinces. Exploiting cross-sectional variation in individual characteristics and preferences, we find evidence for heterogeneous feature importance across the income distribution. Specifically, we find that the relative importance of income tax as a determinant of inter-provincial relocation is 33.02% higher for top income quantiles relative to the national aggregate. This heterogeneity in relative importance is shown to increase by 6.95% over time, with income tax importance increasing by 22.47% for top income quantiles, coinciding with hikes in provincial top marginal income tax rates in the early 2010s. We argue that these results provide strong evidence for an increasing prevalence of tax avoidance practices in Canada through the spatial channel of inter-provincial migration, specifically among top-income earners.

# 1 Introduction

## 1.1 Motivation

With increasingly frictionless global and regional economies, and the ensuing high degree of labor market integration, the geographical mobility of labor has become drastically easier. As such, inter-country, and especially intra-country inter-jurisdiction competition for productive labor has been amplified. In parallel, empirical observation of an increase in the frequency of tax schedule reforms in Canada, specifically, tax increases promoting progressivity, has brought the idea of "fiscal attractiveness" subtle political traction in recent times. Since 2001, there have been 38 reforms to top marginal income tax rates across the 10 provinces and 3 territories of Canada, 22 of which were hikes in the top tax rates[1]. Rationally, individuals factor in regional differentials in fiscal programs and tax rates when making their location decisions, seeking to enjoy a greater share of their generated surplus post-taxation; a pure tax avoidance incentive. Hence, regional differentials in tax rates have the potential to significantly affect the geographical allocation of labor and more specifically skilled workers who are believed to be more mobile than the general labor force (Moretti and Wilson 2017). What is left to examine is the extent to which this tax avoidance incentive is present in Canada in response to tax schedule reforms of the past.

## 1.2 Summary of Findings

Using cross-sectional data from the Canadian censuses of 2006, 2011, and 2016, we fit a total of 30 random forests on a binary variable indicating inter-provincial migration in the past 5 years. These 30 random forests amount to 1 random forest for the aggregate sample, 5 random forests on different income quantile sub-samples, and 4 random forests on geographical region sub-samples, all repeated 3 times every year. We then extract sets

[1]These numbers are obtained by web-scraping provincial tax rates from archived provincial tax forms.

of feature importance for various variables, with emphasis on the feature importance of the amount of income tax paid. We observe stark heterogeneity across the income distribution, with the importance of income tax increasing monotonically with income quantiles such that the relative importance of income tax as a determinant of inter-provincial relocation is 33.02% higher for top income quantiles relative to the national aggregate. This heterogeneity increases by 6.95% over time, with income tax importance increasing by 22.47% for top income quantiles. We argue that this increase, coinciding with hikes in provincial top marginal income tax rates in the early 2010s, is evidence of an increasing prevalence of tax avoidance practices in Canada through the spatial channel of inter-provincial migration, specifically among top-income earners. Finally, we find little heterogeneity in the importance of income tax across geographical regions, showing that tax avoidance is a nationwide phenomenon.

## 1.3 Literature Overview & Contribution

Traditionally, the empirical literature on the effect of taxation on the spatial mobility of labor has suffered from a scarcity owing to the confounded nature of individual relocation decisions. A battery of unobservables would have to be accounted for, rendering estimation of the mobility margin difficult. Seminal papers on this topic include Akcigit et al. 2016, Kleven et al. 2013, and Moretti and Wilson 2017, examining the mobility of top-income "superstars" such as investors and European football players. The former two examine mobility on an inter-country level, while the latter examines mobility on an intra-country inter-jurisdiction level, the dimension of concern for our study. While these published works examine the relocation decisions of millionaires or a subset of specialized individuals, our study examines the general labor market mobility of all working labor across various income quantiles. To the best of our knowledge, this topic has yet to be explored through machine learning techniques, outside the work of Tarasyev et al. 2018. Still, the latter's approach is on a macroeconomic level, highlighting regional differences in economic indicators, whereas

our study explores the individual-specific characteristics behind relocation decisions.

