Gender wage gap of comparable workers: an application to Chile, 1992-2009.

Abstract

This paper estimates the gender wage gap for Chile between 1992-2009, but using first time a matching comparison. In order to contribute to the empirical literature, this paper uses a novel technique called Coarsened Exact Matching which imposes the comparison among comparable workers. The results suggest that the wage gap exists, but it is lower than preivous estimations, specially when only comparable workers are considered. This results opens the discussion about how well estimated is the gap when exist a high heterogeneity between male and female workers. The results also show a increment in wage gap from 2000. Finally, onlye the 58% of comparable male workers earns more wage than similar females workers. However, this 58% presents larger differential than its comparable 42% of female workers. This differential is also growing during the last years.

1. Introduction

The Sixth World Economic Forum's Global Gender Gap Report 2011 improves the world ranking of Chile from the 64th (2009) to the 46th (2011) position.[1] While advances exist in health access and education equality, the report still indicates alarming inequalities in the labor market, especially in the gender wage gap. Fuentes, Palma, & Montero (2005) suggest a reduction in this gap between 1990 and 2003, but they still estimated the gap around 23% in 2003. Montenegro (2001) estimates a larger gap in the upper quantiles of the conditional wage distribution with a range between 10%-40%. Recently, Perticara & Bueno (2009) using a unique sample with information of current labor experience estimate a gap between 11% and 18% between 2002 and 2006, but interestingly they suggest that the gap has widened during last years.

These articles offer some advances on the gap estimation, but they also leave pending assignments. The gender wage gap is significant during the last 20 years, but with a heterogeneous behavior across the wage distribution. Moreover, there is disagreement about the behavior of the gap during the last years. Additionally, the results obtained by Perticara & Bueno (2009) show that the estimations differ depending on the empirical approaches, especially with better proxies of labor experience. In summary, these facts suggest that additional research is still necessary for the Chilean case.

This paper fills this gap using matching comparison, as Frolich (2007) and Ñopo (2008), with the objective of comparing the wages of similar female and male workers, a condition that is not satisfied with standard techniques such as Oaxaca-Blinder (OB). Additionally, matching does not require the specification of wage equations, which avoids the strong linearity assumption of wage equations. Finally, matching provides a complete distribution of gender gap instead of a unique average gap for the complete sample.

[1] The report is available at http://www.bahrainedb.com/uploadedFiles/Bahraincom/BahrainForBusiness/Global%20Gender%20Report%202011.pdf

Frolich (2007) and Ñopo (2008) suggest the use of Propensity Score Matching, but new matching techniques are now available. Iacus, Porro, & Stuart (2011) and Ho, Imai, & Stuart (2007) indicate that Coarsened Exact Matching (CEM) owns desirable statistical properties over alternative methods. First, CEM bounds the degree of model dependence and causal effect estimation is fixed ex-ante by the user choice. Second, CEM meets the monotonic imbalance bounding which implies that a variable balance does not affect the balance of other variables. Third, this method is easily implemented and it almost does not consume time in computation in comparison to alternative matching procedures. As far as the author knows, these properties have not been used to estimate the gender gap, yet.

In spite of the benefits of matching comparison, the technique also presents constraints. The first one is the intensive loss of observations. The matching between male and female workers implies a wasting of observations. This generates serious consequences on the estimated gap, especially on its standard deviation for testing process. A second problem is that, after matching, male and female workers should present similar distribution of characteristics. This is achieved when the average and higher moments are similar. However, most literature only presents analysis in the first moment, namely the average, and scarce attention has been focused on the higher moments. Both problems are in the core of matching procedure and affect, seriously, the statistical significance of the estimated gap. This paper considers these constraints and both are incorporated and tested in the estimation.

On balance, this article contributes in four areas: 1) The first estimation of the gender gap using matching comparison for the Chilean case, 2) The first application of CEM for estimating the gender wage gap, 3) A deep robustness analysis of matching techniques in reference to size of the sample after matching and the balance property, and 4) An update of the gender gap of Chile with the incorporation of new estimations for 2009.

Our results support the existence of gender wage gap between 1992 and 2009, but with some differences with early literature. First, when comparable workers are considered, then the gender gap is lower than previous estimations between 1992 and 2003. From this point, the gap estimated by CEM is higher than OB. This first result says that the workers characteristics are fundamental for identifing the gap and the comparison of comparable workers is needed. A second result is that wage gap raises from 2000. By 2000 the gap is estimated around 12%, but it is aroun 19% by 2009. This result is in line with Perticara & Bueno (2009)

CEM also provides a the complete distribution of gender gap. Using a set of comparable workers, the 60% of comparable workers show a gender gap against women, and around of 40% of female workers earn higher wages than comparable male workers. However, the gap against women is much larger than against male workers. This gap has been constantly increased during the period of analysis. Between 1992 and 2009 the gap against women has increased around 3.6 times.

The structure of the paper is as follows. Section 2 proposes a theoretical discussion to make a clear distinction between the Oaxaca-Blinder decomposition and Coarsened Exact Matching. Section 3 describes the data for the period 1992-2009, with emphasis on those elements ignored in the previous literature. Section 4 describes the results and Section 5 provides conclusions.

2. Methodology: Oaxaca-Blinder decomposition versus matching estimator.

The gender wage gap is estimated (mostly) using wage equations *à la* Mincer, but also facing several methodological challenges. The first one is the selection bias generated by those who are not participating in the labor market. The second challenge is the linear function form between characteristics and wages, such as OB suggests (Dolton & Makepeace (1987) and Munro (1988)). Additionally, OB provides an average gap, but it loses information about its complete distribution.

