

The role of firms in the gender earnings gap

Using the Canadian Employer-Employee Dynamics Database (CEEDD) to learn about gender pay disparities

Jiang Li¹ Benoit Dostie² Gaëlle Simard-Duplain³

¹Statistics Canada ²HEC Montréal

³CIDER, University of British Columbia

February 26, 2021

Overview of talk

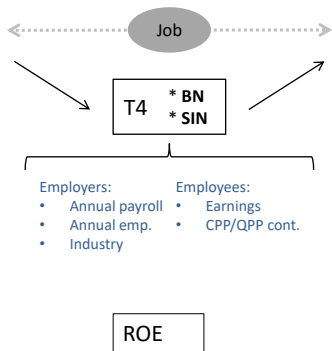
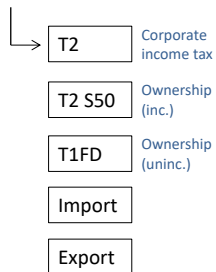
- ▶ Overview of CEEDD
- ▶ Why use CEEDD to study the gender pay gap?
- ▶ The role of firms in the gender pay gap in Canada
- ▶ Opportunities for research

The Canadian Employer-Employee Dynamics Database

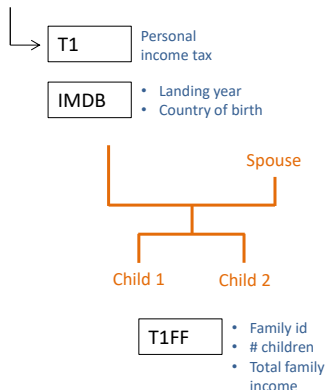
- ▶ Set of **linkable files**
- ▶ Information on **employers and employees**
 - ▶ Employers
 - ▶ E.g., value added, payroll and employment, industry, etc.
 - ▶ Employees
 - ▶ E.g., income from various sources, taxes and transfers, basic demographics (sex, age), family composition, etc.
- ▶ **Longitudinal** structure
 - ▶ Longitudinal employee identifiers
 - ▶ Social Insurance Number, SIN
 - ▶ Longitudinal employer identifiers
 - ▶ Business Number, BN
 - ▶ Allows to identify firm entries and exits, and to follow firms through mergers and acquisitions.

Data sources

Employer (Business Number, BN)



Employee (Social Insurance Number, SIN)



Business Register:
All incorporated and
unincorporated businesses

All tax filers (and their families)
All employees who receive a T4

CEEDD variables (non-exhaustive)

- ▶ Individual-level data
 - ▶ Employment income
 - ▶ Self-employment income
 - ▶ Pension and capital income
 - ▶ Age and sex, marital status
 - ▶ Immigrant status and related variables: landing year, country of birth, knowledge of official language, education.
- ▶ Family-level data
 - ▶ Family identification number
 - ▶ Number and age of children
- ▶ Firm-level data
 - ▶ Industry
 - ▶ Revenues, expenses, assets and liabilities
 - ▶ Employment and payroll, workforce characteristics
 - ▶ Import and export
 - ▶ Business ownership: sex and immigrant status of owner, ownership type, ownership share.
- ▶ Geography
 - ▶ Province

CEEDD coverage

- ▶ For the most part, the data is available starting in 2001 and runs up to two years prior to the current vintage.
 - ▶ Some exceptions; e.g.,
 - ▶ IMDB has information for all non-PRs admitted in 1980 or later.
 - ▶ Import and export data starts in 2010.
- ▶ Population of CEEDD:
 - ▶ Individual tax filers and their families
 - ▶ Employees who receive T4 slips
 - ▶ Incorporated and unincorporated businesses
 - ▶ Business owners:
 - ▶ Owners of unincorporated businesses, if they report self-employment income
 - ▶ Owners of private incorporated businesses

Why use CEEDD to study the gender pay gap?