Literature on the mobility response of labor to taxes within Canada also remains scarce, but many unpublished or outdated works estimate general mobility responses to related factors. Day and Winer 2006 investigate the response of aggregate migration in Canada as a consequence of regional variation in unemployment insurance systems. Moreover, Meyer 1998 examines a similar fiscal migration response, examining the relocation responsiveness of poor individuals to provincial welfare benefits. In both studies, induced migration is found to be modest in magnitude, signaling that the Canadian labor market is generally unresponsive, with a high degree of spatial inertia. While our study seeks to exploit the same inter-provincial variation in income tax schedules and fiscal programs, our research design will only allow us to examine the relative importance of these factors over time, not commenting on the inert nature of Canadian labor.

Our findings contribute to the above literature in three ways. First, it provides a missing machine learning outlook on the estimation of the most important determinants of long-term inter-provincial labor mobility. Second, it adds to the literature on the estimation of the mobility margin and the degree of tax avoidance, conducting heterogeneity analysis across the income distribution, geographically and temporally. Finally, it extends the current outdated work on the mobility of Canadian labor, examining the effect of changes in top marginal income tax rates in the early 2010s, and commenting on the general pervasiveness of tax avoidance practices in Canada.

# 2 Research Design

## 2.1 Data

Our data comes from the census public use microdata files for Canada. These censuses are conducted every 5 years and are released with a lag of 2 years, meaning that a discontiguous set of years will be examined in the study. The latest publicly accessible census is that of

2016, and we limit our analysis only back to 2011 and 2006, beyond which censuses fail to report information on the amount of income tax paid by individuals, which is critical for our analysis. The census for 2011 was replaced by the National Household Survey (NHS), which for our purposes contains the same information. The censuses provide information on an individual's current province of residence and their province of residence five years ago. Moreover, multiple individual and household-specific socioeconomic factors are reported, such as age, ethnicity, income tax paid in the past year, total income, and shelter cost.

| Region | **Aggregate** | Atlantic | Central | Prairie | Pacific |
|---|---|---|---|---|---|
| Income %-ile | 5th | 25th | 50th | 75th | 95th |
| Census 2006 | **3014** | 434 | 972 | 1095 | 513 |
| $N = 129,765$ (15.37%) | **2.32%** | 4.41% | 1.20% | 4.76% | 3.22% |
| | 167 | 555 | 542 | 706 | 923 |
| | 2.81% | 2.27% | 1.75% | 2.34% | 2.85% |
| Census 2011 | **3312** | 564 | 948 | 1212 | 588 |
| $N = 140,059$ (15.79%) | **2.36%** | 5.80% | 1.08% | 4.75% | 3.45% |
| | 189 | 626 | 613 | 825 | 954 |
| | 2.75% | 2.32% | 1.87% | 2.41% | 2.78% |
| Census 2016 | **2408** | 348 | 669 | 951 | 440 |
| $N = 131,131$ (14.09%) | **1.84%** | 3.81% | 0.82% | 3.79% | 2.82% |
| | 121 | 381 | 456 | 568 | 753 |
| | 2.04% | 1.49% | 1.53% | 1.78% | 2.33% |

Table 1: Sample Selection Across Censuses and Inter-Provincial Mobility Counts

*Note:* The first column contains the sample size for each census year after sample selection, with the percentage of remaining observations relative to the initial samples. The second to sixth columns provide counts of inter-provincial mobility across income quantiles and geographical regions. The percentage of each respective group's total sample that relocated provinces over 5 years is displayed below each count.

For each census, sample selection was done, focusing only on individuals that fit the profile of one who is the principal agent in their inter-provincial relocation decision. We select individuals of working age, between the ages of 18 and 64, omitting any dependent children

sampled that do not have a say in their relocation decision, and any retired individuals who are assumed to be highly immobile. Most importantly, we delimit our samples only to "primary household maintainers", a dummy variable recorded in the census, allowing us to abstract from dependent spouses who again are assumed to not have a say in their inter-provincial relocation decision. Finally, we remove individuals who were not residents of Canada 5 years ago, i.e. external immigrants for which we do not have a prior province of residence recorded. These observations would contaminate our findings, not in keeping with the general theme of the study, in which long-term relocation implications and decisions are our concern. A breakdown of this sample selection for each census year is in Table 1.

