Intergenerational Income Mobility in Canada and the United States (DRAFT)

Marie Connolly, Miles Corak and Catherine Haeck*

October 17, 2016

Abstract

In this paper, we investigate differences in intergenerational income transmission in Canada and the United States. We develop new estimates for Canada following the methodology in Chetty et al. (2014a), based on newly linked administrative tax data, which add younger cohorts to Statistics Canada's Intergenerational Income Database. We use a subsample of our data to get a sample of children comparable to Chetty et al.'s: born in 1980-1982, and observed in 2011-2012, at age 29-32. We look at these children and their parents' income, and then compute their rank in the American income distribution. Looking at Canada as a whole, we find rank-rank correlations of about 0.2, compared to 0.34 in the US. To look at geographical patterns, we compute mobility measures at the level of the Canadian Census divisions. Top regions (largely driven by Alberta and its resource boom in those years) show markedly higher levels of mobility than US communities. We use K-means, a machine learning algorithm, to classify areas into different clusters of mobility. We then look at the correlation between our mobility measures and a host of community-level characteristics in the areas where the children in our analysis grew up, both overall and within cluster.

JEL codes: J62, J61, H0, R0

Keywords: social mobility, intergenerational income mobility, geographical patterns, Canada, United States

*Connolly: Université du Québec à Montréal, connolly.marie@uqam.ca. Corak: University of Ottawa, miles.corak@uottawa.ca. Haeck: Université du Québec à Montréal, haeck.catherine@uqam.ca. The authors would like to thank the Fonds de recherche du Québec - Société et culture for their funding (grant 2016-PU-195586) and the Social Analysis and Modelling Division at Statistics Canada, in particular Yuri Ostrovsky and Grant Schellenberg, for making this work possible. Thanks especially to Charles-A. P.-Demontigny for outstanding research assistance, and to Martin Leblond-Létourneau for the occasional invaluable help. We also thank those who provided comments: David Card and Emmanuel Saez. All errors remain our own. The analysis presented in this paper was in part conducted at the Quebec Interuniversity Centre for Social Statistics which is part of the Canadian Research Data Centre Network (CRDCN). The services and activities provided by the QICSS are made possible by the financial or in-kind support of the Social Sciences and Humanities Research Council (SSHRC), the Canadian Institutes of Health Research (CIHR), the Canada Foundation for Innovation (CFI), Statistics Canada, the Fonds de recherche du Québec - Société et culture (FRQSC), the Fonds de recherche du Québec - Santé (FRQS) and the Quebec universities. The views expressed in this paper are those of the authors, and not necessarily those of the CRDCN or its partners.

More than twenty years ago, Card and Freeman started their volume on differences between Canada and the United States by the following observation: "Canada and the United States are as close economically and socially as any pair of countries in the world" Card and Freeman (1993, p. 1). Canadians and Americans also share many opinions and values, such as their definition of what constitutes the "American Dream" or the Canadian "Good Life" and the meaning of equality of opportunity Corak (2009). Yet social mobility is markedly lower in the United States. Estimates of the intergenerational elasticity of income, or the degree to which a child's income as an adult is related to his parents' income, have been found to be much higher in the U.S. (where they range between 0.4 and 0.6) than in Canada (where similar figures are found between 0.2 and 0.3) Corak (2009). Concerns about increasing inequalities inevitably tie in to intergenerational issues and the transmission of advantage, or disadvantage, since more inequality tends to be associated with less mobility. Why is mobility so much lower in the United States than in Canada?

Up until recently, mobility estimates on the U.S. were mostly based on survey data of limited sample size, such as the Panel Study of Income Dynamics, while virtually all the estimates on Canada are based on the cohort of children born in the 1960s from the much larger administrative tax files from Statistics Canada's Intergenerational Income Database (IID) Corak and Heisz (1999). Recent work by Chetty et al. (2014a) and Chetty et al. (2014b) uses newly available administrative federal tax data to provide mobility estimates at a subnational level, the commuting zone, something not possible with survey data. Chetty et al. (2014b) show that rank-rank correlations (between a child's income rank and his parents' income rank) have been relatively stable for children born between 1971 and 1993, while income inequality has increased over time. They further document the wide geographical variations in mobility across the United States, with some areas as mobile as Denmark or Canada. Such a level of disaggregation has not been studied in Canada, nor have figures for more recent cohorts been computed to allow for a direct comparison with Chetty et al. (2014a)'s work.

In this paper, we investigate differences in intergenerational income mobility in Canada and the United States and relate those differences to various observable characteristics. We start by developing new estimates for Canada following the methodology in Chetty et al. (2014a,b), using newly added cohorts to the IID. This renders possible a direct comparison with Chetty et al. (2014a)'s core sample, which consists of children born from 1980 to 1982. Our analysis puts the Canadians in the U.S. distribution, which allows us to assess the mobility of Canadians as if they were part of the larger U.S. market. Our analysis provides multiple measures of mobility, and uses K-means, an unsupervised machine learning algorithm, to identify geographic clusters of mobility, both in Canada alone and in the two countries. We also look at correlations with a host of community-level characteristics in the areas where the children in our analysis grew up. Chetty et al. (2014a) found that factors related to racial segregation, income inequality, the educational system, social capital and family structure were highly correlated with absolute mobility at the commuting-zone level. We study these variables and others that could be relevant in the Canadian context, such as factors related to language, geographical mobility, and the Indigenous population. This is not a causal analysis, but it allows for a compelling portrait of an early cohort of Millennials: how they fare in life in adulthood, and how where they were raised is related to their social mobility. This is an important first step towards understanding mobility, one that will direct researchers' and policymakers' attention to crucial areas.

Our findings confirm that Canada is a more equal and mobile society. The income distributions of parents and children are more compressed in Canada, with fewer households experiencing both extreme poverty, and extreme wealth. At the national level, the correlation between a child's income rank and his or her parents' rank is 0.22, compared to 0.341 in the United States. At the subnational level, just like in the United States, there exists great variation in the mobility measures we computed. Applying a K-means algorithm to classify Canadian Census divisions into one of three groups, we find one large set of Census divisions with average values of mobility, which groups together the largest cities (Toronto, Montreal, Vancouver). Another cluster consists of areas that are very mobile, in which children do really well in terms of income. These areas are concentrated in Alberta, Saskatchewan and Newfoundland, and appear to be linked with the boom in the oil industry at the time (our child outcomes are measured in 2011 and 2012). An interesting question would be to see what happens to these areas in bust years. The final cluster displays very little mobility–though still more than the poorest, least mobile areas of the United States, such as the South–and is predominant in the Northern part of the country, and areas that are less densely populated, as well as a few pockets of lower opportunity in places like rural Quebec. Our machine learning algorithm also helps us see which communities of Canada and the United States are similar in terms of mobility. We find that the oil producing region of the West are grouped with our resource-rich communities, and that the US South is clustered with our North. The two remaining groups are almost split exactly along the border: while similar in terms of their absolute mobility, the communities in the US have both a higher average rank-rank slope (less mobility) and much higher mean and median parental incomes. A number of factors identified by Chetty et al. (2014a) to be highly correlated with mobility in the United States also turn out to be important in Canada: those include factors related to family structure, income inequality, and the labor market. Race is less of an issue in Canada, mainly since the fraction of black residents is considerably smaller there. We uncover an interesting parallel however with the Indigenous population: for our least mobile areas, the fraction of aboriginals appears to be a strong correlate. In light of recent calls for reconciliation (Feir and Hancock 2016; Truth and Reconciliation Commission of Canada 2015), researchers and policymakers should pay attention to our findings to see how they can shed light on an overlooked segment of the Canadian population.

The remainder of this paper is structured as follows. We first present the data we use, followed by the methodology. We then offer a presentation of the findings, and a final section containing a discussion and conclusion. An appendix follows.

1 Data

1.1 Intergenerational Income Database

We base our analysis on Statistics Canada's Intergenerational Income Database, which contains tax records of successive cohorts of Canadian individuals and their parents.[1] The IID was created in the 1990s and featured in several prominent studies of intergenerational issues, either directly related to income transmission (Corak 2006; Corak and Heisz 1999), or to other topics such as the effects of worker displacement (Oreopoulos et al. 2008), the long-run consequences of living in a poor neighborhood (Oreopoulos 2003), or the effect of divorce and parental death on children (Corak 2001). The original cohorts of the IID covered children born between 1963 and 1970. In order to offer a direct comparison to Chetty et al. (2014a), and as part of a broader initiative to study the evolution of social mobility in Canada, we worked with Statistics Canada to add three new cohorts to the IID. The new cohorts are referred to as the 1991, 1996, and 2001 cohorts, based on the year of the match between parents and their offspring.

In the IID, the unit of observation is a parents-children match. Before the introduction of the Canada Child Tax Benefits in 2006, Canadian tax filers did not directly identify their children on their tax return, so a special algorithm had to be designed to link together children and their parents in the tax files. The match is made using the T1 Family File (T1FF), which is a data set of all the T1 records provided by the Canada Revenue Agency to Statistics Canada that has been processed to identify family members. To create the IID for a given cohort, the T1FF from the cohort year is used and all the individuals aged 16 to 19 in that year for which parental information on at least one parent is available are identified as being in the desired cohort. The T1FF for the next year is then used to add the children of the relevant cohort that were not captured in the cohort year, and the process is repeated for the next three years, so that cohort members can be linked to their parents in up to five tax years. Appendix Table A3 gives the breakdown of the years at which the link was made for the cohort of children born in 1980 and 1982 in the IID.

In this paper, we align our analytical sample with Chetty et al. (2014a)'s core sample, which they define as all children born in the 1980-1982 birth cohorts, for whom they identify parents, and whose mean parent income between 1996 and 2000 is strictly positive. We select into our sample all children born in 1980 (coming from the 1996 cohort) and 1982 (coming from the 2001 cohort).

[1] A more detailed description of the data and the sample we use is available in the Data Appendix in section A.

We define parent income as the individual average of the total household pretax income, using the Canada Revenue Agency's definition of total income. The average is taken over the years 1996 to 2000, for which relatively few observations are missing since the parents had to have filed their tax return in 1996 or 2001 to be in our dataset to start with. We use compute child income as the mean of the child's total income for the tax years 2011 and 2012, including the child's married spouse total income figures for the same years if a spouse is reported for those years. We only use married spouses even if a significant fraction of the children in 2011 and 2012 report having a common-law partner, to be closer to Chetty et al.'s definition, who do not have information on common-law spouses (filing separately).

All the dollar figures in the IID T1 files are in Canadian current dollars. We first adjust all figures for inflation by converting to 2012 Canadian constant dollars using the Consumer Price Index (CANSIM Table 326-0021). We then convert to 2012 US dollars using the Organisation for Economic Co-operation and Development's Purchasing Power Parity (PPP) for private consumption between the US and Canadian dollar for the year 2012 (a value of 1.284164 C$ for each US$).[2] For all relevant calculations, we use the IID weights provided by Statistics Canada to obtain estimates representative of the Canadian population belonging to the cohort of interest. Appendix Table C1 provides summary statistics for a number of variables in the sample of children born in 1980 and 1982 in the IID.

We identify the geographic location of a child based on his or her postal code at the time when the parents-child tax files match is done, which occurs during the late teenage years. This is consistent with Chetty et al. (2014a), who use location in 1996 for 96% of their sample. Using the postal code, we recuperate the province and the Census division from 1996 Census geography. The Census division is one of the most stable geographical units, especially in the time frame considered. Unlike metropolitan areas, Census divisions span the whole territory of Canada and thus allow us to compute statistics covering the whole country. Also, Census divisions are not arbitrary: their boundaries reflect local and provincial administrative units, such as counties, regional districts, regional municipalities and other types of provincially legislated areas. There are 288 Census divisions in the 1996 and the 2001 Census geographies (Statistics Canada (1997), Statistics Canada (2003)).

1.2 Other Data Sources

Our main data source is the IID, but we make use of other data sources in the last piece of our analysis, which seeks to establish correlations between our mobility measures and various socioeconomic variables. One such source is the Canadian Census of Population. We use the restricted-access

[2] Retrieved online at http://stats.oecd.org/Index.aspx?datasetcode=SNA_TABLE4

Census microfiles for 1996, which are based on the answers to the long-form questionnaire. The long-form version of the Census is given to one in five household and contains detailed questions on topics such as education, ethnicity, employment, and income; 20% of the long-form respondents are available in the Census microfiles, thus covering around 4% of the Canadian population. With its large sample size, the Census is ideal to get estimates at the Census division level, something that is not easily feasible using other data sources, especially for the less-populated Census divisions. Table C3 in the appendix gives a list of all the non-IID variables we use, along with their source data and a short definition. From the 1996 Census, we extract the following information: high school dropout rate for parents, college graduation rate for parents, manufacturing employment share, teenage labor force participation, fraction living less than 15 kilometers from work, fraction of single mothers, fraction married, fraction divorced, migration inflow, fraction foreign born, fraction black, fraction aboriginal, and finally segregation indices related to the black and the Indigenous populations. From the Census geography, we count the number of native reserves in a Census division. We use the 2001 Census of Population to get population counts by Census division. We use the 2011 Census of the Population to get the oil and mining employment share in 2011, and the 2011 National Household Survey to get the high school dropout rate for children born in 1980 and 1982. We also compute a few measures directly from the IID, such as the Gini coefficient and the Census division departure and arrival rates. Finally, we compare our findings with similar measures for the United States computed at the Commuting Zone level by Chetty et al. (2014a), who provide provide data tables on the Equality of Opportunity Project's website.[3]

2 Methodology

Social mobility is a topic that has been largely covered, in the economics literature as well as in the broader social sciences. Economists have typically focused on the intergenerational elasticity (IGE) of income, which can be expressed as the coefficient β in the following equation, usually estimated by ordinary least squares:

$$ln(Y_{i,t}) = \alpha + \beta ln(Y_{i,t-1} + \varepsilon_{i,t}) \tag{1}$$

where Y is a measure of income, i indexes a particular family, and t and $t-1$ represent the children and the parents' generations, respectively (see Corak 2013, or Black and Devereux 2011). The IGE is useful in that it provides a single measure of mobility (or lack thereof, for high IGEs correspond to situations of low mobility), thus allowing researchers to compare different countries or to plot the relationship between income inequality and mobility, known as the Great Gatsby

[3]http://www.equality-of-opportunity.org/index.php/data

Curve (Corak 2013). However, the IGE cannot capture all aspects of mobility, not simply because of the commonly-made linearity assumption. Transition matrices and other measures, such as transition probabilities, can be used to detect non-linearities. Directional rank mobility and rank-rank correlations (or slopes) have been found to be more stable and robust to variable definitions and sample selection, and also have policy relevance (Chetty et al. 2014a,b; Corak et al. 2014).

In this paper, we provide multiple measures of mobility defined as in Chetty et al. (2014a) to allow for a direct comparison. This includes measures of both relative and absolute mobility. In terms of relative mobility, we compute the IGE as defined in equation 1 and the rank-rank slope ρ_{PR}. The rank-rank slope comes from the estimated coefficient when the child's income rank R_i is regressed on his or her parents' income rank P_i. This can be done at the national level, but also at the subnational level. We use the child's Census division at the time of the match with his or her parents, and estimate a series of regressions, one for each Census division, indexed c. The estimated regression model is the following:

$$R_{ic} = \alpha_c + \beta_c P_{ic} + \varepsilon_{ic}. \tag{2}$$

Using the estimates from equation 2 and following Chetty et al., absolute measures of mobility can be defined as the expected rank of a child who grew up in Census division c with parent at national income rank p:

$$\bar{r}_{pc} = \alpha_c + \beta_c p. \tag{3}$$

In terms of absolute income mobility, we use Chetty et al.'s absolute upward mobility, which is the mean rank of children whose parents are at the 25$^{\text{th}}$ percentile of the national parent income distribution. It corresponds to the predicted value of the child's rank when the parents' rank is equal to 25 using the prediction in equation 3: $\bar{r}_{25,c} = \alpha_c + 25\beta_c$. We also provide the probability of rising from the bottom quintile to the top quintile of the income distribution: P(child in Q5 | parents in Q1). Unlike Chetty et al. however, we do not use measures related to the poverty line, like the percentage of children living above the poverty line. Given differences in measurement between Canada and the United States, and the ambiguous interpretation of the comparison of the two countries' poverty lines, we shy away from using this additional measure.

We do however present quintile transition matrices, which offer another look at mobility. Transition matrices are complementary to the IGE and rank-rank measures because they do not rely on a linearity assumption and especially because the speak to the direction of mobility. In such a matrix, each cell gives the conditional probability of the child having an income in a certain quintile (represented by the rows), conditional on being from a family with income in a give quintile (represented by the columns). Since the quintiles are based on the Canadian national distributions, each column sums to 1. Let $P_{o,d}$ be the probability of moving from origin (parental income quintile) o, indexing the column, to destination d (child income quintile). Two probabilities from the transition matrix are particularly relevant to relay information about mobility: $P_{1,5}$ (P(child in Q5 | parents in Q1)), the "rags to riches" movement, which expresses upward mobility, and $P_{1,1}$ (P(child in Q1 | parents in Q1)), the intergenerational cycle of poverty.

2.1 Assigning Canadians a US rank

We compute rank-rank correlation measures using the Canadian children and parents American ranks, which we assign as follows. We take the PPP-converted total incomes figures described in section 1. We then ask ourselves the rank this income would have in the American income distribution, as described in Chetty et al. (2014a)'s Online Table 2.[4] The table gives, for each centile of the parents or children income distribution, the mean of all parent or children incomes, rounded to the $100. We use as centile cutoffs the midpoints between the two means.

2.2 Clustering using unsupervised learning

Once we have computed our measures of income mobility by Census division, we are interested in classifying our geographical units into a certain number of subgroups based on their mobility. We borrow from the machine learning field in computer science and turn to a clustering method that relies on unsupervised learning: the K-means algorithm (James et al. 2013). The idea here is that we have a certain number of observations (the Census divisions, along with the Commuting Zones when looking at Canada and the United States together), and that we use certain criteria to group them into a pre-specified number of clusters. Clustering finds homogeneous subgroups among our geographical units, which allows us to look at which parts of Canada are more or less mobile, as well as which parts of Canada are similar to the United States in terms of mobility. K-means clustering aims to partition observations into groups for which the within-cluster variation is as small as possible, where the within-cluster variation is defined as the sum of all the pairwise squared Euclidean distances between the observations in the cluster, divided by the total number of observations in that cluster. Formally, and following James et al. (2013), section 10.3.1, the optimization problem that K-means seeks to solve is the following:

$$\underset{C_1,\ldots,C_K}{\text{minimize}} \left\{ \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \right\}, \tag{4}$$

where K is the number of pre-defined clusters, C_k denotes a cluster, i and j denote particular observations within a cluster, and x_j is one of the p features of the data (in our case a mobility

[4]Retrieved online at http://www.equality-of-opportunity.org/index.php/data

measure at the Census division level). We use six such features in this paper: relative mobility as measured by the rank-rank slope, absolute mobility as measured by the predicted rank at the 25^{th} percentile of the parental income distribution ($\bar{r}_{25,c}$), the mean and the median of the parental income, and two probabilities from the quintile transition matrix: the probability that a child with parents from the bottom quintile of the income distribution reaches the top quintile, and that he or she stays at the bottom quintile (P(child in Q5 | parents in Q1) and P(child in Q1 | parents in Q1), respectively). Solving equation 4 turns out to be a difficult task, and an algorithm, implemented using the language R, is used (James et al. 2013, p.388). The algorithm starts by assigning each observation to a random cluster. Then, it computes the cluster centroid (vector of the p feature means for the observations in the k^{th} cluster) and assigns each observation to the cluster whose centroid is closest (where closest is defined using Euclidean distance). The second step is repeated until the cluster assignments are stable. Since this procedure selects a local optimum, the whole process should be repeated a number of times using different initial cluster assignments, and the global optimum can be determined by comparing the value of the objective function in equation 4.

2.3 Correlations

In the final section of our analysis, we compute the correlations between our various mobility measures at the Census division level and various characteristics of the Census division. This allows us to identify factors correlated with mobility at a finely disaggregated level, and to compare the importance of various covariates in Canada and in the United States. While this analysis is not causal, it is a first step towards identifying where differences, small or large, lay, and where to focus more targeted research. All correlations come from univariate linear regressions of our absolute mobility measure on the variables of interest, with the two variables previously standardized to have a mean of zero and a standard deviation of one: that way, the estimated coefficient on the explanatory variable gives the correlation between the two. We do this both overall as well as within clusters, as identified by the K-means algorithm described above.

3 Findings

All the findings presented below use the IID as described in section 1 above: the amounts presented are in 2012 US dollars converted using the PPP, so all figures should be readily comparable to those in Chetty et al. (2014a).

3.1 Income distributions

A few key percentiles of the Canadian and American income distributions, for both children and parents, are presented in Table 1. Looking at parents, the bottom half of their income distributions are fairly similar in the two countries; divergences start to appear at the median and above, with American parents having higher mean incomes at every percentile reported. At the 99th percentile, the mean income for Canadian parents is U$246,760, compared to US$420,100 in the United States. The income distributions of the children born in 1980-1982 do not display the same patterns. For children, the mean income is higher for Canadians at every percentile of their income distribution except the 99th, at which point the mean income for Americans is US$193,300, roughly US$10,000 more than for Canadians.

3.2 National estimates

We start by presenting national estimates, to compare with Figure I and Table I in Chetty et al. (2014a). We begin with a visual inspection in Figures 1 and 2. Figure 1 shows the mean child household income by level of parental income, where each dot represents one percentile of the parental income distribution. Our curve here is flatter than in the US, and our upper percentiles are also lower, showing that the Canadian income distribution is more compressed for both parents and children. In Figure 2, we can see the relationship between log parental income and log child income, the slope of which gives us the IGE. The relationship appears much more linear than for the US, which has a more pronounced S-shape. We can also observe that the fraction with incomes under US$500 is markedly lower than Chetty et al.'s percentage of children with zero income, again speaking to the more compressed income scale in Canada. At the 10th percentile of the parental income US distribution, more than 10% of children report a zero income, while by the corresponding 4th percentile in Canada there are only 5% of children with an income under US$500.

[Insert Figure 1 here.]

[Insert Figure 2 here.]

We move to numerical estimates in Table 2. We present our figures for similar subsamples: male or female children, married or single parents, and fixed age at child birth, where we use both the Canadian median age at child birth and the American one. In these subsamples, the core sample is restricted to children whose parents' age is within a five-year window of the median age (both the mother and the father). Additional subsamples which may be relevant in the Canadian context are whether the child lives in the same province or in a different province in 2011 and 2012 as in the year of the match with his or her parents, and whether the child has filed his or her tax return in French or in English. As in our American counterpart, we play around with the sample and the computation of the child's income to assess the sensitivity of the mobility estimates, in particular the IGE. We start with our main IGE estimate, where we exclude children with an average income under US$500, and use a log-log specification. We use a US$500 cutoff since we believe those ultra low income cases are more likely to be coding errors than to reflect actual low income situations. The social safety net is relatively generous in Canada, so total income, which includes benefits, is not likely to be so low. We test the sensitivity of the IGE estimates to this exclusion by recoding children incomes are under US$500 to US$500 or under US$1000 to US$1000. We do not have the issue of observing zeros since we exclude children for which we have no tax information, and no one reports an income of zero. Our IGE estimates are however sensitive to the treatment of the under US$500 or 1000 observations. On line 1, where under US$500 are excluded altogether, our IGE is 0.252 for the core sample (column (2)). It goes up to 0.3 when we recode under US$500 to 500, and to 0.289 when we do the same for under US$1000. Our value of 0.252 is slightly larger than the Corak and Heisz (1999) value of around 0.2 using the cohort of children born 1963-1970, although the samples are not directly comparable since they looked at father and son pairs. It is also markedly lower than the Chetty et al. (2014a) value of 0.344 and other estimates for the United States, confirming the usual findings of higher mobility in Canada compared the US.

[Insert Table 2 here.]

A more stable relationship, as Chetty et al. (2014a) argue, is the rank-rank correlation. Here we take our cohort of children and parents, use their incomes in US dollars, and assign them a US rank based on Chetty et al.'s marginal income distributions. Our rank-rank slope estimate for Canadians, using their US ranks, is 0.22, compared to 0.341 for the US. Figure 3 illustrates this relationship by showing a binned scatterplot of the mean percentile rank of children by the parental rank. Our series for Canada has a similar shape as Chetty et al.'s for the US, with the bottom tail slightly under the regression line and the top slightly above, except flatter. Both relationships are strikingly linear.

[Insert Figure 3 here.]

Columns (3) to (12) of Table 2 offer the same set of estimates at the national level but for various subsamples of the data. Most subsamples give very similar estimates: the rank-rank slopes are all between 0.2 and 0.23. There is one exception: the subsample of children who have moved to a different province between the time they were matched to their parents in our data in late teenage year, to the time we observe their income in 2011 and 2012. For these kids, who represent 8.2% of our data, the mobility measures are much lower, which in the case of the IGE and the rank-rank measures means that parental income is a weaker predictor or outcomes, or more mobility. This is not so surprising, since young adults who decide to move to a different province might do so to attend a post-secondary education institution that could be better than a more local one, or to take a better job, or they may simply have more motivation or other unobservable characteristics that result in better outcomes and more mobility. We will come back to geographical mobility and its links with income mobility when we look at the correlates of mobility at the Census division level.

