The impact of the occupations and economic activities on the gender wage gap using a counterfactual framework

Dusan Paredes
The impact of the occupations and economic activities on the gender wage gap using a counterfactual framework

Dusan Paredes
Faculty of Economics, Universidad Católica del Norte
Antofagasta, Chile.

Abstract

This paper presents a methodology to estimate the individual gender wage gap as the difference between wages of the women and their counterfactuals defined by Coarsened Exact Matching. If the women show a higher wage than comparable men, then it is called positive gap. Using eight surveys between 1992 and 2009 for Chile, a stable average of 44% of women show positive gap. This group is considered interesting from the policy perspective because they can provide lessons to decrease the negative discrimination observed on women. Additional analysis shows that the occupations such as Managers, Professionals and Technicians and Associated Professionals always increase the positive gap. Finally, the most successful economic activity to increase the positive gap is Finance and Insurance Activities.

Key words: Gender earnings gap, statistical discrimination, occupational sorting.

JEL classification: J01, J20, J31
1. Introduction

The gender wage gap is not only attributable to the discriminated women in the labor market (hereafter negative gap); it also applies for those men who receive a lower wage than women with similar characteristics (hereafter positive gap). Consequently, the gender gap works in both senses and with the same negative incidences on social cohesion, growth and efficiency (Knowles, Lorgelly, & Owen, 2002). However, standard methodologies such as Oaxaca-Blinder generate an average estimation of the wage gap, undoubtedly a valuable metric, but which necessarily reinforces a picture where male workers earn a higher wage than females as a general rule (Ñopo, 2008). With the current availability of new econometric techniques grounded on matching comparison, the average scenario could be complemented with the analysis of the complete distribution, identifying the women with higher or lower wage than comparable men and analyzing the reasons behind this differential. Regrettably, this type of analysis is still absent in the literature of labor economics and in particular for the case of Chile, and more alarming, it is missed in the policy design oriented towards gender equity (Arulampalam, Booth, & Bryan, 2007) (Bonjour & Michael, 2001).

This paper contributes to cover this unexplored feature of the gender gap moving the analysis from the average perspective toward one focus on the gap distribution of Chile. The methodology discovers two groups: men who earn a higher wage than comparable women, and, more interestingly for this paper, women with higher wage than comparable male workers. This second gap, labeled as positive gap, is a interesting group for policy makers because its characterization could provide information to orient the design of effective policy to reduce the wage gap between genders.
The advantage of this approach is not only the identification of both groups, namely positive and negative gap; but rather its usefulness consist of the wide set of analysis to do after the classification. For example, if the researcher specifies the gap as the wage of a woman minus the wage of a comparable man, then the gender gap ranges from negative (negative discrimination for women) to positive differentials (positive discrimination for women). With this information at hand, it is possible to evaluate how the sources highlighted by the literature, such as occupations (Petersen & Morgan, 1995) and economic activity (Liu & Zhang, 2000); affect the transition from a negative gap toward a positive gap.

The individual gap constructed as a subtraction between wages of comparable workers also invites for additional analysis. For example, it ranges from positive to negative values. This situation implies that the explanatory power of these sources could vary along the gap. For example, what type of occupation helps to explain the transition from negative to positive discrimination? Or how do economic activities help to increase the positive discrimination or, at least, decrease the negative one? This scenario fits perfectly with the advantages of using quintile regression. Note that the difference between the quantile approach here proposed and the standard quintile approach, it is to estimate the gender gap along the wage distribution (Garcia, Hernandez, & Lopez-Nicolas (2001), Machado & Mata (2005)). The last case estimates the gap along the wage distribution and it allows discover if the gender gap is wider for high, medium or low wages. Consequently, this paper uses the quantile regression to analyze the gap distribution instead of level wage. Here the idea is how a variable helps to reduce the negative gap or to increase the positive.
The main challenge is to find a method to estimate the distribution of the gender gap, or equivalently, to estimate the gender gap at individual level. The individual gender wage gap is constructed through a combination of two techniques: Coarsened Exact Matching (CEM) and Propensity Score Matching (PSM). CEM finds groups of comparable women for men and its comparability is defined according to an available set of human capital variables (Iacus, Porro, & Stuart, 2011). Each male worker (hereafter known as control group) owns a weight that represents its comparability with the female group (hereafter known as treated group). However, a richer description would be possible if both groups of comparable workers can be reduce to one-to-one comparability where the gender wage gap would be estimated by pairs of observations. Regrettably, CEM only warranties a comparability at group level. The methodology is then pushed forward using a Propensity Score Matching (PSM) that matches workers who already were matched using CEM (Ñopo, 2008). PSM uses the estimated probability of being women to match one male worker with a particular female. The combination of both approaches let us to identify a one-to-one group of male and female workers and then an individual measurement of gender wage gap is estimated.