## 2.2 Methodology

To examine the factors behind the inter-provincial relocation response of individuals, we define our response variable to be a binary variable that measures whether an individual's current province of residence is the same as their province of residence five years ago. We then select 17 individual-specific features to include in our classification models that are most correlated with mobility status. The full list of variables can be seen in more detail in the note of Figure 1. Our classification model of choice is a random forest, fitting a total of 30 random forests, as explained above in section 1.2, and extracting the feature importance of all variables for all years[2]. Crucially, with feature importance being normalized, we maintain the same number of features in all of the random forests, allowing for a comparison of feature importance across models. For each random forest, we cross-validate over the number of estimators, maximum depth, minimum samples per split, and minimum samples per leaf. Finally, the features were all standardized so that the scale of the variables did not interfere with the feature importance calculation. The resulting mobility counts for samples across income quantiles and geographic regions are presented in Table 1.

[2]For conciseness, we omit the comparisons to other classification methods in terms of predictive accuracy, though the random forest ranked top for most sub-samples examined.

# 3 Results

AUCs for each of the 30 random forests are reported in Table 2. Ranging from the low 0.7s to the low 0.8s, the magnitude of the AUCs reflects a good degree of predictive accuracy of our models, with any higher AUCs presumably unattainable given the confounded nature of relocation decisions. Models fitted on income quantile sub-samples seem to marginally overperform models fitted on geographic region sub-samples. This will be relevant later on when discussing the credibility of both the results from our income quantile heterogeneity, and our geographic region heterogeneity analysis.

| Region | Aggregate | Atlantic | Central | Prairie | Pacific |
|---|---|---|---|---|---|
| Income %-ile | 5th | 25th | 50th | 75th | 95th |
| Census 2006 | 0.803 | 0.803 | 0.804 | 0.719 | 0.708 |
| | 0.762 | 0.752 | 0.822 | 0.821 | 0.804 |
| Census 2011 | 0.814 | 0.734 | 0.776 | 0.737 | 0.745 |
| | 0.817 | 0.811 | 0.776 | 0.813 | 0.782 |
| Census 2016 | 0.798 | 0.782 | 0.809 | 0.754 | 0.708 |
| | 0.726 | 0.816 | 0.817 | 0.823 | 0.787 |

Table 2: Random Forest Prediction Accuracy (AUC)

## 3.1 Income Quantile Heterogeneity

The extracted feature importance for each of the 15 random forests over income quantiles for each of the 3 census years can be found in Figure 1. Across all income groups, age and province of work are consistently the most important features in predicting 5-year inter-provincial mobility. This is sensible, given the increased immobility with age, and the greater tendency to relocate given that an individual's work is outside of their province of residence. Factors such as total income and shelter cost are also of high importance across all quantiles. Heterogeneity starts to be observed when examining the next most important

