

Online Appendix

The Insurance Implications of Government Student Loan Repayment Schemes

Martin Gervais
University of Georgia
mgervaisca@gmail.com

Qian Liu
Brock University
qliu@brocku.ca

Lance Lochner
University of Western Ontario
llochner@uwo.ca

December 13, 2022

Appendix A Data Overview

This paper uses data from the Education and Labour Market Longitudinal Platform (ELMLP) provided by Statistics Canada. Our primary data source comes from the Canada Student Loans Program (CSLP) administrative records on student loan disbursement, needs determination, and repayment. The ELMLP also allows us to link the CSLP data to administrative records from the Postsecondary Student Information System (PSIS) which includes additional institutional and individual characteristics, as well as the income and family records from the T1 Family File (T1FF).

A.1 CSLP

The CSLP data provides recipient-level longitudinal information on loan application, disbursement, and repayment by loan year from 2003–2004 to 2015–2016. The CSLP loan year begins on August 1 and ends on July 31 of the following calendar year. The CSLP data also contains information on each borrower’s institution of enrolment and basic demographics.

The main CSLP variables we use are: student loan variables including loan amount, repayment amount, and repayment status; demographics including birth year, gender, and highest education level. We explain our construction of various loan-related variables and efforts to address a few data inconsistencies in Appendix B.

The CSLP data does not contain information on collections or loan rehabilitation after someone enters default. We, therefore, acquired additional summary records from the CSLP to impute payments after default. The details are in Appendix C.

The CSLP data is reported by loan year, but income records in the T1FF are available by calendar year. To match the two, we impute the loan amount, repayment amount, and repayment status by calendar year, given the variables by loan year in the CSLP. See Appendix D for details.

A.2 PSIS

The PSIS provides information on student enrolment, majors, and degree attainment at Canadian public post-secondary institutions by reporting year. A reporting year starts from the day after the end of the institution’s previous winter term, which is usually a date in April, May or June, and ends one year from this start date. The ELMLP includes PSIS data

from 2009–2010 to 2019–2020 for all provinces and territories.¹ The main PSIS variables we use include citizenship status, highest education level, highest degree obtained, last major, and last institution enrolled.

A.3 T1FF

The T1FF is an annual administrative dataset of Canadian tax filers and their families. It includes selected information from income tax records from 1992 to 2018 for the PSIS population. The main T1FF variables we use include after-tax income (excluding capital gains/losses), marital status, whether the individual lives with parents, family size, and province of residence.

Appendix B CSLP Variables and Inconsistencies

This appendix describes the construction of our main variables (loan amount and payments), the data inconsistencies we encounter, and how we address those inconsistencies using the CSLP data. The reference period in this appendix is the loan year. See Appendix D for details on how we construct the variables by calendar year. To facilitate our discussion, we introduce the following variables and their definitions:

balstudyend : Balance of student loan at the end of the study period. It does not include any amount repaid before the end of study.

graceint : Interest accrued over the 6-month non-repayment period following study end date.

paidprin_t : The total amount of principal that has been repaid by end of loan year t .

paidint_t : The total amount of interest that has been repaid by end of loan year t .

outprin_t : The principal amount that is not yet repaid by end of loan year t .

¹While records are available for earlier years, the PSIS did not include some Canadian provinces until 2009. Due to the 6-month grace period before loan consolidation and payment begins, we do not observe full schooling records in the PSIS for many borrowers consolidating in 2009 (or earlier).

B.1 Loan Amount

In the data, borrowers may have multiple loan consolidations due to leaving and returning to school. For each consolidation, the loan consolidation amount is the sum of *balstudyend* and *graceint*, which should be cumulative and reflect the total outstanding balance at the time.

To construct the loan amount variables, we begin with the last consolidated loan and its repayment. In theory, $paidprin_t + outprin_t$ should equal $balstudyend + graceint$ at the year of consolidation if no payments were made prior to consolidation. We can use either combination for the last consolidation amount, denoted L_l , if this equality holds. In many cases, this equality is not respected. One common case is that for borrowers who have multiple loan consolidations, the case officer may find it convenient to close the earlier loan with a “paid in full” status, despite the fact that the loan is not fully repaid, and combine the unpaid portion into the later loan. When the later loan consolidation amount (*balstudyend*) does not include the outstanding amounts from previous consolidations, a discrepancy arises between the repayment record ($paidprin_t + outprin_t$) and the reported last consolidation amount ($balstudyend + graceint$). In this case, we use $paidprin_t + outprin_t$ at the consolidation year for the last consolidation amount L_l .

Another reason $paidprin_t + outprin_t$ sometimes differs from $balstudyend + graceint$ for the last consolidated loan is that $paidprin_t$ includes payments made prior to the last loan consolidation, while *balstudyend* is net of payments made before consolidation. As a result, $paidprin_t + outprin_t > balstudyend + graceint$ at consolidation. In this case, we only consider payments made since last consolidation and set the last consolidation amount $L_l = balstudyend + graceint$ for the last consolidated loan.

In other cases with discrepancies, there is no clear explanation. For these unexplained cases, we exclude from our sample borrowers for whom the absolute discrepancy between $paidprin_t + outprin_t$ and $balstudyend + graceint$ is at least 5% of $balstudyend + graceint$; otherwise, we set $L_l = paidprin_t + outprin_t$ for the last consolidated loan (t is the consolidation year) when the discrepancy is positive but less than 5%.

