Working Paper 19
February 2014

THE REALITIES OF ONTARIO'S PUBLIC SECTOR COMPENSATION
METHODOLOGY APPENDIX

Institute for Competitiveness & Prosperity

METHODOLOGY APPENDIX

THIS DOCUMENT IS A BRIEF OVERVIEW of the methodology used to derive the results in the Institute for Competitiveness & Prosperity's Working Paper 19, *The realities of Ontario's public sector compensation*. This appendix does not cover all the tests performed in the analysis of the public wage premium in Ontario, but includes the most relevant ones.

Statistics Canada surveys were the basis of the study

The study divided employees' compensation into two categories: wages and pension benefits. Although there are other types of compensation, such as bonuses and other benefits, the Institute only had access to data for these two categories.

Detailed wage data came from the Labour Force Survey

Wage data were collected from the Labour Force Survey (LFS) from Statistics Canada. The Institute had access to the micro data files from 1997 to 2012. Some changes to the data, however, were necessary to arrive at the appropriate metrics in this analysis. First, the Institute defined the "representative worker" as an employee with three main characteristics: non-student, single job holder, publicly or privately employed. Second, the main variable for this analysis is the weekly earnings, but this variable was derived from two different variables of the survey: hourly earnings and usual hours worked per week. This transformation was necessary because workers in occupations that do not rely on hourly payments may report an hourly wage which they estimate, and this could lead to distortions. For example, one individual who earns $50,000 dollars per year might divide his yearly wage by 50 weeks, and then by 40 hours usually worked to arrive at an hourly wage of $25. Another individual in the same occupation might divide her $50,000 yearly earnings by 50 weeks, but only 36 hours to arrive at an hourly wage of $27.78. These individuals, however, earn the exact same amount.

Notice that if we multiply the usual hourly wage by the hours usually worked, we get $1000 for both individuals, which takes us back to equal wages for individuals in the same occupation. For individuals with well-defined hourly wages and working schedule, the multiplication of hours worked by hourly wages simply changes the frequency of compensation without altering the results. This method relies on accurate measures of hours usually worked nonetheless, and that is one more reason for defining the representative worker.

Controlling for the different qualifications is crucial for an accurate quantification of the public sector premium. To accomplish that, our study focused on a subset of variables available from the LFS (Exhibit A). The main variable is the "class of worker," which defines whether the employed individual is part of the public or private sector.

Pension plan data at provincial level are scarce

The lack of detailed data was the main obstacle for the pension benefit analysis. Statistics Canada collects pension data through the Pension Plans in Canada Survey (PPIC), which could provide information at the provincial level, but the micro data files are not available to the public. Only some tables from this survey at the national level are available. To complete this analysis, the Institute

Exhibit A The Institute used a set of 22 variables for its study of the public sector wage premium

Variables from LFS, 1997-2012

Variable name	LFS Code
Five-year age group of respondent	age_12
3 largest CMAs (census metropolitan areas)	cma
Class of worker, main job	cowmain
Highest educational attainment	educ90
Number of employees at workplace	estsize
Full-time or part-time main or only job	ftptmain
Final individual or family weight	fweight
Usual hourly wages	hrlyearn
Labour force status	lfsstat
Marital status of respondent	marstat
Multiple or single job holder	mjh
Industry of main job (NAICS 2002/2007)	naics_43
Occupation at main job (NOC-S 2006)	noc01_47
Permanent or temporary job	permtemp
Province of employment	prov
Current student status and type of school	schooln
Sex of respondent	sex
Survey month	survmnth
Survey year	survyear
Job tenure in months	tenure
Usual hours per week at main job	uhrsmain
Union membership status	union

Source: Institute for Competitiveness & Prosperity using *Labour Force Survey* data.

used two sources: the Survey of Labour and Income Dynamics (SLID) and national level data from the PPIC. Given that Ontario, as well as the other reference provinces, represent a large percentage of the Canadian economy, it is fair to assume that the national level numbers are similar to the numbers of these provinces. The definition of representative worker was also applied to SLID data. In the case of nationwide data, the sampled population includes all types of employed people, and the self-employed. Although this is not ideal, as pointed out previously, this is still the best approximation available.

