A GINI-BASED ANALYSIS OF THE DIFFERENCES BETWEEN MEN AND WOMEN IN THE LABOR MARKET OVER TIME

Amit Shelef and Edna Schechtman

Discussion Paper No. 21-02

August 2021

Monaster Center for
Economic Research
Ben-Gurion University of the Negev
P.O. Box 653
Beer Sheva, Israel

Fax: 972-8-6472941
Tel: 972-8-6472286

A Gini-based analysis of the differences between men and women in the labor market over time

Amit Shelef [1*], Edna Schechtman [2]

[1*] Amit Shelef (Corresponding author), *Department of Industrial Management, Sapir Academic College,* D.N. Hof Ashkelon 79165, Israel.

Email: amitshe@mail.sapir.ac.il, Tel: +972-54-777-6321.

[2] Edna Schechtman, *Department of Industrial Engineering and Management, Ben-Gurion University of the Negev*, P.O.B. 653, Beer-Sheva 84105, Israel.

E-mail: ednas@bgu.ac.il, Tel: +972-52-8795937.

A Gini-based analysis of the differences between men and women in the labor market over time

Abstract

The common practices in gender gap studies are twofold: look at summary statistics or use regression analyses. We look at the entire distributions and examine the amount of overlapping of one distribution (women) by the other (men) and the average rank of one group had it been ranked by the other. We combine two research areas: An advanced statistical method (the Gini-based overlapping index) and labor economics. The effects of marital status, age, level of education and the effect of occupations which are mainly dominated by either one of the genders are examined on Israeli data over three decades. Results show that the overlapping between men and women increases along time for both salary per month and hours worked per month, while for payment per hour it stays the same. While there is almost no change over time for the groups of bachelors and married subjects, an increase in overlapping is observed for singles in hours worked and monthly pay, but not in hourly salary. In branches which are dominated by women, the overlapping increases along time. Further, for the lowest examined level of education the overlapping decreases along time while for the highest level of education it increases.

Keywords: Classification; Gender; Gini coefficient; Income inequality; Overlapping.

JEL code: J31, D63.

1. Introduction

In the labor market of most industrialized countries there are laws to enforce equality between women and men. However, it is also evident that the difference between the genders' wages still exists in many countries (Blau and Kahn 2003). Blau and Kahn (1996) indicate that at the end of the 80-ies the ratio between the incomes of females and males was around 80%-90% in the Scandinavian countries, France, Australia and New-Zealand and a bit lower (between 65% to 75%) in other Western European countries and in the United States. One of the reasons for the differences between wages of men and women is the difference in the type of occupation. Important contributions by Bergmann (1974), Macpherson and Hirsch (1995), Addison et al. (2018) and others to the understanding of gender differences in the labor market have been the findings that wages of both women and men are lower in occupations which are mainly dominated by women. There seems to be "female-based" and "male-based" occupations. Killingsworth (1990) states that when an occupation is "female-based", the salaries in general are lower (for both genders).

Christofides et al. (2013) studied the gender wage gap across 26 European countries, using 2007 data from the European Union Statistics on Income and Living Conditions. They used quantile regression and found that the gap is wider at the two ends of the income distribution: at the top (they use the term "glass ceiling") and at the bottom ("sticky floors"). In addition, they studied the subgroup of full-time full-year employees and found that for that subgroup the gap was wider.

One possible explanation for the wage gap is that women devote less time to the job market and do not start their jobs immediately after graduation from college due to responsibilities at home. Polachek (1975) and O'Neill (2003) point out that this fact slows them in getting experience that can be gained only on the job. The issue of raising a family was further studied by Kleven et al. (2018). They found that a long run gap of 20% was created with the arrival of children. The main reasons were hours of work, wage rate and participation in the labor market. The gender gap varies over age. Tyrowicz et al. (2017) used semi-parametric models on data from the GSEP (the German Socio-Economic Panel) for years 1984-2015. They found that the gap increased over lifetime. The effect of the workplace (women in management, organizational practices and flexibility in the workplace) was studied by Abendroth et al. (2017) Bertrand (2017) and Huffman et al. (2017).