Table 3 provides the quintile transition matrix for Canada, in which the quintiles are based on the Canadian distribution, unlike in most of our analysis where we assign US ranks to Canadian parents and children. As seen in Table 3, $P_{1,5}$ for Canada is 13.6%, compared to 7.5% in the United States as reported by Chetty et al. (2014a). The rags to riches probability is 1.8 times larger in Canada, based on the two countries' respective income distribution. An alternative look at the transition probabilities would be to use the US ranks assigned to Canadian parents and children. Here however, the restriction that the columns sum to 1 is lifted. Table 4 presents just such a matrix. $P_{1,5}$ is now 15.9%, more than double the corresponding number for the United States. Even more strikingly, $P_{1,1}$, the intergenerational cycle of poverty, is only 17% for Canada (when using US ranks), compared to a figure double that, 33.7%, for the United States. Based on these transition probabilities, and indeed on all the national-level measures presented thus far, Canada is a society where social mobility is on average higher than in the United States.

[Insert Table 3 here.]

[Insert Table 4 here.]

3.3 Subnational estimates

Next we move on to subnational estimates, where we use the Canadian Census divisions and the American Commuting Zones as the geographical units of analysis. We begin by presenting, in Table 5, our mobility measures computed at the Census division level for the 50 largest Census divisions according to their 2001 population.[5] Canada's top performers, in terms of absolute upward mobility, are predominantly in Alberta, Saskatchewan, and Newfoundland, with places in those three provinces clinching the top 6 spots. Canada outperforms the US at the very top of the absolute

[5]Appendix Table C2 gives the same information for all the Census divisions for which the sample size was large enough to release our mobility estimates.

mobility ladder: 11 Census divisions have a predicted rank for parents at the 25$^{\text{th}}$ percentile higher than the top-performing Commuting Zone in the US, Salt Lake City in Utah. Canada also does not have the same lower tail in its 50 largest Census division: the last spot, Windsor in Ontario, just across the border from Detroit, has a value of ($\bar{r}_{25}$) of 39.7, which would place it at the 32$^{\text{nd}}$ spot in the US, between Las Vegas and Chicago. Detroit, for comparison, is ranked 46$^{\text{th}}$ with a $\bar{r}_{25}$ equal to 37.3.

[Insert Table 5 here.]

It is no surprise that young adults from Alberta, Saskatchewan and Newfoundland were doing good in 2011 and 2012. Those were boom years for these provinces, especially Alberta, mostly due to its resources and the tar sands of the northern part of the province. In 2011 and 2012, Alberta had GDP annual growth rates of 6.4 and 3.9%, respectively, compared with 3.0 and 2.5 for British Columbia, 2.4 and 1.3 for Ontario, 1.9 and 1.0 for Quebec, and 3.1 and 1.7 for Canada overall.[6] Newfoundland's economy was strong in 2010 and 2011, but less so in 2012, which saw a 4.39 decline in GDP–the province is susceptible to resource-linked booms and busts. Unemployment was low in Alberta in 2011 and 2012, with monthly unemployment lows of 4.3% (in August and November 2012) and a high of 5.9% (in January 2011). For Canada as a whole, monthly unemployment was between 7.2% (June, August, November and December 2012) and 7.7% (January to March 2011) over the same period.[7] Provinces with low unemployment also included Saskatchewan and, to a lesser extent, Manitoba, whereas higher unemployment could be found in Quebec, British Columbia and Ontario. Newfoundland had, and has historically had, very high rates of unemployment over that period (between 11.4% and 13.2%).

We map our three measures of mobility in Figures 4, 5, and 6.[8] Those maps allow us to see the geographical distribution of mobility across the country. We observe the patterns from Table 5: Alberta, Southern Saskatchewan, and Newfoundland are top performers in terms of absolute mobility, and northern areas of Quebec, Manitoba, Ontario, and Saskatchewan are at the bottom. Looking at Table 5 and comparing the three maps, two findings emerge. The first is that the ranking of Census divisions according to absolute mobility is not the same as the one based on our other two measures, whether it is the rank-rank slope or the rags-to-riches probability. The second finding is that it would be useful to have a way to synthesize the information contained in the various mobility measures. This is precisely what we will do using the K-means algorithm

[6] Growth rates computed from Statistics Canada CANSIM series 384-0038 (GDP, expenditure-based, 2007 chained dollars, annual growth rates).

[7] Unemployment figures are from Statistics Canada CANSIM series 282-0087 (seasonally adjusted unemployment rate, both sexes, 15 years and over).

[8] Appendix Figures B1, B2, and B3 show the equivalent maps where the United States have been mapped as well using the data from Chetty et al. (2014a)'s Online Data Tables.

in the next subsection. But before, we will assess the correlation between our mobility measures, to try to shed some light on the first finding mentioned above. We compute simple population-weighted correlations between the three measures at the Census division level for Canada or at the Commuting Zone level for the United States. The correlation between $\bar{r}_{25}$ and the rank-rank slope is -0.31 in Canada and -0.61 in the United States; the correlation between $\bar{r}_{25}$ and $P_{1,5}$ is 0.69 in Canada and 0.92 in the United States; and the correlation between the rank-rank slope and $P_{1,5}$ is -0.14 in Canada and -0.67 in the United States.[9] The correlations between the measures are systematically lower in Canada. This could be due to the lower number of units of measurement (the Census divisions), as well as to the smaller population in each of the Census division, which would add noise to our estimates. It could also be due to the fact that we are using US ranks when computing our measures; a comparisons with measures using Canadian ranks would be helpful.

[Insert Figure 4 here.]

[Insert Figure 5 here.]

[Insert Figure 6 here.]

To compare the distributions of our mobility measures across Canada and the United States, we present in Figures 7, 8 and 9 histograms of absolute mobility ($\bar{r}_{25}$), the rank-rank slope, and $P_{1,5}$, respectively, where the histograms are done separately for the two countries and overlaid to allow a direct comparison.[10] Looking at Figure 7, we can see that while the predicted income rank for children with parents at the 25$^{\text{th}}$ percentile of the income distribution has a similar range in the two countries, there are a number of communities with predicted ranks in the 30s that can be found in the United States but not in Canada. Conversely, there are more places with predicted ranks above 50 in Canada than in the US. In Figure 8, the picture for relative mobility, as measured by the rank-rank slopes, is even starker: there is a large peak at around 0.2 in Canada, whereas the mass of Commuting Zones in the United States is in the 0.3 to 0,4 range, 1.5 to 2 times larger (less mobile) than in Canada. Finally, Figure 9, for the conditional probability $P_{1,5}$ from the transition matrices, echoes the histograms of absolute mobility found in Figure 7, with distributions overlapping but a mass of less mobile communities showing up in the US and the reverse being true for Canada.

[Insert Figure 7 here.]

[9] The unweighted correlations are generally smaller for Canada, but quite similar in the United States.

[10] The histograms in Figures 7, 8 and 9 are weighted using population counts from the US and Canadian Census. Appendix Figures B4, B5 and B6 show the unweighted versions.

[Insert Figure 8 here.]

[Insert Figure 9 here.]

3.4 Clustering of subnational units using mobility measures

Canada Next, we group the Canadian Census divisions based on their mobility levels. If we were to divide the country into three groups, how would it look like? We use the K-means clustering method, which requires that a number of clusters be chosen first. We use $K = 3$, to split Canada in three broad groups based on the six mobility measures listed in section 2.2. We ran K-means with random restarts and Figure 10 shows a map of Canada with a typical clustering configuration. The K-means algorithm is agnostic to the meaning of the clusters; it merely finds the grouping that minimizes the squared Euclidean distance between the observations in the cluster. However we can, as researchers, label these groupings according to what we know about the Census divisions that form them. Looking at Figure 10, we use the following labels: Cluster 1 ($k = 1$) is what we will call our "Main" cluster, since it contains the most number of CDs, including the most populated ones: Toronto, Montreal, and Vancouver. Cluster 1 covers more than 23 millions Canadians. It also corresponds to average values of mobility, as we will discuss below. We label Cluster 2 ($k = 2$) with the term "Oil," since it appears to correspond, by and large, to areas where the oil and resource-based sectors were prominent, at least in the years our outcomes were measured (2011 and 2012). The Oil cluster covers more than 4 million people and includes the cities of Calgary and Edmonton in Alberta, St. John's in Newfoundland and Regina and Saskatoon in Saskatchewan. Finally, we call Cluster 3 ($k = 3$) "North," since it is overwhelmingly represented by the Northern areas of the country. This cluster is the least populated, with just under 2 million inhabitants.

[Insert Figure 10 here.]

[Insert Table 6 here.]

Table 6 contains a few statistics at the cluster level. Apart from the number of clusters and the total population, each figure corresponds to the population-weighted average of the Census-division level means of the variables listed. Cluster 2, the Oil areas (in yellow on the map in Figure 10), contains Canada's top performers: it has better averages on five of the six measures used by our clustering algorithm, often by a wide mark. The average rank for children from parents at the 25^{th} percentile of the distribution, our absolute measure of mobility, is close to 53, fully 28 ranks better than their parents, and higher than the *most* mobile place in the United States, Salt Lake

City in Utah, by more than 6 ranks. Likewise, its conditional probabilities are very favorable, with close to 22% of children from the bottom quintile of the parental income distribution reaching the top quintile of their income distribution ($P_{1,5}$), and fewer than 15% remaining in the bottom quintile ($P_{1,1}$). Its average rank-rank slope, however, is slightly larger (indicative of less mobility) than that of our main cluster.

The North cluster, Cluster 3 (in red on the map in Figure 10), is the most disadvantaged and least mobile on all counts. Its average absolute mobility, $\bar{r}_{25}$, is 40.39, which puts it on par with places like Dallas and Austin in Texas. An absolute mobility of that value corresponds to the 11$^{\text{th}}$ percentile of the Canadian distribution of $\bar{r}_{25}$, whereas in the US it would be at the 37$^{\text{th}}$ percentile, showing again how much lower the mobility is south of the border. This fact can also be seen by looking at the average rank-rank slope of Cluster 3: at 0.344, it is both much larger than the averages for clusters 1 and 2 (0.222 and 0.23, respectively), and right about at the United States overall level, 0.341.

Canada and the United States We now consider Canada and the United States simultaneously, and ask our learning algorithm to assign Canadian Census divisions and American Commuting Zones to one of two clusters. The idea here is to see, with $K = 2$, if the border drawn according to the clustering matches the actual border between the two countries. The resulting map can be seen at Figure 11. Split into two, North America is not split along the usual east-west border. Roughly speaking, the Census divisions that were in our Oil cluster for $K = 3$ when considering Canada only are now in the same cluster ($k = 2$) as the largest cities in Canada, Toronto, Montreal, and Vancouver. They are grouped with most of the US West and the western part of the Midwest, as well as significant parts of Texas. Only a few pockets east of the Mississippi are grouped with this cluster. The other cluster ($k = 1$), more than three times more populated, groups together the northern part of Canada with the US South and much of the East Coast, as well as the northern Pacific coast. Table 7 presents average characteristics by cluster, just like Table 6 did for Canada. Panel A corresponds to the $K = 2$ case. Interestingly, Cluster 1, the more populated, fares worse in terms of the four mobility measures, but better on the parental income measures. Average mean and median parental incomes are 88,770$ and 61,291$, respectively, compared to 72,747$ and 52,241$ in the other cluster.

[Insert Figure 11 here.]

[Insert Table 7 here.]

Increasing the number of clusters might be more informative and allow a more refined categorization of North American communities. We chose to go with four clusters, and present the map thus generated in Figure 12. In this $K = 4$ clustering, the most populated cluster, $k = 3$, covers most of the US Northeast, Midwest, West Coast, along with Houston, TX, Central Florida and a few other Commuting Zones. The only Canadian Census division part of this cluster is Division No. 15 in Central Saskatchewan (pop. 78K), which contains the city of Prince Albert. The Canadian Oil cluster is now grouped in cluster $k = 1$ with a collection of communities along the central-west part of the country, in Montana, the Dakotas, Wyoming, Nebraska, Oklahoma, Kansas and parts of Texas. Just like the Canadian Oil Census divisions, some of these Commuting Zones have strong employment in the oil and gas extraction industries. In cluster $k = 4$, the American South is grouped with the Canadian North, an area we just identified with lower mobility. The last cluster, $k = 2$, overlaps with the Main cluster in Canada when $K = 3$, plus a few areas scattered across the United States: Miami, FL, Las Vegas, NV, El Paso and Corpus Christi, TX, to name a few.

[Insert Figure 12 here.]

It does appear that there are similarities as well as fundamental differences between the two countries. Both share areas of high mobility and affluence (Cluster 1, Western areas, related to the oil industry) and of low mobility and poverty (Cluster 4, the US South and Canadian North). The latter are also marked by large African American populations, in the United States, and by large Indigenous populations in Canada. However, the rest of the communities are almost split exactly along the border. Cluster 3, the main US cluster, contains only one small community of Canada. Cluster 2, the main Canadian cluster, covers more than 24 million people in Canada out of a population of 29.3 million (in 2001), or more than 82%, and only 8.9 million Americans, out of a population of 281 million (in 2000), or 3%. As seen in Panel B of Table 7, the two main clusters have a similar level of absolute mobility, but Cluster 3, which is virtually only in the US, has a much higher rank-rank slope that Cluster 2, which contains most of Canada's large urban areas, as well as markedly higher average measures of parental income. The mean parental income for Cluster 3 is just under 100,000$, compared to just above 50,000$ for the Main Canadian cluster. The rags-to-riches probability, $P_{1,5}$, has similar averages in the two clusters, but the intergenerational cycle of poverty, $P_{1,1}$, is much lower in Canada, with a probability of 21.19% compared to 33.13% in the United States. The US thus is defined by higher average and median incomes, a higher correlation between parental and children income ranks, and a stronger cycle of poverty. In the next subsection, we will investigate a number of factors that correlate with social mobility.

3.5 Correlations

In the last section of our analysis, we investigate factors that are correlated with mobility. This is not a causal analysis. This is a first step towards understanding what seems to matter for mobility, to help guide our future research on the matter. We do our analysis at the Census division level, taking our subnational estimates presented above and seeing how they correlate with a host of other community-level factors. This follows work done by Chetty et al. (2014a), presented in their Figure VIII. We align our variable definitions with theirs as much as possible, but unfortunately are unable to find sources for a number of interesting variables. Statistics Canada provides the average tuition fees for undergraduate and graduate degrees, as well as violent crime rates, and the Council of Education Ministers provides information on the average student-teacher ratios, but the smallest geographical level for all those is the province, which does not give us enough variation to conduct a meaningful analysis. The variables we use are constructed using the 1996 Census of Population or the IID; both have sufficient sample sizes to get reliable Census-division level estimates. Table 8 shows the univariate correlations we computed between absolute mobility ($\bar{r}_{25}$) and each variable individually. This is done for Canada as a whole, as well as within each cluster as assigned by the K-means algorithm ($k = 3$) and described in the previous subsection. In this table, the correlations are weighted using the 2001 population of the Census divisions; Appendix Table C5 presents the unweighted results.

[Insert Table 8 here.]

We find a few factors that have a relatively strong correlation with upward mobility. Like in Chetty et al. (2014a), family structure appears to be important, with a negative correlation of -0.48 with the fraction of single mothers, of -.3 with the divorce rate, and 0.3 with the fraction married. The correlation with the fraction of single mothers is the largest we find, which was also the case for Chetty et al., though our number is smaller than theirs (-0.48 compared to -0.61 for the population-weighted figure). The same is true for most of our correlations, the largest ones being in the 0.25 to 0.35 range in absolute value, compared to 0.5 to 0.6 and even more for Chetty et al. Perhaps the smaller number of observations that we have in Canada, as counted both by the number of Census divisions and the population within each Census division, make our data noisier than in the United States. However when we consider only Clusters 1 and 2, then correlations with fraction of single mothers is more in line with that of the United States.

The fraction of black residents also seems to matter for mobility, with a weighted correlation of -0.3, though that is likely driven by the Census divisions of the Greater Toronto Area, which exhibit the largest fractions of black residents of the country (0.08 for Toronto itself, and 0.66 for Peel, which includes both Mississauga and Brampton). While race may be an issue, it clearly does not affect the Canadian population at the same magnitude as in the United States, where a large number of communities have shares of black residents above 40%, especially in the South. Intriguingly, our indices of segregation, which we have computed using Census Tracts, do not yield correlations of the expected sign: we get that more segregation is linked to more mobility, unlike what Chetty et al. find. This is something that merits further investigation.

The fraction foreign born is significantly and negatively correlated with mobility, with an overall correlation of -0.22, which is much larger than the figure found for the United States, at -0.02. As mentioned above for race, this large correlation may be related to disproportionate weight for the large urban centers in our calculation, since those also receive the largest shares of immigrants. The unweighted correlation with fraction foreign born is reduced to -0.096. Areas that see a larger inflows of migrants (regardless of place of birth) are areas that appear to be more mobile, and this relationship is especially strong for Cluster 2, our Oil cluster. Geographic mobility, as measured by the fraction of residents moving from one area to another, appears to be generally good for income mobility, with positive correlations found everywhere.

The fraction of aboriginal residents in a Census division is not statistically correlated with absolute mobility, at least when looking at Canada overall. To investigate the issue further, we present in Figure 13 scatter plots of the fraction aboriginal (on the x-axis) and absolute mobility (on the y-axis), where each observation is a Census division. We also group the Census divisions by cluster. The differences are clear. In the Main cluster, the correlation is positive, but becomes slightly negative when we exclude the top three communities in terms of their fraction aboriginal. In the Oil cluster, the correlation is negative but not statistically different from zero, but in the North, the correlation is strongly negative, at -0.52. However, we have some reservations about the figures regarding the fraction aboriginal coming from the Census of Population. As reported by Feir and Hancock (2016), Census enumerators have on occasion been barred from entering Indian reserves, most notably Six Nations 40 in Ontario, the largest reserve in the country, and Akwesasne and Kahnawake near Montreal, Quebec. To use a measure that would not be subject to this bias, we try using the number of reserves at the Census division level, which is more straightforward since we use the Census subdivision type and count up the number designated as reserves. The resulting correlations are however weak. They do come out statistically different from zero when not using the population weights (-0.11), since areas with a high number of reserves are also sparsely populated.

[Insert Figure 13 here.]

The manufacturing share, teenage labor force participation rate and Gini coefficient are all relatively important for mobility, and related to it in a similar way as in the United States. Areas with a larger manufacturing share, a larger Gini coefficient (indicative of more income inequality) and fewer teenage workers all tend of have lower mobility. Our findings confirm the existence of the so-called Great Gatsby curve, which we depict in Figure 14. Areas with more income inequality, which we measure as the Gini coefficient computed on the parental income distribution for the children in our core sample, also tend to have less income mobility. This is true for our three clusters, and is especially strong in the North. The fraction of the population with a short commute, which we could only define as having to commute less than 15 kilometers as opposed to less than 15 minutes in the United States due to data availability, is overall not correlated with absolute mobility. It does however appear to be an important correlate of mobility within our Main cluster, with a correlation of 0.375 there. This correlation vanishes when we do not use the population weights, meaning it the larger number must be related to the effect of bigger cities, where work opportunities within 15 kilometers are probably more common.

We investigate the relation between absolute mobility and the fraction of workers in the oil and mining industries, which is something Chetty et al. (2014a) did not do, so we cannot compare our correlations to theirs. Figure 15 plots the Census divisions by their mobility and fraction working in oil, using the NAICS 2007 2-digit industry code 21. For Canada as a whole, correlation between the two is a very large 0.596, and it is at its highest in the Oil cluster.

[Insert Figure 14 here.]

[Insert Figure 15 here.]

To summarize our findings by cluster, we highlight which factors are most strongly correlated with mobility for each cluster. For the Main cluster, Cluster 1, factors related to family structure, fraction of black residents, migration, income inequality and short commutes are the top correlates. For the Oil cluster, Cluster 2, family structure, migration, manufacturing share, teenage labor force participation rate and income inequality are the most important. For the North cluster, Cluster 3, family structure is again important–though in a lesser measure–, as well as the fraction of aboriginal residents, income inequality, and migration, particularly the departure rate of the Census divisions: children growing up in places from which a larger fraction of them have moved appear to be doing better than those where fewer people have moved. This is linked to geographical mobility: individuals willing and able to relocate tend to improve their situation. This relationship is pictured in Figure 16.

[Insert Figure 16 here.]

4 Discussion and Conclusion

In this paper, we contribute to the literature on intergenerational income mobility by computing a set of estimates using Canadian data that are directly comparable to those from Chetty et al. (2014a) for the United States. This allows for a straightforward comparison of mobility between the two countries. As previously reported in the literature, Canada is, as a whole, a more mobile society than the United States is. Estimates of intergenerational elasticity of income are respectively 0.252 and 0.344 for Canada and the United States; those for rank-rank correlations are 0.22 and 0.341; those for the rags-to-riches probability from the transition matrix are 13.6% and 7.5%. Moving on to subnational estimates, we divided Canada into communities based on Canadian Census divisions, which are somewhat smaller than American Commuting Zones, but represent meaningful administrative entities or groups of entities in Canada. We find that most of our measures have a similar range as in the United States, but that the Canadian distribution over geographical units is generally shifted towards more mobility compared to the United States.

We then use K-means, a machine learning algorithm, to classify communities in Canada, and then throughout Canada and the United States, into clusters based on six measures of income mobility. Looking at Canada only, we split the country in three groups: the main group, which covers major cities such as Toronto, Montreal, and Vancouver, is moderately mobile. We find a second cluster of highly mobile communities, especially along measure of the predicted rank for children from parents at the 25^{th} percentile of their income distribution. This cluster is mainly found in Alberta, southern Saskatchewan and Newfoundland, areas characterized by a strong oil industry, at least during the boom years of 2011 and 2012, when child outcomes are measured. Finally a third cluster consists mainly of northern areas, along with a scattering of communities with lower income mobility, for example in Quebec and in Manitoba.

When we consider the two countries jointly and ask our algorithm to classify areas in four distinct groups, we get that the oil producing regions of the US West are grouped with our previously identified highly mobile areas in Canada. The Canadian North is grouped with the low-mobility US South. The remainder of the two countries is remarkably split along the border, with areas in the US being less mobile but with higher parental incomes.

We finally present a series of correlations between a number of variables and our mobility measures using the Canadian data. We find that some of the same variables that were found to be important correlates of mobility in the United States are also relevant for Canada, such as variables related to family structure, income inequality and the labor market. We also highlight interesting parallels between the African-American population in the US South and the Indigenous population of the Canadian North.

More needs to be done. For this paper, more details on the sample used and the sensitivity of the analysis to some specifications need to be provided. Also, more should be done to relate our findings to the literature, and to offer meaningful discussion of our results. For future research, open questions remain as to why Canada as a whole, as well as the core of Canada (our Main cluster) displays more mobility than the core areas of the United States. One possible avenue to examine would be the relation between the business cycle and the booms and busts of the economy and how they relate to intergenerational transmission of income. The years we considered for child outcomes, 2011 and 2012, were very good years for Alberta and the oil-industry communities. What happened to mobility in the following years, when Alberta entered recession in 2015? Perhaps one difference between Canada and the United States is that there is a greater redistribution of wealth in Canada, which would smooth out total income across the income distribution via transfers and benefits. Nevertheless, this paper offers a contribution by being the first to describe mobility for a younger cohort of Canadians, thus allowing a direct comparison with Canada's neighbors to the South.

References

Black, Sandra E. and Paul J. Devereux (2011). "Recent Developments in Intergenerational Mobility." In: *Handbook of Labor Economics*. Ed. by David Card and Orley Ashenfelter. Vol. 4, Part B. Elsevier. Chap. 16, pp. 1487–1541.

Card, David and Richard B Freeman, eds. (1993). *Small Differences That Matter: Labor Markets and Income Maintenance in Canada and the United States*. University of Chicago Press.

Chen, Wen-Hao, Yuri Ostrovsky, and Patrizio Piraino (2016). "Lifecycle Variation, Errors-in-Variables Bias and Nonlinearities in Intergenerational Income Transmission: New Evidence from Canada." In: *Labour Economics* forthcoming.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez (2014a). "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." In: *Quarterly Journal of Economics* 129.4, pp. 1553–1623.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner (2014b). "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." In: *American Economic Review* 104.5, pp. 141–47.

Cook, Katherine and Andrew Demnati (2000). "Weighting the Intergenerational Income Data File." In: *Social Survey Methods Division, Statistics Canada, mimeo.*

Corak, Miles (2001). "Death and Divorce: The Long-Term Consequences of Parental Loss on Adolescents." In: *Journal of Labor Economics* 19.3, pp. 682–715.

— (2006). "Do poor children become poor adults? Lessons from a cross country comparison of generational earnings mobility." In: *Research on Economic Inequality*. Ed. by In John Creedy and Guyonne Kalb. Vol. 13. Amsterdam: Elsevier.

— (2009). *Chasing the same dream, climbing different ladders: Economic mobility in the United States and Canada*. Economic Mobility Project, Pew Charitable Trusts.