While these points have been discussed in the literature[2], scarce attention has been paid to the implicit *counterfactual* assumption imposed by OB. OB computes how much men would earn by using the shadow prices for female workers, assuming the existence of potential women with similar characteristics as *counterfactual* men:

$$\bar{w}^0 - \bar{w}^1 = \bar{\beta}^{\bar{0}}\overline{x^0} - \bar{\beta}^{\bar{1}}\overline{x^1} = \bar{\beta}^{\bar{1}}\left(\overline{x^1} - \overline{x^0}\right) + (\bar{\beta}^{\bar{1}} - \bar{\beta}^{\bar{0}})\overline{x^0}$$

where $\bar{\beta}^{\bar{0}}$ and β^1 are estimated coefficients for men and women, respectively, and the average characteristics are $\overline{x^1}, \overline{x^0}$. This counterfactual $\beta^1 x^0$ could not exist for some workers and combinations of male characteristics cannot be found for female workers (Barsky, Bound, Charles, & Lupton (2002) and Black, Haviland, Sanders, & Taylor (2004)).

For example, a mining worker who is willing to spend seven days working at 4,500 meters above sea level is a fact hardly replicated by a female worker. Ñopo (2008) labels this problem as *comparability support misspecification* and, as far as the author knows, their implications on gender gap have not been studied for the Chilean case. This paper proposes to cover this gap, following the pioneer works by Frolich (2007) and Ñopo (2008) who suggest the matching comparison as an alternative to deal with the bias of OB decomposition. The next section describes the theoretical decomposition proposed by this technique and more advanced references are available in the previously discussed articles.

a. Non-parametric decomposition.

Let w be a random variable called wage and X an n-dimensional vector with worker characteristics. The functions $F^M(\cdot)$ and $F^F(\cdot)$ identify the cumulative distributions of individual

[2] For wage equations and Heckman corrections, see Paredes & Riveros (1994), Fuentes, Palma, & Montero (2005) and Perticara & Bueno (2009). For an analysis of complete wage distributions, see Montenegro (2001).

characteristics X, conditional on being male or female. The probability functions are described by $dF^M(\cdot)$ and $dF^F(\cdot)$. The probability measured $\mu^F(S)$ indicates the probability of a set S under the distribution $dF^F(\cdot)$ and this measure is analogue for females. The expected value of the wages, conditional on characteristics and gender, are specified by $E[w|M] = g^M(X)$ and $E[w|F] = g^F(X)$, then:

$$E[w|M] = \int_{S^M} g^M(X)\, dF^M(\cdot)$$

$$E[w|F] = \int_{S^F} g^F(X)\, dF^F(\cdot)$$

where S^M and S^F are the support of characteristics for male and female workers. A gross measure of wage gap is defined by

$$\Delta = \int_{S^M} g^M(X)\, dF^M(\cdot) - \int_{S^F} g^F(X)\, dF^F(\cdot)$$

This stage is crucial for understanding how matching comparison overcomes the weakness of the OB approach. It assumes that only a portion of the support is overlapped between male and female workers, defined by $S^M \cap S^F$. The non common support for males and females workers is $\overline{S^M}$ and $\overline{S^F}$, respectively. Using the common and uncommon support, plus the expected valued defined with $\mu^F(S)$ and $\mu^M(S)$, the gender gap can be decomposed by:

$$\begin{aligned}\Delta = & \left[\int_{\overline{S^F}} g^M(X)\frac{dF^M(x)}{\mu^M(\overline{S^F})} - \int_{S^F} g^M(X)\frac{dF^M(x)}{\mu^M(S^F)}\right]\mu^M(\overline{S^F}) \\ & + \int_{S^M \cap S^F} g^M(X)\left[\frac{dF^M(x)}{\mu^M(S^F)} - \frac{dF^M(x)}{\mu^F(S^M)}\right](x) \\ & + \int_{S^M \cap S^F} [g^M(X) - g^F(X)]\frac{dF^M(x)}{\mu^F(S^M)} \\ & + \left[\int_{S^M} g^F(X)\frac{dF^F(x)}{\mu^F(\overline{S^M})} - \int_{\overline{S^M}} g^F(X)\frac{dF^F(x)}{\mu^F(\overline{S^M})}\right]\mu^F(\overline{S^M})\end{aligned}$$

Which, Ñopo (2008) designs by

$$\Delta = \Delta_M + \Delta_x + \Delta_0 + \Delta_F$$

This decomposition is similar to OB, but without the strong imposition of a functional form. The portion Δ_M is the part of the gap explained by differences in characteristics between two groups, those matched and those which are not. The same definition applies for Δ_F. The portion Δ_x is the typical explained gap in the OB decomposition, the difference that depends on the human capital in male and female workers. The key component is Δ_0, which is a portion of the gender gap that cannot be attributed to differences in characteristics. This measure, labeled Δ_0, is a clean estimation of gender gap because is an unexplained portion of wage differential when they are comparable workers.

3. **Data**

This paper uses the National Socioeconomic Characterization Survey (CASEN) between 1992 and 2009.[3] These data have been previously used by Montenegro (2001) and Fuentes, Palma, & Montero (2005), but some changes are incorporated here. Fuentes, Palma, & Montero (2005) only use the years of education as independent variable in the hourly wage equation, ignoring other important controls such as occupation, personal characteristics or economic sector. Furthermore, these works do not discuss how the missing data were processed and how the influence of outliers was corrected. Additionally, CASEN contains weights to expand the sample at population level, but earlier paper do not make reference to this point neither. Likewise, Montenegro (2001) also uses CASEN (1990 to 1998), but similar problems are detected. This author only works with years of education and experience but additional controls are not considered. The inference is affected because the population weights are, again, not considered.

In order to improve the previous estimations, this paper aims to improve these shortcomings. Only workers between fourteen and sixty-four years old, who report a positive wage and who live in an urban area, are considered[4]. The hourly wage is used, but an introductory discussion regarding monthly wage is also provided. The workers with less than 30 hours per week were dropped to avoid the distortion of temporary contracts. Military workers and those who do not specify occupations or economic sector are dropped from the sample[5]. The occupation and economic sector are considered as control variables. Any worker with missing data is dropped. The

[3] The CASEN 1990 is also available, but this survey was sampled just before the return to democracy (1988-1989) and several economic sectors, such as public administration, contain anomalous values.