Role of firms in the gender pay gap increasingly recognized:

- ▶ Slowed down convergence in the economic outcomes of women and men since the 1990s (in Canada and elsewhere).
 - ▶ E.g., Kunze (2018); Olivetti and Petrongolo (2016).
- ▶ Change in factors associated with gap : from human capital differences to differences in occupations, industries, firms.
 - ▶ E.g. Blau and Kahn (2017); Olivetti and Petrongolo (2016); Pelletier et al. (2019); Schirle (2015).
- ▶ Large body of research focused on role of firms.
 - ▶ Women and men sort into firms that vary in pay and non-pay attributes.
 - ▶ At a given firm, women and men differ in their ability or willingness to access better paid position.
 - ▶ E.g., Amuedo-Dorantes and De la Rica (2006); Barth, Kerr, and Olivetti (2017); Bertrand, Black, Jensen, and Lleras-Muney (2019); Datta Gupta and Eriksson (2012); Fortin (2019); Javdani (2015); Jewell, Razzu, and Singleton (2020); Pendakur and Wookcock (2010).

Why use CEEDD to study the gender pay gap?

Different options to study role of firms, all with (dis)advantages:

- ▶ Surveys of individuals (employees)
 - ▶ E.g., Labour Force Survey, Longitudinal and Intl Study of Adults.
 - ▶ Information on employment, occupation and industry, and other features of work like on-the-job training, skills used at work, etc.
- ▶ Survey of employers (and employees)
 - ▶ E.g., Workplace and Employee Survey (1999-2006).
 - ▶ Information on HR practices, workplace performance, technology use, and some worker characteristics.
- ▶ Employer-employee linked administrative data.
 - ▶ E.g., CEEDD.
 - ▶ Strengths:
 - ▶ Very large sample size
 - ▶ Quantity and quality of available variables
 - ▶ Longitudinal and ongoing
 - ▶ Less sensitive to non-response issues (incl. attrition)
 - ▶ Limitations:
 - ▶ Some limitations of available information. E.g., hours and wages not available separately in Canadian admin records (employees); formal information on HR practices, training, etc. (employers).

The role of firms in the gender pay gap in Canada

“What is the role of firm-specific pay policies on the gender earnings gap in Canada?” (with Jiang Li and Benoit Dostie)

- ▶ Based on approach developed in Card, Cardoso, and Kline (2016) (henceforth CCK2016).
- ▶ Start from simple wage equation with worker and firm fixed effects.
- ▶ Estimate model separately for women and men.
- ▶ Use an Oaxaca-style decomposition to interpret **differences in firm premiums** of women and men.
 - ▶ Women and men may be unequally distributed across firms with different premiums.
 - ▶ **Sorting effect**
 - ▶ Women and men may capture different shares of firm surplus.
 - ▶ **Bargaining effect**

Preview of results

- ▶ Firm premiums explain **nearly one quarter** of earnings gap.
- ▶ Differential sorting of women and men in high-paying firms accounts for **approximately half** of that contribution.
- ▶ Firm-level pay disparities explain the other half.
 - ▶ Within-firm differences play a relatively large part compared to other countries.
- ▶ Substantial variation over the life-cycle, by parental and marital status, and across provinces.

Context

A lot of the literature that looks at the role of firms exploits linked employer-employee data. One such approach is CCK2016:

- ▶ Portugal, 2002-2009, private sector employees.
- ▶ Firm premiums account for 20.9% of the gender wage gap.
 - ▶ Sorting effect: 71.4% of firm contribution;
 - ▶ Bargaining effect: 30.6 % of firm contribution.

Subsequent results for other countries:

- ▶ Wide range of estimates for **total firm contribution**
 - ▶ Low range:
 - ▶ 8% in France (Coudin, Maillard, and Tô 2018)
 - ▶ 11% in 1990s West Germany (Bruns 2019)
 - ▶ Mid range:
 - ▶ 26% in 2000s West Germany (Bruns 2019)
 - ▶ 31% in Italy (Casarico and Lattanzio 2019)
 - ▶ 35% in Estonia (Masso, Meriküll, and Vahter 2020)
 - ▶ High range:
 - ▶ 50% in Chile (Cruz and Rau 2017)
 - ▶ 60% in Brazil (Morchio and Moser 2019)
- ▶ Similar estimates of **relative importance of sorting**
 - ▶ Sorting generally accounts for $\sim \frac{3}{4}$ of firm contribution.