— (2013). "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." In: *Journal of Economic Perspectives* 27.3, pp. 79–102.

Corak, Miles and Andrew Heisz (1999). "The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data." In: *The Journal of Human Resources* 34.3, pp. 504–533.

Corak, Miles, Matthew J. Lindquist, and Bhashkar Mazumder (2014). "A comparison of upward and downward intergenerational mobility in Canada, Sweden and the United States." In: *Labour Economics* 30. Special Section articles on What determined the dynamics of labour economics research in the past 25 years? edited by Joop Hartog and and European Association of Labour Economists 25th Annual Conference, Turin, Italy, 19-21 September 2013 Edited by Michele Pellizzari, pp. 185–200.

Feir, Donna and Robert LA Hancock (2016). "Answering the Call: A Guide to Reconciliation for Quantitative Social Scientists." In: *Canadian Public Policy* 42.3, pp. 350–365.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani (2013). *An Introduction to Statistical Learning, with Applications in R*. 6th ed. Springer.

Oreopoulos, Philip (2003). "The Long-Run Consequences of Living in a Poor Neighborhood." In: *The Quarterly Journal of Economics* 118.4, pp. 1533–1575.

Oreopoulos, Philip, Marianne Page, and Ann Huff Stevens (2008). "The Intergenerational Effects of Worker Displacement." In: *Journal of Labor Economics* 26.3, pp. 455–483.

Statistics Canada (1997). *1996 Census Handbook Reference*. Tech. rep. Catalogue No. 92-352-XPE.

— (2003). *2001 Census Handbook Reference*. Tech. rep. Catalogue No. 92-379-XIE.

— (2016). *Annual Income Estimates for Census Families and Individuals (T1 Family File). Family Data. User's Guide*. Tech. rep. 13C0016. Income Statistics Division.

— (undated). "Weights." unpublished mimeo.

Truth and Reconciliation Commission of Canada (2015). *Honouring the truth, reconciling for the future: Summary of the final report of the Truth and Reconciliation Commission of Canada*.

5 Figures

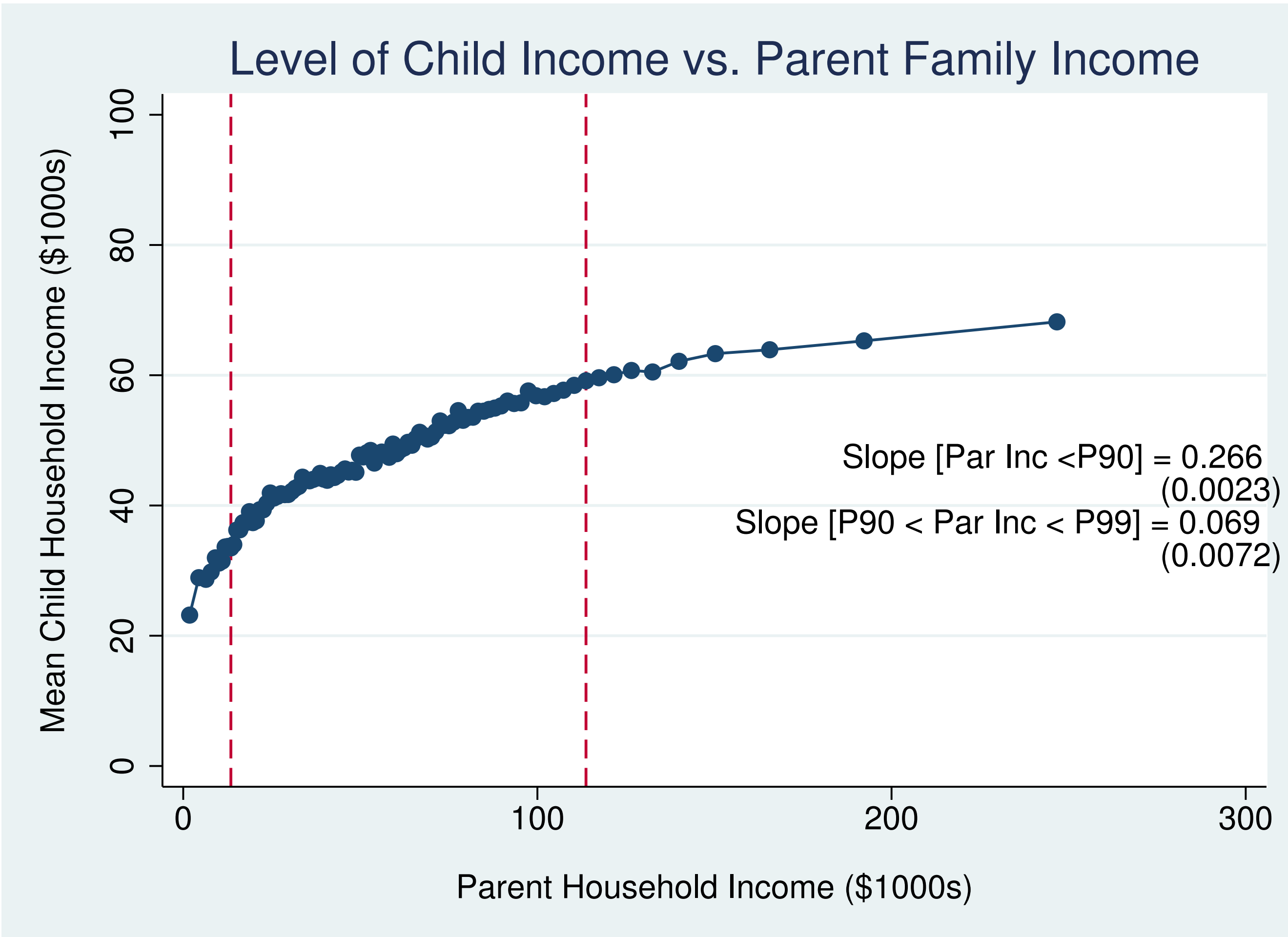

Figure 1: Relation between Children's and Parents' Incomes in Levels

Source: Authors' calculations based on the IID
Note: This figure presents a nonparametric binned scatterplot, where each dot represents one rank of the Canadian parental income distribution. The mean level of parental household income is on the x-axis and the mean level of children household income is on the y-axis. All dollar amounts are in 2012 US$, corrected for inflation in Canadian dollars and converted using the PPP for 2012. The two vertical dashed lines correspond to the 10^{th} and 90^{th} percentiles of the parental income distribution. The top 1% bin is not shown. We report OLS regression coefficients where child income is regressed on parental income using the IID microdata, separately for the bottom 90% of the parental income distribution and for parental incomes between the 90^{th} and 99^{th} percentiles. Standard errors are in parentheses.

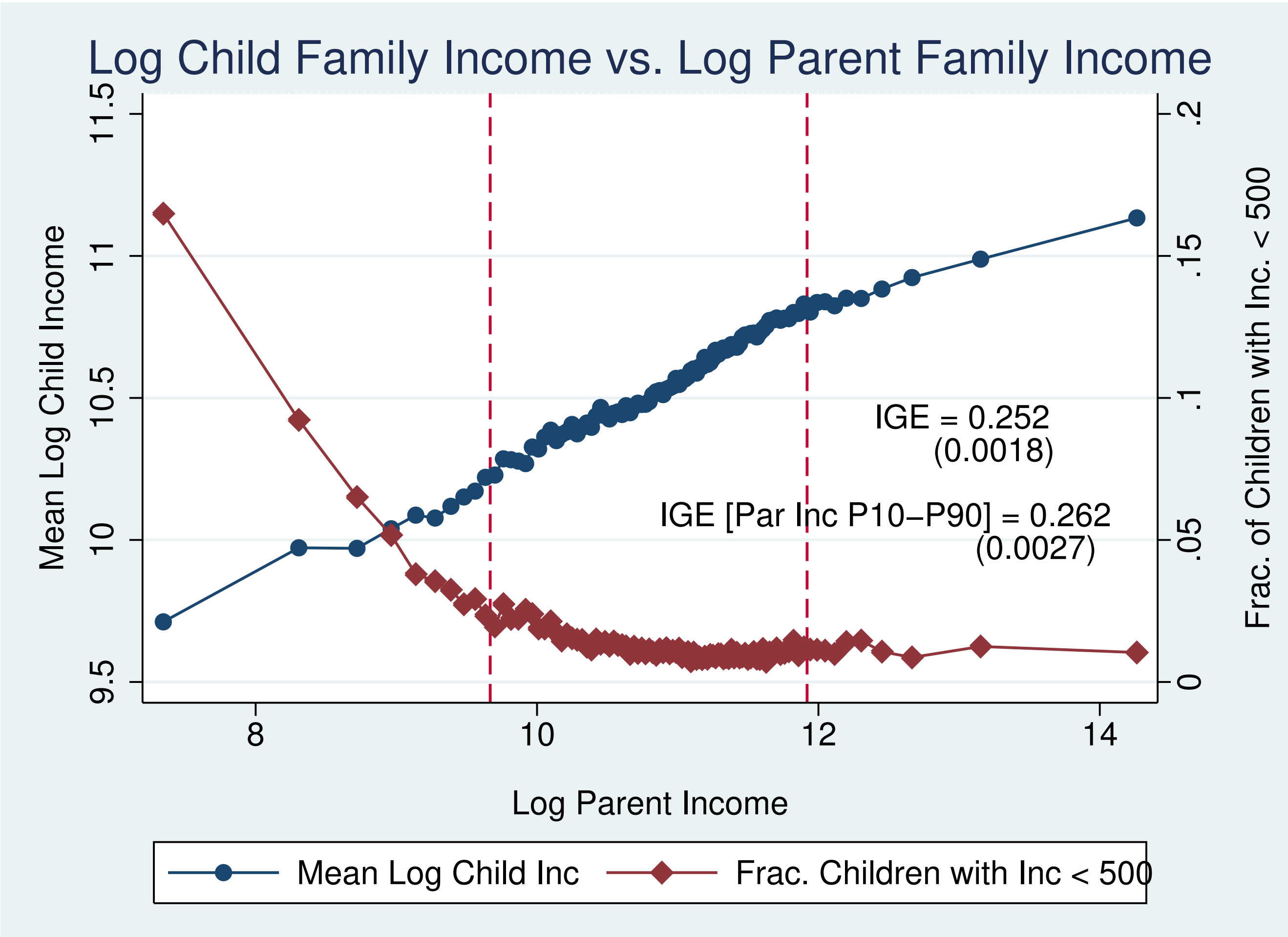

Figure 2: Relation between Children's and Parents' Incomes in Logs

Source: Authors' calculations based on the IID

Note: This figure presents nonparametric binned scatterplots, where each dot represents one rank of the Canadian parental income distribution. The mean log of parental household income is on the x-axis, the mean log of children household income is on the left y-axis, and the fraction of children with income under \$500 on the right y-axis. All dollar amounts are in 2012 US\$, corrected for inflation in Canadian dollars and converted using the PPP for 2012. The two vertical dashed lines correspond to the 10^{th} and 90^{th} percentiles of the parental income distribution. We report OLS regression coefficients (intergenerational elasticities) where log child income is regressed on log parental income using the IID microdata, separately for all parental income distribution and for parental incomes between the 10^{th} and 90^{th} percentiles. Standard errors are in parentheses.

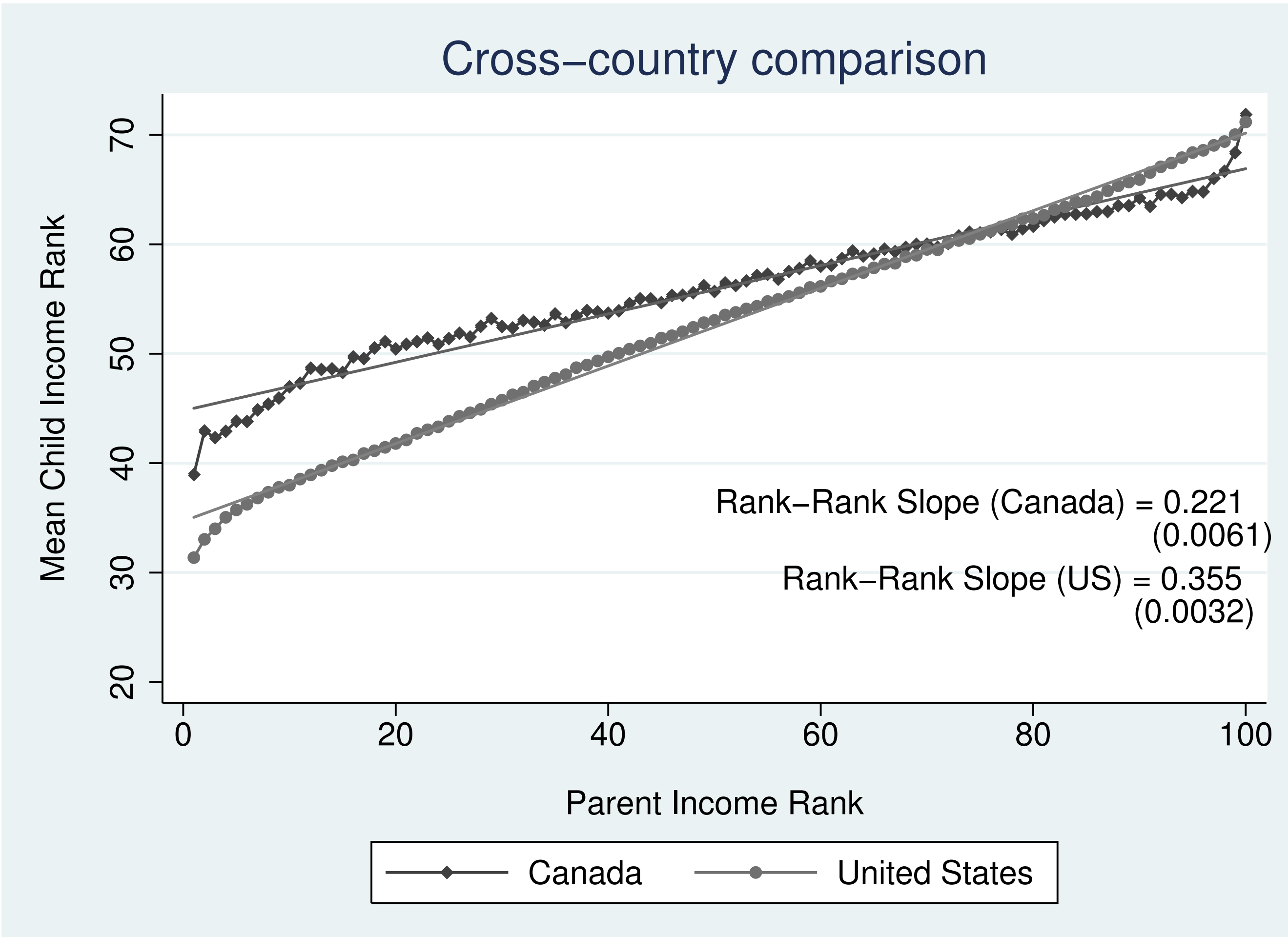

Figure 3: Cross-country Comparison of Rank-Rank Slopes

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figure presents nonparametric binned scatterplots, where each dot represents one rank of the US parental income distribution. The series for the United States comes from Chetty et al. (2014a), the one from Canada is computed using the IID microdata. The parental household income rank is on the x-axis, the mean rank of children household income is on the y-axis. All dollar amounts are in 2012 US$, corrected for inflation in Canadian dollars and converted using the PPP for 2012. The slopes and linear fit lines for Canada are estimated using an OLS regression where the US rank of child income is regressed on US rank of parental income using the IID microdata. Standard errors are in parentheses.

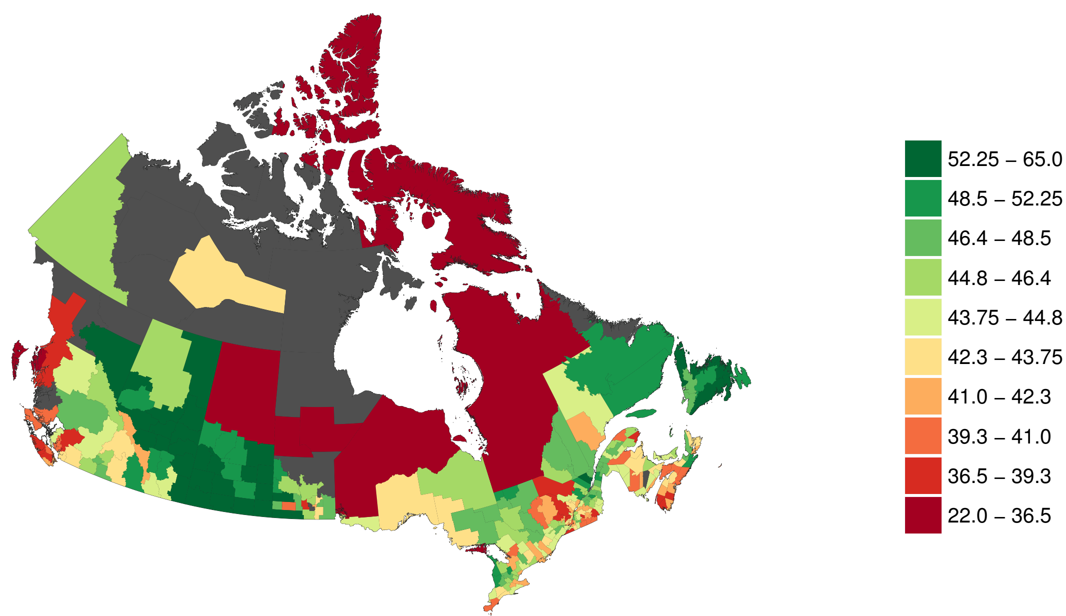

Figure 4: Absolute Upward Mobility: Mean Child Rank for Parents at 25^{th} Percentile

Source: Authors' calculations based on the IID

Note: This figure shows a heat map of $\bar{r}_{25}$, an absolute mobility measure, where each observation is a Census division in Canada. Ranks are based on the US distributions. More mobile places are in green; those with lower mobility appear in red. The grey areas represent places where there were not enough observations in the data to release estimates.

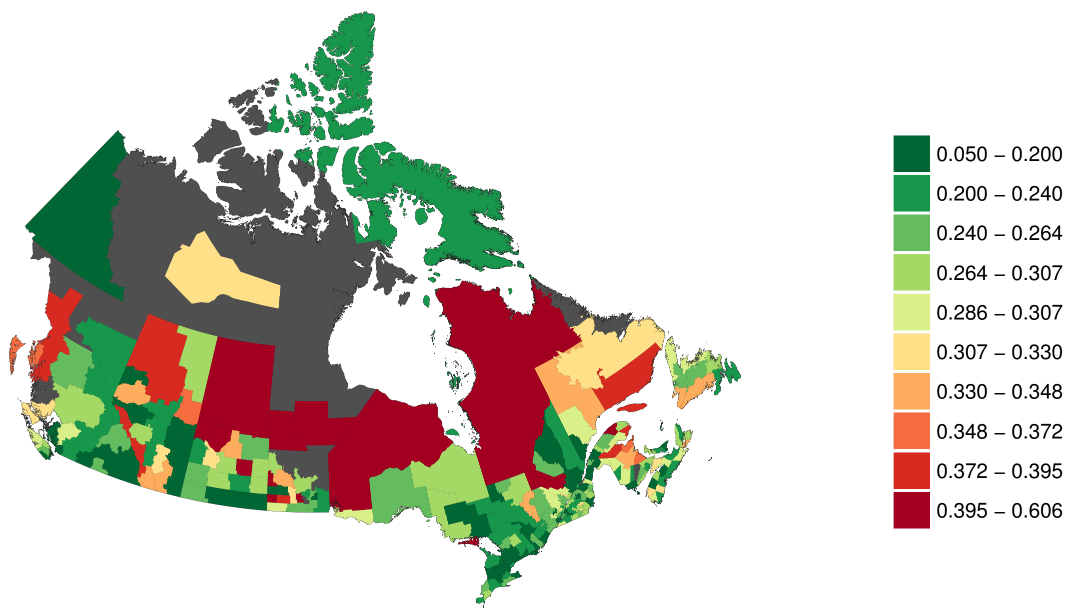

Figure 5: Relative Mobility: Rank-rank Slope

Source: Authors' calculations based on the IID

Note: This figure shows a heat map of the rank-rank slope, a relative mobility measure, where each observation is a Census division in Canada. Ranks are based on the US distributions. More mobile places are in green; those with lower mobility appear in red. The grey areas represent places where there were not enough observations in the data to release estimates.

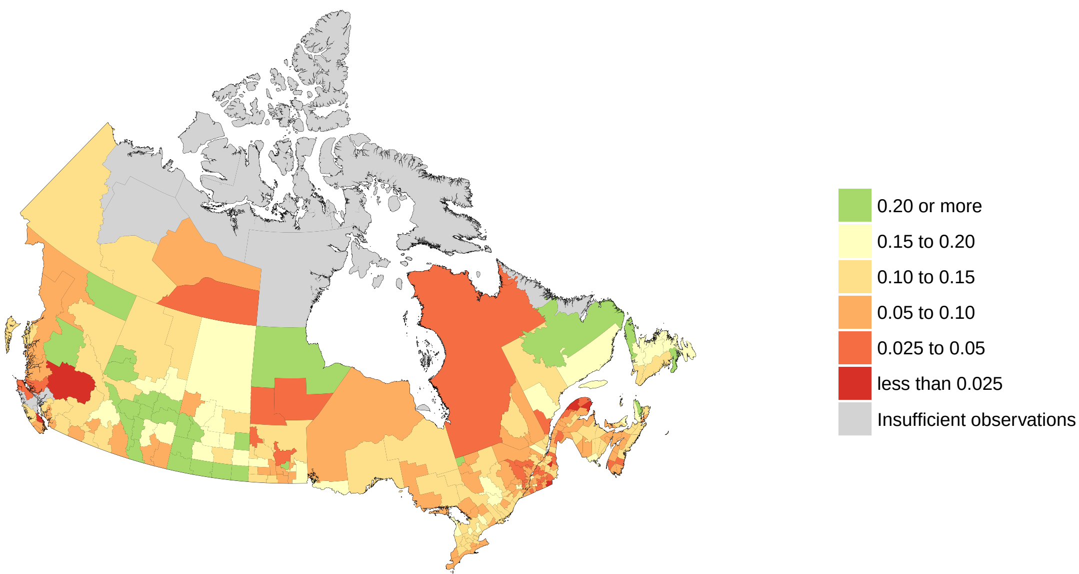

Figure 6: Rags to Riches: P(Child in Q5|Parents in Q1)

Source: Authors' calculations based on the IID
Note: This figure shows a heat map of $P_{1,5}$, the probability that a child from parents in the bottom quintile of the parental income distribution reaches the top quintile of the children's income distribution, where each observation is a Census division in Canada. Ranks are based on the US distributions. More mobile places are in green; those with lower mobility appear in red. The grey areas represent places where there were not enough observations in the data to release estimates.

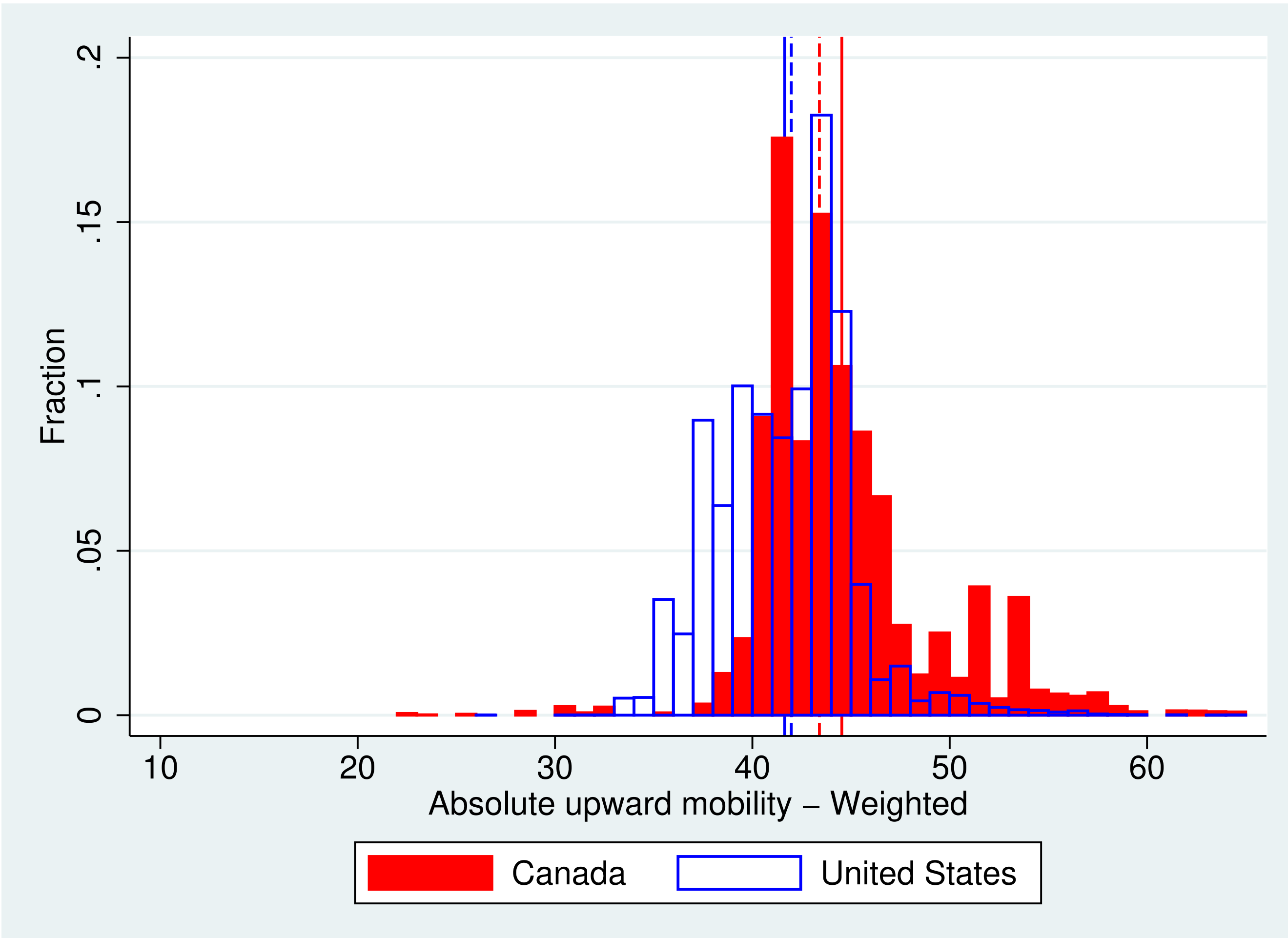

Figure 7: Histogram of Absolute Mobility by Country

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figure shows the histogram of $\bar{r}_{25}$, an absolute mobility measure, where each observation is a Census division in Canada or a Commuting Zone in the United States. The fractions on the y-axis are weighted using population counts for the 2001 Census in Canada and 2000 Census in the United States. The solid/dashed vertical lines represent the weigthed mean/median of the measure on the x-axis for Canada (in red) and the United States (in blue).