The variables available from SLID are similar to the LFS ones, but from 1999 to 2010. The two surveys differ in that SLID follows the same individuals for a number of years, while the LFS is more of a cross-section of the population – the LFS also maintains some of the individuals constant, but only for a few months at a time.

Two main variables from SLID were used to determine coverage ratios and employee contributions: pension plan with current job (penpln1) and registered pension plan contributions (rppc42). The first variable determines whether or not an individual has a pension plan provided by the employer, while the second one provides information on the dollar value of the employee's contribution. This survey does not provide data on employers' contribution. This information was approximated by using the national data from PPIC.

The pension benefit analysis relied on simpler methods, such as calculating ratios and averages, but to evaluate the public sector wage premium more complex methods were necessary.

The Institute calculated the public sector wage premium using econometric methods

To calculate the public sector wage premium accurately, the Institute developed an econometric model that explains individuals' wages based on a variety of wage-generating. By adding these qualifications, the public sector wage premium can be isolated and its value accurately estimated, since we are controlling for the other characteristics. The basic equation was estimated using the Ordinary Least Squares (OLS) method (Exhibit B).

Each variable in this equation, except for job tenure, entered the model as a binary variable: the variables assume values of zero or one depending on the characteristics of the individuals. For example, the "class of worker" variable (cowmain) equals zero if the individual works in the private sector, and it equals one if the individual works in the public sector. This means the coefficient $ß_1$ represents the public sector wage premium. Some of the variables have more than one category. "Highest education attainment," for instance, has 6 categories, which means that variable actually branches out into 5 binary variables, with the sixth category being the base value.[a] Lastly, the variable "job tenure," which represents an approximation for experience, is measured in months, and is not a binary variable.

While the basic equation was used to estimate the overall public sector premium, an extended version of the model was used to quantify the premium for each of the selected occupations (Exhibit C). The equation above was modified to include the interaction between "class of worker" and "occupation at main job" (noc01_47). With this equation, the resulting public sector premium

Exhibit B The basic equation quantifies the overall public sector premium

BASIC EQUATION

$$\log(hrlyearn_i \cdot uhrsmain_i) = \alpha_i + \mathbf{ß_1} \cdot \mathbf{cowmain}_i + ß_2 \cdot sex_i + ß_3 \cdot union_i + ß_4 \cdot educ\,90_i + \\ + ß_5 \cdot noc01\_47_i + ß_6 \cdot ftptmain_i + ß_7 \cdot age\_12_i + \\ + ß_8 \cdot marstat_i + ß_9 \cdot \log(tenure_i) + ß_{10} \cdot permtemp_i + \\ + ß_{11} \cdot estsize_i + ß_{12} \cdot survmnth + \varepsilon_i$$

Source: Institute for Competitiveness & Prosperity.

Exhibit C The extended equation provided us with the occupation-specific premiums

EXTENDED EQUATION

$$\log(hrlyearn_i \cdot uhrsmain_i) = \alpha_i + \mathbf{ß_1} \cdot \mathbf{cowmain_i} + \mathbf{ß_2} \cdot (\mathbf{cowmain}_i \times \mathbf{noc01\_47_i}) + \\ + ß_3 \cdot sex_i + ß_4 \cdot union_i + ß_5 \cdot educ90_i + ß_6 \cdot noc01\_47_i + \\ + ß_7 \cdot ftptmain_i + ß_8 \cdot age\_12_i + ß_9 \cdot marstat_i + \\ + ß_{10} \cdot \log(tenure_i) + ß_{11} \cdot permtemp_i + ß_{12} \cdot estsize_i + \\ + ß_{13} \cdot survmnth + \varepsilon_i$$

Source: Institute for Competitiveness & Prosperity.

a This would technically change the basic equation, but we ignored that to simplify the notation.

becomes the sum of two coefficiets β_1 and β_2, with β_2 depending on the occupation category of the individual. The coefficient β_1 becomes the premium for the base occupation, which was the senior managers' category.[b] Just like the basic equation, the extended model was estimated using the OLS method.