The second stage is focused on the empirical distribution of the individual gender gap and its explanatory variables. The literature suggests, at least, two potential sources to explain this gap. The first one is the occupational segmentation where some occupations with higher wages contract workers with profiles concentrated in a particular gender (Rosenfeld & Kalleberg (1990) and Kilbourn (1994)). The economic activities work in a similar way as occupational segmentation. Economic activities such as mining or construction are strongly dominated by male workers because they own valuable
characteristic for this sector (Fields & Wolff, 1995). In order to do not lose the rich information provided by the gap distribution, these factors are evaluated using quintile regression. This strategy let us to evaluate how, let say the type of education, varies its marginal impact to explain the case of women with low, medium and higher gap. This exercise is repeated for both sources, but with special emphasis on how these factors help to move from a negative toward a positive gap.

The methodology is applied to the data of Chile, but where the applicability can be extended to other contexts. The results confirm the main hypothesis of the paper: the average gender wage gap hides an interesting portion of the gap operating in both directions. Between 1992 and 2009, the proportion of women with lower wage than comparable men is around of 56%. Interestingly, 44% of women reflect the opposite pattern. These results open the discussion about what factors should be related with this positive gap. Additionally, 20 percent of men earn between one and two times the average wage of women, but only the 5 percent of women earn the same ratio over male wages. Moreover, an extreme negative gap of women earn between 2 and 3 times lower wages than comparable men. This same ratio simply does not exist for positive gap.

With respect to the successful occupations, they are Managers, Professionals and Technicians and associated professionals. All of them show a positive and significant impact on positive gap. In other words, any women with some of these occupations presents higher probabilities of belong to the positive gap group. Among these occupations, Professionals show increasing impacts along the complete distribution. This implies that Professionals help to reduce the negative gap, but it also increases the positive gap. Managers only present major differentials for the 44% of women who
already enjoy a positive gap. This occupation does not show statistical significance for reducing the negative gap. Additionally, *Technicians and associated professionals* show a lower impact than previous occupations and its magnitude does not change along the distribution. Finally, the most successful economic activity is *Finance and Insurance Activities* with positive impacts on the total distribution. Other sectors such as *Mining, Construction, Electricity, gas and water supply* and *Transport and communications* present significant and positive impact only for more of the 15% of total women. *Agriculture, forestry and fishing, Manufacturing* and *Wholesale and retail trade* do not show any significant impact.

The Section 2 details the CEM and PSM methodology and Section 3 provides the description of data. Section 4 develops quintile regression approach. Section 5 contains the conclusions and policy remarks.

2. Why Chile?

The economic context of Chile is an interesting case study to analyze the gender wage gap. Chile is one of the most successful economies in Latin America and its performance is continuously cited as an example for the region. The Chile’s average per capita GDP growth of 4.1% during 1991-2005 and it was significantly higher than average world growth rate during the same period. The period 1985-1997 called the *gold age*, obtained the annual average GDP growth rate of 7.1% and per capita income was duplicated (De Gregorio 2005). While this success in economic growth has been widely discussed, the consequences on gender inequality at national level are still unexplored for the same period.

---

1 (Paredes & Riveros, 1994) provide a long term view of the gender wage gap, but it is only estimated for the Metropolitan Region of Santiago. Additionally, they just over between 1958 and 1990.
Some exceptions are Fuentes, Palma, & Montero (2005) who suggest a reduction of the gap between 1992 and 2003, while Montenegro (2001) indicates a persistent and stable gender gap between 1990 and 1998. Perticara & Bueno (2009) examines the gender gap between 2002 and 2006 and they detect that this gap has widened during the last years. Beyond the interesting information provided by these papers, they have no covered the length between 1990 and 2010 and they only provide parital perspective of the complete serie. This lack of discussion does not permit to analyze how the gender inequality has reacted with the recognized economic performance of Chile. This reason motivates to study the Chilean case and to construct a series of gender gap for the longest available period.

A second reason is related with the homogeneity of empirical approaches to understand the case. Fuentes, Palma, & Montero (2005), Montenegro (2001), Paredes & Riveros (1994) and Perticara & Bueno (2009) ground their researchers on Oaxaca-Blinder decomposition (OB). This technique estimates an useful average gender wage gap, but it also loses important information. First, OB does not make a warranaty about the comparability among men and women. Second, OB imposes a linear functinoal form for the wage equation and third; OB does not gives information about the individual gap. This scenario domnated by just one technique is perfect to introduce new techniques. Econometric literature has developed interesting methodologies derived from matching comparison (Frolich, 2007) (Ñopo, 2008). These new methodologies could also help to understand the Chilean case and new information could be added to evaluate how the wage gap has evoluted during the last twenty years.
3. **Methodology**

   **a. Coarsened Exact Matching and Propensity Score Matching.**

   The econometric strategy of this paper is divided in two sections. The first one is the one-to-one comparability of workers. The second one is focused on quantile regression and the role of occupations and economic activity. The first part is pursued with a mixed process grounded on CEM and PSM (Paredes D., 2012). Consider \( n \) workers sampled from a population of \( N \) observations, such that \( n < N \). For the observation \( i \) an indicator \( T_i = 1 \) is referred when the worker is a woman and 0 otherwise. The variable outcome is labeled as \( Y_i(1) \) when \( T_i = 1 \) and \( Y_i(0) \) otherwise. The observed outcome is \( Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i) \). Regrettably, the counterfactual outcome \( Y_i(1) \) when \( T_i = 0 \) (or \( Y_i(0) \) when \( T_i = 1 \)) is unobservable. This problem impedes to build the gap directly from \( Y_i(1) \) and \( Y_i(0) \). However, the counterfactual could be enforced if two workers with similar characteristics are identified in the sample. Let \( X = \{X_1, X_2, \ldots, X_k\} \) be a set of attributes, proxies of human capital for this case, observed for the sample of size \( n \). The CEM overcomes the problem above described because if the control (male workers) and treated (female groups) group are similar in a sense defined by human capital proxies, then they are counterfactuals. Such as Iacus, Porro, & Stuart (2011) suggest, CEM carries out this process in a simple four-stage method:

   b. Begin with the covariates \( X \) and make a copy, which is denoted as \( X^* \).