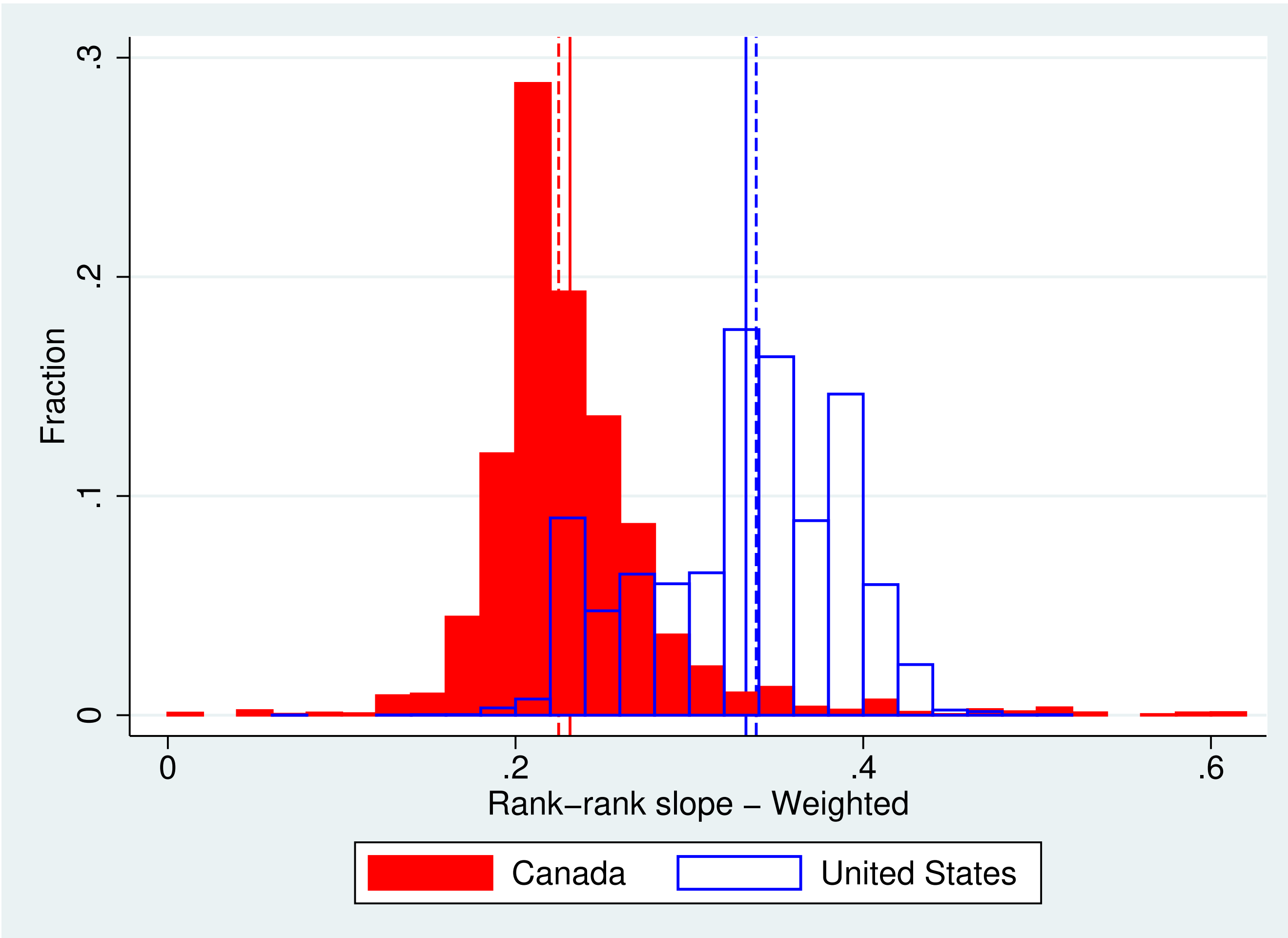

Figure 8: Histogram of Relative Mobility by Country

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figure shows the histogram of the rank-rank slope estimates, a relative mobility measure, where each observation is a Census division in Canada or a Commuting Zone in the United States. The fractions on the y-axis are weighted using population counts for the 2001 Census in Canada and 2000 Census in the United States. The solid/dashed vertical lines represent the weighted mean/median of the measure on the x-axis for Canada (in red) and the United States (in blue).

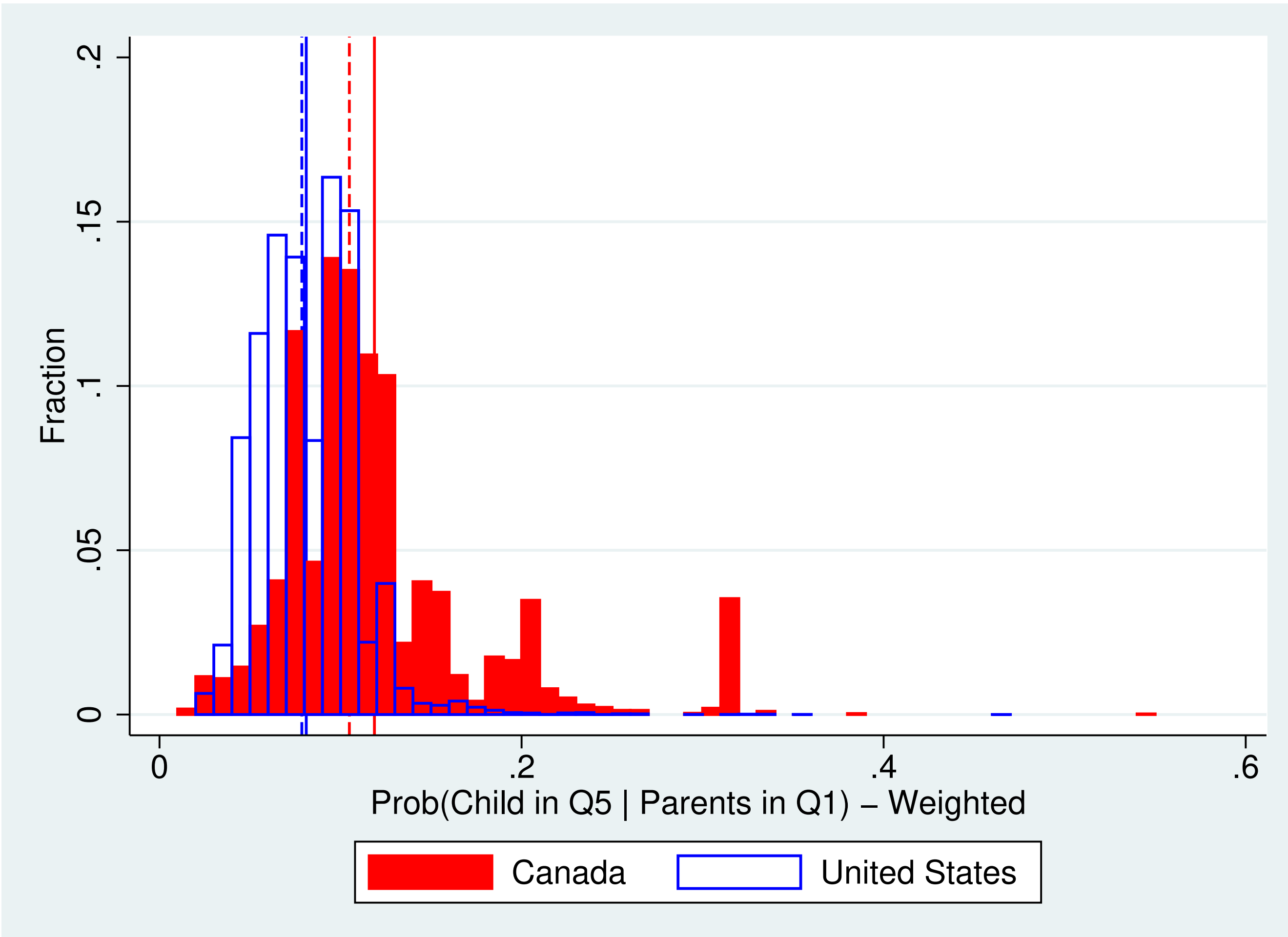

Figure 9: Histogram of P(Child in Q5|Parents in Q1) by Country

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figure shows the histogram of P(Child in Q5|Parents in Q1), with the quantiles based on the US income distributions, where each observation is a Census division in Canada or a Commuting Zone in the United States. The fractions on the y-axis are weighted using population counts for the 2001 Census in Canada and 2000 Census in the United States. The solid/dashed vertical lines represent the weighted mean/median of the measure on the x-axis for Canada (in red) and the United States (in blue).

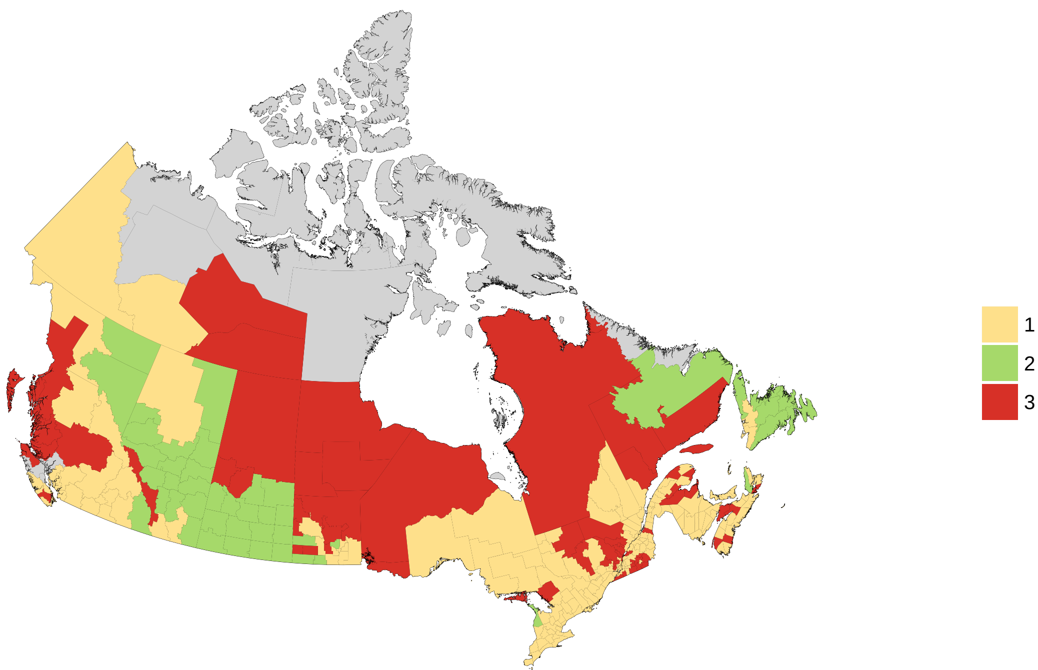

Figure 10: Canadian Clusters

Source: Authors' calculations based on the IID

Note: This figures presents the classification of Canadian Census divisions into 3 clusters based on the K-means algorithm using $K = 3$ and the following six measures: relative mobility as measured by the rank-rank slope, absolute mobility as measured by the predicted rank at the 25$^{\text{th}}$ percentile of the parental income distribution ($\bar{r}_{25,c}$), the mean and the median of the parental income, the probability that a child with parents from the bottom quintile of the income distribution reaches the top quintile, and the probability that he or she stays at the bottom quintile (P(child in Q5 | parents in Q1) and P(child in Q1 | parents in Q1), respectively).

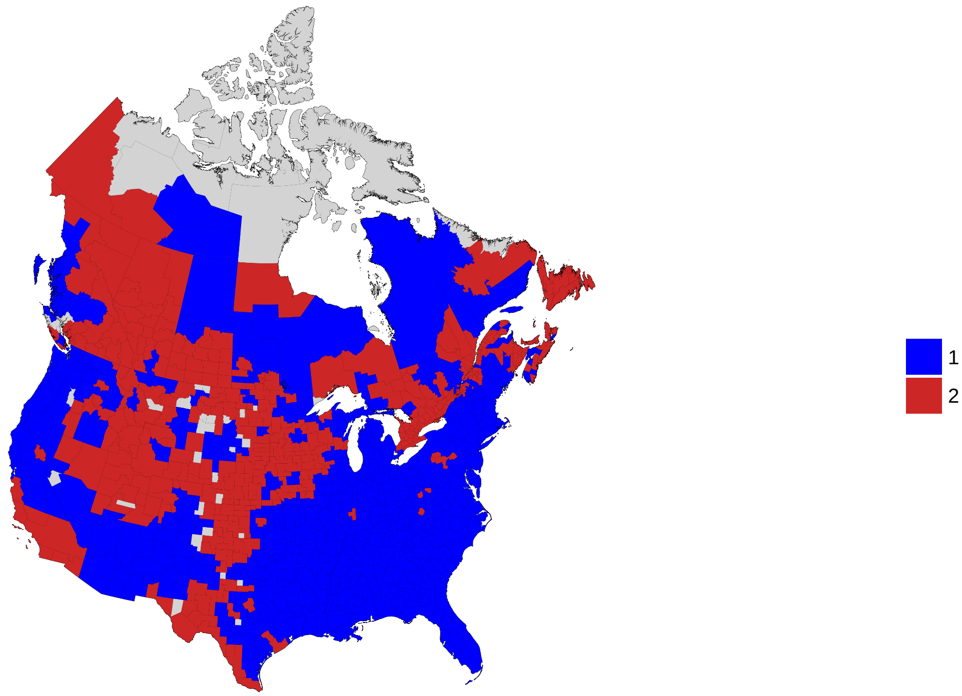

Figure 11: North American Clusters, $K = 2$

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figures presents the classification of Canadian Census divisions and American Commuting Zones into 2 clusters based on the K-means algorithm using $K = 2$ and the following six measures: relative mobility as measured by the rank-rank slope, absolute mobility as measured by the predicted rank at the 25$^{\text{th}}$ percentile of the parental income distribution ($\bar{r}_{25,c}$), the mean and the median of the parental income, the probability that a child with parents from the bottom quintile of the income distribution reaches the top quintile, and the probability that he or she stays at the bottom quintile (P(child in Q5 | parents in Q1) and P(child in Q1 | parents in Q1), respectively).

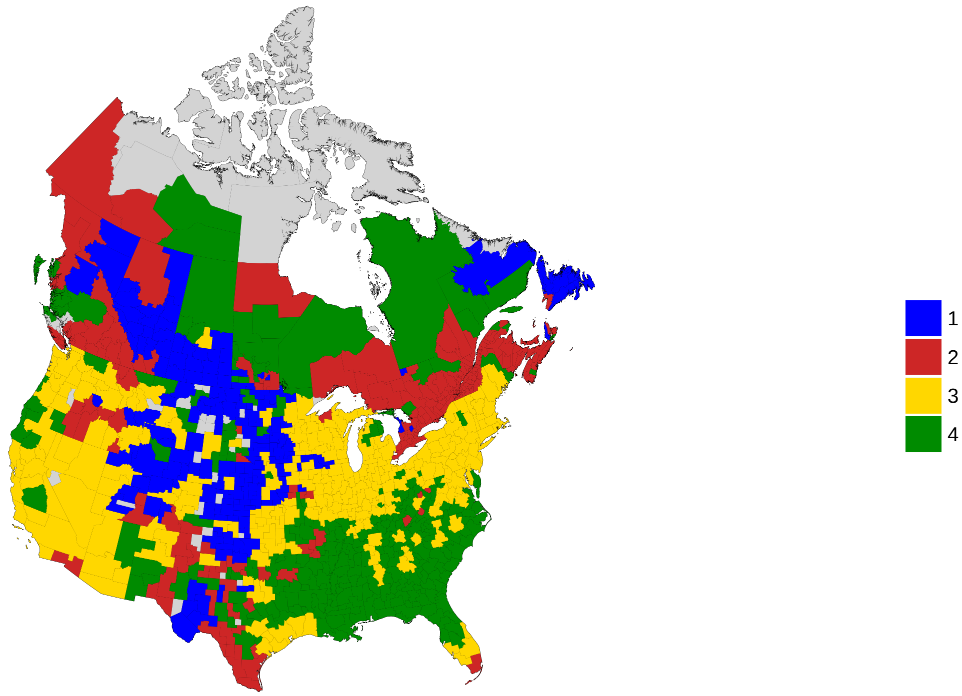

Figure 12: North American Clusters, $K = 4$

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figures presents the classification of Canadian Census divisions and American Commuting Zones into 4 clusters based on the K-means algorithm using $K = 4$ and the following six measures: relative mobility as measured by the rank-rank slope, absolute mobility as measured by the predicted rank at the 25$^{\text{th}}$ percentile of the parental income distribution ($\bar{r}_{25,c}$), the mean and the median of the parental income, the probability that a child with parents from the bottom quintile of the income distribution reaches the top quintile, and the probability that he or she stays at the bottom quintile (P(child in Q5 | parents in Q1) and P(child in Q1 | parents in Q1), respectively).

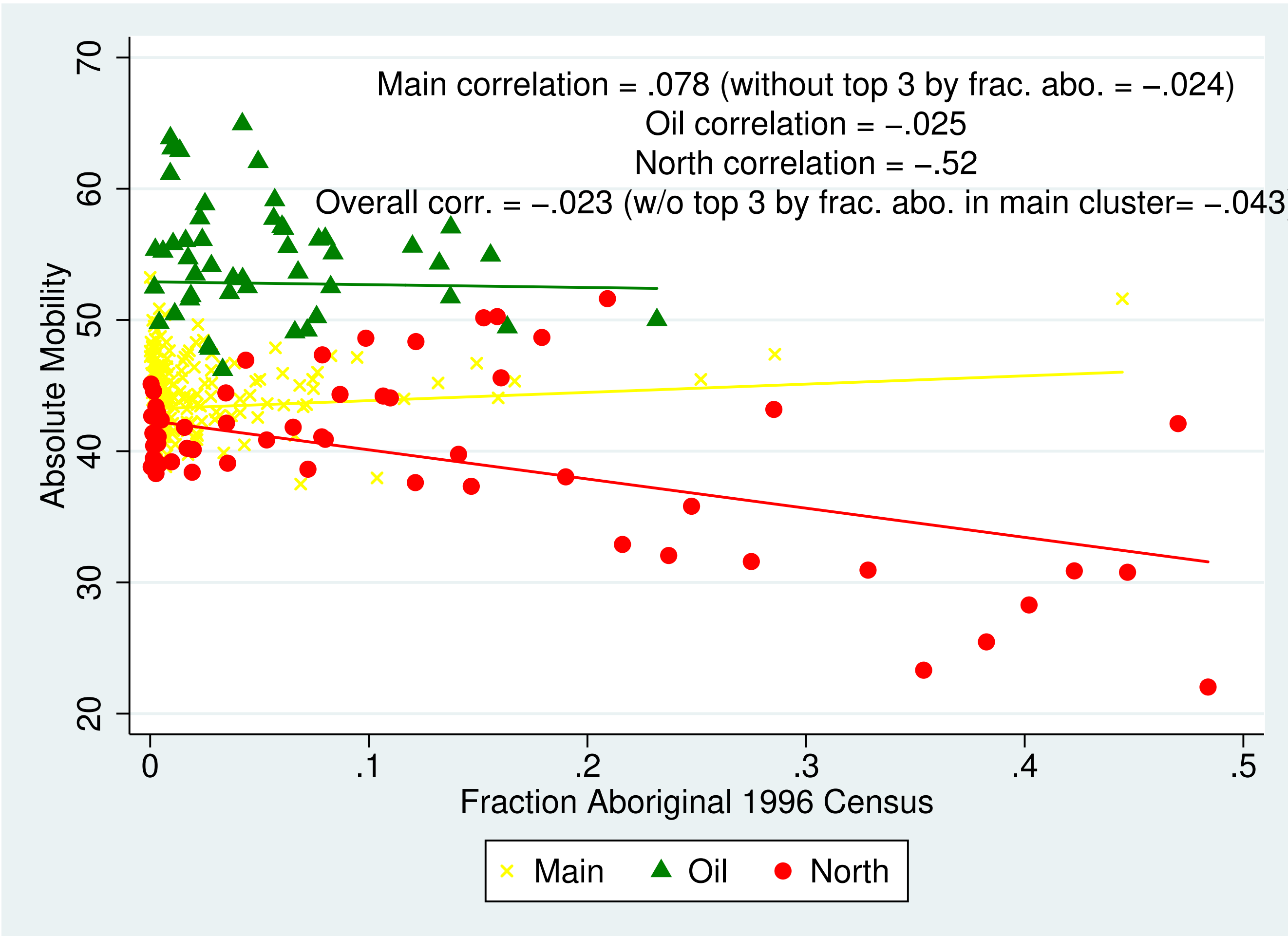

Figure 13: Correlation with Fraction Aboriginal

Source: Authors' calculations based on the IID and on the 1996 Census of Population
Note: This figures presents correlations between absolute mobility at the Census division level ($\bar{r}_{25,c}$) and the fraction of aboriginal residents. Each dot represents one Census division. The lines are the best linear fits by cluster. The correlations and the best fit lines are computed using the 2001 Census population as weights. The three clusters are based on the K-means algorithm using $K = 3$ on all Canadian Census divisions, as described in the text. Correlations are also recomputed excluding the top three Census divisions in terms of fraction aboriginal in Cluster 1, since results are sensitive to their inclusion.