Because our objective was to calculate the premium for each reference province and occupation, the multiple interactions needed to achieve that with one equation would complicate the interpretation of the coefficients. To overcome that, the Institute applied the equation to the data available for each of the provinces individually. Although more time consuming, this strategy simplifies the interpretation of the results, without altering them.

In addition, the LFS data are technically a cross-section of the Canadian economy. This means that each year has a different set of individuals responding to the survey, despite the fact that some of them remain in the pool for six months of the survey rotation. Nevertheless, the Institute evaluated the equations above in two ways: individual years and aggregated years. The individual year strategy required us to evaluate the equations for each year independently from 1997 to 2012 to get a time series of coefficients for the public wage premium. The aggregate version included every year available in one calculation, extending the model to include year binary variables. Moreover, to account for possible variations over the months of data collection, the Institute also included month binary variables. As expected, however, the effect of these variables was not statistically significant, which means there is no bias in the results depending on the month used for analysis.

The model's R-squared, or goodness-of-fit coefficient, was on average around 63 percent, and the number of observations was around 65,800. Although the R-squared can be deceiving in certain situations, the numbers we found indicate that the model has a good explanatory power for the compensation of public and private sector workers.

Other tests reinforce the Institute's findings

The Institute tested different versions of the basic and extended equations to evaluate how persistent the results were. These extensive tests help reinforce the results found in the analysis.

Some reports use industry variables, rather than occupation, to control for other wage-generating characteristics not captured by other variables. The Institute followed the same strategy to compare the results. In general, the public sector premium increases by roughly 0.8 percentage points after controlling for industry – with and without the inclusion of occupation categories. There is, however, one problem with this method. The industry classification provided by LFS includes public administration at federal, provincial, and local levels in individual categories. Controlling for the industry variable makes the interpretation of the public wage premium complicated, since part of the effect is being captured by the public administration categories. That is, by including a second variable that already differentiates between public and private sector, namely "industry of main job," part of the effect we want to estimate is controlled for, making the interpretation of the coefficient β_1 uncertain. Therefore, the Institute believes that using occupation categories is more appropriate for this analysis than industry categories.

Another important test involved distinguishing the results by Census Metropolitan Areas (CMAs). The LFS data provides data differentiated across 4 categories: Toronto, Montreal, Vancouver, and other CMA or non-CMA. After controlling for that, the Institute found an increase in the public sector wage premium, but this increase was not uniform over time. Moreover, this premium for Toronto was lower than the general premium by 1 percentage point for most years in the analysis.

The last important test we performed was to compare the results of applying the equations above to the reported hourly wage, instead of the calculated weekly wages. One disadvantage of using the weekly wages is that we are not controlling for possible differences in working hours between the two sectors. By using hourly wage, we control for the hours worked, but we also open the possibility of misreporting, or artificially calculating hourly wages, as previously explained, particularly for the selected occupations, which are salaried positions.

b The choice of base occupation does not affect the results.

The results show that there is a considerable increase in the public sector wage premium if we use hourly wages as the dependent variable (Exhibit D). In 2012, for example, the premium practically doubles by using the hourly wages. Nonetheless, the trends and conclusions remain the same. That is, regardless of the choice of dependent variable, the wage premium increased considerably over the years, particularly after 2002.