   c. Coarsen \( X^* \) according to user-defined cut points or CEM’s automatic binning algorithm.

   d. Create one stratum per unique observation of \( X^* \), and place each observation in a stratum.
e. Assign these strata to the original data, $X$, and drop any observation whose stratum does not contain at least one treated and one control unit.

These steps generate a smaller sample of observations $n^*$ such that $n^* = n_0 + n_1$ where $n_0$ and $n_1$ are comparable groups of male and female workers. Additionally, several control observations can belong to the same stratum, but some of them present a larger weight in the comparison process if they share more characteristics with the treated observations. Unfortunately, CEM only conforms two comparable groups, but the individual gap between $n_0$ and $n_1$ cannot be computed. A second matching on $n_0$ and $n_1$ is carried out using a Propensity Score Matching (Ñopo, 2008). This procedure summarizes the $X$ vector in a one-dimensional variable $P(X)$, which is the predicted probability of being a woman. This $P(X)$ is constructed with a Logit function and each treated (woman) is matched with a control (man) with the most similar $P(X)$. This process finally warranties a set of one-to-one comparable workers and an individual gender wage gap is easily estimated as the difference in hourly wages. Once the comparability one-to-one is warranted, then the quintile regression is carried out. The complete codes are available in Stata® and they can be requested to the author. In particular, the code `psmatch2` and `cem` are freely available in Stata for PSM and CEM, respectively.

b. Quintile regression

The explanatory power of each dimension is tested along the gap distribution using quantile regression. The quantile regression is part of the Extreme Estimators and

---

2 The method considers 1 for women and 0 for men.

3 This naive difference is possible only after matching by human capital variables and with the assumption that unobservable are independent of the gender variable. Otherwise, this measurement is highly biased by other no considered factors.
their estimates are obtained from the minimization of an objective function determined by a portion $q$ of the total distribution of the dependent variable. This minimization is obtained using a linear system such as the Simplex Method. A detailed description of the method is found in Cameron and Trivedi (2010) and here the main discussion is focused on the empirical exercise.

The theoretical specification of the $q$ th conditional expectation of the gap $g$ given a set of regressors $x$ is specified as $Q_q(g_i, x_i) = x'_i \beta_q$. Note that $\beta_q$ is obtained according to the $q$ quintiles required by the researcher. For this exercise, the quantile regression let to identify, for example, the effect of education on the observed gap, but this effect is differentiated for low and high values of gap. This strategy is worthy given that occupation segmentation or sectorial mix can be presented in some portions of the total distribution. Clearly, the simultaneous consideration of 8 years and quantiles become the presentation of results a fuzzy exercise. For this reason the 8 years are collapsed in a unique database and the quantile regressions are estimated on the collapsed data.

c. Data

The data are obtained from the Chilean National Socio-Economic Characterization Survey CASEN (Encuesta de Caracterizacion Socio Economica Nacional). This survey contains a module of labor market with variables such as the wage, worked hours, education, age, marital status, head household position, occupations, economic activity and geographical region$^4$. While CASEN is available from mid 80’s, the lack of consistency in some variables impedes to understand the data set as dynamic.

$^4$ Chile is divided in fifteen regions from 2007, but they were thirteen before 2007. In order to be consistent with earlier CASEN’s, the analysis is carried out with thirteen regions.
series of surveys. This lack of consistency has been improved with the new versions of CASEN, but the link among surveys is still a complex process. This paper makes this consistency and CASEN’s for 1992, 1994, 1996, 1998, 2000, 2003, 2006 and 2009 were appended. Each survey was carefully checked in order to make a consistency among the different waves. This process does not provide a panel data because the observations are different by year, but at least the dynamic evolution of the gender gap is incorporated.

The hourly wage is computed at real pesos of December 2008\(^5\). Additional variables are years of education, potential experience\(^6\) and its quadratic version, marital status and condition of head of household. The sample selected does not consider rural workers given that its payment structure is conditioned by informal conditions and CASEN does not provide the needed information to control this influence (Reardon, 1997). The temporary workers are also deleted. The selected workers are between 14 and 64 years and they work more than 30 hours per week. Additionally, workers of military forces are also deleted. A description of the human capital proxies is displayed in the Table 1.\(^7\)

<table>
<thead>
<tr>
<th>Table 1: Averages of Human Capital proxies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hourly wage</td>
</tr>
<tr>
<td>Years of Educ.</td>
</tr>
<tr>
<td>Head HH</td>
</tr>
<tr>
<td>Married</td>
</tr>
<tr>
<td>Hourly wage</td>
</tr>
<tr>
<td>Years of Educ.</td>
</tr>
</tbody>
</table>

\(^5\) 1 dollar = 500 pesos aprox.