B.2 Payments

To calculate annual payments, P_t , we use $paidprin_t$ and $paidint_t$: $P_t = P_t^{prin} + P_t^{int} = (paidprin_t - paidprin_{t-1}) + (paidint_t - paidint_{t-1})$, where P_t^{prin} (P_t^{int}) denotes payment

towards principal (interest) in loan year t . We favor using $paidprin_t$ over $outprin_t$ to calculate principal payments, because the change in $outprin_t$ includes not only borrower payments but also amounts paid by the government. When a borrower is in RAP Stage 2, for example, $outprin_t$ declines even when the borrower makes no payment; however, $paidprin_t$ does not change in this case, reflecting the borrower’s own payment amount.

Because a separate record is created whenever the loan status changes or at the end of a loan year, there may be multiple values for $paidprin$, $paidint$, and $outprin$ in any given loan year t . The highest values for $paidprin$ and $paidint$ for each loan year t are used as the end-of-year amounts (i.e., $paidprin_t$ and $paidint_t$), while the lowest value of $outprin$ for each year is used as the end-of-year amount (i.e., $outprin_t$).

Issues with $paidprin_t/paidint_t$

Because $paidprin_t$ and $paidint_t$ are both cumulative amounts, they should not decrease over time. However, we observe significant declines in $paidprin_t$ and $paidint_t$ for many borrowers from British Columbia (BC) from the 2010–11 loan year to the 2011–12 loan year. Conversations with CSLP staff indicate that these cases are associated with BC becoming an integrated province in 2012, which led to known loan reconciliation issues in their records.

To solve this problem, we use $outprin_t$ when appropriate. For those in standard repayment (not on RAP), we set $P_t^{prin} = outprin_{t-1} - outprin_t$, since all payments should be made by borrowers themselves. If $paidint_t - paidint_{t-1}$ is also problematic, we impute the interest payment P_t^{int} using the fact that payments are made towards interest first, in which case we should generally expect that

$$paidint_{t-1} - paidint_{t-2} = outprin_{t-2} \times r, \quad (\text{B1})$$

and

$$paidint_t - paidint_{t-1} = outprin_{t-1} \times r, \quad (\text{B2})$$

for any interest rate r . Using equation (B1) to substitute in for r in equation (B2), we impute interest payments as follows:

$$P_t^{int} = paidint_t - paidint_{t-1} = (paidint_{t-1} - paidint_{t-2}) \times \frac{outprin_{t-1}}{outprin_{t-2}}. \quad (\text{B3})$$

Fortunately, this imputation provides a very good approximation to interest payments for cases in other years/provinces that were unaffected by the 2012 BC loan integration process.

For those with missing $paidint_{t-2}$ or $outprin_{t-2}$ (e.g., consolidate in $t - 1$ so do not have $t - 2$ data), we assume that $r = 5.5\%$ and use equation (B2) directly:²

$$P_t^{int} = paidint_t - paidint_{t-1} = outprin_{t-1} \times r. \quad (\text{B4})$$

If the borrower is on RAP (not in Stage 2 or permanent disability, PD), we set $P_t^{prin} = outprin_{t-1} - outprin_t$. Because the government covers interest when the borrower is on RAP, we assume that $P_t^{int} = 0$ if $paidint_t < paidint_{t-1}$, (i.e., $P_t^{int} = \max\{paidint_t - paidint_{t-1}, 0\}$).

If the borrower is on RAP Stage 2 or RAP-PD, when the government covers both interest and principal amounts, we assume $P_t^{prin} = 0$ and $P_t^{int} = \max\{paidint_t - paidint_{t-1}, 0\}$. This reflects the fact that the vast majority of borrowers enrolled in RAP make zero payments.

There are some additional cases (for borrowers not from BC in 2011–2012) for which $paidprin_t$ or $paidint_t$ decreases. In most cases $paidprin_t$ ($paidint_t$) decreases by less than 5% of the consolidation amount, in which case we adjust it by setting $paidprin_t$ ($paidint_t$) equal to the maximum of its current or any earlier reported values, so it is non-decreasing. This ensures that annual payments are non-negative. We exclude from our sample borrowers who have a decrease in $paidprin_t$ that is 5% or more of the consolidation amount.

In total, we exclude approximately 10% of our sample due to inconsistencies in the consolidation amount or in $paidprin_t$.

Appendix C Collections and Loan Rehabilitation after Default

When a loan enters default, it is “Returned to Government” from CSLP’s third-party Service Provider, and the Canada Revenue Agency (CRA) takes responsibility for loan collection. When a loan is rehabilitated, loans are sent back to the Service Provider for additional repayment. Unfortunately, loan histories in our data are broken whenever a default occurs, so we cannot track collections by the CRA or repayments in rehabilitation for individual borrowers. We, therefore, impute payments following any default using summary data on repayment in collections and in rehabilitation provided to us by the CSLP.

These summary data contain the following statistics based on all borrowers who entered

²The floating interest rate for student loan is 5.5% (i.e., prime rate plus 2.5%) from 2011 to 2014.

default or rehabilitated their loan between April 1, 2010 and March 31, 2011:

$\Phi_{Coll,t}(L_D, T_D)$: Fraction of student loan debt entering default, L_D , in year T_D collected from CRA in year t since default.

$\Phi_{Rehab,t}(L_D, T_D)$: Fraction of student loan debt entering default, L_D , in year T_D rehabilitated in year t since default.