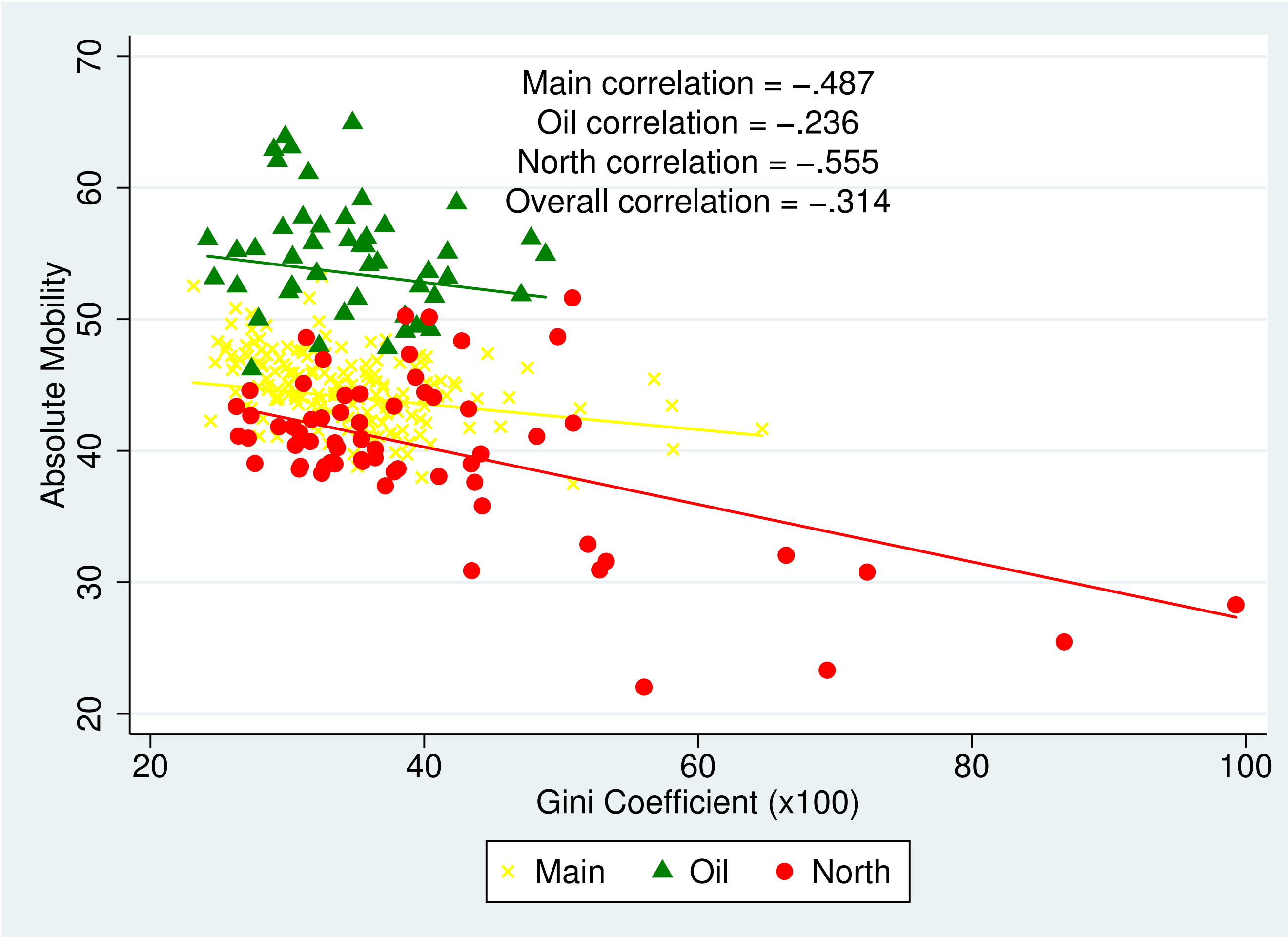

Figure 14: Correlation with Gini Coefficient: The Great Gatsby Curve

Source: Authors' calculations based on the IID
Note: This figures presents correlations between absolute mobility at the Census division level ($\bar{r}_{25,c}$) and the Gini Coefficient, as computed using the core sample of the IID. Each dot represents one Census division. The lines are the best linear fits by cluster. The correlations and the best fit lines are computed using the 2001 Census population as weights. The three clusters are based on the K-means algorithm using $K = 3$ on all Canadian Census divisions, as described in the text.

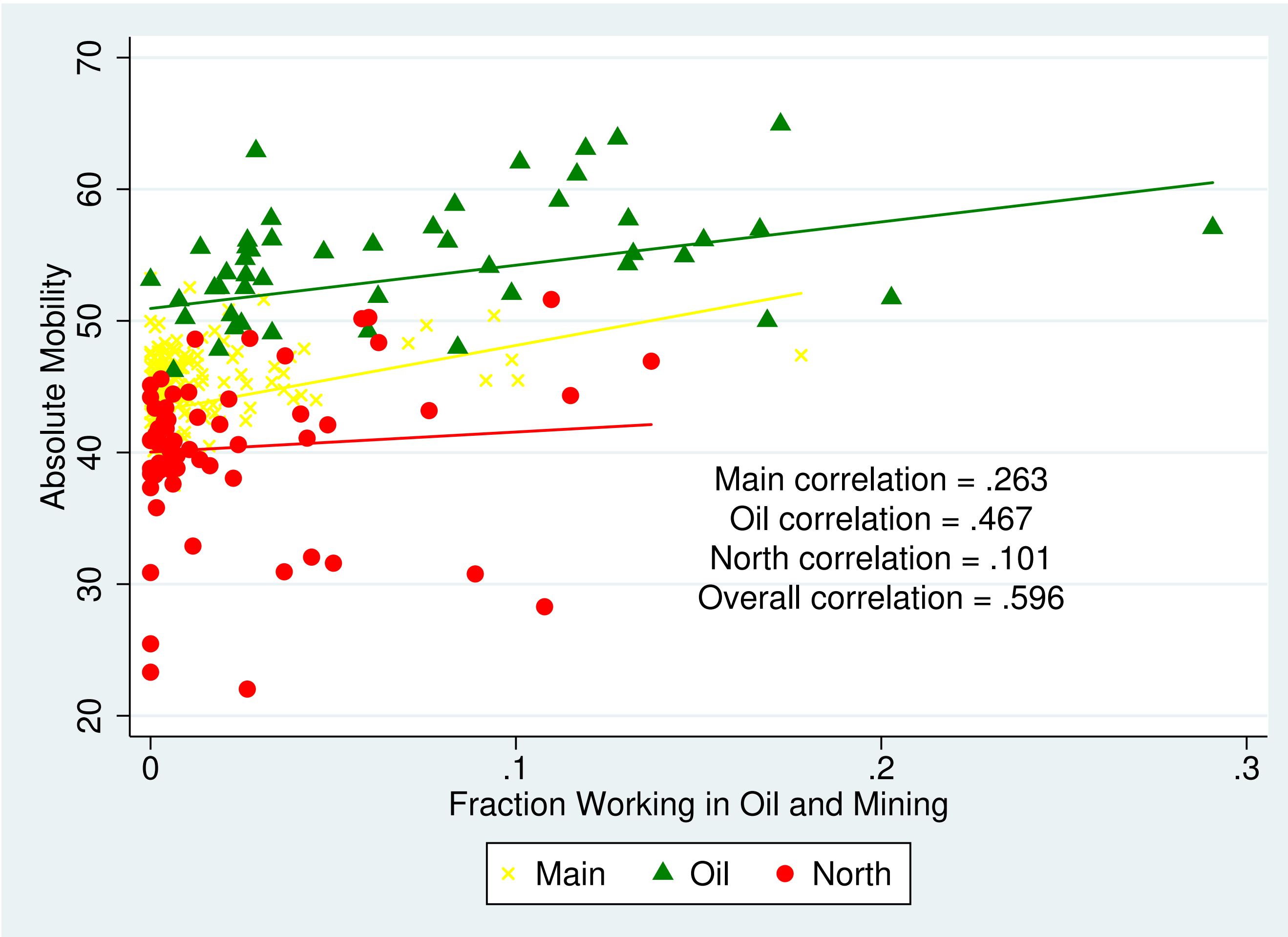

Figure 15: Correlation with Fraction of Workers in Oil and Mining Industries

Source: Authors' calculations based on the IID and on the 2011 National Household Survey Census division profiles
Note: This figures presents correlations between absolute mobility at the Census division level ($\bar{r}_{25,c}$) and the fraction of workers with jobs in the oil and mining industries, as computed using the 2011 National Household Survey from Statistics Canada and looking at the fraction of the total labor force population aged 15 years and over working in the mining, quarrying, and oil and gas extraction (NAICS 2007 2-digit code 21). Each dot represents one Census division. The lines are the best linear fits by cluster. The correlations and the best fit lines are computed using the 2001 Census population as weights. The three clusters are based on the K-means algorithm using $K = 3$ on all Canadian Census divisions, as described in the text.

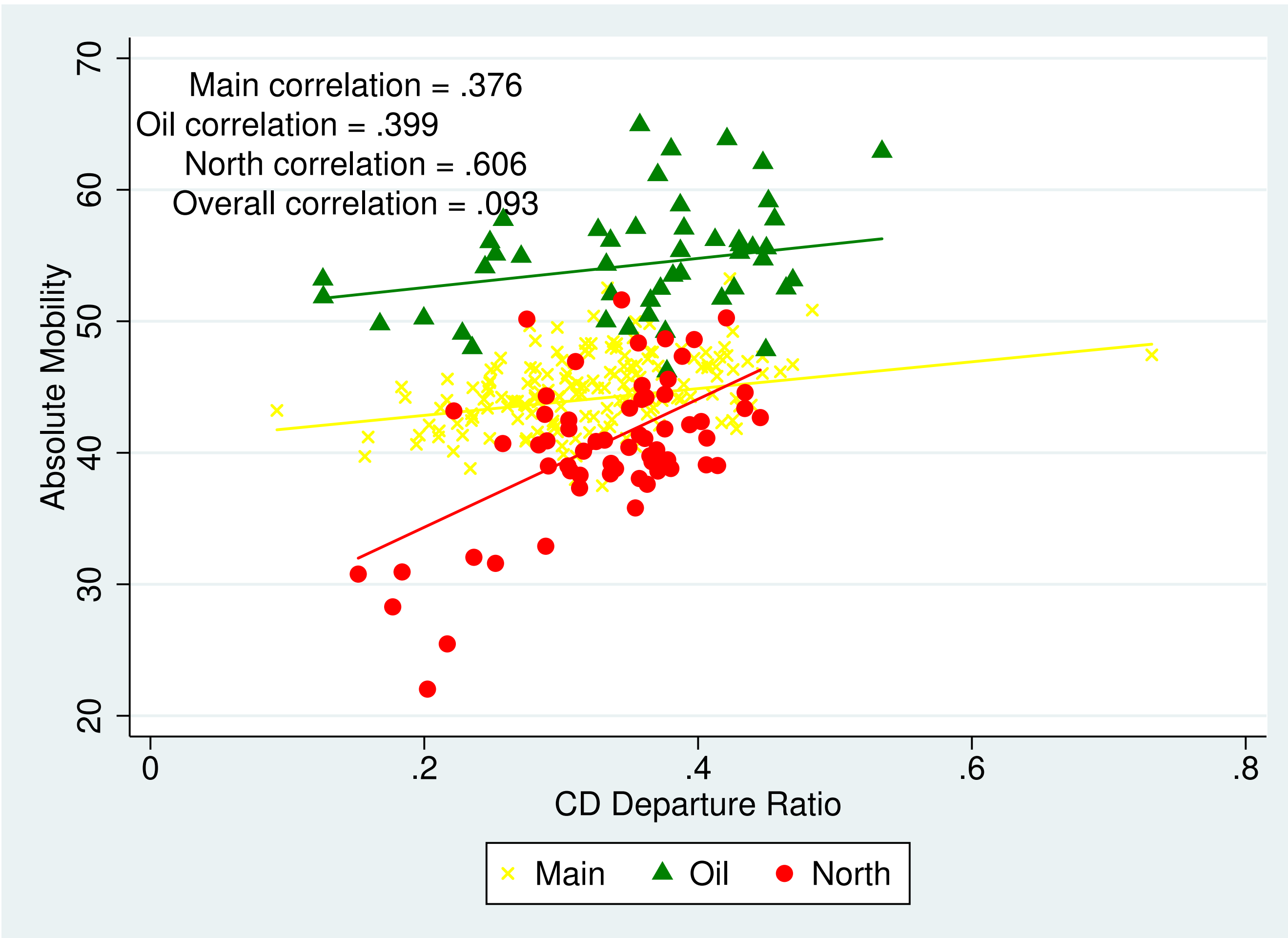

Figure 16: Correlation with CD Departure Ratio

Source: Authors' calculations based on the IID
Note: This figures presents correlations between absolute mobility at the Census division level ($\bar{r}_{25,c}$) and the Census division's departure ratio, as computed by taking the ratio of the children from our cohort who have moved out of their origin Census division by 2011-2012 to the total number of children living in that Census division at the time of the match, using the core sample of the IID. Each dot represents one Census division. The lines are the best linear fits by cluster. The correlations and the best fit lines are computed using the 2001 Census population as weights. The three clusters are based on the K-means algorithm using $K = 3$ on all Canadian Census divisions, as described in the text.

6 Tables

Table 1: Percentiles of the Canadian and US Income Distributions

	Parents		Children	
Percentile	Canada	United States	Canada	United States
1	1,810	1,700	960	-43,800
5	9,070	9,200	5,730	0
10	13,450	15,000	9,650	2,300
20	22,590	24,900	17,860	11,000
50	52,890	59,500	39,980	34,600
80	89,750	107,900	80,140	74,400
90	113,730	144,500	108,270	99,900
95	140,000	194,300	132,000	125,300
99	246,760	420,100	183,920	193,300

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This table presents mean family income by percentiles of the Canadian and American income distributions for parents and children born 1980-1982. All income figures are in 2012 US dollars, converted using the PPP.

Table 2: Intergenerational Mobility Estimates at the National Level

Child's outcome	Parental income definition	(1) Chetty et al. (2014a)	(2) Core sample	(3) Male children	(4) Female children	(5) Married parents	(6) Single parents	(7) Fixed age at child birth, Can.	(8) Fixed age at child birth, US	(9) Live in same province	(10) Live in different province	(11) File in English	(12) File in French
1. Log family income	Log family	0.344	0.252	0.245	0.260	0.235	0.234	0.254	0.261	0.262	0.147	0.25	0.236
(excluding <$500)	income	(0.0004)	(0.0018)	(0.0024)	(0.0028)	(0.0023)	(0.0045)	(0.0035)	(0.0035)	(0.002)	(0.0058)	(0.0021)	(0.0037)
2. Log family income	Log family	–	0.300	0.312	0.287	0.289	0.296	0.291	0.301	0.301	0.154	0.300	0.273
(recoding <$500 to $500)	income		(0.0021)	(0.003)	(0.0031)	(0.0027)	(0.0053)	(0.0041)	(0.0041)	(0.0023)	(0.0058)	(0.0025)	(0.0041)
3. Log family income	Log family	0.618	0.381	0.416	0.343	0.372	0.390	0.357	0.370	0.369	0.162	0.384	0.297
(recoding zeros to $1)	income	(0.0009)	(0.0038)	(0.0057)	(0.0049)	(0.0049)	(0.0091)	(0.0068)	(0.007)	(0.0039)	(0.0071)	(0.0044)	(0.0059)
4. Log family income	Log family	0.413	0.289	0.298	0.278	0.277	0.283	0.399	0.288	0.281	0.291	0.300	0.273
(recoding <$1K to $1K)	income	(0.0004)	(0.0020)	(0.0028)	(0.0029)	(0.0025)	(0.0049)	(0.0109)	(0.0025)	(0.0038)	(0.0038)	(0.0025)	(0.0041)
5. Family income	Family	0.341	0.220	0.220	0.221	0.201	0.234	0.224	0.232	0.229	0.128	0.214	0.207
US rank	income US rank	(0.0003)	(0.0013)	(0.0018)	(0.002)	(0.0016)	(0.0043)	(0.0026)	(0.0026)	(0.0015)	(0.0047)	(0.0016)	(0.0026)
6. Family income	Family inc. US	0.339	0.212	0.215	0.207	0.197	0.233	0.212	0.221	0.218	0.127	0.207	0.220
US rank	rank 1999-2003	(0.0003)	(0.0012)	(0.0017)	(0.0018)	(0.0015)	(0.004)	(0.0024)	(0.0024)	(0.0014)	(0.0042)	(0.0014)	(0.0027)
7. Family income	Top parental	0.312	0.191	0.196	0.185	0.168	0.187	0.191	0.198	0.197	0.110	0.186	0.204
US rank	income US rank	(0.0003)	(0.0012)	(0.0017)	(0.0018)	(0.0014)	(0.0033)	(0.0024)	(0.0024)	(0.0014)	(0.0042)	(0.0014)	(0.0027)
8. Individual income	Family	0.287	0.250	0.258	0.249	0.238	0.281	0.251	0.255	0.258	0.167	0.242	0.271
US rank	income US rank	(0.0003)	(0.0015)	(0.0022)	(0.0021)	(0.0018)	(0.0049)	(0.003)	(0.003)	(0.0017)	(0.0056)	(0.0018)	(0.0031)
9. Individual earnings	Family	0.282	0.201	0.220	0.191	0.189	0.219	0.202	0.206	0.207	0.117	0.19	0.217
rank (excluding < $500)	income US rank	(0.0003)	(0.0017)	(0.0023)	(0.0023)	(0.002)	(0.0054)	(0.0032)	(0.0032)	(0.0018)	(0.006)	(0.0019)	(0.0034)
Number of observations		9,867,736	543,780	284,700	259,080	427,340	90,060	142,200	142,975	441,845	44,625	404,305	118,605

Source: Authors' calculations based on the IID. Column (1) is from Chetty et al. (2014a), Table I, Column (1)

Note: Each pair of figures corresponds to a coefficient estimate from a separate univariate OLS regression and its standard error in parentheses, where the child's outcome is regressed on the parental income as defined. Each column from (2) to (12) refers to a different sample of the IID, as described in the header. In columns (7) and (8), the sample is restricted to children whose both parents' age fall within a five-year window of the median age at child birth, based on age at child birth in Canada and the United States, respectively. In columns (9) and (10), the sample is respectively restricted to children who live in the same province or a different province in 2011 and 2012 than they did at the year of the link with their parents. In columns (11) and (12), the sample is split into whether the child filed his or her tax return in English or in French. The ranks are the US ranks assigned to Canadian children and their parents based on the marginal distribution figures in Chetty et al. (2014a)'s Online Data Table 2. Dollar figures are in 2012 US dollars.

Table 3: National Quintile Transition Matrix based on Canadian Quintiles

	Parent quintile				
Child quintile	1	2	3	4	5
1	30.00	20.90	17.30	14.80	13.50
2	23.00	23.00	20.70	17.80	14.40
3	18.30	21.40	21.90	21.00	17.70
4	15.10	18.80	21.40	23.40	22.70
5	13.60	15.90	18.60	23.10	31.70

Source: Authors' calculations based on the IID
Note: Each cell reports the percentage of children in the quintile of the children's Canadian income distribution referred to in the row header, for children whose parents are in the quintile of the parental Canadian income distribution referred to in the column header. By construction, each column sums to 100%.

Table 4: National Quintile Transition Matrix Based on US Quintiles

	Parent US quintile				
Child US quintile	1	2	3	4	5
1	17.00	9.20	7.20	6.10	5.80
2	21.90	17.50	14.50	12.10	10.30
3	22.10	23.20	21.50	19.20	16.80
4	23.10	28.20	30.00	29.50	27.60
5	15.90	21.90	26.80	33.00	39.40

Source: Authors' calculations based on the IID
Note: Each cell reports the percentage of children in the quintile of the children's US income distribution referred to in the row header, for children whose parents are in the quintile of the paren tal US income distribution referred to in the column header. In this transition matrix, each column does not have to sum to 100% since the US quintiles are assigned to Canadians based on their income levels.

Table 5: Intergenerational Mobility in the 50 Largest Census Divisions

Rank	Census division	CD code	Largest city	Absolute mobility $\bar{r}_{25}$ (1)	$P_{1,5}$ (2)	Rank-rank slope (3)	Population (4)
1	Division No. 8	4808	Red Deer	54.12	22.17	0.226	150,430
2	Division No. 11	4811	Edmonton	53.19	20.41	0.218	964,145
3	Division No. 6	4806	Calgary	51.84	31.23	0.204	1,012,305
4	Division No. 6	4706	Regina	50.23	14.97	0.242	216,160
5	Division No. 1	1001	St. John's	49.79	18.15	0.218	240,245
6	Division No. 11	4711	Saskatoon	49.07	19.35	0.274	234,145
7	Sudbury	3553	Sudbury	47.97	15.50	0.263	153,560
8	Wellington	3523	Guelph	46.49	13.08	0.161	184,840
9	Le Fjord-du-Saguenay	2494	Saguenay	46.44	8.81	0.218	164,810
10	Central Okanagan	5935	Kelowna	46.40	12.94	0.189	145,950
11	York	3519	Markham	46.31	15.31	0.194	725,665
12	La Vallee-du-Richelieu	2457	Chambly	46.03	8.95	0.19	118,630
13	Halton	3524	Oakville	45.66	11.42	0.176	372,410
14	L'Outaouais	2481	Gatineau	45.61	9.33	0.256	224,755
15	Quebec	2423	Quebec	45.01	7.77	0.225	501,845
16	Waterloo	3530	Waterloo	45.01	12.22	0.203	433,875
17	Peel	3521	Mississauga	44.95	12.30	0.183	985,565
18	Thompson-Nicola	5933	Kamloops	44.76	10.71	0.225	118,665
19	Division No. 2	4802	Lethbridge	44.35	9.48	0.336	132,110
20	Ottawa-Carleton	3506	Ottawa	44.21	12.53	0.213	763,790
21	Lambton	3538	Sarnia	44.17	9.18	0.259	125,560
22	Durham	3518	Oshawa	43.99	14.07	0.197	502,905
23	Peterborough	3515	Peterborough	43.55	11.22	0.222	123,605
24	Middlesex	3539	London	43.40	11.66	0.207	398,560
25	Les Moulins	2464	Terrebonne	43.39	10.00	0.209	109,415
26	Thunder Bay	3558	Thunder Bay	43.38	11.46	0.264	149,150
27	Laval	2465	Laval	43.30	11.23	0.218	339,005
28	Greater Vancouver	5915	Vancouver	43.22	9.28	0.203	1,967,475
29	Cape Breton	1217	Cape Breton	42.98	11.48	0.288	107,880
30	Westmorland	1307	Westmorland	42.97	6.98	0.281	122,405
31	Fraser Valley	5909	Abbotsford	42.69	11.50	0.21	233,850
32	Algoma	3557	Sault Ste. Marie	42.57	8.24	0.268	117,200
33	Therse-De Blainville	2473	Blainville	42.46	8.68	0.248	129,110
34	Simcoe	3543	Barrie	42.22	12.46	0.201	372,325
35	Hastings	3512	Belleville	42.16	11.43	0.234	124,420
36	Hamilton-Wentworth	3525	Hamilton	42.12	10.76	0.231	484,385
37	Roussillon	2467	Saint-Constant	41.97	6.98	0.24	137,195
38	Champlain	2458	Longueuil	41.74	9.03	0.226	308,955
39	Toronto Metropolitan	3520	Toronto	41.67	10.42	0.225	2,456,805
40	Francheville	2437	Trois-Rivieres	41.59	7.25	0.284	135,535
41	Frontenac	3510	Kingston	41.58	10.68	0.252	135,410
42	Niagara	3526	St. Catharines	41.34	9.07	0.19	404,590
43	Brant	3529	Brant	41.34	8.94	0.234	116,755
44	Division No. 11	4611	Winnipeg	41.20	11.49	0.262	612,165
45	Capital	5917	Victoria	41.20	5.77	0.202	320,710
46	Sherbrooke	2443	Sherbrooke	41.06	6.51	0.293	137,940
47	Halifax	1209	Halifax	40.64	10.49	0.238	355,945
48	Montreal	2466	Montreal	40.11	7.89	0.249	1,782,830
49	Nanaimo	5921	Nanaimo	39.87	2.12	0.273	125,550
50	Essex	3537	Windsor	39.72	7.48	0.241	371,085

Source: Authors' calculations based on the IID

Note: This table presents mobility measures for the 50 largest Census divisions in Canada by population, based on the 2001 Canadian Census of Population. The Census divisions are ranked in descending order according to their absolute upward mobility ($\bar{r}_{25}$, column (3)). $P_{1,5}$ in column (2) is expressed in percentage. See text for details on the mobility measures and sample. The population in column (4) refers to the population of the Census division from the 2001 Canadian Census of Population.

Table 6: Average Characteristics by Cluster, Canada

Cluster ID (k)	Number of CDs	Population	Absolute mobility	Rank-rank slope	$P_{1,5}$	$P_{1,1}$	Parental income Mean	Parental income Median
1–Main	174	23,298,565	43.32	0.222	10.35%	19.73%	49,914	39,431
2–Oil	48	4,396,540	52.82	0.230	21.85%	14.92%	69,394	55,978
3–North	65	1,939,160	40.39	0.344	7.60%	30.20%	40,727	31,664

Source: Authors' calculations based on the IID
Note: This table presents average characteristics by cluster, as generated by the K-means algorithm described in the text using $K = 3$ on the Canadian Census divisions. The average figures are computed by taking means at the Census division level of the variable indicated, and computing a weighted average where the weight is the population of the Census division from the 2001 Canadian Census of Population. Total population and number of CD per cluster are not weighted. Dollar figures are in 2012 US dollars.

Table 7: Average Characteristics by Cluster, North America

Cluster ID (k)	Number of CDs/CZs	Population	Absolute mobility	Rank-rank slope	$P_{1,5}$	$P_{1,1}$	Parental income Mean	Parental income Median
Panel A: $K = 2$								
1	562	242,859,858	40.92	0.347	7.53%	33.95%	88,770	61,291
2	423	67,824,113	45.45	0.236	11.78%	23.79%	72,747	52,241
Panel B: $K = 4$								
1	170	9,449,319	52.54	0.243	19.24%	18.22%	69,467	56,750
2	256	33,005,972	43.17	0.236	9.73%	21.19%	53,006	38,807
3	276	212,168,698	42.02	0.328	8.44%	33.13%	96,891	66,469
4	283	56,059,982	38.94	0.369	5.98%	34.93%	62,959	44,749

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This table presents average characteristics by cluster, as generated by the K-means algorithm described in the text using $K = 2$ (Panel A) or $K = 4$ (Panel B) on the Canadian Census divisions and American Commuting Zones. The average figures are computed by taking means by cluster at the Census division or Commuting Zone level of the variable indicated, and computing a weighted average where the weight is the population of the Census division or Commuting Zone from the 2001 Canadian Census of Population or the 2000 American Census. Total population and number of CD/CZ per cluster are not weighted. Dollar figures are in 2012 US dollars.