[4] According to Gill (1991), the inclusion of rural population generates a high distortion in the estimation of the gender gap. This paper follows this suggestion in order to avoid the problems derived from self-employer and temporary contracts, which are typical in rural areas.

[5] Wages in the military service are not traded on standard labor markets .

outliers were detected running regressions by year using the hourly wage as a dependent variable and years of education, experience, occupation, economic sector, marital status and household head as independent variables. Observations with standard residuals greater than 3 (absolute value) were deleted. Additionally, *dfbeta* tests were also estimated for each regression, but they do not discard any observation. A general set of descriptive statistics is reported in Table 1

<<INSERT TABLE 1 AROUND HERE>>

4. **Results**

The results are divided into three sections. The first section replicates the, already, existent literature estimating the gap with two dependent variables and using different controls. This exercise allows us to determine the type of wage and the set of controls to be used. The second section uses this choice to estimate the gender gap, but using CEM and some cautions with the fast reduction of observations is discussed. Finally, once the advantage generated by CEM is established, the distribution of gap is analyzed.

a. **Dependent and control variables**

A previous step for the analysis is the use of monthly or hourly wages as dependent, although some elements support the hourly version. For example, most women work in part time jobs, such as those in the agricultural sector, where the labor journey measured by hours is more common than monthly contracts. According to CASEN 1992-2009, the agriculture employees represent between 12-16 percent of the total workforce during the last 20 years. This scenario implies that monthly wages overestimate the real gender gap because the wage for male workers would be larger than females due to a labor journey effect. In order to provide information of this decision, the Table 1 shows the estimations of national gender gap using both wages.

<<INSERT TABLE 2 AROUND HERE>>

The first and second rows show the national gender gap using the average of hourly and monthly wages. As expected, monthly gender gap is higher than estimated hourly gap. Nevertheless, this average comparison is still misleading because the gap could be explained by differences in human capital (Bayard, Hellerstein, Neumark, & Troske, 2003). The third and fourth rows show the gap using the unexplained differential estimated through Oaxaca-Blinder decomposition, but controlling by years of education, potential experience and its quadratic version. The gender gap is larger for both wages in comparison with the first and second rows. This result is explained because the years of education of women are higher than men (see Table 1), then a double discrimination process exists: lower wages for women and less recognition of human capital.

Differences in occupations, economic sector or marital status also explain the wage differential between genders. The fifth and sixth rows add controls such as occupation, economic sector, household head and marital status in the Oaxaca-Blinder decomposition. The gender gap is smaller to previous exercises and men workers earn, on average, 18 percent more than women workers for the complete period. Additionally, the last specifications show the extremely slow reduction, smaller than 4 percent in the gender gap between 1992 and 2009.

Summarizing these points, the hourly wage is selected as a proper dependent variable. The controls must be incorporated in order to identify the gender gap. The occupation or economic sectors are wage variation that should not be attributed to gender-gap. This hypothesis is supported by the estimations and the reduction of gender gap that is much lower when the human capital and economic sector are discounted. This analysis is similar to the previous literature of the Chilean case with an Oaxaca-Blinder perspective. However, the workers could not be comparable and these estimations would be still biased. Next step moves forward the comparison of similar workers.

b. **Assumptions for a correct matching comparison.**

According to Abadie & Imbens (2007), matching comparison identifies and consistently estimates the treatment effect (gender gap for our exercise) if and only if:

1. The treatment M or F is independent of conditional on $X = x$.
2. $c < \Pr(M|X = x) < 1 - c$, for some $c > 0$

The first condition establishes the core of a perfect matching: after controlling by human capital variables X, the gender gap is an effect, perfectly attributable to the gender (F, M) and there is no correlation between X and the treatment indicator $M|X$. The second one establishes that the assignment probability, namely the probability of being a female in our case, is far away from the extreme probabilities 0 and 1. In spite of the simplicity of these statements, Iacus, Porro, & Stuart (2011) discusses that both theoretical constraints are not generally imposed by empirical analysis. For example, Mizala et. al. (2011) propose an interesting matching approach for identifying wage gap between public and private workers in some Latin America countries, but they do not discuss these assumptions. This lack of discussion could also, dangerously, affect the bias of the results .

Only for methodological purposes, the discussion starts with the assumption 2. This assumption requires a functional form for $\Pr()$. Guo & Fraser (2010), such as in most literature, recommend using a logistic regression. I set the dependent variable as 1 for women and 0 for men, while the X set is the same as those used for long hourly wage equations above.[6] After estimating the predicted probabilities, the routine selects only those predicted probabilistic values above 0.1 and

[6] The complete code is written in Stata© and it is available upon author request.

below 0.9, or equivalently, and a $c = 0.1$ is imposed. This pruning warranties that only those comparable observations, in a $\Pr()$ sense, are candidates for a future matching comparison.

<<INSERT TABLE 3 AROUND HERE>>

Table 3 shows how the number of observations changes after the consideration of assumption 2. For example, the number of male workers completely dominates the number of female workers for any region in any year. According to the assumption 2, we do not need those observations with extreme probabilities of being assigned to the treatment. Using the trimming rule, I take out those observations and the new distribution is shown by "Man PS" and Women PS". The main change is related to the distribution of Man PS, now I do not need the complete set of male workers and only a subset of them are candidates to be comparable to women. This imposition is one of the main challenges of exact matching, namely, the process for finding matched workers, because the male and female workers are already similar in a propensity score measurement.