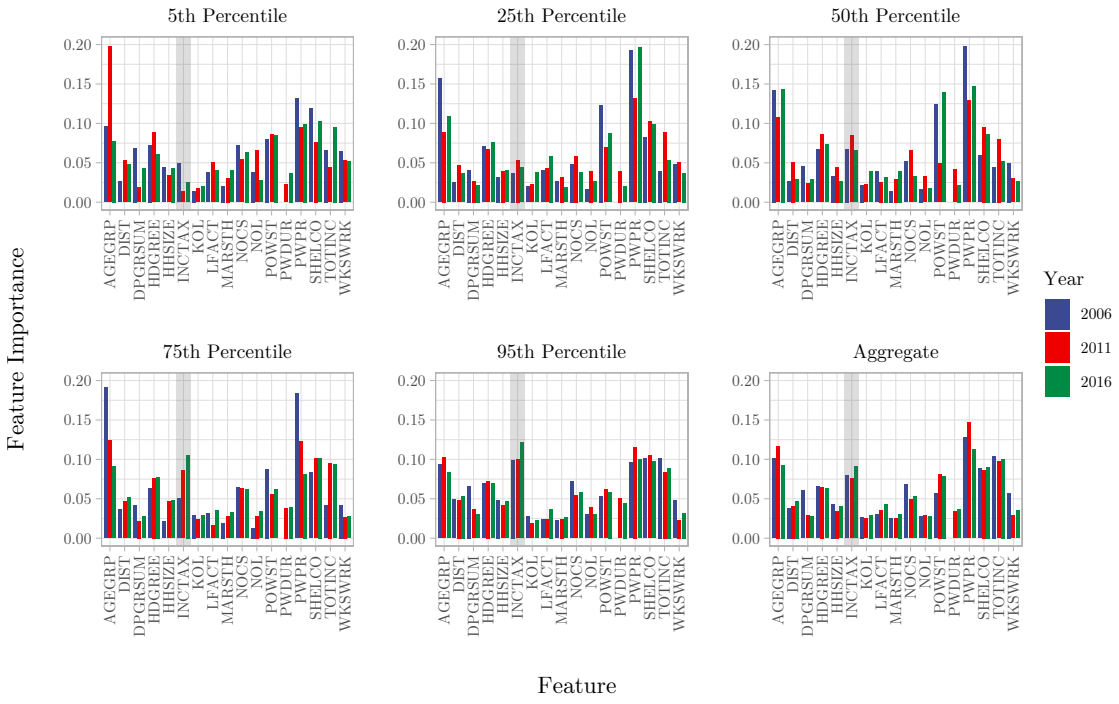


Figure 1: Heterogeneity in Feature Importance Across Income Quantiles

*Note:* The 17 abbreviated features from left to right are: age group, distance from home to work, ethnicity, highest educational degree, household size, income tax paid in the past year, knowledge of official languages, labor force status, marital status, occupational classification, knowledge of non-official languages, place of work status, commuting duration, place of work province, shelter cost, total income, weeks worked in the past year.

feature, namely income tax, which has greater importance for mobility starting with the 50th income percentile. What is even more striking is the evolution of this heterogeneity over time. Figure 2 displays this temporal relationship more clearly. On an aggregate level, the importance of income tax increased by 14.51%, though this is carried by increases for the 75th and 95th percentiles, with 108.24% and 22.47% increases from 2006 to 2016 respectively. We interpret this increase in feature importance as an increase in the prevalence of tax avoidance practices among members of the top income quantiles. For the lower income percentiles, income tax feature importance increased by -49.28%, 18.28%, and -2.79% for the 5th, 25th, and 50th percentiles respectively.