The statistical analyses employed in studies in the area of wage equations are based mainly on regression analyses, where the explanatory variables are identified and their weights (i.e., regression coefficients) are estimated. Our focus is different. We focus on the changes over time, as reflected by parameters which compose the Gini coefficient. That is, we base our analyses on the decomposition of the Gini coefficient. Two new parameters are used in our analyses. The first one is the overlapping index Oij, which measures the amount of overlapping of distribution i by the distribution j. The second one, which is more intuitive, is Fij. Fij is the average rank of observations coming from distribution i, had they been ranked within distribution j. For example, if all women's wages are lower than all men's wages, then the average rank of the women, had they been ranked within the men would be the smallest rank (1). However, if the two distributions are similar then the ranks of women within men would be similar to the ranks of women within themselves, that is: the average rank would be 0.5. The main contribution to the literature is that the paper suggests a different angle to analyze the data using a relatively (relative to OLS that is frequently used) unexplored statistical method. The main criteria we use are the amount of overlapping of one distribution (women) by the other (men) and the average rank of incomes of one group had it been ranked by the incomes of the other.

More specifically, we use the Gini measure of inequality as the basic statistical tool and decompose the Gini coefficient of income of the entire population of males and females in Israel in three years representing three decades: 1997, 2005 and 2015 by gender in three different ways: according to monthly wages, hourly wages and the number of hours worked per month. The objective of the study is to investigate the changes over time for each variable. In order to obtain a deeper understanding, we start with the entire sample, and then look at different age groups, marital statuses, levels of education and types of occupation.

The structure of the paper is as follows: in section 2 we briefly review the method of analysis. Section 3 is devoted to the description of the data, which is a case study – the Israeli experience - and to the empirical results, and section 4 concludes.

2. A brief review of the method of analysis

The methodology used in this research is called ANOGI (ANalysis Of GIni). Generally, ANOGI is similar to ANOVA (ANalysis Of VAriance), and hence the name. It is based on deomposing the Gini coefficient (rather than the variance which is the basic building block of ANOVA). The basic idea is to divide the population into subgroups in several different ways and try to identify which division separates the incomes of men and women in the best way. The methodology is detailed in the Appendix. (For a full description of the methodology, the derivation and the properties of the parameters and estimators involved see Frick et al. (2006)). The Gini coefficient can be expressed in more than a dozen ways (Yitzhaki 1998). The formula of the Gini coefficient used in this paper is (Lerman and Yitzhaki 1984, 1989):

$$G = \frac{2\operatorname{cov}(Y, F(Y))}{\mu} \ ,$$

which is twice the covariance between the income Y and its cumulative distribution $F(Y)$, standardized by mean income μ.

When the subgroups are exclusive the Gini can be decomposed into the between and within contributions, as was recently mentioned in Mussard and Mornet (2018). If the subgroups are not exclusive, it turns out that Gini is not additively decomposable by population sub-groups. As demonstrated by Frick et al. (2006), the Gini decomposition according to population sub-groups offers a parameter (called the overlapping coefficient) which quantifies the amount of overlapping between the distributions.

We note that while other ways of comparing groups are based on a few parameters of the distributions (i.e., means, medians, variances), the proposed method is based on the entire distributions. The decomposition and the statistical properties of the estimators are described in Frick et al. (2006) and in the Appendix.

The basic tool of ANOGI is the following decomposition of the Gini of the entire population, denoted by G_u:

$$G_u = \sum_{i=1}^{n} s_i G_i O_i + G_b = \sum_{i=1}^{n} s_i G_i + \sum_{i=1}^{n} s_i G_i (O_i - 1) + G_{bp} + (G_b - G_{bp}), \qquad (1)$$

where s_i denotes the share of group i in the overall income, O_i is the overlapping index of subpopulation i with the entire population (explained in the Appendix), G_b measures the between-group inequality and G_{bp} is Pyatt's between-group Gini (Pyatt 1976). An additional parameter of interest, to be used in the analysis below, is the average rank of members of group i, had they been ranked within group j (Fij). This is part of the overlapping index, and is more intuitively interpreted. Fii (average rank of members of group i within itself) is equal to 0.5. Fij smaller than 0.5 means that group i lies to the left of group j.

3. Data Description and results

The data consists of the Public Use Files (PUF) of the Israeli family income surveys for the calendar years 1997, 2005 and 2015. The data covers the entire population that live in Israel, except those who live in small villages or institutions and includes ages 25-64, which covers 95 percent of the Israeli population. The sample is a sample of dwellings with each dwelling having an equal probability of being included in the sample. In addition, a weight is attached to each observation. The weights are determined by the central bureau of statistics in order to insure that the sample represents the population and to handle biases that may be caused by non-response. The sample is composed of individual employees with wages and hours worked per month.