Table 8: Correlations between Absolute Mobility and Various CD Characteristics

	All Census divisions		Cluster 1 Main		Cluster 2 Oil		Cluster 3 North	
Variable	corr.	SE	corr.	SE	corr.	SE	corr.	SE
Fraction of single mothers	-.4844	.0522	-.6879	.0548	-.6428	.1129	-.2291	.1257
Divorce rate	-.2988	.057	-.4096	.0693	-.2204	.1438	.1103	.1283
Fraction married	.3003	.0569	.2731	.0731	.3536	.1379	.3872	.119
HS dropout rate, parents	.0507	.0592	.0994	.0759	.3907	.1357	-.1935	.1236
College graduation rate	-.2156	.0583	-.3487	.0712	-.4957	.1281	.0752	.1287
HS dropout rate, children	.1008	.0591	-.0259	.0758	.4691	.1302	-.5267	.1083
Fraction black	-.3019	.0565	-.3897	.0702	-.199	.1445	.0021	.126
Segregation (black)	.2491	.0574	.4079	.0696	.4574	.1311	.0887	.1255
Fraction foreign born	-.2238	.0582	-.2862	.0728	-.1327	.1461	.0103	.1291
Migration inflow (last year)	.1746	.0588	.2367	.0739	.3809	.1363	-.0458	.129
Migration inflow (5 years)	.1296	.0592	.2655	.0733	.3651	.1373	.021	.1291
CD departure rate	.0929	.0593	.3734	.0705	.3992	.1352	.5984	.1014
CD arrival rate	.1649	.0591	.0917	.0766	.3343	.139	.2309	.1243
Fraction aboriginal	-.0229	.0592	.0777	.076	-.0252	.1474	-.52	.1076
Segregation (aboriginal)	.1767	.0583	.2942	.0729	.307	.1403	.2207	.1229
Number of reserves	-.0232	.0595	-.0188	.0758	.1176	.1464	-.0318	.1272
Manufacturing share	-.3472	.056	-.0106	.0761	-.2233	.1437	.1017	.1284
Oil and mining share, 2011	.5962	.0476	.2625	.0736	.4675	.1303	.1011	.1253
Teenage labor force part. rate	.2984	.0565	.1913	.0748	.4683	.1303	.0207	.126
Fraction with commute under 15km	.0085	.0592	.375	.0707	.1004	.1467	.1075	.1253
Gini coefficient	-.3145	.0562	-.487	.0666	-.2357	.1433	-.5551	.1048

Source: Authors' calculations based on the IID and additional data sources (see subsection 1.2)
Note: Each correlation comes from a univariate regression of standardized values of absolute mobility ($\bar{r}_{25}$) on the characteristic listed in the first column. The first set of regressions is estimated using the full sample of Canadian Census divisions, whereas the following three are done separately by cluster, as assigned using the K-means algorithm with $K = 3$. All figures are weighted using the population of the Census division from the 2001 Canadian Census of Population.

A Data Appendix

A.1 The Intergenerational Income Database

Overview The Intergenerational Income Database (IID) was created by Statistics Canada. It consists of Canadian administrative tax records in which parents-children units are identified and followed over time, allowing the study of the intergenerational transmission of income. The link between children and their parents is done when children are aged 16 to 19 and living at home with one of their parents, following the algorithm described in Corak and Heisz (1999). The original database contains the cohorts of 1982, 1984 and 1986, as referenced by the year of the match, and consists of tax records from 1978 to 1995 for both parents and children, when available. An update to the IID added additional years of tax files, through 2008 (see Chen et al. (2016) for a recent paper using this IID update). More recently, additional cohorts were added to the IID, now allowing researchers to track the evolution of social mobility in Canada. The new cohorts are referred to as the 1991, 1996, and 2001 cohorts, again based on the year of the match between parents and their offspring. The original cohorts covered children born between 1963 and 1970; the new cohorts covers those born 1972 to 1985. Appendix Table A1 shows the various cohorts of the IID along with the years of the match, the birth years of the children covered, the number of observations, both weighted and unweighted, along with the size of the cohort coming from the Census of the Population.

Table A1: Intergenerational Income Database Cohorts

Cohort	Birth years	IID count	IID weighted count	Census population count	Ratio weighted to Census
1982	1963 to 1966	1,183,614	1,517,127	1,898,160	0.799
1984	1965 to 1968	1,124,849	1,517,126	1,696,555	0.894
1986	1967 to 1970	1,155,248	1,517,127	1,603,920	0.946
1991	1972 to 1976	1,102,855	1,484,566	1,510,560	0.983
1996	1977 to 1980	1,166,879	1,558,393	1,570,605	0.992
2001	1982 to 1985	1,350,222	1,634,646	1,596,290	1.024

Source: Authors' calculations based on the IID and Statistics Canada's Census of Population (1986, 1991, 1996 and 2001)
Note: This table shows the unweighted and weighted counts of children by cohort, as well as the year of birth and the population count from the Census. The 1986 Census is used for the 1982, 1984 and 1986 cohorts; otherwise the Census year matches the cohort year. The last column shows the coverage rate, as measured by the ratio of the weighted IID count to the Census population count. The weighted count use the IID weights.

Matching algorithm In the IID, the unit of observation is a parents-children match. Before the introduction of the Canada Child Tax Benefits in 2006, Canadian tax filers did not directly identify their children on their tax return, so a special algorithm had to be designed to link together children and their parents in the tax files. The match is made using the T1 Family File (T1FF), which is a data set of all the T1 records provided by the Canada Revenue Agency to Statistics Canada that has been processed to identify family members. The process involves using the reported spousal Social Insurance Number on the T1 form–which covers both married partners and common-law ones–to identify couples, and looking up names and addresses of tax-filing children. Statistics Canada (2016) contains detailed information on the formation of the T1FF and its evolution over the years. To create the IID for a given cohort, the T1FF from the cohort year is used and all the individuals aged 16 to 19 in that year for which parental information on at least one parent is available are identified as being in the desired cohort. The T1FF for the next year is then used to add the children of the relevant cohort that were not captured in the cohort year, and the process is repeated for the next three years, so that cohort members can be linked to their parents in up to five tax years. Appendix Table A2 shows the distribution of the year of the link, by cohort. In earlier cohorts, 38 to 46% of the children in the IID were matched on the first year. This figure went up to 62% for the 2001 cohort, perhaps as a result of more credits being available to people only when they file their tax reports.

Coverage rate and weights Coverage rate, as measured by the number of children in the database divided to the relevant population estimate from the Census of Population, ranges from 62% to 85%, and from 80% to 102% once the IID weights are used, as seen in Appendix Table A1. For a match to occur, a child needs to be filing a tax return for at least one of the years during which the matching process described above is done, as well as using the same address as at least one of his or her tax-filing parents. Hence while the universe of all tax filers is available, not all children born in the relevant years for the cohort are part of the IID. To correct for the undercoverage of the Canadian population age cohorts, Statistics Canada developed sets of weights to be used to compute estimates representative of the population when using the IID. Cook and Demnati (2000) and Statistics Canada (undated) explain the creation of the weights, respectively for the earlier (1982, 1984, 1986) and later (1991, 1996, 2001) cohorts. For the earlier cohorts, the weights take into account gender, the first two characters of the postal code, and the parents' total market income split into 11 income classes. For the later cohorts, an additional step is added to deal with adult tax filers who could be linked to their parents in the three T1FFs preceding the cohort year (and who end up not being linked in the IID).

Database structure The IID is comprised of multiple types of files. The Family File (one per cohort) contains one line per child in the cohort, and contains information coming from the T1FF of the year linked (which can be the year of the cohort or up to four years after). Variables include the child and his or her parents' dates of birth, the family composition, the child's gender, the number of children in the family, the year linked, and the postal code.

Although the links are made based on SINs, for confidentiality reasons the SINs of the children and their parents are replaced with unique and meaningless case numbers. Those case numbers are used to extract the relevant information from the annual T1 files, which contain the year-specific tax return variables for the children and their parents. Each observation in the T1 files corresponds to one individual for one given year. Variables from the T1 include the marital status, the language in which the tax filer reported his or her information to the Canada Revenue Agency, the 6-character alphanumeric postal code, the spouse's case number (derived from his or her SIN, as for the children and their parents). The T1 files also include all of the income information coming from the tax return: various income sources such as earnings, self-employment net income, interest and investment income, dividends, other employment income, rental income, and the different government benefits. The last type of files included in the IID contains geographical information. These contain the various geographical identifiers that can be derived from the postal code, and relate to the last Census year. For example, the postal codes from tax years 1997 through to 2001 can be linked to files containing information based on the 1996 Census geography, to the 2001 Census geography for the years 2002 to 2006, and so on.

Notes Note that due to the way the IID algorithm works, individuals who immigrated to Canada after the age of 19 are not part of the data, even if their later tax records would be part of the T1FF. Also, while the text here refers to children and their parents, no biological link can be established from tax data, and the "parents" really should be thought of as male and/or female household heads. For example, a 16 year old girl reporting on her tax return the same address as her mother and stepfather, even if her biological father is alive and filing his taxes, and perhaps even sharing custody, would show up in our data as having her stepfather as a "father." Another quirk of the IID is that the later cohorts are based what are called the master tax files, which for a given tax year contain all the tax records filed to the Canada Revenue Agency by the regular tax filing deadline of April 30 of the year following the tax year. The master tax files do not contain the tax return information for late filers. The master tax files are regularly updated with the information from the late filers and from any revisions to the on-time filers' returns; the updated files are called the historical files. The original cohorts from the 1980s, as well their updated versions (with the tax data through to 2008), are based on those historical files. A private communication with someone from Statistics Canada's Social Analysis and Modelling Division reported that only about 5% of all tax filers are late filers and that biases resulting from their exclusion are negligible. Furthermore, since the T1 files are of administrative nature and contain information reported by tax filers, there can be inconsistencies from year to year: different reported dates of birth, gender or SIN, for example. The T1FF deals with this by comparing the same individual from year to year and creating longitudinal date of birth, gender and SIN variables, which are equal to the most commonly reported value of the variable in the yearly T1 files. Note that the IID case numbers are based on the longitudinal SINs.

A.2 Analytical Samples and Variable Definitions

For the current paper and for comparability's sake, we align our analytical sample with Chetty et al. (2014a)'s core sample, which they define as all children born in the 1980-1982 birth cohorts, for whom they identify parents, and whose mean parent income between 1996 and 2000 is strictly positive. We select into our sample all children born in 1980 (coming from the 1996 cohort) and 1982 (coming from the 2001 cohort). Due to the age at which the children are matched to their parents (16 to 19 at the first year the match is attempted), and the tax years that are used to conduct the initial match (1996 and 2001), the IID does not contain individuals born in 1981. We drop the handful of observations for which the longitudinal year of birth is not equal to the year of birth from the Family File, which is taken from the T1 of the year linked, and which is the basis of our chosen age group. Appendix Table A3 shows the distribution of the years linked for analytical sample. Children born in 1980 are from the 1996 cohort, and linked from 1996 to 2000; children born in 1982 are from the 2001 cohort, and linked from 2001 to 2005. The majority of the observations in our sample (90%) are linked to their parents at age 20 or below.

Parent income We define parent income as the individual average of the total household pretax income, using the Canada Revenue Agency's definition of total income. From 1982 onwards, this refers to: Canada/Quebec Pension Plan benefits, capital gains/losses calculated, dividends (taxable grossed up), earnings from T4 slips including commissions, interest and investment income, Old Age Security pension, other employment income, other income, pension and superannuity income, rental income, self-employment net income (from business, commission, farming, fishing or professional), and employment insurance benefits. In 1986 tax (GST and FST, now HST) credits were introduced; in 1998 limited partnership income and Registered Retirement Savings Plan income; in 1992 Net Federal supplements, Social Assistance payments and Workers' Compensation payments; in 1996 Guaranteed Income Supplements; more followed in later years.

We take each parent's total income for the tax years 1996 to 2000, a potential of up to five years for each parent. If we cannot find a parent's record in the T1 file, we assign a value of zero for that year's income. We then take the mean of the sum of the father's and mother's total income, that is, we add up all the income figures and divide by 10 for a two-parent family or by 5 for a single-headed family. If we observe a parental death, we do not assign a value of zero; we instead adjust the denominator to only take into account the years when the parent was alive. If do not observe at least one of the parents in one year in the T1 files, meaning we never have any measure of income, we drop the child from the sample. This only happens on very rare occasions, given that to be in our dataset a parent must appear in the T1 files at least once between 1996 and 2005, depending on the cohort and the year linked.

Note that if the child is recorded as having two parents in the year linked, but the parents separate or divorce at a later date, we keep adding the individual incomes of the mother and father as defined in the year linked. Also, since our children born in 1982 are from the 2001 cohort, two "parents" can be identified in the year linked (2001 or later) but traced back in time even if they were not yet forming a household. Suppose a 2001-cohort child's mother and stepfather start filing taxes using the same address in 2000, then in 2001 they would both be considered the child's parents. This means that when computing the mean parental income from 1996 to 2000, the mother and stepfather were actually not in the same household from 1996 to 1999. Appendix Table C1 shows that 75% of children grew up in households with married parents, 5$ with common-law partners as parents, and the balance, 20%, with single parents.

For our main parental income variable, we restrict our sample to children for whom the average individual parent income from 1996 to 2000 is US$500 or higher. There are 9,337 observations with a parental income under US$500, or 1.67% of the sample. This is the average of total income–not earnings–, including benefits, over five years, for one or two parents. Given the generous social net in Canada, these ultra-low income situations are likely to be coding errors and not representative of a low-income situation.

Child income Following Chetty et al. (2014a)'s definition, we use compute child income as the mean of the child's total income for the tax years 2011 and 2012, including the child's married spouse total income figures for the same years if a spouse is reported for those years. We only use married spouses even if a significant fraction of the children in 2011 and 2012 report having a common-law partner, to be closer to Chetty et al.'s definition, who do not have information on common-law spouses (filing separately). If we cannot find a child in both the 2011 and the 2012 T1 files, meaning his or her income information is completely missing for the years that we are interested in, we drop that individual from the sample. As for parental income, we drop from our sample observations with child income under US$500. We also try specifications where figures under US$500 are assigned a value of US$500 (see below for details on conversion from Canadian to American dollars), or where we do the same for figure under US$1000 (recoding to US$1000).

We can see in Appendix Table C1 that 30% of children are married in 2011 and 32% are in 2012.

CPI and PPP conversions All the dollar figures in the IID T1 files are in Canadian current dollars. We first adjust all figures for inflation by converting to 2012 Canadian constant dollars using the Consumer Price Index (CANSIM Table 326-0021). We then convert the Canadian constant dollars to 2012 US dollars using the Organisation for Economic Co-operation and Development's Purchasing Power Parity (PPP) for private consumption between the US and Canadian dollar for the year 2012 (a value of 1.284164 C$ for each US$).[11]

US ranks Children income US ranks and parental income US ranks were assigned to Canadian children and their parents based on the PPP-converted total incomes figures described above. Chetty et al. (2014a)'s article comes with online data tables, among which is a table providing the national marginal income distributions by centile (Online Table 2).[12] The table gives, for each centile of the parents or children income distribution, the mean of all parent or children incomes, rounded to the $100. We use as centile cutoffs the midpoints between two means. For example, the mean child income for the 10th and 11th centiles are $2,300 and $3,300, respectively. The midpoint between the two is $2,800, and we use that as the cutoff between the 10th and 11th centiles.

Weights All the relevant computations use the IID weights to produce figures that correspond to population estimates. The weight variable used is a1w_t1ff3. For children born in 1980, the weights are taken as provided. For those born in 1982, we rescaled the weights so that their sum is equal to the sum of the 1980 weights. This was done because the two sets of children come from two different IID cohorts, and we wanted to give each cohort the same weight. Our findings are not sensitive to the rescaling of the 1982 weights.

Geography We base our geography variables on the postal code of the year linked, which occurs when the child is 16 to 19 in the first year (up 20 to 23 in the fifth year). Using the postal code, the 1996 Census geography is added to the data. Even though the children born in 1980 and 1982 come different cohorts and different years linked, which means that some would be closer to the 2001 Census than the 1996, we use the 1996 geography for all of the observations in our analytical sample to avoid issues of comparability over time. The two geographical variables that we use are the province or territory and the Census division. Apart from the creation of the Nunavut territory in 1999 (previously part of the Northwest Territories), the provincial and territorial boundaries do not change over time. The Census division is one of the most stable

[11] Retrieved online at http://stats.oecd.org/Index.aspx?datasetcode=SNA_TABLE4

[12] Retrieved online at http://www.equality-of-opportunity.org/index.php/data

geographical units too, especially in the time frame considered. Unlike metropolitan areas, Census divisions span the whole territory of Canada and thus allow us to compute statistics covering the whole country. Also, Census divisions are not arbitrary: their boundaries reflect local and provincial administrative units, such as counties, regional districts, regional municipalities and other types of provincially legislated areas. There are 288 Census divisions in the 1996 and in the 2001 Census geographies (Statistics Canada (1997), Statistics Canada (2003)). Appendix Table A4 shows the population statistics by Census division and by Commuting Zone, the US geography used. In terms of population, the average Commuting Zone is more than three times the zone of the average Census division, which makes sense since the total population of the United States is ten times that of Canada, and there are roughly 2.5 times more Commuting Zones than Census Divisions.

Table A2: Distribution of Year Linked

Year of the link	Count	Weighted count	Distribution of year linked
Panel A: 1982 Cohort			
1982	515,880	637,721	0.45
1983	218,040	278,389	0.19
1984	204,600	269,951	0.18
1985	139,130	187,846	0.12
1986	105,960	143,220	0.09
Total	1,183,610	1,517,127	
Panel B: 1984 Cohort			
1984	419,460	551,304	0.38
1985	249,790	336,634	0.23
1986	264,830	364,401	0.24
1987	127,470	177,652	0.12
1988	63,300	87,135	0.06
Total	1,124,850	1,517,126	
Panel C: 1986 Cohort			
1986	513,910	673,211	0.44
1987	278,320	363,255	0.24
1988	177,990	233,396	0.15
1989	132,480	176,612	0.11
1990	52,550	70,653	0.05
Total	1,155,250	1,517,127	
Panel D: 1991 Cohort			
1991	515,050	688,077	0.38
1992	214,630	289,931	0.16
1993	195,360	263,295	0.14
1994	122,870	167,419	0.09
1995	54,950	75,844	0.04
Total	1,102,860	1,484,566	
Panel E: 1996 Cohort			
1996	533,340	696,760	0.46
1997	279,690	373,771	0.24
1998	203,600	277,216	0.17
1999	105,180	147,023	0.09
2000	45,070	63,624	0.04
Total	1,166,880	1,558,393	
Panel F: 2001 Cohort			
2001	842,400	1,001,706	0.62
2002	215,380	263,858	0.16
2003	183,860	229,798	0.14
2004	73,060	93,475	0.05
2005	35,520	45,809	0.03
Total	1,350,220	1,634,646	

Source: Authors' calculations based on the IID
Note: This table shows the unweighted and weighted counts of children by year of the link between their tax record and that of their parents, by cohort. The weighted count use the IID weights.

Table A3: Distribution of Year Linked, 1980-1982 Cohort

Year of the link	Observations (unweighted)	Observations (weighted)	Age at the time of the link
Panel A: Born in 1980			
1996	49,890	64,286	16
1997	65,160	86,333	17
1998	80,110	108,726	18
1999	55,130	77,556	19
2000	18,190	25,864	20
Panel B: Born in 1982			
2001	244,560	302,291	19
2002	17,110	22,198	20
2003	17,230	22,611	21
2004	7,510	9,929	22
2005	4,970	6,627	23
Total	559,860	726,420	18.6

Source: Authors' calculations based on the IID
Note: This table shows the distribution of the year linked (the year the match between parents and children was made in the T1 Family File) for the children in our sample, along with their birth year and age at the time of the link. Observations counts are presented, as well as weighted observation counts, using IID sample weights. The Total row shows total counts and the average age at the time of the link (which is the same whether the sample weights are used or not).

Table A4: Population Statistics by Census Division and Commuting Zone

Statistic	Canada	United States
Geographical unit	Census Division	Commuting Zone
Number of units	287	741
Mean	103,255	379,787
Median	37,965	103,842
Min	1,315	1,193
Max	2,456,805	16,393,360
P1	5,710	2,407
P5	10,195	6,745
P10	13,625	11,487
P25	21,400	38,384
P75	80,805	289,849
P90	164,810	803,201
P95	398,560	1,533,306
P99	1,782,830	4,642,561
Total population	29,634,265	281,421,906

Source: Authors' calculations using the population figures come from Statistics Canada's 2001 Census of Population for Canada, and from the Census Bureau's 2000 Census for the United States, as reported by Chetty et al. (2014a)'s Online Data Tables

Note: This table presents statistics regarding the population of Canadian Census divisions and American Commuting Zones.

B Appendix Figures

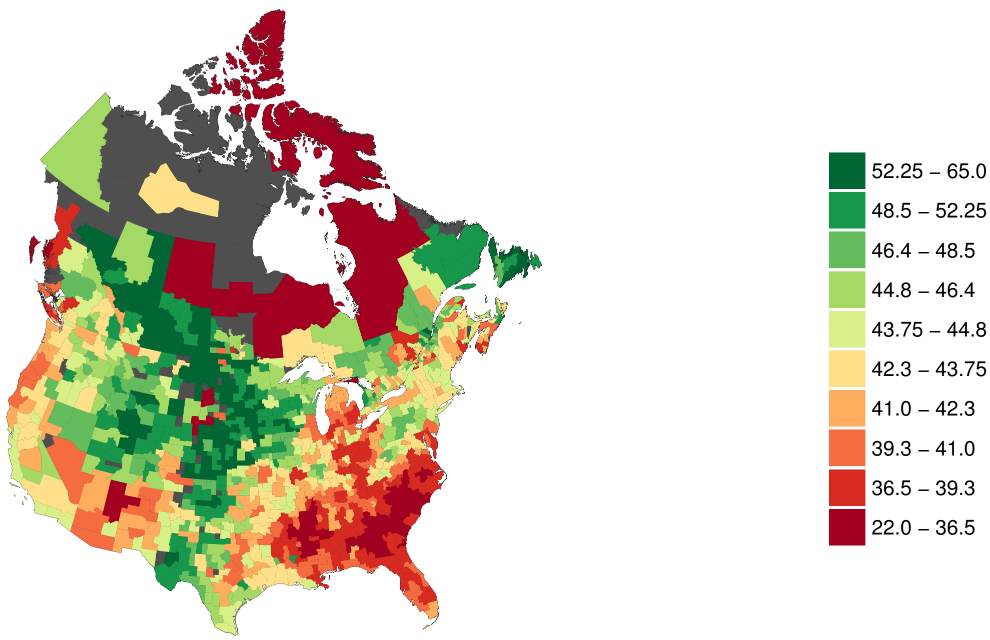

Figure B1: Absolute Upward Mobility: Mean Child Rank for Parents at 25^{th} Percentile (North America)

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figure shows a heat map of $\bar{r}_{25}$, an absolute mobility measure, where each observation is a Census division in Canada or a Commuting Zone in the United States. Ranks are based on the US distributions. More mobile places are in green; those with lower mobility appear in red. The grey areas represent places where there were not enough observations in the data to release estimates.

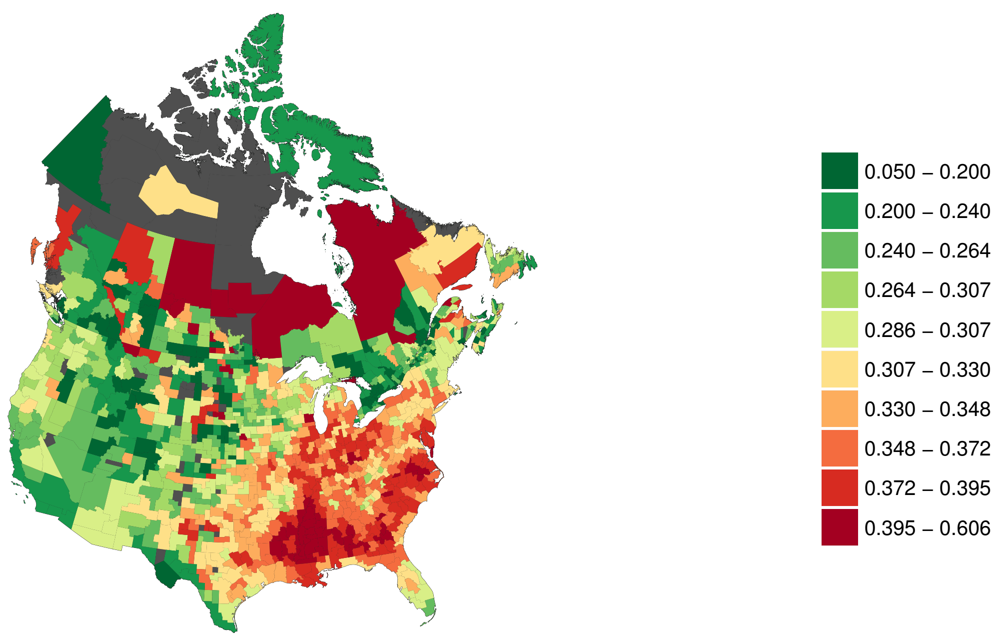

Figure B2: Relative Mobility: Rank-rank Slope (North America)

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figure shows a heat map of the rank-rank slope, a relative mobility measure, where each observation is a Census division in Canada or a Commuting Zone in the United States. Ranks are based on the US distributions. More mobile places are in green; those with lower mobility appear in red. The grey areas represent places where there were not enough observations in the data to release estimates.