Once the assumption number 2 is warranted, I return on the assumption number one. This assumption is strongly related to the balance of covariates before and after matching. At this point, two decisions must be considered: 1) How does the researcher measure the balance property?, and 2) what are the variables considered for matching process? The concept balance is not just a similarity in average, let us say education years, between male and female workers. High orders of the complete empirical distribution of covariates, histograms or quintiles must be also considered (Blackwell, Iacus, King, & Porro, 2009). I capture this balance using the L_1 statistics proposed by Iacus, Porro, & Stuart (2011). Instead of considering only a mean test, this statistic considers the proximity of multivariate histogram. The statistic is:

$$L_1(f,g) = \frac{1}{2}\sum_{l_1 \dots l_k} \left| f_{l_1 \dots l_k} - g_{l_1 \dots l_k} \right|$$

The procedure builds a crosstab of k covariates for treated (women) and control (men) group. The frequencies for treated $f_{l_1 \dots l_k}$ and control $g_{l_1 \dots l_k}$ are used to estimate the balance measure. If L_1 is close to zero, then the balance property is satisfied. Otherwise, the matching procedure does not satisfy the first assumption discussed. Formally, I have three samples of workers: 1) the complete sample of workers, 2) the sample with the propensity score imposition, 3) the sample with the CEM procedure and propensiy score imposition. For this last group, I apply the CEM procedure using the same set of variables of the logistic regression in the previous figure in order to maintain a fair comparison between two methodologies. If the matched frequencies are f^m and g^m for CEM matching, f^{ps} and g^{ps} for trimming using propensity score and, f and g for the total sample, the expected relationship is:

$$L_1(f^m, g^m) < L_1(f^{ps}, g^{ps}) < L_1(f, g)$$

Following the literature, the imposition of $c = 0.1$ should also reduce $L_1(f^{ps}, g^{ps})$, but less than $L_1(f^m, g^m)$. A second consideration with CEM is the metric of the variables. For example, a continuous variable makes it harder for finding an exact matching. Iacus, Porro, & Stuart (2011) suggest to use any information to categorize the continuous variable, in order to make easier the CEM procedure. For example, the Chilean education system splits the education system in primary (12 years), secondary or high school (16 years) and college or university (20 years or more). In order to facilitate the coarsened procedure, the variable years of education are categorized using these blocks. The next section provides a detailed discussion about what variables must be considered for CEM.

<<INSERT FIGURE 1 AROUND HERE>>

Figure 4 shows the years on the horizontal. The line labeled "Total" represents the original imbalance in the empirical distribution of covariates for both groups. The imbalance is extremely high, between 0.91 and 0.95, supporting the danger of a direct comparison between female and male workers such as Oaxaca-Blinder methodology. The second line, labeled PS, represents the L1 statistics when the observations with $0.1 < \Pr(\quad) < 0.9$ are dropped. Clearly, PS gradually reduces the imbalance, but it is still bounded between 0.86 and 0.90. The most significant reduction is obtained by CEM procedure where the L1 statistic is reduced between 0,64 and 0,71. These results support the idea that the gender gap could be highly biased due to the wrong comparison of workers.

c. **A comparison of regional gender gap using the Oaxaca-Blinder and CEM.**

Figure 2 shows the gender wage gap using Oaxaca-Blinder and CEM. In the case of Oaxaca-Blinder, the results are similar to those reported in the last row of Table 1. To make a fair comparison, CEM uses the same set of independent variables of OB decomposition. CEM produces a complete empirical distribution of wages between male and female workers, while OB only provides an average gap. Only for explanatory reasons, the complete distribution is summarized with the weighted average of the gap and it is the benchmark for OB.

<<INSERT FIGURE 2 AROUND HERE>>

Figure 2 shows interesting patterns. First, the results are different than those provided by Fuentes, Palma, & Montero (2005). As was discussed in the data section, this paper includes additional considerations such as inclusion of occupation, economic sector and some personal characteristics in the wage equation plus a treatment of outliers and the consideration of population weights. The first difference is that the OB, here reported, is lower than the same set of. This difference is probably due to the additional variation provided by occupation and economic sector. In the earlier literature, these are omitted variables in the Mincer equation and the bias rises. A second difference is the trend across the years. While Fuentes, Palma, & Montero (2005) suggest a gap of

28% for 2003, the new OB only reaches to 14%, putting in evidence the overestimation. Finally, while the authors indicate a reduction of the gap, around 30 points between 1990-2003, the OB reported estimation shows a reduction not higher than 5 points.

However, the most interesting results emerge when the gap is estimated only among comparable workers. CEM reveals a lower gap, between 1992 and 2000, than the OB decomposition. Moreover, the gender gap of comparable workers is considerably lower than those reported by Montenegro (2001) and Fuentes, Palma, & Montero (2005). Nevertheless, the most interesting pattern appears between 2000 and 2009. Such as Perticara & Bueno (2009) suggest, the gender gap is widening during the last ten years. The gap moves from 11% to 20% in the selected period, supporting the bias hidden when the comparability is ignored. This increment is significant and the gender gap increased around 8 points between 2000 and 2009.

Summarizing, the standard methodology used by the literature provides a biased picture of the gender gap in Chile. In contrast to Montenegro (2001) and Fuentes, Palma, & Montero (2005), this paper shows that the comparison of comparable workers reveals the bias behind the OB decomposition. The gender gap between comparable workers is lower among who are not. However, an alarming increment in gender gap is observed from 2000, as in the previous literature. Setting the advantages of CEM on OB, the paper turns the analysis into the advantage provided by the empirical distribution of gender gap as an output of CEM.

d. **A comparison of regional gender gap using the Oaxaca-Blinder and CEM.**

Figure 3 lets us analyze the gender gap using the complete distribution on the horizontal axis and the gender gap in the vertical axis. This plot contains the same series for the eight years under consideration. An interesting result shows that, on average, the gender wage gap exists for almost the 58% of the total comparable workers, but obviously, not all the women are subject to gender gap. This result is one of the advantages derived from matching comparison because the researcher is able to analyze the complete distribution of gaps. That implies that even if some female workers earn higher wages than male workers, its proportion is just around the 42% of the total workers. Additionally, while the gender gap of males against women is around 750 and 2,900 pesos for the complete period, the gap of women against men is bounded just between 600 and 2,000 pesos. This result shows that, even though both genders show wage gaps, the gap against women is much lower than against men and this result is stable during the set of years.

