The Effect of Cohort Size on Youth Earnings*

Louis-Philippe Morin†
University of Ottawa

April, 2011

Abstract

In this paper, I use data from both the Canadian Labour Force Surveys (LFS), and the 2001 and 2006 Canadian Censuses to estimate the impact of an important labour supply shock on high-school graduates earnings. The abolition of Ontario's Grade 13 generated a double cohort of high-school graduates that simultaneously entered the Ontario labour market, generating a large and sudden increase in the labour supply. This provides a rare occasion to measure the impact of cohort size on earnings without worrying about the supply shock being confounded with unobserved trends—a recurring problem in the literature. The census findings suggest that the effect of the supply shock is statistically and economically important, depressing weekly earnings by 5 to 9 percent. This effect is indeed important, especially if we consider that it is estimated two years after the double cohort. The LFS results suggest that the immediate impact of the supply shock—about six months after high school graduation—is larger in magnitude (between 14 and 25 percent).

Keywords: Labour Supply Shock, Youth.

JEL classification: J10, J20, J21.

*I would like to thank Michael Baker, Jean-Thomas Bernard, Pierre Brochu, Rob McMillan and Aloysius Siow for their helpful comments. I also thank Nyamekye Asare and Sergei Shibaev for their research assistance. The data used in this study were accessed on site at the Carleton, Ottawa, Outaouais Local Research Data Centre (COOL RDC). This study reflects the views of the author, not of Statistics Canada.

†Contact info: Louis-Philippe Morin, Department of Economics, University of Ottawa, Ottawa, Ontario, Canada, K1N 6N5. Telephone: (613) 562-5800 ext. 4867. Email: lmorin@uottawa.ca.

1 Introduction

Economists have studied the effects of cohort size on youth economic outcomes extensively following the entrance of baby boomers onto the labour market and the associated worsening of the youth labour market situation. Since cohort size does not vary substantially from one year to the next, studies (e.g. Welch (1979); Berger (1985, 1989); Macunovich (1999); and Korenman and Neumark (2000)) have focused on long term (typically 8-25 years) variations in cohort size.[1] One problem with this strategy is that it is hard to isolate cohort size effects from other unobserved trends which unrelated to demographics. This could explain why, in the 1980s, the situation of youth in the United States worsened while demographic conditions should have improved it (Blanchflower and Freeman 2000).

The 1997 Ontario secondary school reform allows me to shed light on how well the labour market can absorb a sudden influx of workers. In particular, this reform provides a rare occasion to measure the impact of cohort size on youth earnings without having to worry about the supply shock being confounded with unobserved trends. Following the abolition of Grade 13, two cohorts of high school graduates simultaneously entered the labour market in 2003, creating a large and sudden youth labour supply increase. Compared to 2001, the number of high school graduates increased by more than 30 percent in 2003.

The Ontario supply shock can, in terms of its intensity, be compared to an immigration shock. In his seminal 1990 paper, Card looked at the inflow of Cubans immigrants into Miami following the 1980 'Mariel Boatlift.' This study was followed by three studies looking at the impact of massive immigrant inflows, caused by political crises, into France, Portugal and Israel, respectively. In France and Portugal, those inflows were the result of the loss of colonies, while those into Israel followed the dissolution of the USSR (Hunt (1992); Carrington and de Lima (1996); and Friedberg (2001)). These studies found the supply shock effects to be very small or non-existent. Only Hunt (1992) found a modest effect of the influx of immigrants on native workers.

Although helpful in understanding the effect of immigration inflows on local labour markets, these studies cannot shed light on the potential effects of exogenous increases of local workers. In particular, it is possible that local workers and immigrants are poor substitutes. One advantage of the supply shock studied in this paper is that it is composed of potential workers *almost identical* to what would be referred to in the immigration literature as 'native workers.' This study can

[1] See Korenman and Neumark (2000) for an extensive review of the literature on cohort size and youth labour markets. See also Brunello (2010).

therefore shed light on the capacity of the labour market to absorb supply shocks without having tastes, skills, or preferences playing any confounding role in the determination of the outcome of interest.

I use two sources of information to estimate the impact of the double cohort of youth earnings. First, I use the 2001 and 2006 Canadian Census master files. The Canadian Censuses are very useful to estimate the effect of cohort size for (at least) two reasons: 1) the richness of the data makes it possible to get a measure of weekly earnings—something that is crucial if we are interested in the effect of cohort size on the price of labour—, and 2) it is the largest Canadian data set available to researchers. The large sample size makes precise estimation possible, even for very small subsample of the Canadian population (e.g. Ontario high school graduates born in 1984). The second source of data used in this paper consists of the 2002 and 2004 Labour Force Survey (LFS) master files. The LFS contains rich information on individuals' labour market conditions (e.g. hourly wage), and by observing individuals shortly after the double cohort, it allows me to estimate an 'immediate' impact of the double cohort on young workers.

My results suggest that a supply shock like the one created by the double cohort can significantly impact the labour market. The Census results suggest that the Ontario double cohort decreased weekly wages of workers who recently graduated from high school by between 5 and 9 percent. The magnitude of estimated impact of the supply shock increases as the control group (used in the regression estimation) is further away in age to the treatment group, indicating that workers similar in age to the double-cohort graduates may have been affected by the supply shock as well. The double cohort also affected the likelihood to be working full time and full year. By taking this last finding into account, I estimate the (lower and upper) bounds of the supply shock effect on wages to be -3 and -16.5 percent. Finally, the Census findings are corroborated by the LFS results which indicate that the immediate (six months after the shock) impact of the double cohort was to depress wages by 14 to 25 percent.

The next section describes the Ontario double cohort and its potential consequences for the estimation of the cohort size effect. I describe the two sources of data used in this paper in Section 3. The estimation strategy is presented in Section 4. Section 5 presents the findings from the Census data followed with the findings from the LFS. Section 6 looks at the impact of the supply shock on employment, and its consequences for the estimation results presented in Section 5. Section 7 concludes.

2 The Ontario Double Cohort and Labour Supply

In 1997, the provincial government of Ontario introduced an important reform to its secondary school system. The centerpiece of this reform was the compression of the curriculum from five to four years. It brought the length of Ontario's secondary school curriculum into line with most surrounding provinces. Starting in 1999, students would now be expected to graduate from high school after four years (after Grade 12) instead of five.[2] An inevitable consequence of this reform was that, in 2003, both the first cohort from the new curriculum and the last cohort from the old curriculum graduated from high school in the same year, creating a double cohort of high school graduates. Since students graduate from secondary school almost simultaneously across the province, one would expect the labour supply shock caused by the double cohort to be important and concentrated within a short time span.