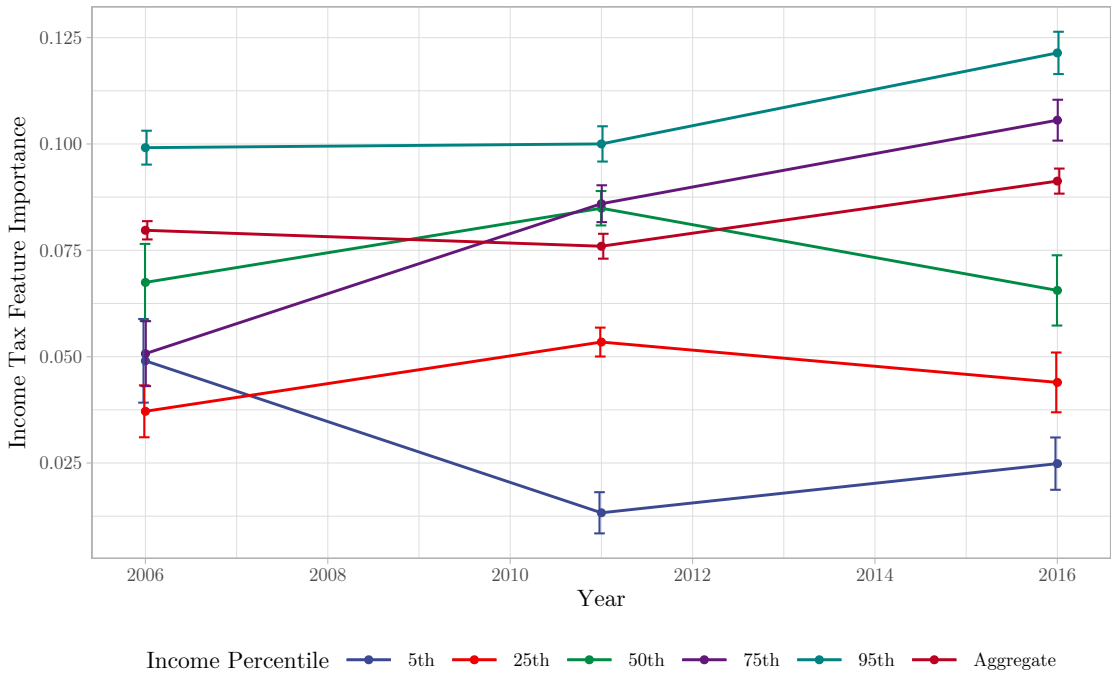


Figure 2: Income Tax Feature Importance Over Time

These changes can be contextualized by referring to the nationwide increases in top marginal income tax rates that occurred in the early 2010s. Most notably, Quebec announced an increase in its top rate from 20% to 24% starting 2011 and 24% to 25.75% starting 2015, Ontario from 11.16% to 13.16% starting 2013, British Columbia from 14.7% to 16.8% starting 2014, and Alberta from its 10% flat rate to 15% starting 2016. Thus, with a greater tax burden on the top income quantile, incentives to relocate provinces grew. This is demonstrated by greater feature importance for income tax, observed with the increase for the 95th percentile in 2016 relative to the limited change prior to 2011, plausibly suggesting increased tax avoidance. Of course, this should not be interpreted as a causal relationship, but rather that one of the channels behind this 33.02% heterogeneity in income tax feature importance relative to the national aggregate is this widespread increase in income tax rates.

## 3.2 Geographic Region Heterogeneity

For completeness, we repeat this analysis for heterogeneity over geographical regions,