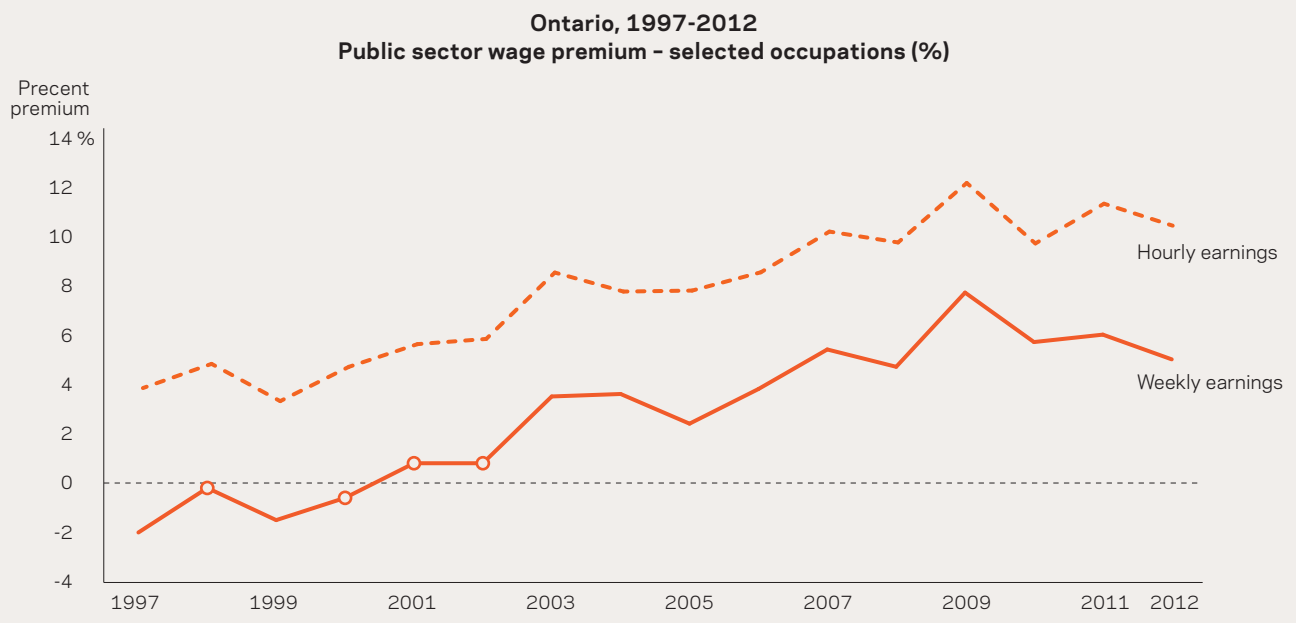
For the different occupations, the rank-dependent premium is also true when using hourly wages as dependent variable: managerial occupations still have negative premiums, but they are less negative, while clerical and administrative occupations have even larger wage premiums (Exhibit E).

The difference between the original and adjusted premiums comes from persistently lower hours worked in the public sector than in the private sector. On average, usual hours worked per week are 5 percent lower in the public than in the private sector, and this difference is around 9 percent for senior managers. This means that part of the wage difference between the sectors is reduced by longer hours worked by private sector workers.

Apart from the problem of artificially reported hourly wages, it is hard to determine whether hourly wages are appropriate because hours worked in the sectors might not be flexible. That is, although it could be possible for public sector workers to increase their work hours, the extra hours could fall under overtime hours with different payment structure, which would alter the premium. In addition, it is possible that private sector workers cannot decrease their working hours. This means the accurate evaluation of the wage premium is the one using the weekly earnings, and differences in hours count as part of the "compensation premium." The most precise evaluation, however, is to think of the weekly earnings premium as the lower bound, and the hourly earnings premium as the upper bound for the public sector premium. To avoid overestimating the premium and the possible misreporting of hourly wages, the Institute chose to report in the study the more conservative measure. Regardless of the choice of dependent variable, it is reassuring to find that the overall trends and conclusions are still the same, which increases the confidence in the findings and the recommendations provided.[c]

c The most noticeable difference is in the "specialist managers" category, which show positive premiums when hourly wage is the dependent variable. Nonetheless, this category still shows wage premium growth over time, with negative values in the first half of the period analyzed and positive values in the later years.