$\theta_{Rehab,\tau}(L_R)$: Fraction of the amount rehabilitated, L_R , repaid during year τ since rehabilitation.

These fractions are based on 6 bins for L_D or L_R : \$1–3,000; \$3,001–6,000; \$6,001–10,000; \$10,001–15,000; \$15,001–20,000; and \$20,001 or more. T_D reflects 1, 2, 3, 4, and 5+ years. Payment rates are available annually for 8 years after default or rehabilitation. We impute payments in year $t = 1, \dots, 8$ since default, $P_t(L_D, T_D)$, as a function of the loan amount at the time of default, L_D , and year of default (since consolidation), T_D , based on the following:

$$P_t(L_D, T_D) = L_D \left[\Phi_{Coll,t}(L_D, T_D) + \sum_{\tau=1}^t \Phi_{Rehab,t-\tau+1}(L_D, T_D) \theta_{Rehab,\tau}(L_D) \right]. \quad (C5)$$

The first term in brackets reflects the fraction of debt initially defaulted on that is collected from CRA each post-default year t , while the second term reflects the fraction of debt initially defaulted on that is collected through rehabilitation in post-default year t . The latter is calculated as the sum of the fraction of debt in default that was rehabilitated each year up to post-default year t multiplied by the fraction of that rehabilitated debt that is repaid in post-default year t .³

See Tables C1–C3 for $\Phi_{Coll,t}(L_D, T_D)$, $\Phi_{Rehab,t}(L_D, T_D)$, and $\theta_{Rehab,\tau}(L_R)$ as provided by the CSLP.

³This approach implicitly assumes that borrowers who rehabilitate their student loans do not exit rehabilitation and that they hold similar loan balances when they enter rehabilitation as they had when they initially defaulted. This and other reasonable assumptions have little impact on post-default payment amounts P_t , since loan rehabilitation rates, $\Phi_{Rehab,t}(L_D, T_D)$, are extremely low (less than 3% for nearly all (L_D, T_D) combinations).

Table C1: $\Phi_{Coll,t}(L_D, T_D)$

		L_D					
		\$1–3,000	\$3,001–6,000	\$6,001–10,000	\$10,001–15,000	\$15,001–20,000	\$20,001 or more
$t = 1$	$T_D = 1$	24.6%	12.5%	8.0%	6.4%	5.5%	3.7%
	$T_D = 2$	28.1%	16.1%	9.7%	7.8%	6.6%	5.0%
	$T_D = 3$	30.8%	17.9%	11.6%	8.6%	7.4%	5.8%
	$T_D = 4$	31.9%	17.0%	12.4%	8.9%	6.5%	4.9%
	$T_D = 5+$	27.9%	16.3%	11.0%	7.9%	6.7%	3.0%
$t = 2$	$T_D = 1$	15.5%	10.3%	7.9%	5.8%	4.8%	3.9%
	$T_D = 2$	15.4%	13.0%	9.7%	7.4%	8.0%	4.8%
	$T_D = 3$	15.0%	13.3%	11.1%	7.3%	6.3%	5.2%
	$T_D = 4$	14.2%	13.0%	11.1%	7.9%	7.4%	4.7%
	$T_D = 5+$	11.9%	11.0%	9.4%	7.1%	4.0%	5.0%
$t = 3$	$T_D = 1$	10.1%	9.5%	7.3%	5.7%	5.1%	3.7%
	$T_D = 2$	10.9%	11.2%	8.8%	7.6%	7.5%	4.6%
	$T_D = 3$	12.7%	9.4%	10.2%	8.2%	6.4%	4.5%
	$T_D = 4$	9.2%	10.4%	9.7%	8.4%	6.9%	5.8%
	$T_D = 5+$	9.3%	8.9%	7.8%	6.6%	4.8%	4.2%
$t = 4$	$T_D = 1$	8.0%	8.0%	7.0%	5.7%	5.1%	3.7%
	$T_D = 2$	8.6%	9.5%	7.7%	6.6%	6.9%	4.7%
	$T_D = 3$	9.4%	10.2%	9.4%	6.7%	6.9%	4.7%
	$T_D = 4$	8.8%	7.8%	8.2%	7.5%	5.2%	5.6%
	$T_D = 5+$	6.5%	6.4%	6.3%	6.4%	4.8%	4.1%
$t = 5$	$T_D = 1$	7.6%	7.1%	6.5%	5.4%	5.1%	3.7%
	$T_D = 2$	7.8%	7.3%	7.3%	7.0%	7.0%	4.6%
	$T_D = 3$	6.0%	8.1%	7.8%	6.7%	6.2%	5.5%
	$T_D = 4$	5.4%	7.1%	7.6%	5.9%	5.5%	5.4%
	$T_D = 5+$	5.4%	5.8%	5.4%	6.0%	3.7%	4.1%
$t = 6$	$T_D = 1$	3.0%	4.2%	4.1%	4.0%	4.5%	3.5%
	$T_D = 2$	2.7%	5.0%	5.4%	5.4%	5.6%	4.4%
	$T_D = 3$	3.9%	5.5%	4.8%	5.9%	5.0%	4.6%
	$T_D = 4$	2.6%	4.8%	5.7%	6.0%	6.4%	4.8%
	$T_D = 5+$	3.8%	3.6%	4.1%	4.9%	4.3%	3.8%
$t = 7$	$T_D = 1$	1.1%	2.4%	2.5%	3.3%	3.1%	3.3%
	$T_D = 2$	1.6%	3.0%	3.8%	4.2%	5.2%	4.0%
	$T_D = 3$	2.3%	3.0%	3.6%	4.1%	4.9%	3.9%
	$T_D = 4$	2.6%	4.0%	3.8%	3.7%	4.5%	4.4%
	$T_D = 5+$	2.0%	2.7%	3.2%	2.8%	3.6%	2.7%
$t = 8$	$T_D = 1$	0.7%	1.4%	2.0%	2.6%	3.3%	3.0%
	$T_D = 2$	1.3%	2.1%	2.7%	3.3%	4.2%	3.1%
	$T_D = 3$	0.8%	2.0%	2.2%	3.9%	4.5%	3.8%
	$T_D = 4$	0.9%	2.3%	3.6%	2.8%	4.8%	3.5%
	$T_D = 5+$	1.6%	2.0%	2.2%	3.1%	3.5%	2.5%