Figure 1 shows the number of recent high school graduates[3] aged 17 to 19 between 1999 and 2006 for Ontario and the Rest-of-Canada (henceforth RoC). The number of recent graduates jumped by 34.1 percent (from 91,291 to 122,406) between 2001 and 2003 in Ontario, while only increasing by 0.6 percent in the RoC over the same period.[4] The drastic contrast in growth rates in recent high school graduates, combined with an economic climate of stability in Canada over this period, will allow me to clearly identify the effect of an increase in cohort size on youth earnings.[5]

3 Data

In order to estimate the impact of the double cohort on youth earnings, I combine information from the Canadian Labour Force Survey (LFS), and the 2001 and 2006 Canadian Censuses (long form survey). Both sources of information will complement each other as the Censuses contain a very large number of observations, while the LFS contains detailed labour force information, and allows one to observed graduates shortly after having graduated from high school.[6] The large sample size

[2] For studies looking at the impact of the Ontario secondary school reform on academic performance of college-bound students, see Krashinsky (2006) and Morin (2010). Krashinsky (2009) looks at the return to Grade 13. See King et al. (2002, 2004, 2005) for details about the the reform.

[3] Recent high school graduates are individuals who had graduated from secondary school at the time of the first Labour Force Survey interview, but who were attending secondary school in the previous March.

[4] The fact that the growth rate for the RoC increases in 2005 should not have a significant impact on the estimation of the cohort size effect on Ontario graduates since I concentrate on individuals who were not enrolled in school in 2005. If anything, it would bias the cohort size parameter estimator downward.

[5] Between 2000 and 2005, the average real GDP growth rates for Ontario and Canada were 2.3 and 2.5 percent, respectively. Source: Statistics Canada Table 384-0002.

[6] This strategy has been used before by Lemieux and Milligan (2008) for estimating the effect of social assistance on a variety of labour market outcomes (e.g. employment and annual earnings). They use the LFS to complement

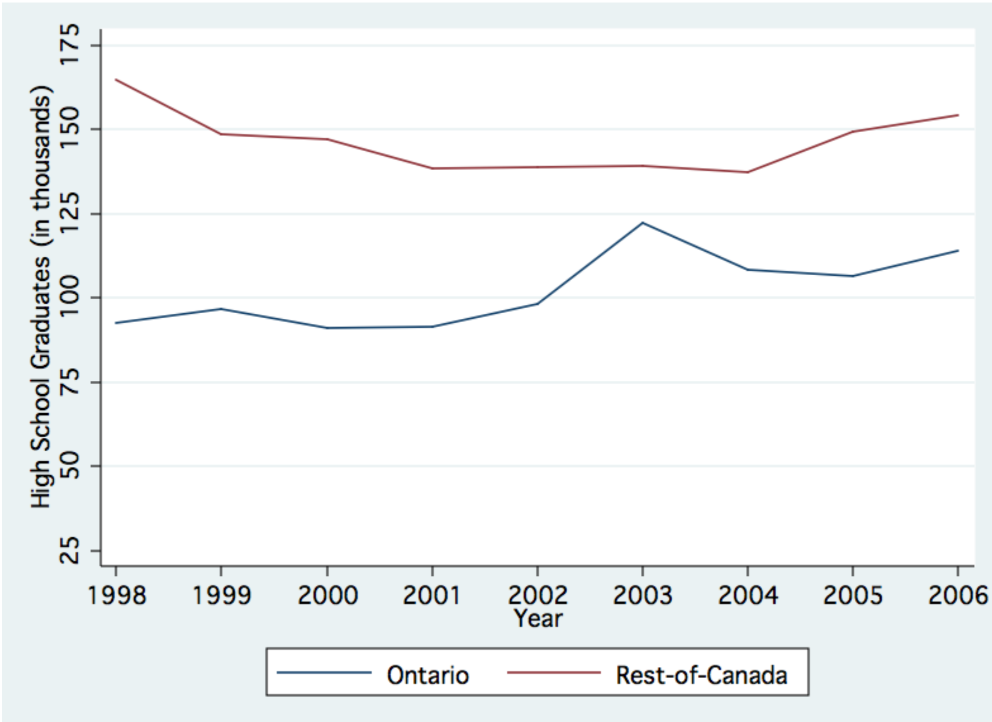

Figure 1: Number of New High School Graduates per Year

of the Censuses will prove to be very helpful as the population of interest (i.e. 2003 Ontario high school graduates who did get post-secondary education) represents a small fraction of the Canadian population.

3.1 Census Data

The main findings of this paper are based on the Canadian Census master files. The 2001 and 2006 Census long-form questionnaires target approximately 20 percent of the Canadian households.[7] There are many advantages to using the census master files when looking at the impact of the double cohort on youth earnings. First, the census master files are the largest Canadian data sets available to researchers containing both detailed information on the respondents' earnings and education level. Since the main effect of the double cohort should be concentrated on a small

their Census results for the exact same reasons: 1) The large sample size of the Census data allows them to study a small subsample of the Canadian population, and 2) Since the LFS is conducted monthly Lemieux and Milligan (2008) observe individuals soon before, and soon after a policy change affecting social assistance.

[7] The 2001 and 2006 long-form questionnaires are available at http://www.statcan.gc.ca/imdb-bmdi/instrument/3901_Q2_V2-eng.pdf and http://www.statcan.gc.ca/imdb-bmdi/instrument/3901_Q2_V3-eng.pdf, respectively.

fraction of the Canadian population (i.e. Ontario high school graduates born in 1984 and 1985), the size of the census could be crucial to compute any meaningful statistics.

Second, the long-form questionnaire is rich enough in terms of individuals' labour market activities to get a measure of one's price of labour. For example, it contains information on the labour force status, the number of weeks worked last year (e.g. 2005 for the 2006 Census), whether the individual mainly worked full- or part-time during these weeks, and their annual wages and salaries for the last year. Third, the master files contain the year of birth of the individuals, and not simply the age on the day on the survey. Since, Ontario uses December 31st as cutoff date to determine when a child can enroll in primary school, it is straightforward to identify who is expected to have graduated from high school in 2003, and most importantly, who should be a Grade 12 graduate (as opposed to Grade 13). Grade 12 graduates are excluded from the analysis, in order to avoid having the effect of the cohort size being confounded with a potential (lack of) Grade 13 effect; Grade 12 students might have a different level of human capital than Grade 13, including them in the analysis might bias the results.[8] Finally, the Census also contains information on gender, educational attainment, visible minority status, immigrant status, marital status, the province of residence (now, one year ago, and five years ago), and workers' industry sector. This information will be used to identify the treatment and potential control groups (e.g. using age, province of residence, and educational attainment) and as controls in the regression analysis (e.g. gender, immigrant status).

The main variable of interest is the (log of) weekly wages earned in the year prior to the census. Annual wages (i.e. gross wages and salaries before deductions) are adjusted using the provincial consumer price indices to be expressed in 2000 dollars, and divided by the number of weeks worked in the year prior to the census to represent weekly wages.

The goal of the paper is to estimate the impact of the increased cohort size on wages (i.e. the price of labour). Therefore, I make a series of restrictions to help the identification of this impact. First, I avoid having education playing any role in the wage determination by discarding Grade 12 graduates, and by focusing on individuals with a high school diploma, but no further schooling. I further concentrate the analysis to full-time[9] workers as is done in studies where the number of

[8] Krashinsky (2009) estimates the return to Grade 13 using the Survey of Labour and Income Dynamics (SLID). His OLS estimates suggest that Grade 13 graduates earned on average 10 percent more than Grade 12 graduates while the IV estimates are significantly larger (i.e. 19 percent), one year after graduation. The effect disappears two years after graduation, except for individuals earning less than the median wage for which the return is around 5 percent.