<<INSERT FIGURE 3 AROUND HERE>>

A second result is the evolution of the gap for similar percentiles between the first and the last year in order to provide a dynamic picture. Figure 4 shows the distribution of plots the gender gap for 1992 and 2009. Additionally, the relative gap (gap 1992 divided by gap 2009) is also incorporated. The first pattern is how the gap, against women or men, has increased between 1992 and 2009. For the gap against women, the size of the gap for 2009 is around 3.6 times the gap for the same percentile in 1992. The relative comparison is constant for most part of the series when the gender gap against women is identified. In the opposite case, the relative gap shows a higher volatility. For example, the relative gap is decreasing for the right tail of the gender gap distribution. In other words, the positive premium for women over men workers rapidly decreases across the upper tail of wage gaps.

<<INSERT FIGURE 4 AROUND HERE>>

Summarizing, a set of interesting patterns appears. The gender gap exists for almost 60% of the comparable workers: 6 out of 10 comparable couples show a negative gender effect against women. However, the women also show an unexplained positive wage variation against men, but its proportion is just around 40%. The relative gender gap has increased between 1992 and 2009 for the complete distribution. The gap against women or men has increased around 3.7 between both series, but its magnitude varies across the distribution. Discrimination against women has increased more than against men between 1992 and 2009.

5. **Conclusions**

This paper contributes, with new empirical insights, about how to estimate the gender gap using comparable workers and applying these ideas to the Chilean context. The results provide a new set of estimation oriented to improve the understanding of the Chilean case and some details about the gap of Oaxaca-Blinder are also discussed.

The results suggest a significant difference between Oaxaca-Blinder decomposition and the proposed Coarsened Exact Matching. When the data set incorporates a proper consideration of controls variables and weights, the Oaxaca-Blinder estimated previously by the literature differs from the estimation provided by this paper. My estimations show that gender gap is lower than in Montenegro (2001) and Fuentes, Palma, & Montero (2005), especially in the last years. In 2009, Oaxaca-Blinder estimates the gender gap around 16%, while Fuentes, Palma, & Montero (2005) reaches around 23%. The differences are attributed to the improvements in the treatment of outliers and the consideration of additional controls and population weights.

Despite the improvements, the Oaxaca-Blinder incorporates a theoretical bias when no comparable workers are considered. This paper proposes the Coarsened Exact Matching and remarkable differences are estimated. First, CEM estimates and average gender gap much lower than Oaxaca-Blinder during most of the period. This implies that some male workers who earn high wages are not comparable and they should not be considered for wage gap. CEM estimates also support the previous results provided by Perticara & Bueno (2009) regarding the widening of the gender wage gap since 2000. Between 2000 and 2009, the gap grew from 12% until 20%. This result also supports the conclusion derived from Sixth World Economic Forum's Global Gender Gap Report 2011 and additional attention of policy maker must be paid to the labor markets.

CEM provides a missed dimension of the wage gap distribution. 60% of comparable workers show a gender gap against women, and around of 40% of female workers earn higher wages than comparable male workers. However, the gap against women is much larger than against male workers. Worryingly, the gender gap has widened from 2000. This increment is bigger than those estimated by Oaxaca-Blinder. In 2000 the gender gap was around 12% on average, but it is around 20% in 2009. The situation is not different when the gap distribution is analyzed. Between 1992 and 2009 the gap against women has increased around 3.6 times.

This paper shows the necessity of additional discussion about the gender gap in Chile, as well as the scarce discussion about a fair comparison of wages. Moreover, the scenario is much more complex than the previous literature stated and new evidence is provided to analyze the potential reasons behind the increment in the gender gap. New empirical directions also must be directed with a strong focus on the comparability problem.

TABLES AND FIGURES

Table 1: Descriptive statistics for CASEN 1992-2009

	1992	**1994**	**1996**	**1998**	**2000**	**2003**	**2006**	**2009**	**1992**	**1994**	**1996**	**1998**	**2000**	**2003**	**2006**	**2009**
	Man								Women							
Hourly wage	647	972	1125	1360	1321	1565	1618	2062	472	745	877	1095	1090	1317	1362	1816
Years education	10,11	10,34	10,67	10,91	11,02	11,34	11,26	11,46	10,66	11,02	11,33	11,61	11,63	12,00	11,93	12,27
Experience	20,17	20,58	20,12	20,52	21,21	21,18	21,68	22,30	18,06	18,52	18,52	18,67	19,49	19,90	20,43	20,69
Household head	0,68	0,68	0,68	0,69	0,70	0,67	0,60	0,60	0,16	0,18	0,17	0,20	0,21	0,22	0,25	0,28
Marital Status	0,64	0,65	0,63	0,63	0,62	0,57	0,51	0,50	0,39	0,42	0,41	0,41	0,41	0,40	0,37	0,35
Participation																
Agric. and Fishing	0,08	0,08	0,07	0,08	0,09	0,08	0,08	0,08	0,04	0,03	0,03	0,03	0,04	0,04	0,05	0,04
Mining	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,04	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Manufacture	0,22	0,20	0,20	0,19	0,18	0,18	0,18	0,14	0,17	0,14	0,13	0,11	0,11	0,10	0,11	0,08
Utilities	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Construction	0,15	0,15	0,15	0,13	0,14	0,14	0,16	0,15	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Retails	0,17	0,18	0,18	0,18	0,17	0,18	0,17	0,18	0,24	0,24	0,25	0,25	0,24	0,25	0,26	0,26
Transp. and Com.	0,11	0,11	0,11	0,13	0,12	0,12	0,12	0,12	0,03	0,03	0,03	0,03	0,03	0,04	0,04	0,04
Gov and bussiness serv.	0,05	0,06	0,07	0,08	0,08	0,08	0,08	0,09	0,06	0,08	0,08	0,09	0,10	0,08	0,08	0,09
Public administration	0,17	0,18	0,17	0,18	0,18	0,18	0,16	0,19	0,45	0,46	0,45	0,47	0,46	0,48	0,45	0,47
Force manager	0,06	0,06	0,05	0,06	0,06	0,05	0,04	0,02	0,05	0,05	0,05	0,06	0,06	0,07	0,04	0,02