\(^6\) Even when the proxy for experience is not the best one recommended by the literature, CASEN does not provide a better variable. The potential experience is estimated as age – years of education – 6.

\(^7\) A complete description of occupations and economic activities is provided in the appendix.
CASEN, as most surveys, does not control a priori by the existence of outliers. In order to avoid the effect of outliers, a Mincer equation was estimated for each year using as dependent variable the hourly wage and the set of independent variables such that years of education, experience and its quadratic version, occupations, economic sectors and regional fixed effects that indicate where the worker is located. This fixed effect helps to control by any idiosyncratic effect attributable to the local market where this worker is competing. After the eight regressions (one for each CASEN), all the observations with standard error or student error higher than 3.5 were deleted. A DfBeta test is also applied, but this process does not delete any additional observation. The detected outliers are described on Table 2 and they do not go beyond the 1% per year.

Table 2: Outliers and total sample

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CASEN</td>
<td>28731</td>
<td>32793</td>
<td>27857</td>
<td>37118</td>
<td>40881</td>
<td>43817</td>
<td>49707</td>
<td>44584</td>
</tr>
<tr>
<td>Outliers</td>
<td>338</td>
<td>36</td>
<td>295</td>
<td>274</td>
<td>391</td>
<td>393</td>
<td>495</td>
<td>512</td>
</tr>
<tr>
<td>Observations</td>
<td>28393</td>
<td>32757</td>
<td>27562</td>
<td>36844</td>
<td>40490</td>
<td>43424</td>
<td>49212</td>
<td>44072</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Source: Author's calculation using CASEN 1992-2009. Sample only includes wage positive, more than 30 hours worked per week, urban area, military excluded and no missing in the independent variables used to run regressions above described

After this preparation of the date base, the last row of Table 2 shows the total number of available observations used for building Table 1. These observations are the base line for next section.

d. Results
a. Coarsened Exact and Propensity Score Matching

The results section is divided into two parts: the construction of individual gap and the quantile regression. The Table 1 displays the number of male and female workers who are receiving a positive hourly wage between 1992 and 2009. The first step consists of applying the CEM to warrant that only comparable workers be analyzed. A natural assumption behind of this selection is that male and female workers are different a priori in human capital proxies and for this reason the researcher needs to assure the comparability. To evaluate the necessity of this assumption, the statistic $L_1$ is proposed. This statistics compares the empirical histogram of multiple variables between two different groups. If the statistics $L_1$ goes up 1, then both histograms are different, but if it goes to zero, then both groups are similar.

**Figure 1: L1 statistics before and after CEM.**

![L1 statistics](image)

Source: Author's calculation using CASEN 1992-2009, The independent variables are years of education, potential experience and its quadratic version, marital status, status of head of household, and its quadratic version plus the region where the worker is located.
The Figure 1 supports the relevance of CEM. The male and female workers in the original sample are highly different for the complete period of years. By 1992, 1994 and 1996, the $L_1$ is around 0.75 and its minimum is around 0.60 by 2009. That implies that both genders are heterogeneous in the considered variables, but its heterogeneity decreases across the years. The reduction is expected because women have increased the participation in the labor market and also incremented their level of education during the considered period (Perticara & Bueno, 2009). After CEM, they are more comparable for the complete period. Each year shows a considerable reduction of the $L_1$ statistics supporting the idea that workers are more comparable after CEM. CEM discards observations such that: 1) Female workers who own characteristics not presented in male workers, or 2) Male workers who own characteristics not presented by female workers.

The summary of these groups are displayed by Table 3:

Table 3: Controls and treated observations generated by CEM.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Male</td>
<td>19416</td>
<td>22532</td>
<td>18736</td>
<td>24579</td>
<td>27050</td>
<td>28634</td>
<td>32319</td>
<td>28176</td>
</tr>
<tr>
<td>Total Female</td>
<td>8977</td>
<td>10225</td>
<td>8826</td>
<td>12265</td>
<td>13440</td>
<td>14790</td>
<td>16893</td>
<td>15896</td>
</tr>
<tr>
<td>Male Matched</td>
<td>11981</td>
<td>14995</td>
<td>11918</td>
<td>15961</td>
<td>19871</td>
<td>22205</td>
<td>26988</td>
<td>24668</td>
</tr>
<tr>
<td>Female Matched</td>
<td>7979</td>
<td>9198</td>
<td>7833</td>
<td>11107</td>
<td>12381</td>
<td>13650</td>
<td>15949</td>
<td>14979</td>
</tr>
<tr>
<td>%Male</td>
<td>0.62</td>
<td>0.67</td>
<td>0.64</td>
<td>0.65</td>
<td>0.73</td>
<td>0.78</td>
<td>0.84</td>
<td>0.88</td>
</tr>
<tr>
<td>%Female</td>
<td>0.89</td>
<td>0.9</td>
<td>0.89</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
<td>0.94</td>
<td>0.94</td>
</tr>
</tbody>
</table>


The first and second rows of Table 3 show the decomposition by gender of the third row in Table 2. They are the total of observations before CEM. The third and fourth row of Table 3 shows the number of matched observations after CEM. For example, 62 percent of total male workers share similar characteristics with women, but 89 percent of females are statistically similar to male workers. The heterogeneity between row one and two is summarized by $L_1$ labeled as Ex-Ante in Figure 1, while the heterogeneity between
rows three and four is labeled as Ex-Post. The trend of $L_1$ is supported by the trend of the number of matched observations: the proportion of comparable workers increases along the period.