Notes: These statistics are based on data provided by CSLP from newly defaulted loans established in Departmental Accounts Receivable System (DARS) from April 1, 2010 to March 31, 2011.

Table C2: $\Phi_{Rehab,t}(L_D, T_D)$

		L_D					
		\$1-3,000	\$3,001-6,000	\$6,001-10,000	\$10,001-15,000	\$15,001-20,000	\$20,001 or more
$t = 1$	$T_D = 1$	1.5%	1.7%	1.3%	1.7%	2.1%	2.1%
	$T_D = 2$	2.0%	2.3%	2.1%	2.5%	2.5%	3.4%
	$T_D = 3$	1.2%	1.7%	2.1%	3.3%	1.7%	2.4%
	$T_D = 4$	0.6%	1.7%	1.8%	2.1%	2.0%	2.8%
	$T_D = 5+$	0.3%	1.2%	1.9%	2.8%	2.8%	0.9%
$t = 2$	$T_D = 1$	1.9%	2.0%	1.7%	2.6%	2.3%	1.9%
	$T_D = 2$	1.7%	1.7%	2.1%	3.1%	2.1%	2.6%
	$T_D = 3$	1.2%	1.3%	1.4%	2.2%	2.9%	2.7%
	$T_D = 4$	0.1%	1.3%	1.6%	1.9%	2.3%	1.9%
	$T_D = 5+$	0.4%	1.1%	0.7%	1.2%	2.0%	0.9%
$t = 3$	$T_D = 1$	1.0%	1.3%	1.1%	1.1%	1.6%	1.5%
	$T_D = 2$	1.1%	1.2%	1.3%	1.4%	1.2%	1.1%
	$T_D = 3$	0.5%	0.7%	0.6%	1.2%	1.5%	1.1%
	$T_D = 4$	0.3%	0.6%	0.5%	1.5%	1.8%	2.1%
	$T_D = 5+$	0.2%	0.1%	0.8%	0.4%	1.1%	2.0%
$t = 4$	$T_D = 1$	0.5%	1.0%	1.1%	0.6%	1.4%	1.4%
	$T_D = 2$	0.3%	0.6%	0.8%	0.7%	1.4%	1.9%
	$T_D = 3$	0.7%	0.6%	0.4%	0.4%	1.5%	1.4%
	$T_D = 4$	0.5%	0.2%	0.3%	0.2%	0.7%	0.6%
	$T_D = 5+$	0.0%	0.3%	0.2%	0.3%	0.5%	0.5%
$t = 5$	$T_D = 1$	0.3%	0.5%	0.8%	1.0%	0.5%	0.8%
	$T_D = 2$	0.2%	0.7%	0.3%	1.3%	0.5%	0.4%
	$T_D = 3$	0.2%	0.7%	0.7%	0.4%	0.7%	1.5%
	$T_D = 4$	0.1%	0.3%	0.4%	0.4%	0.5%	0.6%
	$T_D = 5+$	0.4%	0.0%	0.2%	0.6%	0.6%	0.0%
$t = 6$	$T_D = 1$	0.3%	0.3%	0.5%	0.6%	0.9%	0.9%
	$T_D = 2$	0.3%	0.2%	0.4%	0.8%	1.2%	0.8%
	$T_D = 3$	0.0%	0.1%	0.4%	0.6%	0.7%	0.4%
	$T_D = 4$	0.1%	0.0%	0.3%	0.5%	0.6%	0.8%
	$T_D = 5+$	0.1%	0.0%	0.6%	1.2%	0.7%	1.0%
$t = 7$	$T_D = 1$	0.1%	0.4%	0.4%	0.8%	0.7%	0.9%
	$T_D = 2$	0.1%	0.4%	0.4%	0.7%	1.8%	1.0%
	$T_D = 3$	0.1%	0.4%	0.7%	0.5%	1.1%	0.5%
	$T_D = 4$	0.5%	0.4%	0.4%	0.7%	0.8%	2.1%
	$T_D = 5+$	0.1%	0.1%	0.4%	0.7%	0.9%	0.2%
$t = 8$	$T_D = 1$	0.4%	0.2%	0.3%	0.7%	0.2%	0.8%
	$T_D = 2$	1.0%	0.5%	0.3%	0.3%	0.8%	0.7%
	$T_D = 3$	0.0%	0.2%	0.3%	0.5%	0.5%	0.6%
	$T_D = 4$	0.0%	0.1%	0.3%	0.3%	0.6%	0.8%
	$T_D = 5+$	0.0%	0.3%	0.2%	0.4%	0.9%	0.4%

Notes: These statistics are based on data provided by CSLP from newly defaulted loans established in Departmental Accounts Receivable System (DARS) from April 1, 2010 to March 31, 2011.