Professionals	0,06	0,07	0,08	0,08	0,08	0,09	0,08	0,09	0,10	0,11	0,10	0,12	0,12	0,13	0,11	0,16
Technicians and associate prof.	0,06	0,07	0,08	0,08	0,09	0,09	0,07	0,09	0,07	0,09	0,11	0,11	0,09	0,11	0,12	0,14
Office workers	0,06	0,07	0,07	0,08	0,07	0,07	0,07	0,06	0,16	0,18	0,18	0,20	0,19	0,19	0,17	0,16
Service and sales workers	0,11	0,11	0,12	0,11	0,11	0,10	0,11	0,15	0,24	0,22	0,24	0,21	0,22	0,21	0,24	0,24
Skilledagric, forestry and fishery	0,03	0,03	0,03	0,03	0,04	0,04	0,04	0,03	0,01	0,00	0,01	0,01	0,01	0,01	0,01	0,00
Craft and related trades workers	0,28	0,24	0,24	0,23	0,22	0,24	0,25	0,22	0,09	0,06	0,05	0,04	0,05	0,05	0,05	0,03
Operators and assemblers	0,13	0,15	0,16	0,16	0,16	0,16	0,16	0,15	0,02	0,04	0,03	0,03	0,03	0,02	0,03	0,02
Unskilled workers	0,21	0,19	0,17	0,17	0,17	0,16	0,18	0,18	0,28	0,25	0,23	0,23	0,22	0,22	0,22	0,23
In labor market	19394	22537	18778	24611	27096	28672	32356	28210	8968	10225	8832	12272	13452	14795	16902	15902

Source: Estimation of the author using CASEN.

Table 2: Gender wage gap for monthly and hourly wage

	1992	1994	1996	1998	2000	2003	2006	2009	Δ 1992-2009
Montly	41,8	32,0	35,2	31,8	27,8	26,8	24,0	19,1	-22,8
Hourly	37,1	30,5	28,2	24,2	21,2	18,8	18,9	13,6	-23,6
SMMonthly	44,4	38,0	41,9	39,6	34,2	35,4	32,5	28,3	-16,1
SMHourly	40,3	37,1	35,6	32,4	27,7	27,8	27,6	23,0	-17,3
LMMonthly	22,7	17,6	21,1	17,6	20,1	20,0	17,4	19,3	-3,4
LMHourly	20,9	20,5	17,8	13,4	16,5	14,9	15,7	16,2	-4,7

Source: Estimation by the author using CASEN. The observations are weighted at population level.

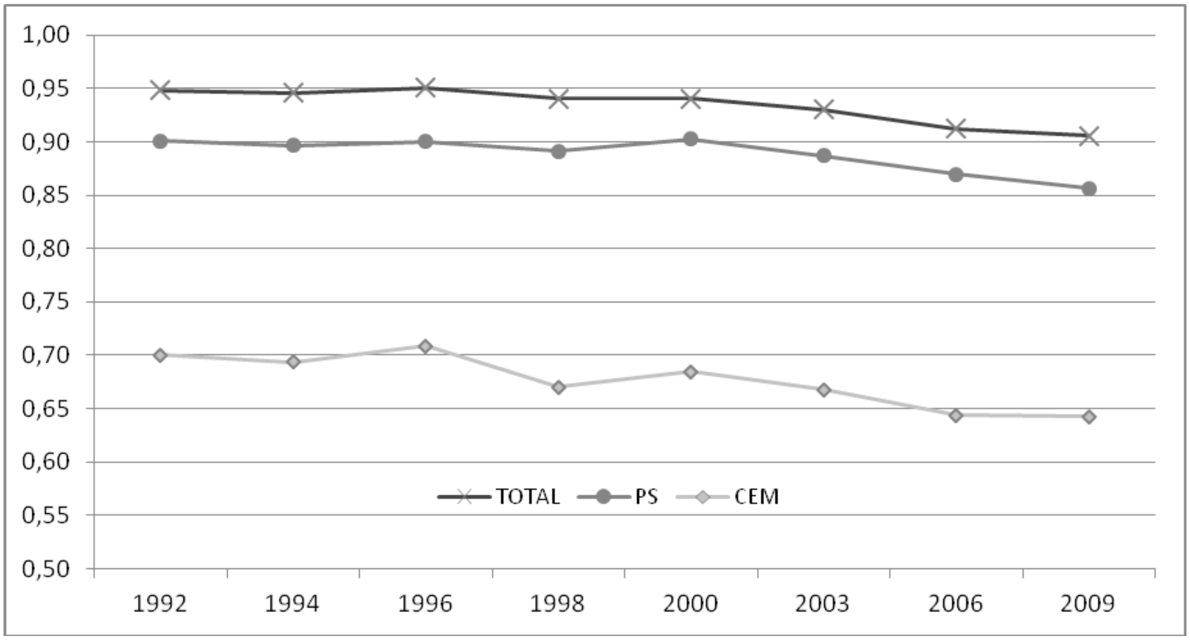
Table 3: Observations before and after PS trimming

	Before PS trimming		After PS trimming	
	Men	Women	Men	Women
1992	19,394	8,968	8,632	7,979
1994	22,537	10,225	10,274	9,230
1996	18,778	8,832	8,817	7,862
1998	24,611	12,272	12,010	10,988
2000	27,096	13,452	13,981	12,198
2003	28,672	14,795	15,648	13,587
2006	32,356	16,902	19,508	16,142