As previously discussed, CEM warranties two groups of comparable workers, but they do not permit to estimate a one-to-one gap. This problem is solved using Propensity Matching Estimator. Using the *counterfactuals* identified by CEM, a logit model is estimated where the dependent variable is 1 for women and 0 for men for each year. The predicted probability, called propensity score, is used to match a woman with the closest propensity score in the control group. The Table 4 helps to understand the methodology.

### Table 4: Propensity Score Matching

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>11981</td>
<td>14995</td>
<td>11918</td>
<td>15961</td>
<td>19871</td>
<td>22205</td>
<td>26988</td>
<td>24668</td>
</tr>
<tr>
<td>Treated</td>
<td>7979</td>
<td>9198</td>
<td>7833</td>
<td>11107</td>
<td>12381</td>
<td>13650</td>
<td>15949</td>
<td>14979</td>
</tr>
<tr>
<td>Av. CEM</td>
<td>-191.5</td>
<td>-278.69</td>
<td>-245.89</td>
<td>-215.72</td>
<td>-150.62</td>
<td>-205.71</td>
<td>-176.97</td>
<td>-222.12</td>
</tr>
<tr>
<td>Av. PSM</td>
<td>-157.8</td>
<td>-181.61</td>
<td>-176.68</td>
<td>-188.77</td>
<td>-142.32</td>
<td>-183.3</td>
<td>-109.63</td>
<td>-181.11</td>
</tr>
<tr>
<td>Gap CEM</td>
<td>-0.19</td>
<td>-0.24</td>
<td>-0.21</td>
<td>-0.17</td>
<td>-0.13</td>
<td>-0.15</td>
<td>-0.13</td>
<td>-0.15</td>
</tr>
<tr>
<td>Gap PSM</td>
<td>-0.16</td>
<td>-0.16</td>
<td>-0.15</td>
<td>-0.15</td>
<td>-0.12</td>
<td>-0.14</td>
<td>-0.08</td>
<td>-0.12</td>
</tr>
</tbody>
</table>

Source: Author’s calculation using CASEN.

The first and second rows show the number of control and treated observations identified by CEM (see Male and Female matched in Table 3). The third row shows the average difference between wages of these comparable workers, which is negative for the complete period indicating a gender gap against women. After CEM, each woman is one-to-one matched with a man using the propensity score. The new average gap is displayed in the fourth row. There are some small differences between average gaps of CEM and PSM because PSM uses only a subset of the sample generated by CEM. However, this number is used only to show the idea of both procedures, but the emphasis is focused on

---

8 The complete set of estimations is omitted by space constraints, but they can be requested to the author.
the complete distributions of counterfactuals identified by PSM and CEM. In order to compare the advantages derived from comparison of comparable workers versus a raw wage comparison, the average gaps for CEM and PSM are compared with the gap reported on table 1. The gap for year $j$ is estimated as:

$$gap_j = \frac{(\bar{w}_{women} - \bar{w}_{men})}{\bar{w}_{women}}$$

where the average wages are estimated for the raw average wage (Table 1), CEM and PSM (Table 4). The temporal tendency is displayed in Figure 2.

Figure 2: Average gender gap

Source: Author’s calculation using CASEN

The first result is the evident reduction of the raw gender gap from 1992 until 2009, showing a reduction of almost 25%. However, this reduction is biased because it uses workers with heterogeneous characteristics. When the homogeneity is imposed by CEM or PSM, the picture is different. The gap is not as high as the raw average reports. Particularly, the gap is lower than expected when similar workers are considered. Second,
and maybe more interesting, both gaps, CEM and PSM, show an incredible persistency and its evolution has been almost constant during the period. According to the PSM, the gap has been reduced just by the 4% between 1992 and 2009. This result shows the advantages of matching comparison compared to averages comparison. Moreover, the comparable workers provide a ignored dimension about the gender gap in Chile: the gender gap has been extremely stable during the last 20 years.

Next step is the analysis of the distribution of comparable workers provided by PSM. Once control and treated observations are matched, the individual gap is simply estimated as the difference of the hourly wage between a female (treated) and its matched male (control). For each year, an empirical distribution of gender gap is estimated. This distribution ranges from negative gaps for women with lower wages than male to positive gaps in favor of women. The interpretation of the eight years distributions in a unique graph is a tedious exercise for the reader. Consequently, a contour plot of the kernels for each year is displayed in Figure 3, where the variable considered is the ratio between the wage difference of wages and the average wage of women.
Figure 3: Gender gap distribution for comparable workers between 1992 and 2009

Source: Author’s calculation using CASEN

The horizontal axis represents the percentile of the gap distribution, but it is slightly modified for a better comprehension. Instead of using gap in Chilean pesos, this gap is normalized using the average wage of women. The left side of the horizontal axis represents those women with a wage almost 3 times lower than comparable men (negative gap). In contrast, the right side of the horizontal axis contains those women who earn significantly higher wages than men. Finally, the periods are represented by the vertical axis. The color of surface represents the ratio between the gap for a particular percentile and the average wage of women.