[9] The Canadian Census and the Labour Force Survey define working full time as working 30 hours or more a week.

hours worked is not perfectly observed (e.g. Katz and Murphy (1992), Card and Lemieux (2001), and Boudarbat, Lemieux and Riddell (2010)). In order to focus on high school graduates who had fully entered the labour market, I restrict the sample to individuals who did not go to school, and worked 48 weeks in the year prior to the census.[10] Finally, I discard individuals with weekly wages of less than $75 in 2000 dollars.[11]

3.2 LFS Data

The January Labour Force Surveys complement the Census data as it allows me to concentrate on the very narrow group of individuals who should be most affected by the reform, Grade 13 graduates who entered the labour market a few months following the double cohort. I rely on the January surveys for two reasons. First, since we only know the age of respondents in the LFS—as opposed to their year of birth in the Census—we can only disentangle Grade 13 from Grade 12 graduates in January. Hence, in January 2004, Grade 13 graduates should be 19 years old, while Grade 12 should be 18 years old. Second, full-time workers observed in January occupy regular jobs as opposed to a mix of regular and summer jobs for months immediately following usual high school graduation dates. Labour supply for summer jobs might be only driven by demographics (e.g. the number of individuals aged between 15 and 19) and not on schooling attainment.[12]

Aside from allowing me to observe double-cohort graduates only a few months after their graduation, the LFS offers another advantage over the Census data. The LFS has information about workers' hourly wages, giving me a direct measure of the price of labour. Like the Census, the LFS contains information on gender, educational attainment, marital status, the province of residence, and workers' industry sector, which will be used to identify the treatment and potential control groups and as controls in the regression analysis. Although there is no information about race or immigrant status in the LFS prior to 2006, the Census results suggest that the inclusion of these personal characteristics does not affect the estimated cohort effect. I restrict the LFS sample to individuals who are not enrolled in school, that have a high school diploma (but no further schooling), and work full time (30 hours or more a week). I discard individuals with hourly wages less

[10] The results obtained from looking at individuals who worked 26 weeks or more, or 39 weeks or more (three quarters of the year) are very similar to the ones presented in this paper.

[11] The same restriction is used by Boudarbat, Lemieux and Riddell (2010). The cutoff roughly represents half of the minimum wage on a 30-hour week. A similar restriction is also used in Katz and Murphy (1992).

[12] If this is the case, the labour supply shock for 2003 summer jobs should be close to null, and wages for summer jobs should not be affected as much as wages for permanent jobs.

than \$2.5 in 2000 dollars, which is consistent with the weekly-wage cutoff of \$75. The Appendix presents more details on the data construction and restrictions.

Finally, I use the August LFS to get estimates of the cohort size (see Figure 1).[13] Between the months of May and August, LFS respondents aged 15 to 24 are asked if they were in school in the previous March. Since students can graduate from secondary school during the summer, the August survey has the advantage of including many recent graduates, giving a better picture of the expected increase in labour supply to come.[14] In this paper, a cohort of graduates is composed of 17, 18 and 19 year-old individuals who had graduated from high school when first interviewed by the LFS, and who were full-time students in March of the same year in a secondary school institution.[15] I estimated cohort sizes using recent high-school graduates aged 17 to 19 to include both Grade 12 and Grade 13 graduates.

4 Estimating the Impact of Cohort Size on Earnings

Basic economic theory predicts that a positive supply shock should negatively affect wages. We would therefore expect to observe lower wages for individuals who were part of the double cohort as compared to a more 'normal' cohort of high school graduates, after controlling for other factors affecting individual wages. A major difficulty faced by researchers is that other types of shocks—unrelated to cohort size—can occur around the time of the cohort-size increase. This is especially true when observing individuals over long periods of time. Here, the short time span over which individuals are observed (five years in the case of the Census data, and two years in the case of the LFS data), and the magnitude of the cohort size increase should mitigate this difficulty.

Although Ontario's economy grew at a steady pace (and did not experienced any significant downturn) in the early 2000's, there are two (potential) demand shocks that must be accounted for when trying to identify the cohort-size effect: a demand shock that affects all Ontario workers, and one that only affects young graduates across but across all of Canada.[16]

[13]The Minister of Education and Training grants diplomas at any time during the year to students who have successfully completed the necessary secondary school requirements. Hence, there is no specific month where all eligible students graduate from high school. Nevertheless, most students complete Ontario Secondary School Diploma (OSSD) requirements by the end of the spring.

[14]The August LFS is the last one of the year which specifically asks the respondent whether she was enrolled in high school in March of the present year, allowing me to differentiate recent high school graduates from the previous year's graduates.

[15]Each household interviewed in the LFS remains in the sample for six consecutive months. Questions about educational attainment are only asked during the first interview.

[16]Ontario's real GDP grew at an average rate of about 2.3 percent a year between 2000 and 2005. It experienced slightly slower growth rates in 2001 and 2003, but so did the Rest of Canada.

The identification strategy in this study is to disentangle these two types of shocks (affecting all Ontario workers and affecting all young Canadian workers) from the labour supply shock following the double cohort, using both workers from Ontario, who were presumably not affected by the supply shock, and recent high school graduates from other provinces as controls. In particular, I use triple-difference estimation which essentially compares wage gaps between a control group (e.g. experienced workers) and recent high-school-graduate workers across provinces and across time. Shocks specifically affecting recent high school graduates across Canada can be controlled for by comparing wages of Ontario recent secondary school graduates to wages of similar workers in other provinces. Demand shocks affecting Ontario can be captured by comparing wages of recent high-school graduate Ontario workers to wages of other Ontario workers, who should not be affected by the increase in cohort size, at least in the short run, but who should be affected by demand shocks.[17] Triple-difference estimation allows me to control for these two types of shocks simultaneously. After controlling for the potential effect of labour market conditions unrelated to the double cohort and personal characteristics, differences in outcomes of inexperienced workers before and after the double cohort should be due to the increase in the number of recent high school graduates.

The implementation of a triple-difference estimation is straightforward. The difficulty in the estimation comes from choosing a group of workers affected by demand shocks in a similar fashion to recent high school graduates while not being affected by the supply of this type of labour. The next sub-section presents details about the triple-difference estimation technique and different control groups used to estimate the effect of a supply shock on wages.