Table C3: $\theta_{Rehab,\tau}(L_R)$

	L_R					
	\$1-3,000	\$3,001-6,000	\$6,001-10,000	\$10,001-15,000	\$15,001-20,000	\$20,001 or more
$\tau = 1$	16.6%	8.5%	4.9%	2.9%	6.5%	5.0%
$\tau = 2$	13.3%	8.4%	6.6%	6.0%	3.1%	8.1%
$\tau = 3$	12.9%	7.5%	6.4%	6.4%	5.5%	4.6%
$\tau = 4$	14.2%	9.7%	6.3%	5.8%	4.7%	4.8%
$\tau = 5$	14.1%	7.7%	5.4%	6.9%	6.9%	5.3%
$\tau = 6$	14.3%	8.6%	8.2%	5.8%	7.2%	3.4%
$\tau = 7$	14.9%	10.0%	6.9%	4.7%	5.2%	6.6%
$\tau = 8$	12.0%	4.7%	7.6%	4.8%	7.0%	3.2%

Notes: These statistics are based on data provided by CSLP from loans that were rehabilitated from Departmental Accounts Receivable System (DARS) from April 1, 2010 to March 31, 2011.

Appendix D From Loan Year to Calendar Year

The CSLP data is organized by loan year, which begins on August 1 and ends on July 31 of the following calendar year. Appendix B explains our construction of loan amounts and payments by loan year. To link the CSLP data to income records in the T1FF tax file by calendar year, we first create measures of payments, repayment status, outstanding debt amount, and scheduled payments at the monthly level, then aggregate monthly measures to the calendar year. This appendix describes this process.

D.1 Payments and Status

The CSLP records report when a borrower pays off his/her loan, enters default, or declares bankruptcy. We also observe the total number of months a borrower was enrolled in RAP during each loan year, but we do not know the monthly status of the loan. We, therefore, impute loan status and payment for each month $k \in \{1, 2, 3, \dots, 12\}$ of the loan year (where $k = 1$ represents August of calendar year t and $k = 12$ represents July of the following calendar year). We then aggregate the status and payments within calendar years.

We begin by setting all monthly status measures to ‘fully repaid’ and payments $p_{t,k} = 0$ after a loan is fully repaid.

Next, consider the case of default.⁴ For any month following a default, the loan status is set to ‘default’ and monthly payments are given by $p_{t,k} = P_t(L_D, T_D)/12$, where $P_t(L_D, T_D)$ is the annual post-default payment amount defined by equation (C5). Since default in Canada follows 9 months of consecutive missed payments, we set $p_{t,k} = 0$ for the 9 months leading up to and including the month of default.⁵

Student debts are, in rare cases, expunged through bankruptcy. We, therefore, set loan status to bankruptcy and payments $p_{t,k} = 0$ for all months after a bankruptcy.

Because our data do not report monthly payment amounts or detailed loan status measures, it is not always possible to determine whether borrowers are making regular loan payments (including short periods of delinquency that do not end in default) or enrolled in RAP during any given post-consolidation month. Aside from lengthy periods of delinquency that lead to default, there is no way to pinpoint months in which no payment is made. As we explain below, our imputations implicitly assume that repayments are smoothed across months of a

⁴Appendix C describes our imputation of annual payments post-default.

⁵The loan status remains in repayment until the default occurs.

year when borrowers are in repayment, accounting for differences in payment amounts when borrowers are on RAP vs. making regular payments (including delinquency).

To impute monthly loan payments and RAP status, we begin by defining the variable $repay_{t,k}$, which is set to 0: from the month after full repayment onward, during the 9 months leading up to a default and thereafter, or from the month after bankruptcy onward. Otherwise, we set $repay_{t,k} = 1$, reflecting that the borrower is in some form of repayment (including RAP or short periods of delinquency). It is these months for which we need to impute payment amounts. While the data do not report monthly RAP status, they do report the total number of months on RAP for each loan year ($monthsrap_t$) along with the monthly affordable payment ($affordpay_t$) when an individual is on RAP. We use these measures to impute monthly RAP status and payment amounts for all months since consolidation in which $repay_{t,k} = 1$.

Our imputation exploits the fact that repayment assistance under RAP (as well as Interest Relief, the pre-cursor to RAP) is administered in 6 month increments to help determine whether a borrower is on RAP for a specific month. We develop an algorithm that first imputes RAP status during each month for which $repay_{t,k} = 1$, then imputes payment amounts during each of these periods.

The algorithm begins by defining a monthly RAP status measure, $RAP_{t,k}$, and initializing this measure to zero for every post-consolidation month. We then update this status measure each month using different rules depending on whether RAP status can be directly inferred from the sequence of $monthsrap$.

First, there are months when it is clear whether or not an individual is on RAP based on the time sequence for $monthsrap$. (For example, if $monthsrap_{t-1} = 0$, $monthsrap_t = 4$, $monthsrap_{t+1} = 2$, and $monthsrap_{t+2} = 0$, then the individual must be on RAP for the 4 months at the end of loan year t and the 2 months at the beginning of loan year $t + 1$, but not for other months.) We set $RAP_{t,k} = 1$ for months when it is clear that an individual is on RAP given the sequence for $monthsrap$, while $RAP_{t,k}$ will remain equal to 0 for months when it is clear that an individual is not on RAP.