The gap equals to zero is a critical value because it represents the limit between both types of gap. The portion of women with negative gap occupies the largest portion of the surface. This portion is consistent during the complete period, bounding 56 percent
of comparable workers. It is noteworthy how the extreme stability of the gap distribution differs from the decreasing raw gap observed in Figure 2. The average picture indicates a continuous reduction of the gender gap between 1992 and 2009, however the distributional approach indicates that the same proportion of women earn a lower wage than comparable men for the complete period. Undoubtedly, the distribution approach contributes with a new perspective for analysis the case, especially from the policy perspective. Simultaneously, the positive gap is also observed for the complete period and it covers 44% of comparable workers.

The tails of the distribution is another feature hidden by the average perspective. For example, almost 13% (blue and red areas) of women show a negative gap almost 2 times the female average wage, but this same ratio of 1 exist only for the 6% women with positive gap. Note that this implies that 13% of women earn between 200% and 300% less than their counterfactuals, a number far away from the raw gap observed in Figure 2. A similar ratio for women with positive gap does not exist. These features seem to indicate than the average gender gap detected for Chile is a description of the complete distribution, but important features of the gender wage gap cannot be observed. In this sense, the distributional approach seems to open a new room for alternative discussions related with the equity and wages.

b. Distribution of gender gap with a quintile regression approach

Once the empirical distribution of gender gap is estimated, the complete distribution of the gender gap is the new dependent variable. The distribution is divided into two groups: those women who earn a higher wage than comparable workers and those with a lower wage. Each one of these groups owns interesting information for
policy makers, especially for understanding the sources that affect the positive gap. Obviously, each couple of workers, namely a woman and its counterfactual, presents a different gap. As Figure 1 shows, between the percentile 15 and 85 (purple and green area) there is a portion of gap that is not greater than 1 times the average women wage. This fact implies an overrepresentation of low individual gaps. Considering this scenario, the econometric technique must be chosen to capture the features of the complete distribution and the attraction of low gaps must be controlled. A summarized picture of the gap for the eight years is displayed in Figure 4.

**Figure 4: Empirical distribution of gender gap**

The total distribution of the gap indicates that almost a 57% of the women show a negative gap against comparable men. In the quantile regression the position of the observation is fundamental for interpreting the output. Each graph contains a vertical line on the percentile 57 and a horizontal line indicating the coefficient equal to 0. To the left of the vertical line are the women with negative discrimination and to the right are the women with positive gap. The horizontal line indicates if the variable pushes the gap toward a positive or negative magnitude.

**Source:** Author’s calculation using CASEN
The first source is labeled as occupation segmentation. This analysis indicates how the gender gap change given that the women participate in a specific occupation. The quantile regression captures 9 occupations with 8 dummy variables and the category *Elementary occupations* used as benchmark plus controls as education, experience, and marital status among other variables.\(^9\) The eight graphs are presented using the coefficient in the vertical axis and the percentile on the horizontal axis.

The results for occupations are showed on Figure 5. The first group, namely *Managers*, always show a positive impact on the individual gap: this occupation always helps the women to move toward a positive gap. However, its impact is much higher for those women with extreme positive gap. From the gap equals to zero, this occupation shows an exponential positive effect. This occupation is highly clustered in the public sector such as legislator, a sector that imposes policies of equity between genders. Given that woman own, in average, a higher education (see Table 2), a scenario with a hard regulation on gender parity generates a positive significant impact on gender gap. Note how this occupation *always* helps to increment the positive gap, a different situation than the variables previously discussed (education or experience).

The second group with a permanent positive effect is the *Professionals*, but with a lower impact relatively to *Managers*. The impact of this occupation is positive and increasing along the percentiles with a clear linear structure instead of exponential rate. This feature implies that *Professionals* always help to increase the positive gap and it does not depend of the position inside the distribution. For example, *Managers* has an effect over 1000 pesos only for the 30\% of the sample, while *Professionals* have the same

\(^9\) A complete description is detailed on the appendix.
effect for the 60% of the sample. From this perspective, Professionals seems to be a more fair and equal process to reduce the negative discrimination gap that Managers.

Technicians also help to reduce the negative gap, but its impact is much lower than Managers and Professionals. The impact is under 1000 pesos affect almost the 90 percent of the sample. Curiously, the impact of Technicians is extremely constant along the distribution with an exception on the right tail. This fact indicates that Technicians is not an occupation destined to increase the positive wage gap and its role is secondary in the equity process.