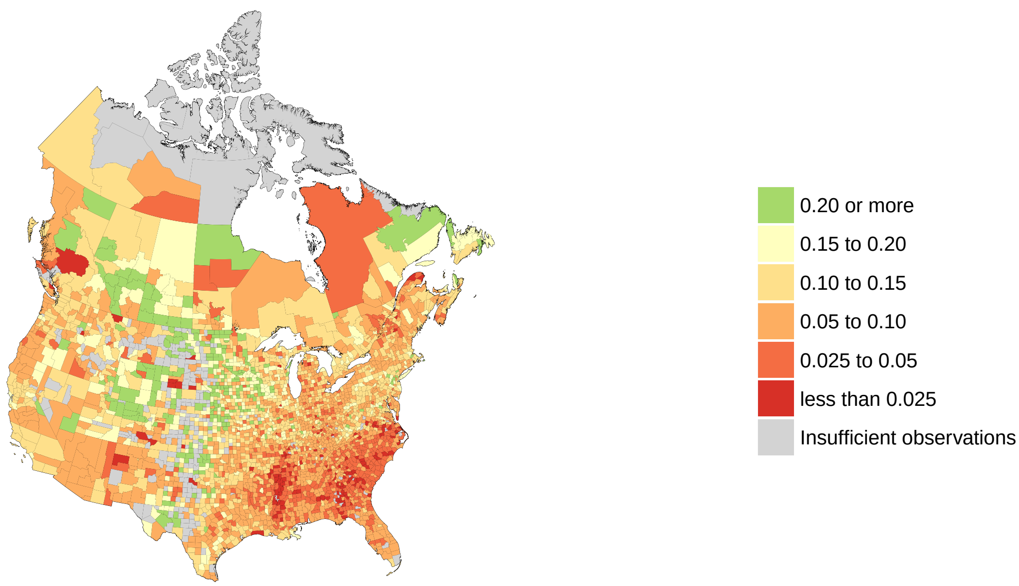

Figure B3: Rags to Riches: P(Child in Q5|Parents in Q1) (North America)

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figure shows a heat map of $P_{1,5}$, the probability that a child from parents in the bottom quintile of the parental income distribution reaches the top quintile of the children's income distribution, where each observation is a Census division in Canada or a Commuting Zone in the United States. Ranks are based on the US distributions. More mobile places are in green; those with lower mobility appear in red. The grey areas represent places where there were not enough observations in the data to release estimates.

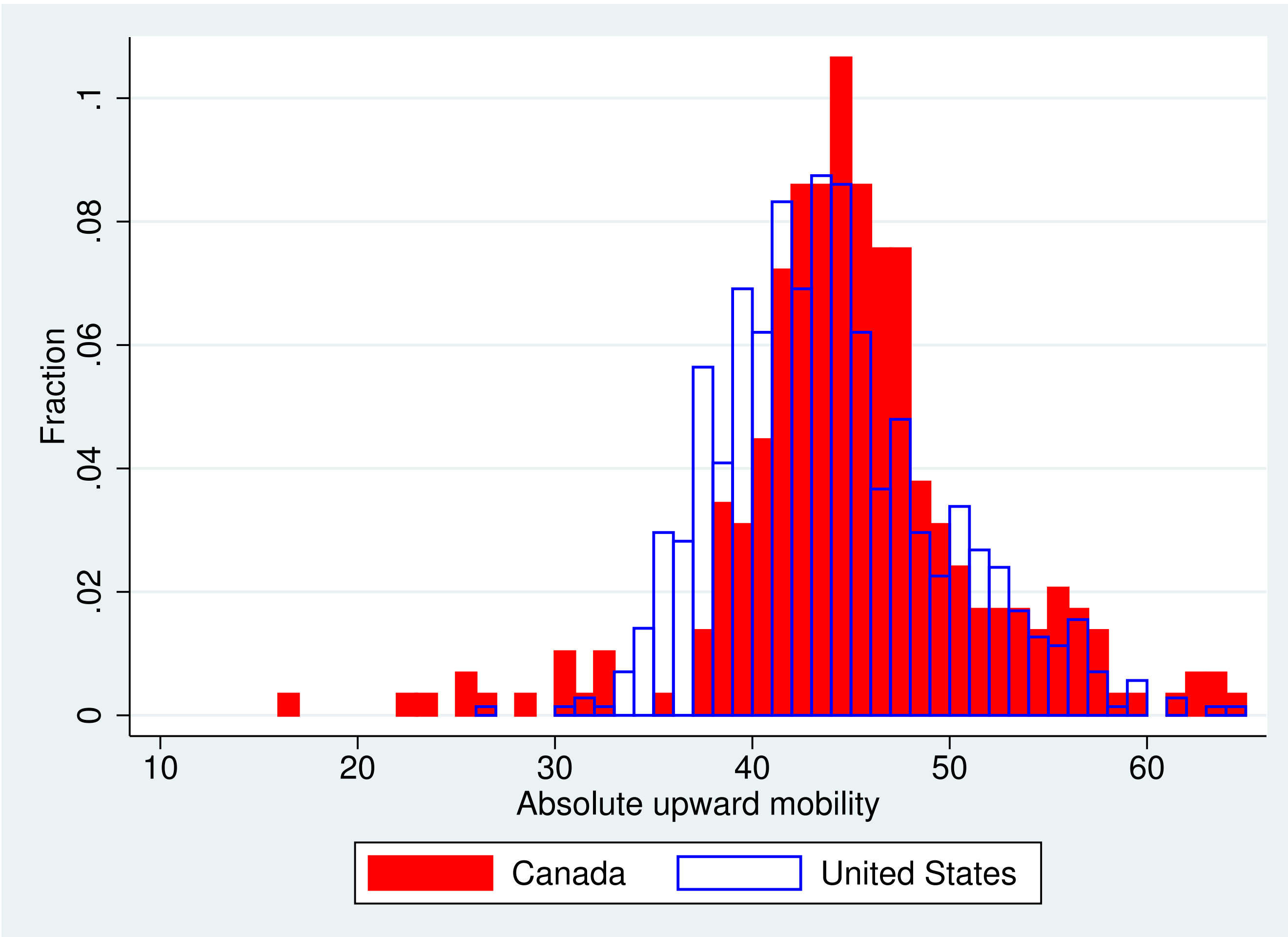

Figure B4: Histogram of Absolute Mobility by Country (Unweighted)

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figure shows the histogram of $\bar{r}_{25}$, an absolute mobility measure, where each observation is a Census division in Canada or a Commuting Zone in the United States.

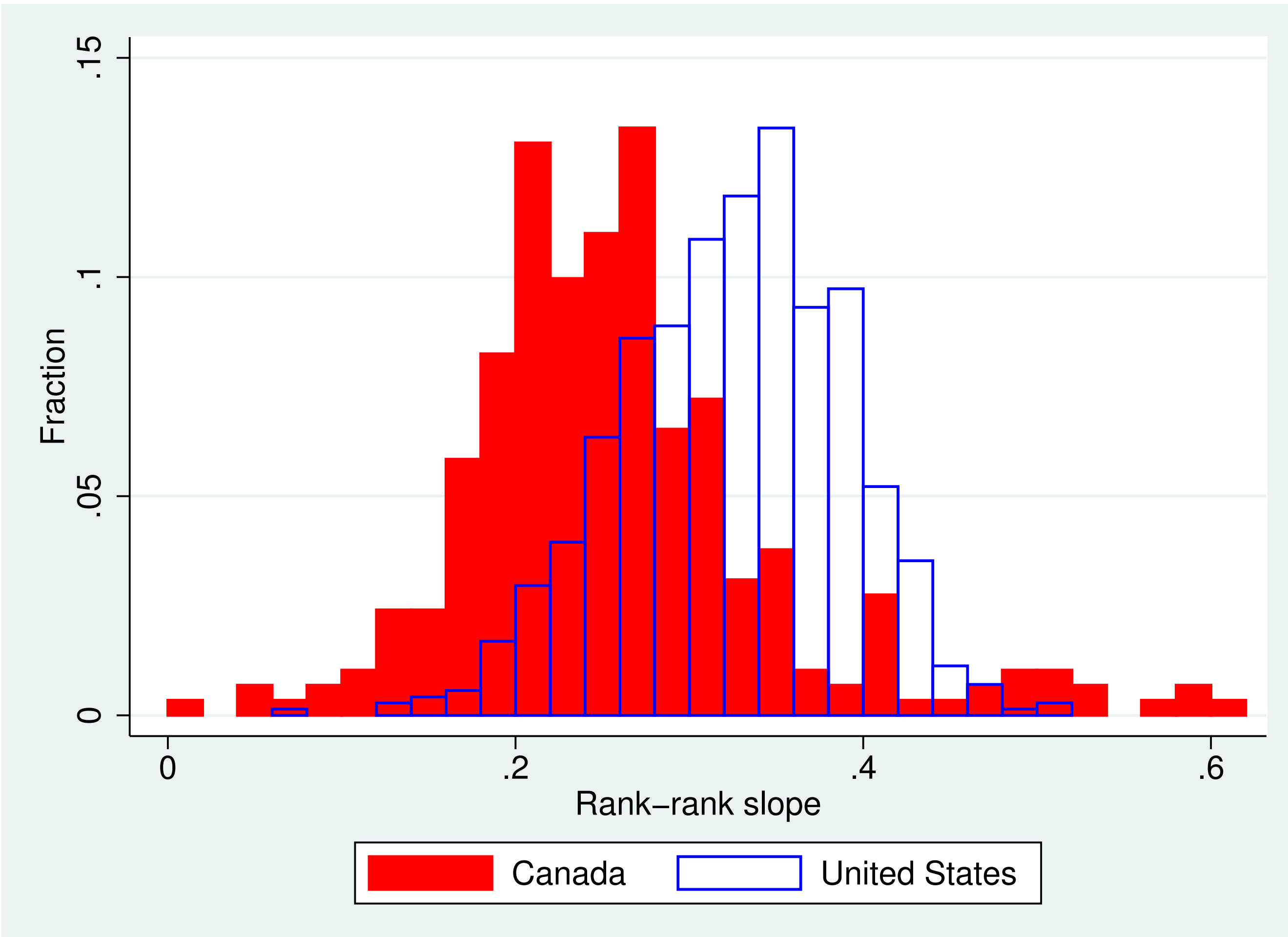

Figure B5: Histogram of Relative Mobility by Country (Unweighted)

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figure shows the histogram of the rank-rank slope estimates, a relative mobility measure, where each observation is a Census division in Canada or a Commuting Zone in the United States.

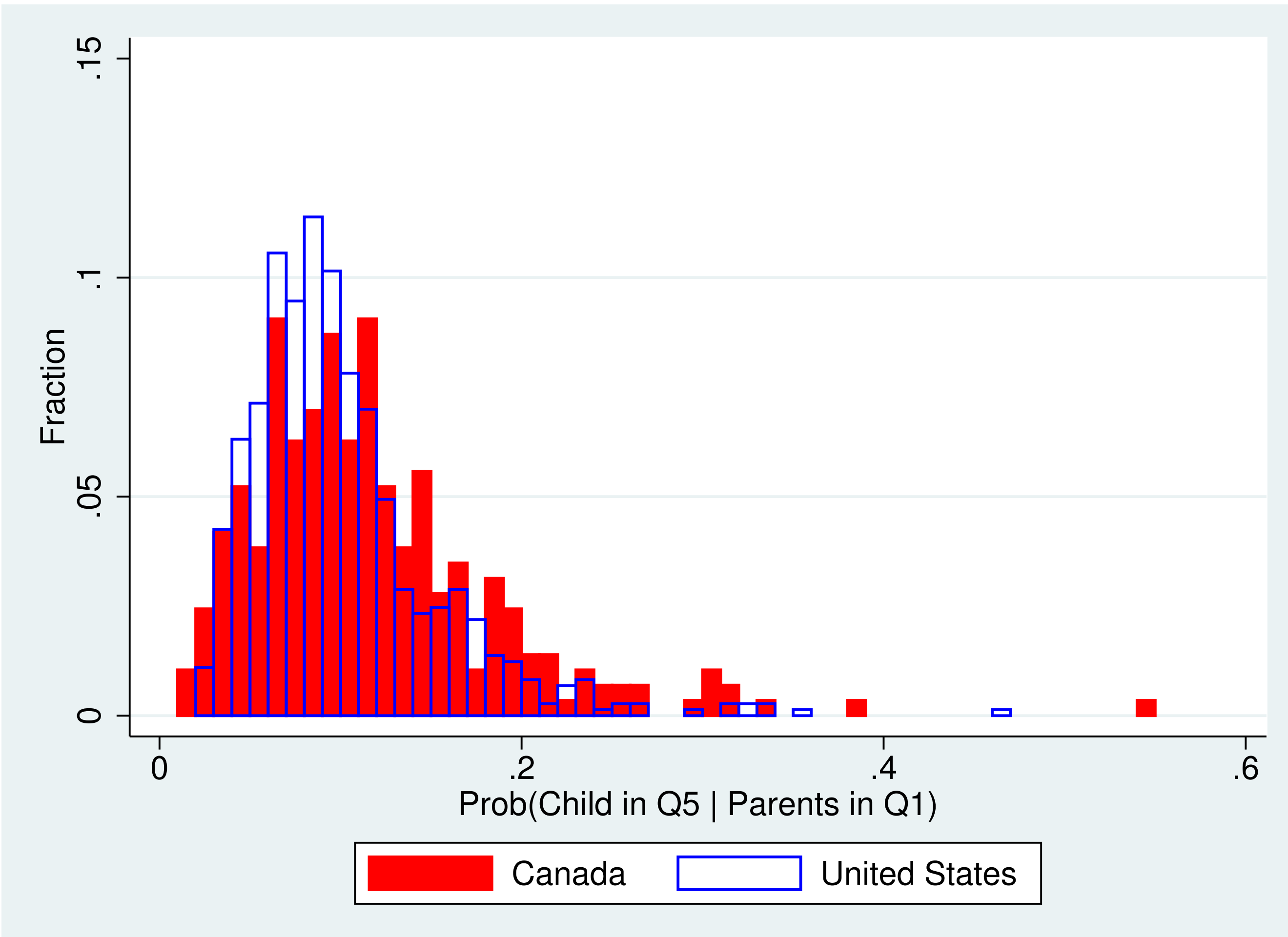

Figure B6: Histogram of P(Child in Q5|Parents in Q1) by Country (Unweighted)

Source: Authors' calculations based on the IID and on Chetty et al. (2014a)'s Online Data Tables
Note: This figure shows the histogram of P(Child in Q5|Parents in Q1), with the quantiles based on the US income distributions, where each observation is a Census division in Canada or a Commuting Zone in the United States.

C Appendix Tables

Table C1: Summary Statistics, IID Children Born 1980 and 1982

Variable	Mean	SD	Median	N
Fraction female	0.487	0.500	–	619,696
Child age in 2011	30.000	1.000	30	619,696
Father's age in 1996	45.042	6.371	44	543,661
Mother's age in 1996	42.472	5.799	42	575,059
Father's age at child birth	30.037	6.295	29	543,661
Mother's age at child birth	27.466	5.732	27	575,059
Fraction of parents married	0.752	0.432	–	619,696
Fraction of parents common-law partners	0.049	0.215	–	619,696
Fraction of single-headed parental households	0.198	0.399	–	619,696
Fraction of single-headed parental households headed by female	0.693	0.461	–	106,894
Fraction of children married in 2011	0.303	0.459	–	619,696
Fraction of children married in 2012	0.328	0.470	–	619,696
Parental income (CAN$)	80,863	124,350	66,129	619,696
Parental income (CAN$), excl. <500	82,581	125,048	67,524	608,385
Top individual parental income	59,833	109,815	48,238	619,696
Child family income, not including common-law partner's income (US$)	47,134	52,781	35,387	619,696
Child family income, including common-law partner's income (US$)	49,980	48,947	42,740	619,696
Fraction children with income under US$500	0.108	0.310	–	619,696
Child individual income (USD)	30,540	30,002	26,756	619,696
Child individual earnings (USD)	26,206	27,389	21,973	619,696
Parental income rank (Canada)	50.500	28.866	50.5	608,385
Parental income rank (US)	45.111	26.443	45	608,385
Child income rank (Canada)	50.500	28.866	50	553,677
Child income rank (US)	59.705	25.214	63	553,677
Child income rank, not including common-law partners (Canada)	50.500	28.866	50	552,676
Child income rank, not including common-law partnerss (US)	54.738	25.074	55	552,676

Source: Authors' calculations based on the IID
Note: This table presents summary statistics for the sample of children born in 1980 and 1982 in the IID. All figures presented are weighted using the IID weights. The income figures in Canadian dollars are in 2012 dollars, correcting for inflation using the CPI; those in US dollars are in 2012 US dollars, converted using the PPP. See text for details.

Table C2: Intergenerational Mobility in All Census Divisions

Census division	CD code	Largest city	Absolute mobility $\bar{r}_{25}$ (1)	$P_{1,5}$ (2)	Rank-rank slope (3)	Population (4)
Division No. 1	1001	St. John's	49.79	18.15	0.218	240,245
Division No. 2	1002	Marystown	55.36	24.53	0.214	24,255
Division No. 3	1003	Harbour Breton	52.51	14.17	0.346	19,260
Division No. 4	1004	Stephenville	46.52	14.87	0.352	22,015
Division No. 5	1005	Corner Brook	48.46	19.24	0.264	40,050
Division No. 6	1006	Gander	50.45	15.57	0.255	35,835
Division No. 7	1007	Clarenville	52.49	18.98	0.282	36,990
Division No. 8	1008	Lewisporte	55.22	17.59	0.301	41,755
Division No. 9	1009	St. Anthony	56.11	30.08	0.297	19,915
Division No. 10	1010	Labrador City	50.01	23.89	0.314	27,765
Kings	1101	Montague	45.75	15.09	0.189	19,060
Queens	1102	Charlottetown	43.93	11.56	0.204	70,365
Prince	1103	Summerside	44.70	9.57	0.277	43,960
Shelburne	1201	Shelburne	41.08	6.33	0.218	16,090
Yarmouth	1202	Yarmouth	38.81	7.74	0.306	26,520
Digby	1203	Digby	39.19	4.81	0.333	19,245
Queens	1204	Queens	38.40	4.17	0.321	11,590
Annapolis	1205	Middleton	42.09	6.50	0.165	21,470
Lunenburg	1206	Lunenburg	41.25	12.01	0.235	47,005
Kings	1207	Kentville	42.51	5.76	0.281	58,135
Hants	1208	East Hants	42.73	7.31	0.204	40,175
Halifax	1209	Halifax	40.64	10.49	0.238	355,945
Colchester	1210	Truro	40.12	6.90	0.305	48,785
Cumberland	1211	Amherst	40.60	6.19	0.328	31,715
Pictou	1212	Pictou	43.66	14.42	0.215	46,250
Guysborough	1213	Guysborough	50.85	10.56	0.3	9,720
Antigonish	1214	Antigonish	47.37	14.95	0.218	19,390
Inverness	1215	Port Hawkesbury	47.82	19.22	0.271	19,665
Richmond	1216	Chapel Island 5	42.13	7.88	0.35	10,125
Cape Breton	1217	Cape Breton	42.98	11.48	0.288	107,880
Victoria	1218	Wagmatcook 1	43.63	20.26	0.208	7,865
Saint John	1301	Saint John	41.09	11.21	0.283	75,195
Charlotte	1302	St. Stephen	44.06	18.38	0.196	27,020
Sunbury	1303	Oromocto	42.28	13.51	0.234	25,710
Queens	1304	Minto	46.96	13.37	0.042	11,635
Kings	1305	Quispamsis	43.96	11.58	0.252	63,890
Albert	1306	Riverview	44.17	5.69	0.212	26,465
Westmorland	1307	Westmorland	42.97	6.98	0.281	122,405
Kent	1308	Carleton	45.47	10.06	0.368	30,970
Northumberland	1309	Miramichi	42.42	10.11	0.345	50,160
York	1310	Fredericton	40.88	6.41	0.304	86,435
Carleton	1311	Kent	43.88	6.88	0.23	26,895
Victoria	1312	Drummond	44.26	12.00	0.308	20,915
Madawaska	1313	Madawaska	44.24	6.93	0.296	34,850
Restigouche	1314	Campbellton	40.22	9.65	0.402	35,410
Gloucester	1315	Bathurst	42.92	7.73	0.359	81,760
Les Îles-de-la-Madeleine	2401	Cap-aux-Meules	47.21	13.31	0.187	12,575
Pabok	2402	Chandler	38.79	1.93	0.403	19,175
La Côte-de-Gaspé	2403	Murdochville	45.12	2.58	0.267	18,270
Denis-Riverin	2404	Sainte-Anne-des-Monts	39.46	1.88	0.42	12,495

Continued on next page

Table C2 – Continued from previous page

Census division	CD code	Largest city	Absolute mobility $\bar{r}_{25}$ (1)	$P_{1,5}$ (2)	Rank-rank slope (3)	Population (4)
Bonaventure	2405	Bonaventure	45.82	6.65	0.311	17,990
Avignon	2406	Carleton-sur-Mer	44.20	10.44	0.298	14,965
La Matapédia	2407	Amqui	47.93	6.12	0.243	19,395
Matane	2408	Matane	44.92	4.90	0.264	22,065
La Mitis	2409	Mont-Joli	44.26	8.68	0.277	18,700
Rimouski-Neigette	2410	Rimouski	46.57	8.49	0.19	50,980
Les Basques	2411	Trois-Pistoles	47.29	8.73	0.181	9,585
Rivière-du-Loup	2412	Saint-Antonin	47.68	5.70	0.263	31,040
Témiscouata	2413	Saint-Louis-du-Ha! Ha!	46.45	3.14	0.202	21,910
Kamouraska	2414	Kamouraska	49.23	7.48	0.134	21,870
Charlevoix-Est	2415	La Malbaie	47.76	10.62	0.24	16,385
Charlevoix	2416	Baie-Saint-Paul	52.55	11.87	0.078	12,725
L'Islet	2417	L'Islet	46.60	10.79	0.089	18,940
Montmagny	2418	Montmagny	45.12	4.32	0.306	22,875
Bellechasse	2419	Saint-Henri	48.31	11.28	0.17	28,900
L'Île-d'Orléans	2420	Sainte-Famille	53.24	7.31	0.102	6,705
La Côte-de-Beaupré	2421	Boischatel	49.98	16.28	0.162	20,570
La Jacques-Cartier	2422	Lac-Beauport	46.34	6.10	0.181	26,380
Québec	2423	Québec	45.01	7.77	0.225	501,845
Les Chutes-de-la-Chaudière	2425	Levis	47.21	5.18	0.204	78,275
La Nouvelle-Beauce	2426	Sainte-Marie	49.53	12.90	0.135	25,350
Robert-Cliche	2427	Beauceville	44.94	1.31	0.27	18,285
Les Etchemins	2428	Lac-Etchemin	47.63	2.85	0.264	17,300
Beauce-Sartigan	2429	Saint-Georges	43.90	8.13	0.277	46,950
Le Granit	2430	Frontenac	44.45	11.51	0.175	21,400
L'Amiante	2431	Thetford Mines	45.93	11.41	0.257	42,255
L'Érable	2432	Plessisville	48.02	5.63	0.155	23,200
Lotbinière	2433	Saint-Apollinaire	48.03	12.31	0.265	26,330
Portneuf	2434	Portneuf	48.52	13.57	0.174	43,900
Mékinac	2435	Saint-Tite	44.58	4.33	0.307	12,570
Le Centre-de-la-Mauricie	2436	Shawinigan	42.38	4.10	0.313	63,215
Francheville	2437	Trois-Rivières	41.59	7.25	0.284	135,535
Bécancour	2438	Saint-Pierre-les-Becquets	44.45	3.95	0.252	18,620
Arthabaska	2439	Victoriaville	42.04	3.79	0.242	62,670
Asbestos	2440	Asbestos	42.68	5.59	0.324	14,230
Le Haut-Saint-François	2441	Cookshire-Eaton	38.80	4.63	0.286	21,020
Le Val-Saint-François	2442	Windsor	41.12	4.89	0.287	27,690
Sherbrooke	2443	Sherbrooke	41.06	6.51	0.293	137,940
Coaticook	2444	Coaticook	40.96	2.15	0.269	16,220
Memphrémagog	2445	Magog	39.32	6.83	0.246	41,200
Brome-Missisquoi	2446	Cowansville	40.41	4.66	0.287	44,830
La Haute-Yamaska	2447	Granby	40.70	2.96	0.237	77,535
Acton	2448	Acton Vale	39.03	4.82	0.271	14,830
Drummond	2449	Drummondville	42.45	6.67	0.218	86,030
Nicolet-Yamaska	2450	Nicolet	41.81	3.88	0.276	22,830
Maskinongé	2451	Louiseville	42.49	8.20	0.245	22,665
D'Autray	2452	Lavaltrie	41.37	6.97	0.263	37,580
Le Bas-Richelieu	2453	Sorel-Tracy	45.13	15.16	0.285	49,200
Les Maskoutains	2454	Saint-Hyacinthe	44.63	9.93	0.228	77,025
Rouville	2455	Marieville	46.54	15.80	0.158	29,490