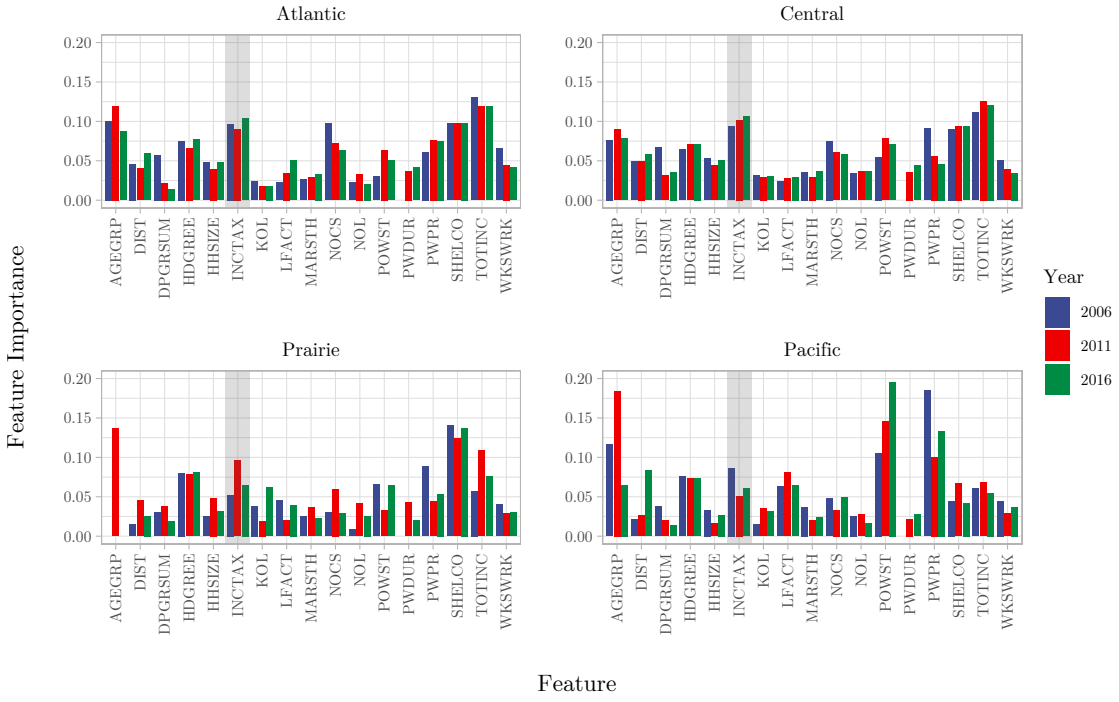


Figure 3: Heterogeneity in Feature Importance Across Geographical Regions

*Note:* The 17 abbreviated features are displayed in the note below Figure 1. Atlantic refers to Newfoundland and Labrador, New Brunswick, Nova Scotia, and Prince Edward Island; Central refers to Quebec and Ontario; Prairie refers to Manitoba, Saskatchewan, and Alberta; Pacific refers to British Columbia.

displaying the set of extracted feature importance in Figure 3. Magnitude-wise, we observe a rough parity between all regions in terms of the importance of income tax, with slightly lower levels for Prairie and Pacific, and gradual increases in the feature importance of income tax over time for all regions except for Pacific. Otherwise, we observe minimal heterogeneity in what factors are the most important for inter-provincial relocation decisions, except for place of work status for Pacific recording a larger importance compared to other regions. From this, we conclude that tax avoidance pervasiveness, observed specifically among top-income quantile members, is not endemic to a specific region. This conclusion has distributional implications, in which we don't expect the distribution of provincial fiscal revenues to be skewed towards a selection of provinces due to inter-provincial relocation.

# 4 Conclusion

Studying the main determinants of inter-provincial relocation decisions of labor in Canada, we find evidence for heterogeneity in feature importance across the income distribution. The relative importance of income tax as a determinant of inter-provincial relocation is 33.02% higher for top income quantiles relative to the national aggregate. This heterogeneity increases by 6.95% over time, with income tax importance increasing by 22.47% for top income quantiles, coinciding with hikes in provincial top marginal income tax rates in the early 2010s. We interpret these results as strong evidence for the increasing prevalence of tax avoidance practices in Canada through the spatial channel of inter-provincial migration. As mentioned above, our results should not be interpreted causally, but are still a strong indication that the nationwide increases in provincial top marginal income tax rates are one of the possible mechanisms behind this interpreted increase in tax avoidance prevalence. Our results are also silent on the absolute magnitude of tax avoidance prevalence, only commenting on its evolution, and its relative magnitude across income quantiles or geographic regions.

Policy-wise, our results signal that provinces should pay attention to the degree of their "fiscal attractiveness". While we find no heterogeneity in income tax importance across regions, implying minimal distributional implications, our analysis is simply a snapshot of the mobility margin in the immediate future post-tax reforms. It is plausible that long-standing and sustained differentials in provincial marginal income tax rates could induce irreversible changes to the fiscal well-being of provinces. Still, our study should be extended to involve some sort of revenue analysis prior to making any concrete policy recommendations. Moreover, a large proportion of top-income earners are also investors, who are subject to other capital gains, dividends, and property taxes at the provincial level. Accounting for these in a more holistic analysis of inter-provincial tax differentials induced mobility is imperative. This would require richer data, with the census only reporting income taxes paid.

# 5 References


Akcigit, Ufuk, Salomé Baslandze, and Stefanie Stantcheva. 2016. "Taxation and the International Mobility of Inventors." *American Economic Review* 106 (10): 2930–2981.

Day, Kathleen M., and Stanley L. Winer. 2006. "Policy-Induced Internal Migration: An Empirical Investigation of the Canadian Case." *International Tax and Public Finance* 13 (5): 535–64.

Kleven, Henrik Jacobsen, Camille Landais, and Emmanuel Saez. 2013. "Taxation and International Migration of Superstars: Evidence from the European Football Market." *American Economic Review* 103 (5): 1892–924.

Meyer, Bruce D. 1998. "Do the Poor Move to Receive Higher Welfare Benefits?" *National Bureau of Economic Research*.

Moretti, Enrico, and Daniel J. Wilson. 2017. "The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists." *American Economic Review* 107 (7): 1858–1903.

Tarasyev, Alexandr A., Gavriil A. Agarkov, and Seyed Iman Hosseini. 2018. "Machine Learning in Labor Migration Prediction ." *AIP Conference Proceedings*.