4.1 Estimation Strategy

The triple-difference estimation strategy is represented in a regression framework by the following equation:

$$ln(w_{igpt}) = \eta_{gp} + \lambda_{gt} + \phi_{pt} + \beta(DC_t \times Youth_g \times ON_p) + \boldsymbol{X_{igrt}\gamma} + \varepsilon_{igpt} \quad (1)$$

where i represents an individual, g a group of workers (e.g. recent high school graduate), p a province, and t represents time. $ln(w_{igpt})$ is the log of the weekly, or hourly wages, depending on the specification. $Youth_g$ is a dummy variable equal to 1 if the individual is a recent high school

[17] See Shimer (2001) for evidence on the impact of young workers on unemployment and labour force participation.

graduate, and 0 otherwise. DC_t is a dummy variable equal to 1 if the individual is observed after the double cohort, 0 otherwise, while ON_p is equal to 1 in the individual resides in Ontario. Therefore, the $DC_t \times Youth_g \times ON_p$ term represents the 'treatment' group: Ontario high school graduates who entered the labour market following the double cohort. β captures the effect of the double cohort on youth earnings. η_{gp}, λ_{gt}, and ϕ_{pt} allow for the possibility that 1) the groups of workers have been affected differently by (demand) shocks across time (e.g. between 2000 and 2005 when using the Census data); 2) the average wage might differ across worker groups and that this difference might differ across provinces; 3) there were province specific shocks across time. Finally, $\boldsymbol{X_{igpt}}$ is a vector of personal characteristics (e.g. gender, race, marital status, worker industry sector) that will be used to verify the robustness of my results.

If one believes that any type (e.g. experienced versus inexperienced, or skilled versus unskilled) of labour can be considered (to some extent) as a substitute to another labour type, then there is no perfect control group. Recall that the perfect control group would be affected by demand shocks in a similar way as recent high school graduates, while not being affected by the increase supply of high school graduates. To address the possibility of having a less than perfect control group, I estimate equation (1) using different control groups to see whether the estimates vary significantly from one specification to another. I consider workers with a high school degree—the same level of education as the treatment group—but from different age groups and provinces as potential control groups. The idea is that more experienced workers are less likely to be close substitutes to recent high school graduates, but would still be affected by labour demand shocks. When analyzing the Census data, I divide the workers into six age groups: 21 years old (youth), 25 to 29, 30 to 34, 35 to 39, 40 to 44, and 45 to 49 years old. The choice of the control group could be crucial if the Ontario economy had experienced a major expansion or recession during the years surrounding the double cohort. Fortunately, this is not the case. Ontario did not experience a major recession or boom between 2001 and 2006 but did experience a slower growth rate in 2003 than the rest of the country. Using different control groups will allow me to check whether this slow-down could affect the estimation of the supply shock effect. If demand shocks are not important, one would expect to have similar estimates for the effect of the double cohort on wages of control groups which are not substitutes for high school graduates.

5 Results

Before presenting the results from estimating equation (1), it is worthwhile to present summary statistics on the evolution of the average average weekly wages between 2000 and 2005. Table 1 presents average weekly wages (in 2000 dollars) by age group and region (Ontario versus the RoC) for full-time, full-year workers. The number of observations for each group is presented in square brackets. One can notice an important strength of the Census data: its large sample size. For both the 2001 and 2006 Censuses, I observe more than 2,000 full-time, full-year Ontario workers that are 21 years of age and have a high school diploma. The second striking finding from Table 1 is that the average weekly wages of young Ontario workers actually decreased by 7.4 percent between 2000 and 2005. This is especially surprising given that Ontario's economy expanded at a fairly steady rate over this period. This sharp decrease in wages is by far the most significant among all worker groups considered in Table 1. In the absence of any other shock to Ontario's economy, this drop in wages is indicative of a significant labour-supply effect. Further supporting the (substantial) cohort size shock idea is the fact that, young workers in the RoC saw their wages increased by 3.5 percent (an average annual growth rate of 0.7 percent) which is comparable to the growth rates of most of the worker age groups in the RoC, and of older workers in Ontario. In the RoC, only workers aged 30-34 and 40-44 did not a see (statistically) significant increase in their wages. Interestingly, Ontario workers aged 25-29 also saw their wages decrease between 2000 and 2005 indicating that this group of workers was not totally isolated from the supply shock. More generally, we can see the wage growth rates improve with age in Ontario, suggesting a lower level of substitutability. Note that this conjecture is further supported by the fact that we do not observe this trend in the RoC. Overall, the information found in Table 1 points toward a large impact of the double cohort on wages.

5.1 Census Regression Results

Table 2 presents the regression results from estimating equation (1) using the Census data, and workers aged 25 to 29 as the control group. Specification (1) only includes a set of fixed effects and interaction terms for time, province, and worker-group (η_{gp}, λ_{gt}, and ϕ_{pt} in equation (1)), along with the $DC_t \times Youth_g \times ON_p$ dummy variable.[18] Recall that $DC_t \times Youth_g \times ON_p$ is meant to capture the effect of the double cohort on the wages of young Ontario workers. Specification (2)

[18] The province used as a base (the omitted provincial dummy variable) in equation (1) is Québec.

Table 1: Average Weekly Wages of Full-Time, Full-Year Workers (Census Data)

	Ontario			Rest of Canada		
Weekly Wages	2000	2005	Difference	2000	2005	Difference
Youth	426.28	395.27	-7.4%***	389.75	403.40	3.5%***
	(199.86)	(179.57)		(198.32)	(206.48)	
	[2,210]	[2,110]		[3,590]	[4,095]	
Aged 25–29	645.32	620.25	-3.9%***	591.43	587.10	-0.7%
	(362.04)	(312.97)		(326.54)	(323.28)	
	[15,945]	[15,440]		[19,435]	[21,505]	
Aged 30–34	743.61	734.73	-1.2%	664.04	693.89	4.5%***
	(769.95)	(569.59)		(603.14)	(807.00)	
	[19,245]	[16,685]		[24,485]	[21,720]	
Aged 35–39	797.35	797.01	-0.0%	705.46	728.57	3.3%***
	(589.88)	(761.73)		(447.08)	(643.68)	
	[24,035]	[21,145]		[33,485]	[26,330]	
Aged 40–44	842.13	846.26	0.5%	747.23	756.78	1.3%
	(859.48)	(890.81)		(826.85)	(601.02)	
	[24,445]	[27,520]		[37,205]	[36,920]	
Aged 45–49	859.86	887.11	3.2%***	765.28	799.56	4.5%***
	(689.12)	(959.58)		(554.79)	(929.60)	
	[20,090]	[25,475]		[30,795]	[37,525]	

Notes: The average wages are expressed in 2000 dollars. Standard deviations are in parentheses. The observations are weighted using the Census weights. The number of observations, rounded to a base of 5, are in square brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

adds personal characteristics (i.e. gender, immigrant status, a visible minority indicator, a rural area indicator, and multiple indicators for marital status) to the regression equation. Specification (3) includes industry fixed effects (based on 20 sectors of activity), while specification (4) allows for the personal characteristics to have differential effects and the industry fixed effects to vary across worker groups. All Census regression estimation results presented in in this paper are done using weights—in the case of the Canadian Census, only weighted estimation results can be released to the public.[19]

The results coming out of Table 2 suggest that the double cohort had a statistically and economically significant impact on wages. All else equal, workers from the double cohort are earning on average about 6 percent less than similar workers who were part of a normal double cohort. Adding control variables does not materially affect any of the estimates. In particular, the estimate of the double cohort effect ranges from -7.2 percent when only including basic controls to -5.7 percent when industry fixed effects are included. The only estimate that seems to be affected by the

[19]Unweighted regression results are almost identical.

inclusion of covariates is the Youth coefficient. By including interaction terms between the industry fixed effects and the youth dummy, the coefficient estimate in specification (4) captures the wage gap between younger and older workers in the 'omitted' industry (agriculture, forestry, fishing and hunting).