Continued on next page

Table C2 – Continued from previous page

Census division	CD code	Largest city	Absolute mobility $\bar{r}_{25}$ (1)	$P_{1,5}$ (2)	Rank-rank slope (3)	Population (4)
Le Haut-Richelieu	2456	Saint-Jean-sur-Richelieu	44.23	7.87	0.218	99,630
La Vallée-du-Richelieu	2457	Chambly	46.03	8.95	0.19	118,630
Champlain	2458	Longueuil	41.74	9.03	0.226	308,955
Lajemmerais	2459	Sainte-Julie	46.71	8.84	0.174	99,385
L'Assomption	2460	L'Assomption	43.22	5.14	0.259	103,205
Joliette	2461	Joliette	44.45	9.77	0.233	52,255
Matawinie	2462	Rawdon	39.08	3.73	0.301	42,600
Montcalm	2463	Saint-Roch-de-l'Achigan	38.61	4.60	0.311	38,285
Les Moulins	2464	Terrebonne	43.39	10.00	0.209	109,415
Laval	2465	Laval	43.30	11.23	0.218	339,005
Montréal	2466	Montréal	40.11	7.89	0.249	1,782,830
Roussillon	2467	Saint-Constant	41.97	6.98	0.24	137,195
Les Jardins-de-Napierville	2468	Napierville	43.40	3.58	0.213	22,515
Le Haut-Saint-Laurent	2469	Ormstown	39.00	8.69	0.308	21,595
Beauharnois-Salaberry	2470	Salaberry-de-Valleyfield	44.28	9.28	0.237	58,060
Vaudreuil-Soulanges	2471	Vaudreuil-Dorion	43.94	8.05	0.179	101,290
Deux-Montagnes	2472	Deux-Montagnes	41.12	7.52	0.233	80,805
Thérèse-De Blainville	2473	Blainville	42.46	8.68	0.248	129,110
Mirabel	2474	Mirabel	43.37	4.00	0.243	27,110
La Rivière-du-Nord	2475	Sainte-Sophie	39.00	4.40	0.29	88,750
Argenteuil	2476	Lachute	44.98	19.24	0.302	28,225
Les Pays-d'en-Haut	2477	Saint-Sauveur	41.83	6.57	0.2	30,460
Les Laurentides	2478	Montcalm	38.29	6.44	0.261	37,705
Antoine-Labelle	2479	Mont-Laurier	41.54	7.20	0.264	32,355
Papineau	2480	Thurso	45.85	9.91	0.254	20,090
l'Outaouais	2481	Gatineau	45.61	9.33	0.256	224,755
Les Collines-de-l'Outaouais	2482	Pontiac	44.09	13.81	0.229	35,040
La Vallée-de-la-Gatineau	2483	Maniwaki	40.90	11.77	0.339	18,570
Pontiac	2484	Bristol	47.15	16.04	0.23	14,375
Témiscamingue	2485	Ville-Marie	46.02	9.47	0.233	17,280
Rouyn-Noranda	2486	Rouyn-Noranda	47.04	6.66	0.186	39,265
Abitibi-Ouest	2487	La Sarre	50.39	20.50	0.203	21,690
Abitibi	2488	Amos	49.66	8.06	0.244	24,270
Vallée-de-l'Or	2489	Val-d'Or	46.94	12.83	0.282	41,845
Le Haut-Saint-Maurice	2490	La Tuque	37.32	9.60	0.493	15,715
Le Domaine-du-Roy	2491	Roberval	45.02	9.76	0.252	32,150
Maria-Chapdelaine	2492	Dolbeau-Mistassini	47.65	14.25	0.163	26,415
Lac-Saint-Jean-Est	2493	Alma	48.31	8.34	0.137	50,940
Le Fjord-du-Saguenay	2494	Saguenay	46.44	8.81	0.218	164,810
La Haute-Côte-Nord	2495	Forestville	46.71	3.52	0.111	12,790
Manicouagan	2496	Baie-Comeau	41.82	18.18	0.311	33,370
Sept-Rivières - Caniapiscau	2497	Port-Cartier	44.33	11.99	0.343	38,530
Minganie - Basse-Côte-Nord	2498	Havre-Saint-Pierre	50.16	16.47	0.389	12,220
Nord-du-Québec	2499	Chibougamau	30.94	3.63	0.464	38,475
Stormont, Dundas and Glengarry United Counties	3501	Cornwall	43.89	11.37	0.247	107,545
Prescott and Russell United Counties	3502	Clarence-Rockland	47.66	9.52	0.262	74,980

Continued on next page

Table C2 – Continued from previous page

Census division	CD code	Largest city	Absolute mobility $\bar{r}_{25}$ (1)	$P_{1,5}$ (2)	Rank-rank slope (3)	Population (4)
Ottawa-Carleton	3506	Ottawa	44.21	12.53	0.213	763,790
Leeds and Grenville United Counties	3507	Brockville	44.43	13.62	0.125	95,180
Lanark	3509	Perth	42.79	7.32	0.208	60,955
Frontenac	3510	Kingston	41.58	10.68	0.252	135,410
Lennox and Addington	3511	Loyalist	44.00	10.91	0.186	37,965
Hastings	3512	Belleville	42.16	11.43	0.234	124,420
Prince Edward	3513	Prince Edward	42.16	11.02	0.199	24,360
Northumberland	3514	Hamilton	46.13	9.88	0.166	75,940
Peterborough	3515	Peterborough	43.55	11.22	0.222	123,605
Victoria	3516	Kawartha Lakes	46.59	15.93	0.183	68,460
Durham	3518	Oshawa	43.99	14.07	0.197	502,905
York	3519	Markham	46.31	15.31	0.194	725,665
Toronto Metropolitan	3520	Toronto	41.67	10.42	0.225	2,456,805
Peel	3521	Mississauga	44.95	12.30	0.183	985,565
Dufferin	3522	Orangeville	49.81	16.06	0.052	50,360
Wellington	3523	Guelph	46.49	13.08	0.161	184,840
Halton	3524	Oakville	45.66	11.42	0.176	372,410
Hamilton-Wentworth	3525	Hamilton	42.12	10.76	0.231	484,385
Niagara	3526	St. Catharines	41.34	9.07	0.19	404,590
Haldimand-Norfolk	3528	Norfolk County	44.54	9.69	0.25	103,330
Brant	3529	Brant	41.34	8.94	0.234	116,755
Waterloo	3530	Waterloo	45.01	12.22	0.203	433,875
Perth	3531	Stratford	47.57	12.51	0.153	72,455
Oxford	3532	Woodstock	45.96	13.57	0.175	97,965
Elgin	3534	St. Thomas	43.53	8.39	0.222	80,150
Kent	3536	Chatham-Kent	40.52	8.58	0.274	105,850
Essex	3537	Windsor	39.72	7.48	0.241	371,085
Lambton	3538	Sarnia	44.17	9.18	0.259	125,560
Middlesex	3539	London	43.40	11.66	0.207	398,560
Huron	3540	South Huron	48.74	11.33	0.192	58,700
Bruce	3541	Saugeen Shores	51.57	17.64	0.23	62,940
Grey	3542	Owen Sound	45.95	11.74	0.218	87,670
Simcoe	3543	Barrie	42.22	12.46	0.201	372,325
Muskoka	3544	Huntsville	39.75	8.00	0.261	51,710
Haliburton	3546	Dysart and Others	44.13	9.29	0.275	14,925
Renfrew	3547	Renfrew	46.88	16.21	0.203	93,760
Nipissing	3548	North Bay	45.33	14.59	0.228	81,595
Parry Sound	3549	Nipissing	44.44	9.48	0.24	39,325
Manitoulin	3551	Northeastern Manitoulin and the Islands	32.89	6.87	0.487	12,520
Sudbury	3552	Espanola	47.88	12.30	0.177	22,825
Sudbury	3553	Sudbury	47.97	15.50	0.263	153,560
Timiskaming	3554	Temiskaming Shores	48.29	12.94	0.21	33,995
Cochrane	3556	Cochrane	45.48	13.71	0.277	84,295
Algoma	3557	Sault Ste. Marie	42.57	8.24	0.268	117,200
Thunder Bay	3558	Thunder Bay	43.38	11.46	0.264	149,150
Rainy River	3559	Rainy River	44.06	15.95	0.295	21,875
Kenora	3560	Kenora	32.05	5.61	0.502	61,460
Division No. 1	4601	Alexander	47.27	17.92	0.227	16,340
Division No. 2	4602	Hanover	43.51	6.61	0.252	50,480

Continued on next page

Table C2 – Continued from previous page

Census division	CD code	Largest city	Absolute mobility $\bar{r}_{25}$ (1)	$P_{1,5}$ (2)	Rank-rank slope (3)	Population (4)
Division No. 3	4603	Thompson	45.27	6.00	0.247	41,735
Division No. 4	4604	Lorne	53.13	14.37	0.26	9,815
Division No. 5	4605	Boissevain	54.72	14.39	0.303	13,625
Division No. 6	4606	Virden	48.35	14.90	0.456	9,915
Division No. 7	4607	Brandon	40.86	8.20	0.402	56,215
Division No. 8	4608	North Norfolk	46.70	6.80	0.315	14,430
Division No. 9	4609	Portage la Prairie	37.60	6.10	0.506	22,520
Division No. 10	4610	Macdonald	46.16	6.46	0.248	9,465
Division No. 11	4611	Winnipeg	41.20	11.49	0.262	612,165
Division No. 12	4612	Springfield	47.40	18.43	0.123	19,135
Division No. 13	4613	St. Andrews	43.55	5.18	0.246	41,420
Division No. 14	4614	Rockwood	46.20	26.37	0.221	16,910
Division No. 15	4615	Neepawa	55.56	16.81	0.124	21,575
Division No. 16	4616	Roblin	50.26	14.71	0.526	10,195
Division No. 17	4617	Dauphin	45.20	11.27	0.341	22,530
Division No. 18	4618	Gimli	45.59	2.54	0.277	22,335
Division No. 19	4619	Peguis 1B	22.03	12.02	0.504	15,700
Division No. 20	4620	Swan River	48.61	4.18	0.279	10,860
Division No. 21	4621	The Pas	31.59	4.47	0.528	22,380
Division No. 22	4622	Thompson	28.28	3.65	0.606	34,980
Division No. 23	4623	Churchill	25.47	38.83	0.57	8,970
Division No. 1	4701	Estevan	64.93	33.62	0.138	29,780
Division No. 2	4702	Weyburn	63.08	26.74	0.189	21,080
Division No. 3	4703	Assiniboia	62.90	25.27	0.18	14,580
Division No. 4	4704	Maple Creek	57.76	29.31	0.265	11,540
Division No. 5	4705	Melville	59.14	16.70	0.276	31,885
Division No. 6	4706	Regina	50.23	14.97	0.242	216,160
Division No. 7	4707	Moose Jaw	53.48	19.53	0.268	45,820
Division No. 8	4708	Swift Current	55.80	21.52	0.261	30,315
Division No. 9	4709	Yorkton	53.63	21.29	0.29	36,290
Division No. 10	4710	Wynyard	55.60	16.15	0.492	18,955
Division No. 11	4711	Saskatoon	49.07	19.35	0.274	234,145
Division No. 12	4712	Battleford	52.51	30.22	0.314	23,375
Division No. 13	4713	Kindersley	62.05	25.34	0.227	23,775
Division No. 14	4714	Melfort	56.20	14.51	0.252	38,835
Division No. 15	4715	Prince Albert	49.45	16.57	0.35	78,355
Division No. 16	4716	North Battleford	48.67	18.49	0.465	36,990
Division No. 17	4717	Lloydminster	51.62	5.44	0.432	39,325
Division No. 18	4718	La Loche	30.77	18.45	0.581	31,955
Division No. 1	4801	Medicine Hat	56.04	9.81	0.165	66,675
Division No. 2	4802	Lethbridge	44.35	9.48	0.336	132,110
Division No. 3	4803	Willow Creek No. 26	44.06	20.03	0.346	36,905
Division No. 4	4804	Hanna	63.86	16.68	0.09	11,105
Division No. 5	4805	Strathmore	49.21	18.80	0.329	46,505
Division No. 6	4806	Calgary	51.84	31.23	0.204	1,012,305
Division No. 7	4807	Wainwright	61.13	23.69	0.15	39,660
Division No. 8	4808	Red Deer	54.12	22.17	0.226	150,430
Division No. 9	4809	Clearwater County	56.14	23.25	0.267	19,450
Division No. 10	4810	Lloydminster	58.83	21.01	0.184	81,240
Division No. 11	4811	Edmonton	53.19	20.41	0.218	964,145
Division No. 12	4812	Cold Lake	54.92	19.87	0.375	57,590

Continued on next page

Table C2 – Continued from previous page

Census division	CD code	Largest city	Absolute mobility $\bar{r}_{25}$ (1)	$P_{1,5}$ (2)	Rank-rank slope (3)	Population (4)
Division No. 13	4813	Lac Ste. Anne County	57.12	19.41	0.212	62,910
Division No. 14	4814	Yellowhead County	56.95	10.59	0.148	26,690
Division No. 15	4815	Canmore	41.09	31.33	0.411	33,790
Division No. 16	4816	Wood Buffalo (Fort Mc. Murray)	57.06	12.53	0.268	42,865
Division No. 17	4817	Slave Lake	45.47	14.03	0.386	57,195
Division No. 18	4818	Greenview No. 16	51.73	30.47	0.347	14,135
Division No. 19	4819	Grande Prairie	57.72	21.90	0.183	85,445
East Kootenay	5901	Fernie	52.07	6.40	0.18	55,890
Central Kootenay	5903	Nelson	43.43	11.16	0.201	56,500
Kootenay Boundary	5905	Trail	47.18	13.91	0.26	31,420
Okanagan-Similkameen	5907	Penticton	45.14	7.45	0.158	75,985
Fraser Valley	5909	Abbotsford	42.69	11.50	0.21	233,850
Greater Vancouver	5915	Vancouver	43.22	9.28	0.203	1,967,475
Capital	5917	Victoria	41.20	5.77	0.202	320,710
Cowichan Valley	5919	North Cowichan	38.63	7.11	0.263	71,315
Nanaimo	5921	Nanaimo	39.87	2.12	0.273	125,550
Alberni-Clayoquot	5923	Port Alberni	37.96	14.07	0.307	30,135
Comox-Strathcona	5925	Campbell River	42.93	10.58	0.216	95,585
Powell River	5927	Powell River	43.97	10.45	0.149	19,575
Sunshine Coast	5929	Sechelt	40.48	8.67	0.184	25,450
Squamish-Lillooet	5931	Squamish	37.49	11.78	0.307	32,925
Thompson-Nicola	5933	Kamloops	44.76	10.71	0.225	118,665
Central Okanagan	5935	Kelowna	46.40	12.94	0.189	145,950
North Okanagan	5937	Vernon	45.20	9.77	0.168	72,370
Columbia-Shuswap	5939	Salmon Arm	42.32	18.35	0.258	47,825
Cariboo	5941	Quesnel	47.34	2.22	0.278	65,210
Mount Waddington	5943	Port McNeill	39.76	3.22	0.335	13,085
Central Coast	5945	Bella Bella	23.31	7.02	0.348	3,780
Skeena-Queen Charlotte	5947	Prince Rupert	35.80	13.26	0.376	21,565
Kitimat-Stikine	5949	Terrace	38.04	9.28	0.408	40,720
Bulkley-Nechako	5951	Smithers	43.98	24.32	0.263	40,680
Fraser-Fort George	5953	Fraser-Fort George D	45.93	14.05	0.218	94,855
Peace River	5955	Fort St. John	55.09	10.80	0.22	54,790
Stikine Region	5957	Liard River 3	47.39	9.43	0.113	1,315
Fort Nelson-Liard	5959	Fort Nelson	54.31	54.38	0.196	5,710
Yukon	6001	Whitehorse	45.33	12.09	0.198	28,520
Baffin Region	6104	Fort Simpson	51.62	10.34	0.006	28,625
Keewatin Region	6105	Fort Smith	42.10	4.61	0.41	8,480
Fort Smith Region	6106	Yellowknife	43.18	8.74	0.323	14,325
Inuvik Region	6107	Inuvik	30.88	3.36	0.318	7,530

Source: Authors' calculations based on the IID

Note: This table presents mobility measures for all the Census divisions in Canada (except those with too few observations to divulge due to confidentiality constraints). $P_{1,5}$ in column (2) is expressed in percentage. See text for details on the mobility measures and sample. The population in column (4) refers to the population of the Census division from the 2001 Canadian Census of Population.

Table C3: Source and Definition, Additional Variables

Variable	Definition	Source
Fraction of single mothers	Number of single female households with children divided by total number of households with children	1996 Census[1]
Divorce rate	Fraction of people 15 or older who are divorced	1996 Census[1]
Fraction married	Fraction of people 15 or older who are married and not separated	1996 Census[1]
HS dropout rate, parents	Fraction of individuals aged 25 and above who do not have an high school degree	1996 Census[1]
College graduation rate	Number of individuals with a certificate or a bachelor degree divided by total population	1996 Census[2]
HS dropout rate, children	Fraction of individuals born in 1980 and 1982 who do not have an high school degree	2001 Census[3]
Fraction black	Number of individuals who are black alone divided by total population	1996 Census[1]
Segregation (black)	Multi-group Theil Index calculated at the census-tract level over four groups: White alone, Black alone, Hispanic, and Other, where Black is compared to all other groups	1996 Census[1]
Fraction foreign born	Share of CD residents born outside Canada	1996 Census[2]
Migration inflow (last year)	Migration into the CD in the last year from other CDs divided by CD population	1996 Census[2]
Migration inflow (5 years)	Migration into the CD in the last 5 years from other CDs divided by CD population	1996 Census[2]
CD departure rate	Number of children living in the CD in 1996 that have moved out of the CD by 2011 divided by total number of children living in the CD in 1996	IID core sample
CD arrival rate	Number of children living in the CD in 2011 that were not living in that CD in 1996 divided by total number of children living in the CD in 2011	IID core sample
Fraction aboriginal	Number of individuals who declare to be aboriginal divided by total population	1996 Census[1]
Segregation (aboriginal)	Multi-group Theil Index calculated at the census-tract level over four groups: White alone, Black alone, Aboriginal, and Other, where Aboriginal is compared to all other groups	1996 Census[1]
Number of reserves	Number of Census subdivisions within the CD that are designated as Indian reserves	1996 SGC[4]
Manufacturing share	Number of individuals over 15 years old working in the manufacturing industry divided by the population that are working and are 15 years old or more	1996 Census[2]
Oil and mining share, 2011	Number of individuals over 15 years old working in the mining, quarrying, and oil and gas extraction industries (NAICS 2007 2-digit code 21), divided by the population that are working and are 15 years old or more	2011 National Household Survey[5]
Teenage labor force part. rate	Fraction of children aged 15 to 17 having an income from work	1996 Census[1]
Fraction with commute under 15km	Number of workers that commute less than 15km to work divided by total number of workers. Sample restricted to workers that are 16 or older and not working at home	1996 Census[1]
Gini coefficient	Gini coefficient computed using parents of children in the core sample, with income top-coded at $100 millions in 2012 dollars	IID core sample

Note: This table presents the additional variables used in the correlation analysis and gives their definition and source. 1: refers to restricted-access microfiles from the 1996 Census of Population by Statistics Canada. 2: refers to the dataset-part A of the 1996 Census of Population by Statistics Canada, retrieved online from http://www12.statcan.gc.ca/datasets/Index-eng.cfm (accessed June 21, 2016). 3: refers to restricted-access microfiles from the 2001 Census of Population by Statistics Canada. 4: refers to Standard Geographical Classification (SGC) 1996 by Statistics Canada. 5: refers to the 2011 National Household Survey Census division profiles, Statistics Canada, retrieved online from http://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/prof/index.cfm?Lang=E (accessed October 16, 2016).

Table C4: Statistics by Cluster, Additional Variables

Variable	All	Cluster 1 Main	Cluster 2 Oil	Cluster 3 North
Fraction of single mothers	0.186	0.190	0.166	0.176
Divorce rate	0.0571	0.0579	0.0517	0.0602
Fraction married	0.432	0.433	0.445	0.395
HS dropout rate, parents	0.422	0.403	0.445	0.610
College graduation rate	0.148	0.156	0.133	0.0851
HS dropout rate, children	0.260	0.245	0.303	0.349
Fraction black	0.0203	0.0242	0.00749	0.00247
Segregation (black)	0.0643	0.0563	0.0807	0.124
Fraction foreign born	0.177	0.200	0.115	0.0386
Migration inflow (last year)	0.0549	0.0537	0.0583	0.0629
Migration inflow (5 years)	0.156	0.155	0.155	0.164
CD departure rate	0.247	0.245	0.223	0.326
CD arrival rate	0.259	0.257	0.276	0.251
Fraction aboriginal	0.0235	0.0149	0.0412	0.0870
Segregation (aboriginal)	0.0738	0.0626	0.0914	0.169
Number of reserves	4.120	3.744	4.170	8.529
Manufacturing share	0.155	0.165	0.0882	0.179
Oil and mining share, 2011	0.0140	0.00502	0.0570	0.0237
Teenage labor force part. rate	0.350	0.338	0.413	0.356
Fraction with commute under 15km	0.220	0.226	0.173	0.247
Gini coefficient	41.90	42.48	39.89	39.50
Number of CDs	287	174	48	65
Total population	29,634,264	23,298,564	4,396,540	1,939,160

Source: Authors' calculations based on the IID and additional data sources (see subsection 1.2)
Note: This table presents averages of the additional variables used in the correlations analysis, as well as the count of Census divisions and the total population, for the country overall and by cluster, as generated by the K-means algorithm described in the text using $K = 3$ on the Canadian Census divisions. The average figures are computed by taking means at the Census division level of the variable indicated, and computing a weighted average where the weight is the population of the Census division from the 2001 Canadian Census of Population. Total population and number of CD per cluster are not weighted.

Table C5: Correlations between Absolute Mobility and Various CD Characteristics (Unweighted)

	All Census divisions		Cluster 1 Main		Cluster 2 Oil		Cluster 3 North	
Variable	corr.	SE	corr.	SE	corr.	SE	corr.	SE
Fraction of single mothers	-.4757	.0523	-.4596	.067	-.5342	.1246	-.2734	.1245
Divorce rate	-.3333	.0561	-.3174	.0717	-.167	.1454	.0906	.1289
Fraction married	.3762	.0551	-.1342	.075	.2635	.1422	.4726	.1141
HS dropout rate, parents	.0715	.0591	.2963	.0728	.236	.1433	-.0233	.126
College graduation rate	-.1012	.0592	-.2975	.0722	-.415	.1341	.0496	.1293
HS dropout rate, children	-.2251	.0642	-.1818	.074	.1661	.1454	-.5579	.1065
Fraction black	-.0981	.0589	-.2584	.0737	-.05	.1473	.0359	.1259
Segregation (black)	.2163	.0578	.4161	.0693	.3407	.1386	.1342	.1248
Fraction foreign born	-.0955	.0593	-.345	.0709	-.1353	.1461	-.033	.1294
Migration inflow (last year)	.0162	.0595	-.1389	.075	.0661	.1471	-.1565	.1278
Migration inflow (5 years)	-.0237	.0595	-.195	.0742	.121	.1464	-.0462	.1293
CD departure rate	.3466	.0543	.3734	.0699	.3372	.1388	.5687	.0963
CD arrival rate	.1552	.059	-.0497	.0763	.2645	.1422	.3042	.1215
Fraction aboriginal	-.3088	.0563	.0895	.0759	-.206	.1443	-.554	.1049
Segregation (aboriginal)	.1855	.0582	.227	.0743	.116	.1464	.3248	.1192
Number of reserves	-.1139	.0593	-.1723	.0743	.0808	.147	-.1117	.1271
Manufacturing share	-.2477	.0577	.0593	.0756	-.2934	.141	.1469	.128
Oil and mining share, 2011	.4706	.0522	.2578	.0737	.3861	.136	.1574	.1244
Teenage labor force part. rate	.1813	.0583	-.2002	.0747	.3967	.1353	-.0125	.126
Fraction with commute under 15km	-.0856	.059	.0558	.0761	-.3463	.1383	.109	.1252
Gini coefficient	-.4311	.0534	-.4232	.0691	-.1873	.1448	-.6079	.1

Source: Authors' calculations based on the IID and additional data sources (see subsection 1.2)
Note: Each correlation comes from a univariate regression of standardized values of absolute mobility ($\bar{r}_{25}$) on the characteristic listed in the first column. The first set of regressions is estimated using the full sample of Canadian Census divisions, whereas the following three are done separately by cluster, as assigned using the K-means algorithm with $K = 3$.