Second, there are months when it is unclear (from the sequence of $monthsrap$) whether the borrower is on RAP. For example, if $monthsrap_{t-1} = 0$, $monthsrap_t = 6$, $monthsrap_{t+1} = 0$, we know that the individual is on RAP for 6 months during year t , but we do not know which months. Another example arises when $monthsrap_{t-1} = 0$, $monthsrap_t = 4$, $monthsrap_{t+1} = 8$, and $monthsrap_{t+2} = 0$. In this case, we know that the individual is on RAP for the last 4 months of year t and the first 2 months of year $t + 1$, but we do not know which 6 months

the individual is on RAP for the remaining months of year $t + 1$.⁶

For months during loan year t when we are unsure whether the individual is on RAP, we divide the number of months on RAP by the total number of months for which we are uncertain whether the individual is on RAP, setting

$$RAP_{t,k} = \frac{monthsrap_t - \sum_{m=1}^{12} \mathbb{1}(RAP_{t,m} = 1)}{M_t - \sum_{m=1}^{12} \mathbb{1}(RAP_{t,m} = 1)},$$

where $M_t = \sum_{m=1}^{12} repay_{t,m}$ reflects the total number of months in loan year t that a borrower is in repayment, $\mathbb{1}(\cdot)$ is the indicator function, and $\sum_{m=1}^{12} \mathbb{1}(RAP_{t,m} = 1)$ reflects the number of months that an individual is known to be on RAP during loan year t .⁷

Having determined all monthly $RAP_{t,k}$ status measures, we next impute monthly payments. If the individual has not been on RAP during the loan year (i.e., $\sum_{m=1}^{12} RAP_{t,m} = 0$) or has been on RAP for all months (i.e., $\sum_{m=1}^{12} RAP_{t,m} = M_t$), we assume that monthly payments are constant throughout the loan year:

$$p_{t,k} = \frac{P_t}{M_t}, \text{ if } repay_{t,k} = 1 \text{ and } \sum_{m=1}^{12} RAP_{t,m} \in \{0, M_t\},$$

where P_t is the total amount repaid during loan year t .

If there are uncertain months during a loan year (i.e., months with $0 < RAP_{t,k} < 1$), we first allocate $affordpay_t$ to the months we know are on RAP, then distribute the remaining amount equally across the remaining uncertain months:

$$p_{t,k} = \begin{cases} affordpay_t, & \text{if } repay_{t,k} = 1 \text{ and } \sum_{m=1}^{12} RAP_{t,m} \in (0, M_t) \text{ and } RAP_{t,k} = 1 \\ \frac{P_t - affordpay_t \times \sum_{m=1}^{12} \mathbb{1}(RAP_{t,m} = 1)}{M_t - \sum_{k=m}^{12} \mathbb{1}(RAP_{t,m} = 1)}, & \text{if } repay_{t,k} = 1 \text{ and } \sum_{m=1}^{12} RAP_{t,m} \in (0, M_t) \text{ and } RAP_{t,k} \neq 1. \end{cases}$$

For the few cases with $P_t < affordpay_t \times \sum_{m=1}^{12} \mathbb{1}(RAP_{t,m} = 1)$, we assume $p_{t,k} = P_t/M_t$ for all months of loan year t .

Finally, we report that an individual is on RAP during a calendar year if $RAP_{t,k} > 0$ for any month during that year. We sum $p_{t,k}$ over all months within a calendar year to obtain

⁶Most cases are consistent with periods of 6 consecutive months on RAP; however, we occasionally observe sequences of $monthsrap$ that are not (e.g., $monthsrap_{t-1} = 0$, $monthsrap_t = 2$, and $monthsrap_{t+1} = 0$). These cases are treated in the same way as cases consistent with 6 months of RAP where the precise timing of RAP is uncertain.

⁷In these cases, $RAP_{t,k}$ will be a number between 0 and 1, and $RAP_{t,k}$ reflects the probability that someone is on RAP in month k of year t given a uniform prior over all months of that loan year for which $repay_{t,k} = 1$ but RAP status is not known from the sequence of $monthsrap$.

the total payment for that year.

D.2 Outstanding Debt

Appendix B describes how we construct the outstanding debt amount at the end of the loan year (July 31). Because we only use outstanding debt at the beginning of each calendar year (for the purpose of categorizing individuals based on their current debt), we simply assume that debt declines evenly over the loan year to obtain the outstanding amount at the beginning of January.

After someone has defaulted or declared bankruptcy, we no longer observe their debt, but we continue to include them in our analysis. To impute outstanding debt following default or bankruptcy, we assume $D_{t_c+1} = (1 + r_{t_c})D_{t_c} - P_{t_c}$, where D_{t_c} is the outstanding debt at the beginning of calendar year t_c , r_{t_c} is the annual floating interest rate in year t_c , and P_{t_c} is the payment amount in year t_c .

D.3 Scheduled Payment

The CSLP data provides the scheduled monthly payment at the end of each loan year (i.e., July), which we assume remains constant throughout the corresponding calendar year. When a loan enters repayment, the scheduled monthly payment is calculated based on a pre-agreed repayment schedule, with the majority of borrowers opting for the 114-month (9.5-year) schedule. For many borrowers, the scheduled payment remains constant; however, it can change over time if the loan terms are revised (e.g., the borrower applies to shorten or extend the repayment period).