The rest of occupations: Clerical support workers, Service and sales workers, Skilled agricultural, forestry and fishery workers, Craft and related workers and Plant and machine operators, and assemblers present a very small impact on the gap distribution. They do not show impact over 200 pesos for the complete sample, a small marginal effect compared with the three previous occupations. These results provide and interesting and unknown pattern of the occupation segmentation in the Chilean case. Only three occupations seem to help women to obtain a positive gap: Managers, Professionals and Technicians. The effectiveness of these occupations to increase the positive gap is measured in two dimensions: percentage of sample with a positive impact and the magnitude of the impact. In terms of percentage of the sample the three occupations generate a positive impact for the 100 percent of the sample.
The second source described refers to the role of economic activity on the individual gender gap and the results are in Figure 6. The survey indicates 9 economic sectors, and then the quantile regression includes 8 dummy variables. The Administrative and support activities are fixed as benchmark. The choice of the sector with the main contribution is more complicated than occupations. If success is defined as those economic activities with a positive impact along the complete distribution, then Finance and insurance services is the most successful economic activity. This activity reports a significant increment from the percentile 57 where its contribution to move

---

See appendix for details.
toward a positive gap. However, note how the economic activities always report a lower impact than occupations.

Figure 6: Quantile regression for sectorial mix.

There are other activities with a positive impact for almost the complete distribution. For example, *Mining* owns a positive impact from the percentile 10. The positive impact is almost constant until the gap 0 (percentile 57). The impact rapidly grows from this percentile and for the right tail the marginal change is much higher than any other sector (around 1500 pesos). This pattern exists because the participation of *Professionals* and *Technicians* in *Mining* has grown rapidly from 1998. The Figure 7 shows this pattern where a 12% the *Professionals* in mining sector were women by 1992, but this group represents a 34% by 2009. A similar tendency is appreciated for
Technicians who increased from 2% until 19% by 2009. Such as it was discussed above, these two occupations present a positive impact on gender gap. This fact plus the increasing participation of these occupations in Mining help to understand why this sector is generating the estimated impact.

Figure 5: Participation of Females Professionals and Technicians in Mining sector

Source: Author’s calculation using CASEN

Construction also presents positive impacts from the 20 percentile of the gap distribution. The impact is almost constant around 300 pesos until the gap equals to zero and its marginal variation becomes significant for the right tail. The impact for this tail is around 800, lower than the 1300 pesos for Mining. Electricity, gas and water supply follows a similar pattern with a positive impact above the 20 percent, but its impact is always bounded around 500 pesos.

Transport and communications, similarly to Mining, presents a positive impact from the percentile 10. The variance of this impact is very low (see gray bounds) and its maximum impact is around the last 40% of the sample. In other words, Transport is an important sector for those women who really earn a positive gap, but its help to move
from a negative toward a positive is moderated. The rest of economic sectors do not show a significant impact on the gap.

Summing up, Finance is a sector that always helps to move from a negative toward a positive gap. However, its impact (in pesos) is lower than other sectors. Mining and Transports and communications helps from the percentile 10, but its main contribution is related with the right tail of gender gap. Both sectors seem to explain why the women already earn a significant higher wage than men, but they do not much to make the transition. Construction and Electricity, gas and water supply are also relevant from the 20 percentile, but its monetary impact is moderated in comparison with the other sectors.

e. Conclusions

The empirical distribution of the gender gap is generally missed in the literature of labor economics. This paper contributes, with a simple methodology, to build the empirical distribution using econometric techniques derived from matching comparison literature. The results support the main hypothesis of this work: only a portion of men earns higher wages than comparable women. Between 1992 and 2009, around of 56 percent of women in Chile earn a lower wage than their counterfactuals. However, a group of extreme negative discrimination against women is observed between 1992 and 2006 where almost a 15% of women earn less than 3 times the observed wage in comparable men. The opposite group does not exist for the case of women with positive gap.

After matching comparison, interesting insights are derived from the quantile regression approach. The most successful occupations to make the transition from a
negative toward a positive gap are Managers, Professionals and Technicians. Inside this group, Professionals is the most effective along the complete distribution, while that Managers shows a high impact only for those women with positive gap. Technicians present a much lower impact than the other two occupations, but are still positive.

The economic sector presents a heterogeneous scenario. Finance is the unique sector that helps to the transition from negative to positive in the complete gap distribution. Transport and communications, Electricity, gas and water supply and Construction are also significant sectors, but they play a role only for a portion of the total sample (between 80 and 90 percent). The highest monetary impact is observed in Mining sector where the women present significant impact in the reduction of the negative gender gap. Moreover, this sector presents a significant impact for those positive discrimination gaps with an impact of around 1300 pesos.

This analysis provides interesting insights for additional research and future policy design. First, the gender gap must be computed using comparable workers. When this condition is maintained, then the gap reduction is much lower that raw wage comparison. The policy makers should be concerned about the gender gap of comparable workers because its magnitude in Chile is almost unalterable during the last 20 years. This result is supported by the recommendation of the Sixth World Economic Forum's Global Gender Gap Report 2011 where the main emphasis for the Chilean case is focused on the equity in labor market. With respect to policy design, the human capital increment in women is relevant for helping them to move from a negative toward a positive discrimination, but it is not enough. Undoubtedly, the occupation with the highest contribution to the transition between negative and positive gap which call to be
concerned about the educational policies in Chile. This result should motivate policies oriented to increase the number of women in this condition, specially trying to link the policy oriented to decrease the gender gap that a policy referred to the formation of human capital and education.