Table 2: Census Results Using Workers Aged 25-29 as the Control Group (Weekly Wages for Full-Time, Full-Year Workers)

	(1)	(2)	(3)	(4)
Youth	-0.390***	-0.383***	-0.336***	-0.435***
	(0.013)	(0.013)	(0.012)	(0.151)
DC	-0.003	-0.002	0.004	0.005
	(0.009)	(0.009)	(0.008)	(0.008)
DC × ON	-0.046***	-0.047***	-0.043***	-0.043***
	(0.011)	(0.010)	(0.010)	(0.010)
DC × Youth	0.037***	0.030**	0.025**	0.025**
	(0.013)	(0.013)	(0.012)	(0.012)
DC × Youth × ON	-0.072***	-0.063***	-0.057***	-0.060***
	(0.022)	(0.021)	(0.020)	(0.020)
Province Fixed Effects	Yes	Yes	Yes	Yes
Province × DC Fixed Effects	Yes	Yes	Yes	Yes
Province × Youth Fixed Effects	Yes	Yes	Yes	Yes
Controls for Personal Characteristics	No	Yes	Yes	Yes
Industry Fixed Effects	No	No	Yes	Yes
Personal Characteristics × Youth Fixed Effects	No	No	No	Yes
Industry × Youth Fixed Effects	No	No	No	Yes
R^2	0.09	0.17	0.25	0.25
N	84,330	84,330	84,330	84,330

Notes: Notes: The sample consists of individuals who worked 48 weeks or more during the year prior to the Census and worked full time during these weeks. 'Youth' is an indicator variable that equals 1 if the individual was 21 years of age on January 1st of the Census year. The omitted provincial dummy variable is Québec. All sampled individuals have a high school diploma, but no further schooling. The personal characteristics include: gender, immigrant status, a visible minority indicator, a rural area indicator, and multiple indicators for marital status. The industry fixed effects reflect 20 sectors of activity (based on NAICS). The estimation was done using Census weights. Robust standard errors are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

There are a few other findings worth mentioning. Workers aged 21 earned on average 34 percent less than workers aged 25 to 29 (from specification (3)) in 2000. This wage gap is fairly constant across provinces as only New Brunswick and Alberta have wage gaps that are statistically significantly different from Québec.[20] The wage gap decreased by 2.5 percent between 2000 and

[20]The differences in wage gap are 6.6 and 2.8 percent for New Brunswick and Alberta, respectively.

2005. Finally, the average real weekly wage of workers aged 25-29 did not change in Québec, but did decrease significantly in Ontario by 4.3 percent.

This last finding is interesting as it suggests that, relative to Québec, the Ontario economy slowed down between 2000 and 2005. This is somewhat surprising as both Québec and Ontario experienced stable unemployment rates (for their population aged 15 and over) over this period. When looking at the results presented in Table 2, one has to keep in mind that the control group used in this table is composed of workers very close in age (and in terms of educational attainment) to the double-cohort graduates. It is quite possible that these two types of workers are substitutes in the eyes of employers (see Card and Lemieux (2001)). Therefore, what looks like a slowdown of the Ontario economy could actually be (at least in part) the impact of the increased supply in young workers on workers aged 25 to 29.

Table 3 compares the estimates of the effect of the increased cohort size for different aged-based control groups. The first column reports the results from specification (4) in Table 2. The next columns presents the results from estimating the same specification, but for workers aged 30-34, 35-39, 40-44, and 45-49 respectively. One can clearly see that, as we move from younger to older control groups, the estimated effect of the double cohort increases significantly. When using workers aged 45 to 49, the estimated effect is -9.1 percent. At the same time, one can see that the difference in wage growth rates between Québec and Ontario shrinks as we use older workers as control groups. For both workers aged 40-44 and 45-49, the difference is very close to zero, and is no longer statistically significant. These results supports the idea that similarly educated workers are seen as substitutes with the level of substitutability decreasing as age separating the workers increases.

5.2 LFS Results

The results found using the Census data suggest that cohort size has a significant negative impact on wages. When using workers that arguably have low substitutability with workers aged 21, the estimated effect of the supply shock is around -9 percent. This effect is economically large if we consider that we observe young workers two years after the double cohort. Since there is five years between the two censuses, one could argue that other shocks, unrelated to the double cohort, could be driving the results. Although this is unlikely given the stability of the Canadian economy over this period, the LFS allows me to estimate the immediate impact of the cohort size using two surveys that are only two years apart (the 2002 and 2004 January LFS). The estimation strategy

Table 3: Double Cohort Effect and Aged-Based Control Groups (Weekly Wages for Full-Time, Full-Year Workers)

Age Group	25–29	30–34	35–39	40–44	45–49
DC × Youth × ON	-0.060***	-0.049***	-0.074***	-0.084***	-0.091***
	(0.020)	(0.020)	(0.020)	(0.019)	(0.020)
DC × ON	-0.043***	-0.030***	-0.017**	-0.004	0.002
	(0.010)	(0.010)	(0.008)	(0.007)	(0.008)
Province Fixed Effects	Yes	Yes	Yes	Yes	Yes
Province × DC F. E.	Yes	Yes	Yes	Yes	Yes
Province × Youth F. E.	Yes	Yes	Yes	Yes	Yes
Controls for Personal Characteristics	Yes	Yes	Yes	Yes	Yes
Industry F. E.	Yes	Yes	Yes	Yes	Yes
Personal Characteristics × Youth F. E.	Yes	Yes	Yes	Yes	Yes
Industry × Youth F. E.	Yes	Yes	Yes	Yes	Yes
R^2	0.25	0.29	0.29	0.30	0.32
N	84,330	94,140	117,000	138,095	125,890

Notes: The sample consists of individuals who worked 48 weeks or more during the year prior to the Census and worked full time during these weeks. 'Youth' is an indicator variable that equals 1 if the individual was 21 years of age on January 1st of the Census year. The omitted provincial dummy variable is Québec. All sampled individuals have a high school diploma, but no further schooling. The personal characteristics include: gender, immigrant status, a visible minority indicator, a rural area indicator, and multiple indicators for marital status. The industry fixed effects reflect 20 sectors of activity (based on NAICS). The estimation was done using Census weights. Robust standard errors are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

is exactly the same as the one used with the Census data. The main difference is that the 'Youth' worker group is composed of 2001 and 2003 high school graduates born in 1982 and 1984, as opposed to 1979 and 1984, when using the Census data. By using the January 2004 LFS, I can observe the wages of young workers only a few months after their graduation.

Table 4 presents the regression results from estimating equation (1) using the LFS data, and workers aged 23 to 27 as control group. I present the results from using the log of hourly wages as dependent variable.[21] The four specifications in Table 4 are the same as in Table 2 with the exception that the LFS data do not contain information about race or immigrant status.