After someone has defaulted or declared bankruptcy, we no longer observe a scheduled payment. To impute the scheduled payment by calendar year following default or bankruptcy, we amortize the remaining debt D_{t_c} over a specific period (i.e., 15 years since consolidation) to impute the scheduled payment during calendar year t_c .

Appendix E Merging Data and Sample Restrictions

Our sample begins with individuals who received student loans for full-time undergraduate studies in the CSLP. As described in Appendix B, we exclude from our sample those with an

inconsistency in loan amounts or cumulative payments of 5% or more of the consolidation amount.

We then link the CSLP to the PSIS and the T1FF. The earliest year available in the PSIS for all provinces is 2009–2010. Because there is a 6-month grace period after leaving school but before loan consolidation, we do not observe schooling records in the PSIS for many borrowers who consolidated in 2009 or earlier. To obtain a high linkage rate (about 90%), we limit our sample to those who last consolidated a CSLP loan in 2010. Furthermore, we restrict our sample to those who can be linked to PSIS records by end of 2010 and those with at least 1 year of tax data in the T1FF between 2010 and 2015.

We impose a few additional restrictions on our sample. We exclude borrowers over 30 years old in 2010 when they consolidated their loans. We also exclude borrowers who were enrolled in a post-secondary institution outside of Canada when they last borrowed or who resided outside of Canada during any year from 2010 to 2015 (based on T1FF); these borrowers are less likely to file Canadian taxes. Due to small sample sizes and confidentiality concerns, we exclude borrowers who were attending a post-secondary institution in the territories when they last borrowed.

Based on the PSIS, many borrowers returned to school without borrowing after their last loan consolidation. We exclude those who enrolled in baccalaureate or higher degree programs after 2010, because they are likely to have very different income and repayment trajectories during periods of enrolment.

Finally, our main sample is limited to borrowers who consolidated at least \$5,000 in student loans to ensure that there is potential for loan payments to serve as a source of insurance.

Appendix F Borrower Characteristics

Parts of our analysis condition on a detailed set of borrower characteristics $X_{i,t}$. For any given t , $X_{i,t}$ includes the following indicators: gender, age, marital status (married or not), citizenship (Canadian citizen or permanent resident), whether a borrower lives with parents, number of family members (1, 2, 3, 4, 5, 6+), province of residence (as of December 31), highest type of post-secondary enrolment (baccalaureate, master’s, PhD), highest degree obtained (less than BA, BA, MA or above), field of study (detailed grouping of 40 majors as listed in Table F1), Maclean’s institutional ranking (1–10, 11–20, ..., 41–49, unranked univer-

sities, college).⁸ $X_{i,t}$ also includes interactions of age (in years) with indicators for gender, citizenship, marital status, highest degree, field of study (broad grouping of 17 fields as listed in Table F2), and institutional ranking.

Due to small sample sizes and confidentiality concerns, our age indicators combine individuals ages 18 and 19 in 2010 (and the same two cohorts in subsequent years). We also combine all 3 territories (Northwest Territories, Yukon, and Nunavut) into a single indicator for province of residence.

The PSIS provides several variables categorizing field of study by the Classification of Instructional Programs (CIP) Canada 2016 version, including a primary grouping code (CIPPG), a 2-digit code, and a 4-digit code. We use these variables to construct two new groupings, a broad grouping and a detailed grouping. Tables F1 and F2 list our adjustment and re-grouping based on the PSIS codes for the detailed and broad groupings, respectively. There are a total of 17 categories in the broad grouping, compared to 40 categories in the detailed grouping. There are a few individuals in our sample who majored in two different fields in their last year of PSIS. We set indicators for both fields to one for these individuals.

⁸The ranking is by national reputation of the institution in 2015. See <https://www.macleans.ca/education/unirankings/university-rankings-2015-national-reputational-ranking/>.

Table F1: Field of Study Classification (Detailed Grouping)

PSIS original grouping (CIP 2-digit)	Our adjustment and grouping
01.Agriculture, agriculture operations and related sciences (except cannabis)	Same
03.Natural resources and conservation	Same
04.Architecture and related services	Same
05.Area, ethnic, cultural, gender, and group studies	Same
09.Communication, journalism and related programs	Combine 09 and 10
10.Communications technologies/technicians and support services	
11.Computer and information sciences and support services	Combine CIP 4-digit codes 30.06 (Systems science and theory), 30.08 (Mathematics and computer science), and 30.30 (Computational science) within CIP 2-digit code 30 (Multidisciplinary/interdisciplinary studies) with CIP 2-digit code 11
12.Personal and culinary services (except cannabis)	Same
13.Education	Same
14.Engineering	Combine 14 and 15
15.Engineering technologies and engineering-related fields	
16.Aboriginal and foreign languages, literatures and linguistics	Same
19.Family and consumer sciences/human sciences	Same
21.Pre-technology education/pre-industrial arts programs	Combine 21 with 53
22.Legal professions and studies	Same
23.English language and literature/letters	Same
24.Liberal arts and sciences, general studies and humanities	Same
25.Library science	Same
26.Biological and biomedical sciences	Combine CIP 4-digit codes 30.01 (Biological and physical sciences) and 30.27 (Human biology) within CIP 2-digit code 30 (Multidisciplinary/interdisciplinary studies) with CIP 2-digit code 26
27.Mathematics and statistics	Same
28.Military science, leadership and operational art	Combine 28, 29, and 43
29.Military technologies and applied sciences	