Empirical model

Additive two-way worker-firm effects model (AKM):

$$\ln y_{i,t} = \alpha + \beta X_{i,t} + \theta_i + \psi_{j(i,t)} + \epsilon_{i,t} \quad (1)$$

where:

- ▶ $y_{i,t}$: earnings of worker i in year t
- ▶ $X_{i,t}$: observed characteristics of worker, firm, or job
- ▶ θ_i : fixed effect for worker i
- ▶ $\psi_{j(i,t)}$: fixed effect for firm j , where worker i works in year t
- ▶ $\epsilon_{i,t}$: residual error
 - ▶ Time-varying, firm-wide shocks to surplus
 - ▶ Firm-worker match effects
 - ▶ Transitory wage shocks

Empirical model

Empirical considerations:

- ▶ For this model to produce unbiased estimates of firm fixed effects, **exogenous mobility** must hold (Abowd, Kramarz, and Margolis 1999); that is,

$$E(\epsilon_{i,t} | i, t, j(i, t), X_{i,t}) = 0$$

- ▶ Worker and firm fixed effects in AKM model only identified in **connected set** (Abowd, Creecy, and Kramarz 2002);
 - ▶ Set of firms connected by worker mobility.
- ▶ For each equation, firm effects are only identified relative to a reference set of firms (Abowd, Creecy, and Kramarz 2002).
 - ▶ Particularly important here because we want to compare firm effects across two groups for whom they're estimated separately.
 - ▶ Firm premiums need to be normalized.
 - ▶ For this, we need to a **dual connected set**;
 - ▶ Firms in the connected set, where both men and women work.

Decomposition of firm-level pay premiums

Total firm contribution to the gender earnings gap:

$$E[\psi_{j(i,t)}^M | Men] - E[\psi_{j(i,t)}^W | Women] \quad (2)$$

where:

- ▶ $\psi_{j(i,t)}^M$: firm effects for men
- ▶ $\psi_{j(i,t)}^W$: firm effects for women

It can be decomposed into **bargaining** and **sorting** effects:

$$\begin{aligned} & E[\psi_{j(i,t)}^M | Men] - E[\psi_{j(i,t)}^W | Women] \\ &= \underbrace{E[\psi_{j(i,t)}^M - \psi_{j(i,t)}^W | Men]}_{\text{bargaining effect}} + \underbrace{E[\psi_{j(i,t)}^W | Men] - E[\psi_{j(i,t)}^W | Women]}_{\text{sorting effect}} \end{aligned}$$

- ▶ Bargaining effect: degree to which women obtain a smaller share of the surplus generated by firms than men.
- ▶ Sorting effect: degree to which women are segregated in firms that pay differently, relative to men.

CEEDD, 2001-2015

- ▶ Workers 25-54, employed by incorporated (T2) businesses in business sector (education, health, and public admin excluded).

Employees:

- ▶ Earnings from employment at T2 business.
 - ▶ If more than one job, highest-paid is used.
- ▶ Exclusions:
 - ▶ If primary income source is self-employment.
 - ▶ If earnings below \$18,733.
 - ▶ Minimum wage \times average FT working hours \times 48 weeks (Galarneau and Fecteau, 2014).

Employers:

- ▶ Labour productivity: real value added per employee.
 - ▶ Sum of T4 payrolls and net income before taxes and extraordinary items.
- ▶ Employment: no. employees (PD7).
- ▶ Exclusions:
 - ▶ Businesses with only one employee throughout sample period.
 - ▶ Firms with very low output or with value added below \$100.