Exhibit D Wage premiums increase consistently if hourly wage is used as dependent variable



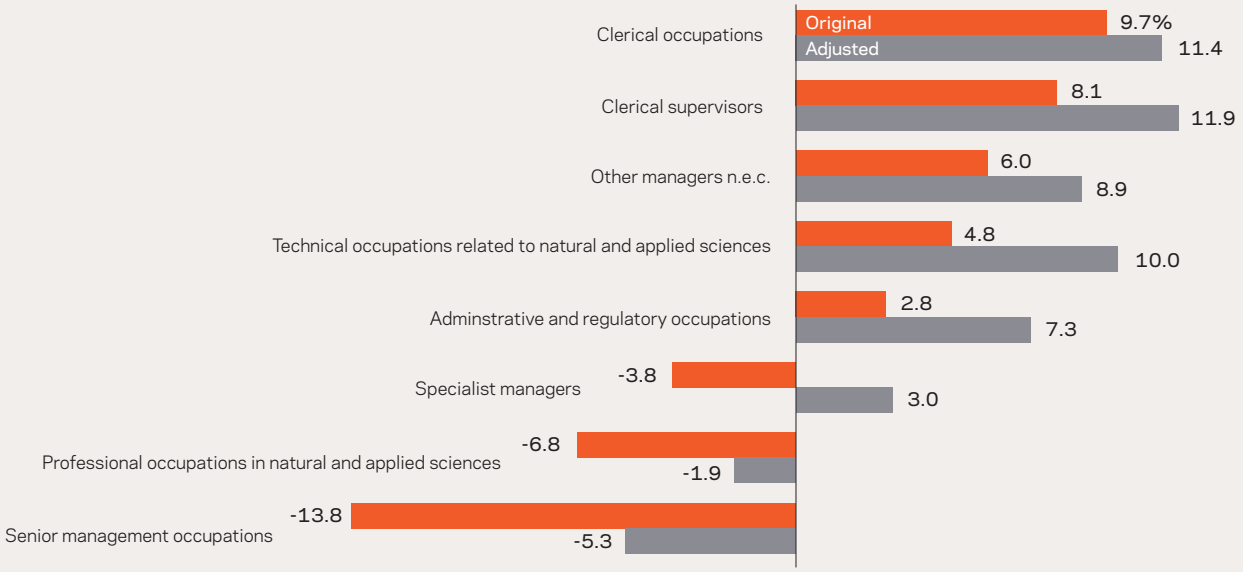
Note: The results are statistically significant at the 5 percent level, except for the data points indicated by hollow circles. In those years, the premiums were indistinguishable from zero at the 5 percent level.
Source: Institute for Competitiveness & Prosperity analysis based on data from the *Labour Force Survey*.

Other variables also affect the public sector premium, but some aspects of wage compensation cannot be directly evaluated. For example, one variable widely described in the literature that cannot be controlled for is "ability." Workers' ability differs considerably, which could not only affect their wages, but also their decisions to get more education. It is possible to speculate that, if ability and education are positively correlated and the public sector has a larger proportion of highly educated workers when compared to the private sector, the wage premium would probably decrease after controlling for ability. Another variable that could possibly decrease the wage premium is immigration status. As explained in the study, there might be an immigrant penalty – intrinsic lower wages experienced by immigrants. If the private sector employs a higher proportion of immigrants, by not controlling for that, the wage premium could be overestimated. Narrowing the data to selected occupations, however, should mitigate part of this issue, but the ideal is to control for that characteristic. The LFS does not have a variable indicating whether or not individuals are immigrants.

Despite some limitations, the Institute believes that the results are very consistent and accurate. Other controls can be added to the analysis, but with little improvement in the accuracy and conclusions for the public sector wage premium.

Exhibit E **Rank-dependent wage premiums still present when using hourly wage as dependent variable**

Ontario, 1997-2012 (average)
Public sector wage premium – original and adjusted (%)



Note: The results are statistically significant at the 5 percent level. Adjusted premium refers to the results holding hours worked constant.
Source: Institute for Competitiveness & Prosperity analysis based on data from the *Labour Force Survey*.

Institute for
Competitiveness & Prosperity

Institute for Competitiveness & Prosperity
ISBN: 978-1-927065-09-9