**Bibliography**


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


Paredes, D. (2012, April). *Working Paper WP2012-05*. Retrieved from Department of Economics: https://2126574539851918994-a-ucn-cl-s-sites.googlegroups.com/a/ucn.cl/wpeconomia/archivos/WP2012-05.pdf?attachauth=ANoY7cq52-bfZ2E-TG3hdfi34TjqnsOfBWp2wkIeNpc4obZa9gJ7g2V18gKkq4C0Buu5Rnq7seTJYIT_QukJZbOrF4OngdPiw8lfVG8_5Mm7udLntYj4HyQmvUz-F-YUa0Y6gb-C01gUMjnLjyHl0Mzgi6VZ4RYd1SY9yb6Bq4S91YuuVHTQkVgINzGthdNQVZ6aw12uaj9n&attredirects=0

**APPENDIX**

*Figure 6: Occupations by year*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Managers</td>
<td>1.44</td>
<td>1.935</td>
<td>1.205</td>
<td>2.049</td>
<td>2.483</td>
<td>2.545</td>
<td>1.952</td>
<td>843</td>
<td>14,452</td>
</tr>
<tr>
<td></td>
<td>5.07</td>
<td>5.91</td>
<td>4.37</td>
<td>5.56</td>
<td>6.13</td>
<td>5.86</td>
<td>3.97</td>
<td>1.91</td>
<td>4.77</td>
</tr>
<tr>
<td>Professionals</td>
<td>1.875</td>
<td>2.32</td>
<td>1.865</td>
<td>2.505</td>
<td>2.786</td>
<td>3.399</td>
<td>3.318</td>
<td>3.32</td>
<td>21,388</td>
</tr>
<tr>
<td></td>
<td>6.60</td>
<td>7.08</td>
<td>6.77</td>
<td>6.80</td>
<td>6.88</td>
<td>7.83</td>
<td>6.74</td>
<td>7.53</td>
<td>7.06</td>
</tr>
<tr>
<td>Technicians and Associate</td>
<td>1.654</td>
<td>2.143</td>
<td>2.096</td>
<td>2.593</td>
<td>2.703</td>
<td>3.207</td>
<td>3.321</td>
<td>4.073</td>
<td>21.79</td>
</tr>
<tr>
<td>Professionals</td>
<td>5.83</td>
<td>6.54</td>
<td>7.60</td>
<td>7.04</td>
<td>6.68</td>
<td>7.39</td>
<td>6.75</td>
<td>9.24</td>
<td>7.20</td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>Agriculture, forestry and fishing</td>
<td>3.04</td>
<td>4.002</td>
<td>2.674</td>
<td>4.08</td>
<td>4.81</td>
<td>5.543</td>
<td>6.955</td>
<td>5.478</td>
<td>36.562</td>
</tr>
<tr>
<td>Mining and quarrying</td>
<td>1.161</td>
<td>1.265</td>
<td>0.873</td>
<td>0.937</td>
<td>1.024</td>
<td>1.064</td>
<td>1.441</td>
<td>1.499</td>
<td>2.964</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.409</td>
<td>0.386</td>
<td>0.317</td>
<td>0.254</td>
<td>0.253</td>
<td>0.245</td>
<td>0.293</td>
<td>0.340</td>
<td>0.360</td>
</tr>
<tr>
<td>Electricity, gas and water supply</td>
<td>0.210</td>
<td>0.309</td>
<td>0.210</td>
<td>0.393</td>
<td>0.351</td>
<td>0.306</td>
<td>0.350</td>
<td>0.471</td>
<td>2.6</td>
</tr>
<tr>
<td>Construction</td>
<td>0.074</td>
<td>0.094</td>
<td>0.076</td>
<td>0.107</td>
<td>0.087</td>
<td>0.070</td>
<td>0.071</td>
<td>0.107</td>
<td>0.86</td>
</tr>
<tr>
<td>Transport and communications</td>
<td>0.015</td>
<td>0.070</td>
<td>0.112</td>
<td>0.104</td>
<td>0.101</td>
<td>0.103</td>
<td>0.107</td>
<td>0.104</td>
<td>0.1056</td>
</tr>
<tr>
<td>Finance and insurance activities</td>
<td>5.538</td>
<td>5.887</td>
<td>5.474</td>
<td>7.082</td>
<td>7.826</td>
<td>8.15</td>
<td>8.711</td>
<td>8.754</td>
<td>57.222</td>
</tr>
<tr>
<td>Administrative and support activities</td>
<td>1.878</td>
<td>1.65</td>
<td>1.527</td>
<td>2.147</td>
<td>2.343</td>
<td>2.122</td>
<td>2.897</td>
<td>2.676</td>
<td>16.756</td>
</tr>
<tr>
<td>Total</td>
<td>28.393</td>
<td>32.757</td>
<td>27.562</td>
<td>36.844</td>
<td>40.49</td>
<td>43.424</td>
<td>49.212</td>
<td>44.072</td>
<td>302.754</td>
</tr>
</tbody>
</table>

Source 1: Author's calculation using CASEN
Source 2: Author’s calculation using CASEN.