Descriptive statistics

Table 2: Descriptive statistics, Employees in CEEDD, 2001-2015

	Overall analysis sample		Connected set		Dual connected set	
	Male (1)	Female (2)	Male (3)	Female (4)	Male (5)	Female (6)
Mean Age	40.0	40.3	40.1	40.5	40.2	40.5
Age at 25-29 (%)	14.0	13.7	13.5	12.9	13.3	12.9
Age at 30-39 (%)	32.9	31.1	33.0	31.3	32.9	31.4
Age at 40-49 (%)	36.4	38.0	36.9	38.7	37.0	38.7
Age at 50-54 (%)	16.7	17.1	16.6	17.2	16.7	17.0
Mean earnings (\$)	66 056	48 032	66 891	48 694	68 704	49 220
Quebec (%)	24.1	23.3	24.4	23.4	24.1	23.4
Ontario (%)	38.9	42.6	39.2	42.8	39.8	43.2
British Columbia (%)	11.4	11.9	11.4	11.8	11.2	11.6
Median firm size	199	308	207	323	307	419
Men (%)	71.3	48.2	71.2	48.3	69.8	50.1
Immigrants (%)	16.0	18.8	15.9	18.7	16.2	18.9
Mean log(VA/PD7)	11.3	11.2	10.8	10.6	11.3	11.2
Person-year obs.	40 853 476	21 564 688	39 572 671	20 738 690	35 979 209	19 640 363
Persons	6 603 544	4 018 592	5 558 251	3 243 861	5 341 050	3 137 873
Firms	484 751	421 625	423 876	356 756	299 973	299 973

Note: PD7 employment is the average number of employees at a firm, calculated from the mean of all non-zero monthly employment submissions from payroll deductions and remittances (PD7). Labour productivity is value added (VA) per PD7 employment.

Exogenous mobility

Several authors find evidence in favour of exogenous mobility assumption. E.g.,

- ▶ Germany (Card, Heining, and Kline 2013)
- ▶ Italy (Macis and Schivardi 2013)
- ▶ Portugal (Card, Cardoso, and Kline 2016)

In CEEDD:

- ▶ Usual check of exogenous mobility difficult to perform/interpret using earnings data and threshold approach.
- ▶ Alternative (CCK2016): compare earnings gains of people who move up the job ladder to losses of those who move down.
 - ▶ Dostie, Li, Card, and Parent (2020) find patterns of earnings changes relatively consistent with exogenous mobility.

Normalization of estimated firm effects

- ▶ We can normalize firm premiums by setting them to zero for low-surplus firms.
- ▶ That is, we assume that low-surplus firms don't share rents with workers (no firm premiums).
 - ▶ **We want to identify a set of low-surplus firms.**
 - ▶ We need a measure of firm surplus: we exploit average value-added per worker, which is in the CEEDD.
 - ▶ We need to estimate the threshold beyond which firms start sharing rents with workers.
- ▶ Estimate following equations and find τ that minimizes MSE:

$$\begin{aligned}\hat{\psi}_{j(i,t)}^M &= \pi_0^M + \pi_1^M \max(0, S_{j(i,t)}^0 - \tau) + \nu_{j(i,t)}^M \\ \hat{\psi}_{j(i,t)}^W &= \pi_0^W + \pi_1^W \max(0, S_{j(i,t)}^0 - \tau) + \nu_{j(i,t)}^W\end{aligned}$$

where $S_{j(i,t)}^0$ is the average value-added per worker in firm j .

Results: sorting and bargaining effects

Table 6: Contribution of firms to the gender earnings gap, by age group

	Gender earnings gap (1)	Mean firm premium (male) (2)	Mean firm premium (female) (3)	Total firm contribution to gap (4)	Sorting effect (5)	Bargaining effect (6)
All	0.268	0.176	0.115	0.061 23%	0.029 48%	0.032 52%
By age group						
25 to 29	0.165	0.163	0.115	0.048 29%	0.016 33%	0.032 67%
30 to 39	0.242	0.175	0.121	0.054 22%	0.022 41%	0.032 59%
40 to 49	0.303	0.180	0.114	0.066 22%	0.035 52%	0.032 48%
50 to 54	0.334	0.180	0.106	0.074 22%	0.041 55%	0.034 45%

Note: Sorting effect is weighted by male premiums and bargaining effect is weighted by female shares. Entries in parentheses are shares of overall gap explained by component in column.

Results: sorting and bargaining effects

Table 6: Contribution of firms to the gender earnings gap, by presence of children

	Gender earnings gap (1)	Mean firm premium (male) (2)	Mean firm premium (female) (3)	Total firm contribution to gap (4)	Sorting effect (5)	Bargaining effect (6)
All	0.268	0.176	0.115	0.061 23%	0.029 48%	0.032 52%
By presence of children						
No children	0.239	0.173	0.116	0.057 24%	0.024 43%	0.033 57%
Children	0.351	0.185	0.113	0.072 21%	0.043 59%	0.030 41%
Children (<6)	0.330	0.177	0.110	0.067 20%	0.038 57%	0.029 43%

Note: Sorting effect is weighted by male premiums and bargaining effect is weighted by female shares. Entries in parentheses are shares of overall gap explained by component in column.