The results from Table 4 corroborate the Census results, suggesting that the supply shock had a significant effect on youth wages. The estimated double-cohort effect is larger in magnitude than when analyzing the Census data, but it is also less precisely estimated. This is not surprising given the smaller sample size in the LFS. Workers from the double cohort earn on average about 23 percent less than similar workers who were part of a normal cohort. The cohort size effect is

[21] Workers aged 23 to 27 in January 2002 or 2004 would be part of the workers aged 25 to 29 in the Census data. I use the log of hourly wages, as it is probably a better measure of the price of labour for young high school graduates, but the results are very similar for weekly wages.

roughly half of the increased in cohort size between 2001 and 2003 (34.1 percent), suggesting that the labour market reacted strongly to the supply shock, at least in the very short run.

Table 4: LFS Results Using Workers Aged 23-27 as the Control Group (Hourly Wages for Full-Time Workers)

	(1)	(2)	(3)	(4)
Youth	-0.401***	-0.380***	-0.364***	-0.551***
	(0.044)	(0.044)	(0.045)	(0.088)
DC	-0.026	-0.038	-0.040	-0.044
	(0.053)	(0.051)	(0.052)	(0.052)
DC × Youth	0.112***	0.115***	0.118***	0.116***
	(0.039)	(0.038)	(0.038)	(0.037)
DC × Youth × ON	-0.219***	-0.223***	-0.229***	-0.232***
	(0.075)	(0.067)	(0.067)	(0.068)
Province Fixed Effects	Yes	Yes	Yes	Yes
Province × DC Fixed Effects	Yes	Yes	Yes	Yes
Province × Youth Fixed Effects	Yes	Yes	Yes	Yes
Controls for Personal Characteristics	No	Yes	Yes	Yes
Industry Fixed Effects	No	No	Yes	Yes
Personal Characteristics × Youth Fixed Effects	No	No	No	Yes
Industry × Youth Fixed Effects	No	No	No	Yes
R^2	0.24	0.31	0.38	0.39
N	1,910	1,910	1,910	1,910

Notes: The sample is composed of full-time workers. 'Youth' is an indicator variable is equal to 1 if the individual is 19 during the LFS reference week. All individuals in the sample have a high school diploma, but no further schooling. The personal characteristics include: gender, a rural area indicator, and marital status. The industry fixed effects are constructed using 9 sectors of activity (based on NAICS). The estimation was done using the LFS weights. Robust standard errors are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5 compares the estimates of the effect of the increased cohort size for different aged-based control groups. Unlike the results coming out of the Census data, the estimates do not show a clear pattern across age groups, and not surprisingly, the estimates also fluctuate more in Table 5 than in Table 3. The estimates fluctuate between -14 and -25 percent. Nevertheless, all estimates are statistically significant and far from zero, indicating that the choice of the control group is not critical. Overall, despite being less precise than the Census results, the LFS results suggest that the double cohort had a significant impact of the youth labour market. In the next section, I look at the impact of the labour supply shock on the probability to be a full-time (and full-year) worker, and its potential consequences for my estimation.

Table 5: Double Cohort Effect and Aged-Based Control Groups (Hourly Wages for Full-Time Workers)

Age Group	23–27	28–32	33–37	38–42	43–47
DC × Youth × ON	-0.232***	-0.253***	-0.207***	-0.143**	-0.173***
	(0.068)	(0.075)	(0.069)	(0.068)	(0.066)
DC × ON	0.039	0.055	-0.053	-0.049	-0.019
	(0.060)	(0.069)	(0.061)	(0.055)	(0.048)
Province Fixed Effects	Yes	Yes	Yes	Yes	Yes
Province × DC F. E.	Yes	Yes	Yes	Yes	Yes
Province × Youth F. E.	Yes	Yes	Yes	Yes	Yes
Controls for Personal Characteristics	Yes	Yes	Yes	Yes	Yes
Industry F. E.	Yes	Yes	Yes	Yes	Yes
Personal Characteristics × Youth F. E.	Yes	Yes	Yes	Yes	Yes
Industry × Youth F. E.	Yes	Yes	Yes	Yes	Yes
R^2	0.31	0.36	0.35	0.42	0.34
N	1,920	1,820	2,235	2,815	3,005

Notes: The sample is composed of full-time workers. 'Youth' is an indicator variable is equal to 1 if the individual is 19 during the LFS reference week. All individuals in the sample have a high school diploma, but no further schooling. The personal characteristics include: gender, a rural area indicator, and marital status. The industry fixed effects are constructed using 9 sectors of activity (based on NAICS). The estimation was done using the LFS weights. Robust standard errors are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

6 Increased Cohort Size and Employment

When interpreting the results coming out of the Census Data, one should keep in mind that I restricted the sample to full-time, full-year (FTFY) workers. It is quite possible that the FTFY status itself might have been affected by the supply shock. In fact, the fraction of FTFY workers among Ontario youth decreased by 1.6 percentage points between 2000 and 2005.[22] This difference is both statistically and economically significant since the fraction of FTFY workers was 19.5 percent in 2000. I investigate the potential impact of the double cohort on the likelihood to be a FTFY worker by estimating (by OLS) equation (1), using a FTFY dummy as dependent variable (instead of the log of wages).

There does not appear to be any change in the likelihood of being FTFY when using individuals aged 25–29 as control group, but a significant difference appears as we move to older control groups. The estimates obtained when I use individuals aged 35–39, 40–44, or 45–49 are in the vicinity of the 1.6 percentage point difference when simply looking at the change in proportions. This is interesting as it suggests the same age-based pattern (for the impact of the supply shock) as the one found when looking for wages. It is quite possible that the FTFY status of workers aged 25–29

[22]This fraction is obtained by dividing the number of 21 year-olds with a high school diploma that work full-time, full-year by the total number of 21 year-olds with a high school diploma.

Table 6: Double Cohort and Full-Time, Full-Year Status

Age Group	25–29	30–34	35–39	40–44	45–49
DC × Youth × ON	0.005	-0.009	-0.015*	-0.018**	-0.012
	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Province Fixed Effects	Yes	Yes	Yes	Yes	Yes
Province × DC Fixed Effects	Yes	Yes	Yes	Yes	Yes
Province × Youth Fixed Effects	Yes	Yes	Yes	Yes	Yes
Controls for Personal Characteristics	Yes	Yes	Yes	Yes	Yes
R^2	0.13	0.15	0.15	0.15	0.15
N	247,930	255,135	291,505	321,385	295,360

Notes: The dependent variable is a dummy variable equal to 1 if the individual worked full-time, full-year in the year prior to the Census. 'Youth' is an indicator variable that equals 1 if the individual was 21 years of age. All sampled individuals have a high school diploma, but no further schooling. The personal characteristics include: gender, immigrant status, a visible minority indicator, a rural area indicator, and multiple indicators for marital status. The industry fixed effects reflect 20 sectors of activity (based on NAICS). The estimation was done using the Census weights for the individuals universe. Robust standard errors are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

was affected by the supply shock, suggesting that there is a certain level of substitution between similarly educated workers given that they are close in age.