2009	28,210	15,902	17,224	15,205

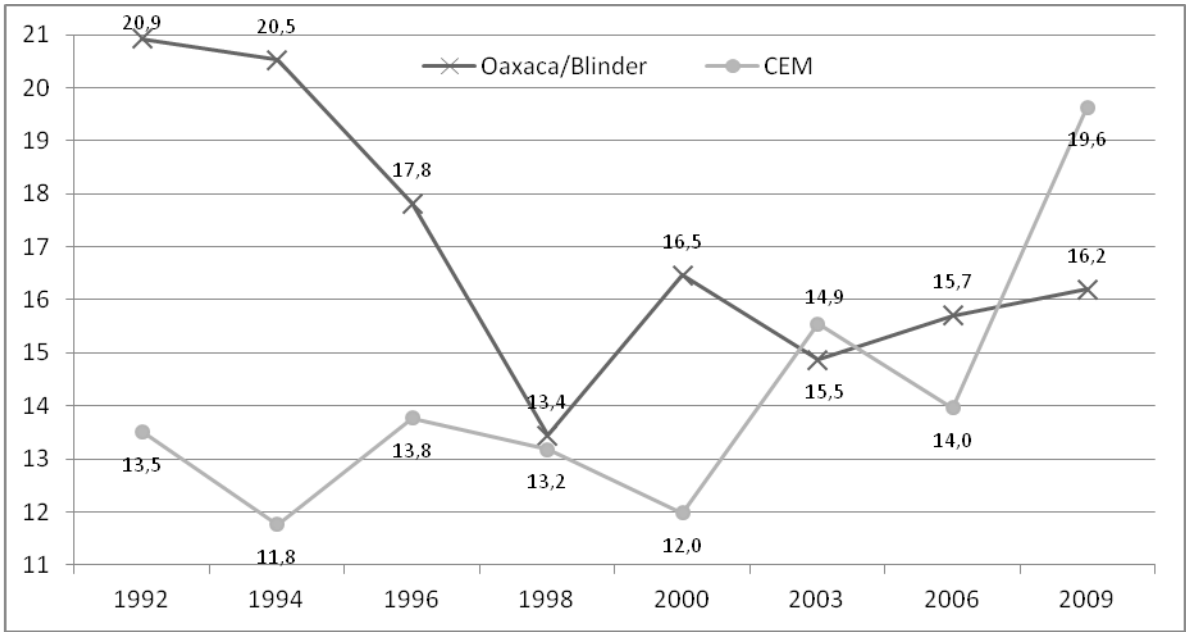
Source: Estimation by the author using CASEN.

Figure : L1 statistics for before and after PS trimming



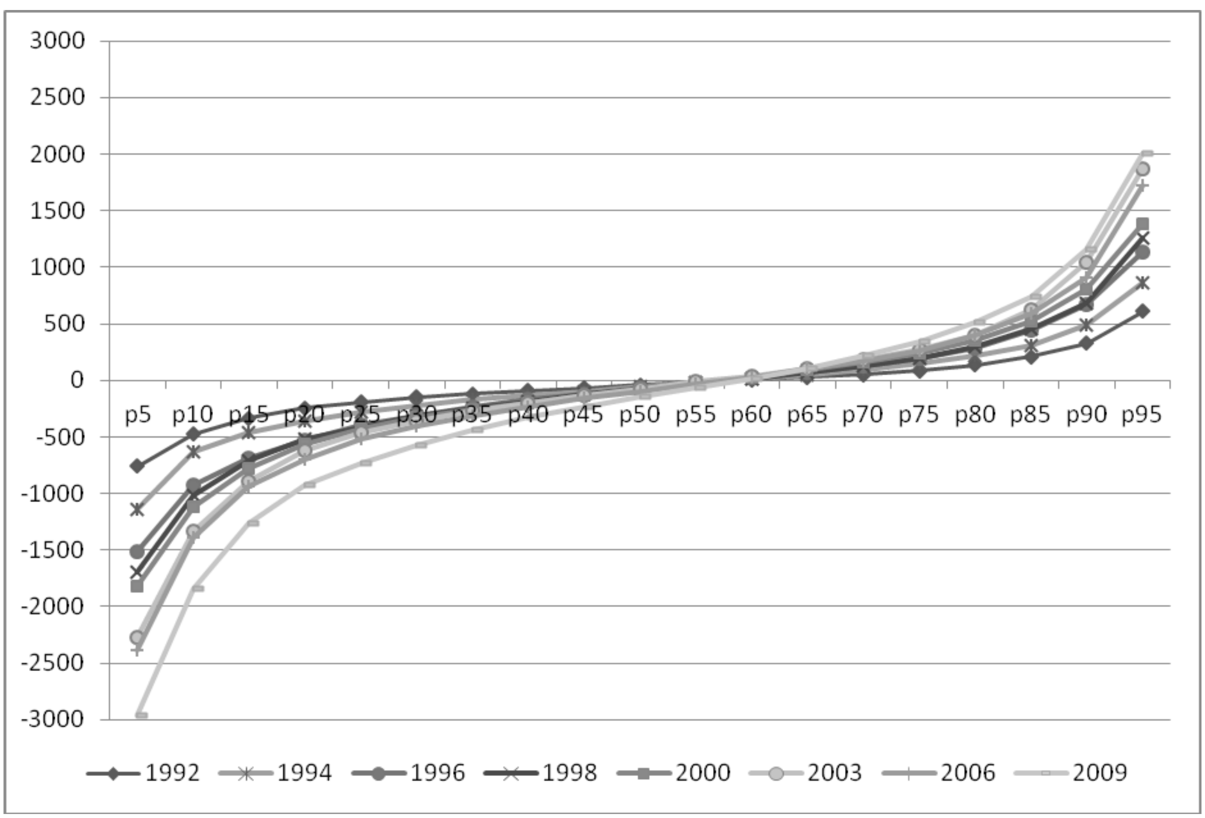
Source: Estimation by the author using CASEN.