Results: sorting and bargaining effects

Table 6: Contribution of firms to the gender earnings gap, by marital status

	Gender earnings gap (1)	Mean firm premium (male) (2)	Mean firm premium (female) (3)	Total firm contribution to gap (4)	Sorting effect (5)	Bargaining effect (6)
All	0.268	0.176	0.115	0.061 23%	0.029 48%	0.032 52%
By marital status						
Married	0.342	0.187	0.119	0.069 20%	0.036 53%	0.032 47%
Common law	0.231	0.155	0.097	0.058 25%	0.025 43%	0.033 57%
Widowed	0.313	0.181	0.108	0.073 23%	0.042 58%	0.031 42%
Divorced	0.230	0.182	0.121	0.060 26%	0.027 46%	0.033 54%
Separated	0.270	0.170	0.108	0.062 23%	0.031 50%	0.031 50%
Single	0.108	0.161	0.116	0.045 42%	0.014 30%	0.031 70%
Single (<30)	0.125	0.162	0.115	0.047 37%	0.015 32%	0.032 68%

Note: Sorting effect is weighted by male premiums and bargaining effect is weighted by female shares. Entries in parentheses are shares of overall gap explained by component in column.

Results: sorting and bargaining effects

Table 6: Contribution of firms to the gender earnings gap, by province

	Gender earnings gap (1)	Mean firm premium (male) (2)	Mean firm premium (female) (3)	Total firm contribution to gap (4)	Sorting effect (5)	Bargaining effect (6)
All	0.268	0.176	0.115	0.061 23%	0.029 48%	0.032 52%
By province						
Quebec	0.210	0.141	0.087	0.054 26%	0.020 37%	0.034 63%
Ontario	0.263	0.185	0.123	0.062 24%	0.029 47%	0.033 53%
BC	0.308	0.166	0.107	0.059 19%	0.035 59%	0.024 41%
QC (ComLaw)	0.224	0.135	0.079	0.056 25%	0.021 38%	0.035 62%

Note: Sorting effect is weighted by male premiums and bargaining effect is weighted by female shares. Entries in parentheses are shares of overall gap explained by component in column.

Results: sorting and bargaining effects

Table 7: Contribution of firms, by province and presence of children

		Gender earnings gap (1)	Total firm contribution (4)	Sorting effect (5)	Bargaining effect (6)
Canada	No children	0.239	0.057 24%	0.024 43%	0.033 57%
	With children	0.351	0.072 21%	0.043 59%	0.03 41%
Quebec	No children	0.184	0.050 27%	0.015 30%	0.035 70%
	With children	0.286	0.066 23%	0.033 50%	0.033 50%
Ontario	No children	0.229	0.058 25%	0.024 41%	0.034 59%
	With children	0.354	0.074 21%	0.044 59%	0.030 41%
BC	No children	0.279	0.055 20%	0.031 55%	0.025 45%
	With children	0.401	0.073 18%	0.050 69%	0.023 31%

Note: Sorting effect is weighted by male premiums and bargaining effect is weighted by female shares. Entries in parentheses are shares of overall gap explained by component in column.

Results: sorting and bargaining effects

Table 8: Contribution of firms, by province and marital status

		Gender earnings gap (1)	Total firm contribution (4)	Sorting effect (5)	Bargaining effect (6)
Canada	Married	0.342	0.069	0.036	0.032
			20%	53%	47%
	Common law	0.231	0.058	0.025	0.033
			25%	43%	57%
	Single	0.108	0.045	0.014	0.031
			42%	30%	70%
Quebec	Married	0.286	0.063	0.029	0.034
			22%	45%	55%
	Common-law	0.224	0.056	0.021	0.035
			25%	38%	62%
	Single	0.072	0.037	0.004	0.033
			51%	11%	89%
Ontario	Married	0.329	0.069	0.035	0.034
			21%	51%	49%
	Common-law	0.200	0.056	0.024	0.032
			28%	43%	57%
	Single	0.085	0.045	0.012	0.033
			53%	27%	73%
BC	Married	0.377	0.064	0.039	0.025
			17%	61%	39%
	Common-law	0.269	0.061	0.036	0.025
			23%	59%	41%
	Single	0.149	0.046	0.023	0.023
			31%	51%	49%