The supply shock not only seems to have decreased wages of FTFY workers, but also the likelihood to be working FTFW. Therefore, the results presented in Table 3 may not be capturing the average treatment effect. Given the absence of an obvious valid instrumental variable, I compute 'worst-case' scenario bounds on the average treatment effect based on Lee (2009).[23] The estimated upper and lower bounds for the average treatment effect on wages are -16.5 and -3.1 percent, respectively. The fact that the lower bound for the effect is negative is somewhat surprising, given that it is computed under a 'worst-case' scenario. That is, I compute the lower bound assuming that the individuals for which the FTFY status was affected by the supply shock would have had the largest wages in the absence of the shock. In practice, this would be very surprising, given the positive correlation between weekly wages and the number of weeks worked found in the Census data. In the end, the estimation results based on 'worst-case' scenarios emphasize the important impact of the supply shock on wages.

[23]Lee (2009) when applying his methodology to the analysis of the Job Corps program trims the treatment group data, as the program is assumed to *positively* affect both wages and the probability to be employed. Since the supply shock is expected to have *negative* impact on both the FTFY status and wages, I trim the 'control' group data instead of trimming the treatment group data. In particular, I trim the data of Ontario 2001 Youth group.

7 Conclusion

For years, economists have been interested in estimating the impact of cohort size on labour market outcomes. Given the small year-to-year variations in cohort size, researchers have typically focused on long-term fluctuations. Doing so introduces an important identification issue as it becomes difficult to separate the effect of cohort size from other unrelated trends—this issue becomes more serious as the studied period lengthens.

This paper studies the effect of the 2003 Ontario double cohort on youth earnings. The double cohort generated a large and sudden influx of workers, making it possible to clearly identify the impact of cohort size on wages. In particular, the short time span over which the supply shock occurred helps resolving the identification problem faced by previous studies looking at cohort size effects.

My results suggest that the double cohort significantly depressed the wages of young workers. The Census results suggest that wages of full-time, full-year workers decreased by 5 to 9 percent due to the supply shock—this effect being estimated two years after the double cohort. Interestingly, the estimated impact of the supply shock becomes more negative as the control group is further away in age to the treatment group. This suggests that workers close in age to the double-cohort graduates were also affected by the supply shock. The Census findings are corroborated by the LFS results, suggesting that the immediate (six months after the shock) impact of the double cohort was to depress wages by 14 to 25 percent. Not only the supply shock affected the wages of full-time, full-year workers, but it also affected the likelihood to be working full time and full year by about 1.5 percentage points. Accounting for this effect on labour market participation, I estimate the impact of the supply shock on wages to be between -3.1 and -16.5 percent (the lower and upper bounds).

An interesting extension to this paper would be to look at the long-term effects of this supply shock. There is growing evidence that entering the labour market during particularly bad labour market conditions (e.g. during a recession) has long-lasting effects on individual economic outcomes. Oreopoulos, Wachter and Heisz (2008), for example, find that the effect of graduating in a recession can affect college graduates wages for up to 10 years. It is unclear whether supply shocks also have similar long-term effects on labour market outcomes. Studies focusing on baby boomers present mixed evidence (e.g. Berger (1985) and Welch (1979)). The small sample size of the LFS data makes it difficult to conclude anything about the long-term effects of the double cohort, but future research could exploit the subsequent Census spells in order to further investigate this question.

References

Berger, Mark C., "The Effect of Cohort Size on Earnings Growth: A Reexamination of the Evidence," *The Journal of Political Economy*, 1985, *93* (3), 561–573.

——, "Demographic Cycles, Cohort Size, and Earnings," *Demography*, 1989, *26* (2), 311–321.

Blanchflower, David G. and Richard B. Freeman, "The Declining Economic Status of Young Workers in OECD Countries," in David G. Blanchflower and Richard B. Freeman, eds., *Youth Employment and Joblessness in Advanced Countries*, University of Chicago Press, 2000, pp. 19–56.

Boudarbat, Brahim, Thomas Lemieux, and W. Craig Riddell, "The Evolution of the Returns to Human Capital in Canada, 1980-2005," *Canadian Public Policy*, 2010, *36* (1), 63–89.

Brunello, Giorgio, "The effects of cohort size on European earnings," *Journal of Population Economics*, January 2010, *23* (1), 273–290.

Card, David, "The Impact of the Mariel Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review*, 1990, *43* (2), 245–257.

—— **and Thomas Lemieux**, "Can Falling Supply Explain The Rising Return To College For Younger Men? A Cohort-Based Analysis," *The Quarterly Journal of Economics*, May 2001, *116* (2), 705–746.

Carrington, William J. and Pedro J. F. de Lima, "The Impact of 1970s Repatriates from Africa on the Portuguese Labor Market," *Industrial and Labor Relations Review*, 1996, *49* (2), 330–347.

Friedberg, Rachel M., "The Impact of Mass Migration on the Israeli Labor Market," *The Quarterly Journal of Economics*, 2001, *116* (4), 1373–1408.

Hunt, Jennifer, "The Impact of the 1962 Repatriates from Algeria on the French Labor Market," *Industrial and Labor Relations Review*, 1992, *45* (3), 556–572.

Katz, Lawrence F and Kevin M Murphy, "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, February 1992, *107* (1), 35–78.

King, Alan J.C., Wendy Warren, Will Boyce, Peter Chin, Matthew King, and Jean-Claude Boyer, "Double Cohort Study: Phase 2 Report," Technical Report, Social Program Evaluation Group, Queen's University 2002.

—— , —— , —— , —— , —— , —— , **and Barry O'Connor**, "Double Cohort Study: Phase 3 Report," Technical Report, Social Program Evaluation Group, Queen's University 2004.

—— , —— , —— , —— , —— , —— , **and** —— , "Double Cohort Study: Phase 4 Report," Technical Report, Social Program Evaluation Group, Queen's University 2005.

Korenman, Sanders and David Neumark, "Cohort Crowding and Youth Labor Markets: A Cross-National Analysis," in David G. Blanchflower and Richard B. Freeman, eds., *Youth Employment and Joblessness in Advanced Countries*, University of Chicago Press, 2000, pp. 57–105.

Krashinsky, Harry, "How Would One Extra Year of High School Affect Academic Performance in University? Evidence from a Unique Policy Change," 2006. Unpublished manuscript.

—— , "How Would One Extra Year of High School Affect Wages? Evidence from a Unique Policy Change," CLSRN Working Paper No. 20 2009.

Lee, David S., "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 2009, *76* (3), pp. 1071–1102.

Lemieux, Thomas and Kevin Milligan, "Incentive effects of social assistance: A regression discontinuity approach," *Journal of Econometrics*, February 2008, *142* (2), 807–828.

Macunovich, Diane J., "The fortunes of one's birth: Relative cohort size and the youth labor market in the United States," *Journal of Population Economics*, 1999, *12* (2), 215–272.

Morin, Louis-Philippe, "Estimating the Benefit of High School for College-Bound Students," CLSRN Working Paper No. 54 2010.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz, "The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates," IZA Discussion Papers 3578 2008.

Shimer, Robert, "The Impact of Young Workers on the Aggregate Labor Market," *The Quarterly Journal of Economics*, 2001, *116* (3), 969–1007.