Figure : Gender wage gap using Oaxaca-Blinder and CEM



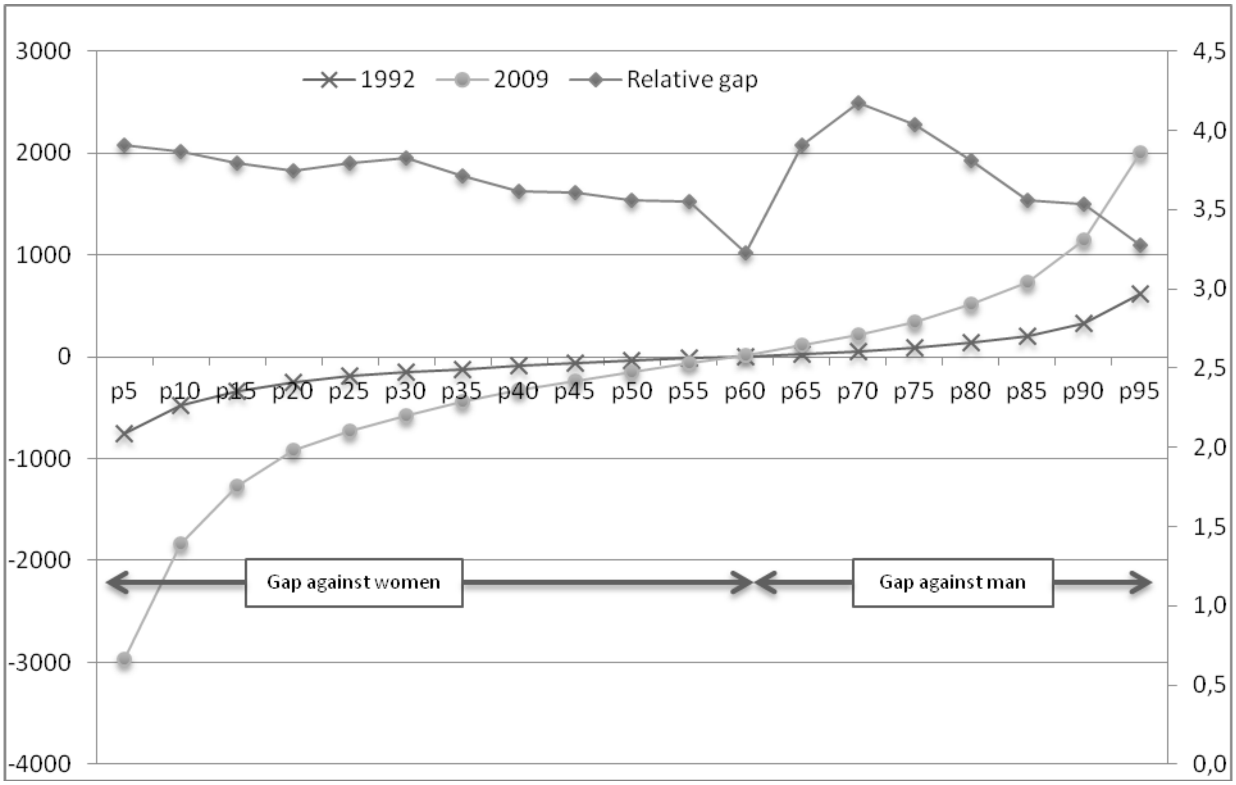
Source: Estimation by the author using CASEN.

Figure : Gender wage gap by percentile using CEM



Source: Estimation by the author using CASEN.

Figure 4: Relative gender wage gap



Source: Estimation by the author using CASEN.

Bibliography

Fuentes, J., Palma, A., & Montero, R. (2005). Discriminacion salarial por genero en Chile: una mirada global. *Estudios de Economia* , 133-157.

Montenegro, C. (2001). *Wage distribution in Chile: Does gender matter? a quantile regression approach.* World Bank, Poverty reduction and economic managment network.

Perticara, M., & Bueno, I. (2009). A new approach to gender wage gaps in Chile. *Cepal Review* , 131-147.

Frolich, M. (2007). Propensity score matching without conditional independence assumption with an application to the gender wage gap in the United Kingdom. *Econometrics Journal* , 359-407.

Ñopo, H. (2008). Matching as a tool to decompose wage gaps. *The review of economics and statistics* , 290-299.

Dolton, P., & Makepeace, G. (1987). Marital status, child rearing and earnings differentials in the graduate labour market. *Economic Journal* , 897-922.

Munro, A. (1988). *The measurement of racial and other forms of discrimination.* University of Stirling.

Ho, D., Imai, K., & Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* , 199-236.

Barsky, R., Bound, J., Charles, K., & Lupton, J. (2002). Accounting for the Black-white wealth gap: a nonparametric approach. *Journal of the American Statistical Association* , 663-673.

Black, D., Haviland, A., Sanders, S., & Taylor, L. (2004). *Gender wage disparities among the highly educated.* University of Syracuse, Center for policy research.

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). Coarsened Exact Matching in Stata. *Stata Journal , 9*, 524-546.

Abadie, A., & Imbens, G. (2007). *Bias corrected matching estimators for average treatment effects.* Working paper, Harvard University.

Paredes, R., & Riveros, L. (1994). Gender wage gaps in Chile. A long term view: 1958-1990. *Estudios de Economia , 21*, 211-230.

Gill, I. (1991). Is there sex discrimination in Chile? Evidence from the CASEN survey. In G. Psacharopoulos, & Z. Tzannatos, *Women's employment and pay in Latin America.* Washington, United States: The World Bank.

Iacus, S., Porro, G., & Stuart, E. (2011). Causal inference without balance checking: coarsened exact matching. *Political Analysis* .

Guo, S., & Fraser, M. (2010). *Propensity score analysis.* SAGE publications.

Bayard, B., Hellerstein, J., Neumark, D., & Troske, K. (2003). New evidence on sex segregation and sex differences in wages from matched employee-employer data. *Journal of Labor Economics* , 887-922.

Blau, F., & Kahn, L. (2003). Understanding International differences in the gender pay gap. *Journal of Labor Economics* , 106-144.