Note: Sorting effect is weighted by male premiums and bargaining effect is

Conclusion

- ▶ Firm premiums explain nearly one quarter of earnings gap.
 - ▶ Sorting and bargaining effects account for near-equal shares of firm contribution.
- ▶ Sorting is most important among married workers.
 - ▶ Muted among single workers.
 - ▶ Substantial differences between married workers and workers in common-law unions, in terms of both firm contribution to gender earnings gap and role of sorting.
- ▶ Substantial variation across regions within Canada, with respect to the relative role of sorting and bargaining.
 - ▶ Importance of bargaining largely driven by Quebec (and Ontario).
 - ▶ In particular: by non-parents and single workers in Quebec.

Opportunities for research

- ▶ Firm-family interactions
- ▶ Business ownership and self-employment
- ▶ Intergenerational applications

Thank you

Table 1: Mean annual earnings in 2015

	Male earnings (1)	Female earnings (2)	Gender ratio (3)
Total	72 556	52 001	0.72
By age group			
25 to 29	50 664	40 020	0.79
30 to 39	66 506	49 264	0.74
40 to 49	82 470	57 215	0.69
50 to 54	85 400	56 597	0.66
By marital status			
Union	80 507	54 288	0.67
Alone	57 421	48 273	0.84
By family status			
Without children	70 721	51 923	0.73
With children	85 539	52 334	0.61

Note: Annual earnings are set to missing if they are less than the rough threshold of \$18,733 threshold (in 2012 real dollars).

Results

Table 3: Summary of AKM estimation results

	Male (1)	Female (2)
Quadratic normalized age / 100	-103.026*** (0,467)	-109.738*** (0,587)
Cubic normalized age / 1,000	988.075*** (32,321)	-1195.573*** (40,774)
Quartic normalized age / 10,000	-1152.255*** (56,387)	-1778.428*** (72,005)
Common law	-0.021*** (0,000)	0.006*** (0,000)
Widowed	-0.062*** (0,001)	-0.016*** (0,001)
Divorced	-0.029*** (0,000)	0.032*** (0,000)
Separated	-0.028*** (0,000)	0.020*** (0,000)
Single	-0.038*** (0,000)	0.033*** (0,000)
Children indicator	-0.002*** (0,000)	-0.001*** (0,000)
Share of children aged <1	0,000 (0,001)	-0.013*** (0,001)
Share of children aged 1-5	0.002*** (0,000)	0,000 (0,001)
<i>N</i>	39 600 000	20 700 000
<i>R</i> ²	0,838	0,836

Note: Married is omitted. Model includes year and province fixed effects. Standard errors are in parentheses. *** indicates $p < 0.01$

Table 4: Summary of estimated two-way fixed effects model

	Male (1)	Female (2)
Standard deviation of ln(earnings)	0.5752	0.499
Number of person-year observations	39 572 671	20 738 690
Summary of parameter estimates		
Number of person effects	5 558 251	3 243 861
Number of firm effects	423 876	356 756
Std. dev. of person effects (across person-yr obs.)	0.438	0.413
Std. dev. of firm effects (across person-yr obs.)	0.192	0.167
Std. dev. of Xb (across person-yr obs.)	0.229	0.220
Correlation of person-firm effects	0.068	0.000
RMSE of model	0.251	0.222
Adjusted R-squared of model	0.810	0.803
Correlation of estimated male-female firm effects ^a		0.599
Inequality decomposition of two-way fixed effects model		
Share of variance of ln(earnings) attributable to:		
Person effects	58.1	68.5
Firm effects	11.1	11.3
Covariance of person and firm effects	3.5	0
XB and associated covariances	11.2	3.9
Residual	16.2	16.4

Note: Summary of estimated two-way fixed effects model from Table 3. ^a Correlation of estimated firm effects for male and female workers across all firms in dual connected set.