Statistics Canada, "2006 Census Dictionary," Statistics Canada Catalogue no. 92-566-XWE 2007.

Welch, Finis, "Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust," *The Journal of Political Economy*, 1979, *87* (5), S65–S97.

A Appendix

Data Construction and Identification of the Treatment Group

I first describe how the data was constructed from the Canadian Census and the Labour Force Statistics and then discuss how the treatment group was identified using the available information.

Census Data

The main outcome variable is the average real weekly wages (in 2000 dollars). This variable is constructed using the WAGES variable from the 2001 and 2006 Censuses and Statistics Canada's provincial consumer price index (CPI).[24] The WAGE variable consists of gross wages and salaries before deductions (e.g. income tax), and It includes commissions and cash bonuses. The WAGE variable is adjusted using the CPI indices to represent wages in 2000 dollars. The WEEKS variable is used to convert the annual wages into weekly wages, and I classify a worker as full- or part-time using the variable FPTIM. An individual is considered to be working full time if she "worked mainly full-time weeks" (i.e. 30 hours or more) in the year prior to the Census (Statistics Canada, 2007).[25] Finally, an individual is considered to be working full time, full year if she worked 48 weeks or more in the year prior to the Census. I restrict the sample to full-time, full-year workers.

Since the outcome of interest is weekly wages in the year prior to the Census, I assign respondents to their province of residence as of June of the previous year (PR1 variable). Individuals that lived out of the country in that year (about 0.8 percent of the sample of individuals aged between 20 and 50) are discarded from the analysis.

Anyone who attended school in the year prior to the Census is excluded. I determine whether someone attended school in the year prior to the Census using the 2001 census ATTENDR and the 2006 census ATTSCHSUM variables. The school attendance indicator variable is equal to one if the individual attended school between September and May prior to the Census, regardless of whether the individual attended to school part-time or full-time. It is not possible to differentiate part-time and full-time attendance in the 2006 Census.

I construct the age of the respondent on January 1st of the Census year using their birth date. Since the last cohort of Ontario's Grade 13 program are expected to be 21 on January 1st 2006, I restrict the sample to individuals aged 21, and individuals aged between 25 and 49.

[24] The provincial annual consumer price indices are from CANSIM's Table 326-0021.

[25] The 30-hour cutoff is used by the Canadian Census to differentiate part- from full-time work. The Census does not contain more information about the average number of hours worked by the individual.

The educational-attainment variable is constructed using the 2001 Census SECGRADR and the 2006 Census SSGRAD variables. It corresponds to the highest educational degree obtained by the individual. A high school graduate in this paper is defined as an individual "with high school certificate or equivalency certificate without further schooling" (Statistics Canada, 2007). Someone is considered to have a bachelor's degree if she has a bachelor's degree, but does not have any other certificate above bachelor. Someone is considered to have a master's degree if she has a master's degree, but does not have any other certificate a doctorate degree or a medical degree (e.g. medicine, dentistry veterinary degrees). Individuals with degrees that do not fall into the categories above are not excluded from the analysis. Note that individuals with a bachelor's or a master's degree were only used for robustness checks.

Some changes to the educational-attainment questions in 2006 make it impossible to have a perfect match between the 2001 and the 2006 educational attainment variables. In particular, unlike the 2001 Census, the 2006 Census does not disentangle high school graduates with further training (but no certificate) from high school graduates without further training. I therefore labeled as high school graduates without further schooling in 2001 high school graduates regardless of whether they have further training, as long as they do not have a certificate above high school diploma. Excluding 2001 high school graduates with further training increases the magnitude of the supply shock by about 2 percentage points. Hence, the estimates presented in this paper could be seen as being on the conservative side.

Finally, I use the Class of Worker variable (COWD) to identify self-employed workers. Self-employed are excluded from the analysis since their wage-setting process is different from paid workers. The number of self-employed is very small, especially among workers aged 21. A detail about COWD that could introduce some measurement error is the fact that the question relates to labour market activity on the month of, instead of on the year prior to the Census. Since excluding these workers does not affect the results, and to be consistent with the LFS, I present results excluding self-employed.

Labour Force Survey Data

The main outcome variable is the real hourly wages (in 2000 dollars). The conversion from current to real wages is done using Statistics Canada's provincial CPI. One significant difference between the Census and Labour Force Survey (LFS) wages is that the LFS wages (HRLYEARN) are observed on January 2002 and 2004. This variable is observed only for employees.

As in the the Census data, an individual is considered to be working full time if she usually works 30 hours or more per week at her main job. I restrict the sample to full-time workers.

I determine whether someone is attending school in the survey month using STUDENT. Full- and part-time students are excluded from the sample.

The LFS does not release the respondent's date of birth. I therefore use age of the respondent on the week of the survey to define my age groups and to identify the treatment group. As will be explained below, the January LFS is the only one allowing me to identify the treatment group. Since the last cohort of Ontario's Grade 13 program is expected to be aged 19 on January 1^{st} 2004, I restrict the sample to individuals aged 19, and individuals aged between 23 and 47 to be consistent with the Census.

The LFS educational-attainment variable is constructed using two variables, EDUCLEV and HSGRAD. A high school graduate in the LFS data is defined as an individual who completed 11 to 13 years of schooling (based on EDUCLEV) and who graduated from high school (HSGRAD). This measure is somewhat cleaner than the Census measure, especially given the fact that it did not change between 2002 and 2004. Someone is considered to have a bachelor's degree if she has a bachelor's degree, but does not have any other certificate above bachelor. Someone is considered to have a master's degree if she has a master's or a PhD degree. The LFS does not differentiate between master's and PhD degrees. Individuals with a bachelor's or a master's degree were only used for robustness checks.

Finally, I exclude self-employed workers from the LFS data using COWMAIN. This variable identifies the class of worker at the respondent's main job.

Identification of the Treatment Group

Due to some differences in the information available in the Census and the LFS differ, I use different strategies to identify the treatment group in these two data sources.

The identification of the treatment group in the Census data is easier since it contains the exact date of birth of the individuals. Since the cutoff birth date for beginning primary school is December 31st in Ontario, one only needs to know the year of birth of an individual to know if she was supposed to be part of the double cohort or not. Graduates from the last Grade-13 cohort are expected to be born in 1984, while graduates from the first cohort of the Grade-12 program should be born in 1985. In order to avoid having the results contaminated by the potential value-added of Grade 13, I exclude Grade 12 graduates.

In the Census data, the treatment group is hence defined as 2006 Census respondents who were: 1) born in 1984, 2) high school graduates, and 3) Ontario residents in 2005. The main analysis is done on full-time and full-year workers (and not enrolled in school).

Since the LFS does not release the date of birth of their respondents, one has to rely on age only. I use the January LFS since it is the only one that allows me to get a good measure of one's date of birth. In January 2004, almost all LFS respondents aged 19 should be born in 1984, corresponding to the birth year of the last cohort of Ontario Grade 13 graduates.

In the LFS data, the treatment group is composed of January 2004 respondents who were both: 1) 19 year old in the survey week, and 2) Ontario high school graduates. The main analysis is done on individuals who worked full time, and did not attend school in January 2004.