30.Multidisciplinary/interdisciplinary studies	Group CIP 4-digit codes 30.13 (Medieval and renaissance studies), 30.21 (Holocaust and related studies), 30.22 (Classical and ancient studies), and 30.29 (Maritime studies) within CIP 2-digit code 30 into a separate category (Multidisciplinary/Humanities)
	Group CIP 4-digit codes 30.10 (Biopsychology), 30.11 (Gerontology), 30.14 (Museology/museum studies), 30.15 (Science, technology and society), 30.17 (Behavioural sciences), 30.20 (International/global studies), 30.23 (Intercultural/multicultural and diversity studies), 30.25 (Cognitive science), 30.26 (Cultural studies/critical theory and analysis), 30.28 (Dispute resolution), 30.31 (Human computer interaction), and 30.33 (Sustainability studies) within CIP 2-digit code 30 into a separate category (Multidisciplinary/Social and behavioural sciences)
	Group CIP 4-digit codes 30.18 (Natural sciences), 30.19 (Nutrition sciences), and 30.27 (Human biology) within CIP 2-digit code 30 into a separate category (Multidisciplinary/Physical and life sciences)
	Combine CIP 4-digit code 30.16 (Accounting and computer science) within CIP 2-digit code 30 with CIP 2-digit code 52 (Business, management, marketing and related support services)
	Combine CIP 4-digit codes 30.01 (Biological and physical sciences) and 30.27 (Human biology) within CIP 2-digit code 30 with CIP 2-digit code 26 (Biological and biomedical sciences)
	Combine CIP 4-digit codes 30.06 (Systems science and theory), 30.08 (Mathematics and computer science), and 30.30 (Computational science) within CIP 2-digit code 30 with CIP 2-digit code 11 (Computer and information sciences and support services)
	Group all other 30 codes into a separate category (Multidisciplinary/Other)
31.Parks, recreation, leisure and fitness studies	Same
32.Basic skills (not for credit)	Combine 32, 33, 34, 35, 36, and 37
33.Citizenship activities (not for credit)	
34.Health-related knowledge and skills (not for credit)	
35.Interpersonal and social skills (not for credit)	

36. Leisure and recreational activities (not for credit)	
37. Personal awareness and self-improvement (not for credit)	
38. Philosophy and religious studies	Combine 38 and 39
39. Theology and religious vocations	
40. Physical sciences	Same
41. Science technologies/technicians	Combine 41 with 46, 47, and 48
42. Psychology	Same
43. Security and protective services	Combine 43 with 28 and 29
44. Public administration and social service professions (except cannabis)	Same
45. Social sciences	Same
46. Construction trades	Combine 46, 47, 48, and 41
47. Mechanic and repair technologies/technicians	
48. Precision production	
49. Transportation and materials moving	Same
50. Visual and performing arts	Same
51. Health professions and related programs (except cannabis)	Same
52. Business, management, marketing and related support services (except cannabis)	Combine CIP 4-digit code 30.16 (Accounting and computer science) within CIP 2-digit code 30 (Multidisciplinary/interdisciplinary studies) with CIP 2-digit code 52
53. High school/secondary diploma and certificate programs	Combine 53 with 21
54. History	Same
55. French language and literature/letters	Same
60. Dental, medical and veterinary residency programs	Same
71. Cannabis-related instructional programs	Same
89. Unclassified instructional program	Same

Table F2: Field of Study Classification (Broad Grouping)

PSIS original grouping (CIPPG)	Our adjustment and grouping
001.Personal improvement and leisure	Combine CIP 2-digit codes 21 (Pre-technology education/pre-industrial arts programs) and 53 (High school/secondary diploma and certificate programs) within CIPPG code 120 with CIPPG code 001
010.Education	Same
020.Visual and performing arts, and communications technologies	Same
030.Humanities	Same
040.Social and behavioural sciences and law	Group CIP 2-digit code 22 (Legal professions and studies) within CIPPG code 040 into a separate category (Legal professions and studies)
	Group all other (i.e., except CIP 2-digit code 22) within CIPPG code 040 into a separate category (Social and behavioural sciences)
050.Business, management and public administration	Same
060.Physical and life sciences and technologies	Same
070.Mathematics, computer and information sciences	Same
080.Architecture, engineering, and related technologies	Group CIP 2-digit code 04 (Architecture and related services) and CIP 4-digit code 30.12 (Historic preservation & conservation) within CIPPG code 080 into a separate category (Architecture and related studies)
	Group CIP 2-digit codes 14 (Engineering) and 15 (Engineering tech. and engineering-related fields) within CIPPG code 080 into a separate category (Engineering and engineering-related fields)
	Group CIP 2-digit codes 46 (Construction trades), 47 (Mechanic & repair tech), and 48 (Precision production) within CIPPG code 080 into a separate category (Trades and related studies)
090.Agriculture, natural resources and conservation	Same
100.Health and related fields	Same
110.Personal, protective and transportation services	Same
120.Other	Combine CIP 2-digit codes 21 (Pre-technology education/pre-industrial arts programs) and 53 (High school/secondary diploma and certificate programs) within CIPPG code 120 with CIPPG code 001
	Group CIP 4-digit code 30.99 (Multidisciplinary/interdisciplinary studies, other) within CIPPG code 120 into a separate category (Multidisciplinary studies, other)
500.Unclassified	Same