The data presented in Table 1 show the decomposition of monthly pay, hourly pay, and hours worked per month side-by-side for the three decades represented by 1997, 2005 and 2015. The jackknife standard errors, calculated according to Yitzhaki (1991), are in parentheses. The overlapping index Oi (also Oij) is the index by Yitzhaki (1994). See formulas A.2 and A.3 in the Appendix. We note in passing that the analyses with and without weights gave similar results. The results reported here are without weights.

Table 1. The decompositions of Gini for the three variables: salary per month, payment per hour and hours worked per month for 1997, 2005 and 2015, age range between 25 and 64. Jackknife standard errors, calculated according to Yitzhaki (1991), are in parentheses.

		Salary per month			Payment per hour			Hours of work per month		
		1997	2005	2015	1997	2005	2015	1997	2005	2015
1	**Sample size (number of observations)**	5406	5794	8659	5406	5794	8659	5406	5794	8659
	Men (i=1)	2895	2921	4270	2895	2921	4270	2895	2921	4270
	Women (i=2)	2511	2873	4389	2511	2873	4389	2511	2873	4389
2	**Proportion (P_i)**									
	Women (i=2)	0.46	0.50	0.51	0.46	0.50	0.51	0.46	0.50	0.51
3	**Overall mean (M_u)**	6204.85	7853.84	10807.22	35.72	44.97	62.03	178.94	175.77	172.39
	Men (i=1)	7705.17	9610.22	13067.17	38.16	47.95	67.54	205.43	201.88	193.47
	Women (i=2)	4475.08	6068.12	8608.54	32.90	41.94	56.68	148.41	149.22	151.89
4	**Share in the total (S_i)**									
	Women (i=2)	0.33	0.38	0.40	0.43	0.46	0.46	0.39	0.42	0.45
5	**Overall Gini (G_u)**	0.416	0.403	0.410	0.381	0.354	0.357	0.208	0.204	0.187
6	**Within group Gini (G_i)**									
	Men (i=1)	0.402	0.386	0.398	0.381	0.358	0.368	0.161	0.157	0.149
	Women (i=2)	0.380	0.385	0.394	0.376	0.346	0.338	0.222	0.215	0.196
7	**Overlapping with entire sample (O_i)**									
	Men (i=1)	0.90	0.90	0.92	0.98	0.98	0.99	0.91	0.92	0.97
	Women (i=2)	0.98	1.01	1.00	1.01	1.01	1.00	0.88	0.89	0.90
8	**Overlapping between groups (O_{ji})**									
	Overlapping index men by women	0.79	0.80	0.85	0.95	0.97	0.98	0.80	0.85	0.93
	Overlapping index women by men	0.97	1.02	1.01	1.03	1.01	0.99	0.78	0.78	0.79
9	**Average ranking within the other group (F_{ji})**									
	Men, if ranked within Women	0.69	0.68	0.65	0.56	0.56	0.56	0.79	0.78	0.77
	Women, if ranked within men	0.31	0.32	0.35	0.44	0.44	0.44	0.26	0.27	0.30
10	**Within-group component (S_i*G_i*O_i)**	0.366	0.363	0.378	0.376	0.350	0.352	0.166	0.164	0.158
	Men (i=1)	0.241	0.214	0.219	0.213	0.189	0.196	0.090	0.084	0.080
	Women (i=2)	0.125	0.149	0.160	0.163	0.161	0.156	0.076	0.080	0.079
11	**Between-group component (G_b)**	0.050	0.040	0.031	0.005	0.004	0.005	0.043	0.039	0.029
12	**Between-group Pyatt (G_{bp})**	0.130	0.113	0.103	0.037	0.033	0.044	0.080	0.075	0.060
13	**Difference G_b-G_{bp}**	-0.080	-0.073	-0.072	-0.032	-0.030	-0.038	-0.037	-0.036	-0.032

Before we start, note that the sample sizes for 1997 and 2005 are 5406 and 5794, respectively, but in 2015 there are 8659 observations, as a result of a change in the Israeli family income survey. The proportion of women in the sample was 0.46 in 1997, rose to 0.5 in 2005 and was about the same (0.51) in 2015. Jackknife standard errors, calculated according to Yitzhaki (1991), are between 0.001-0.0184.

We now turn to look at trends over time. We start with a descriptive analysis. Figure 1 presents the histograms for the three variables over the 3 years under study. As can be seen, the histograms for men and women overlap for the variable "salary per hour" with no visual difference over the years. However, for both salary per month and hours worked, one can notice that the differences in the areas on the right-hand-side of the histogram become smaller in 2015, meaning more overlap and a tendency to close the gap.

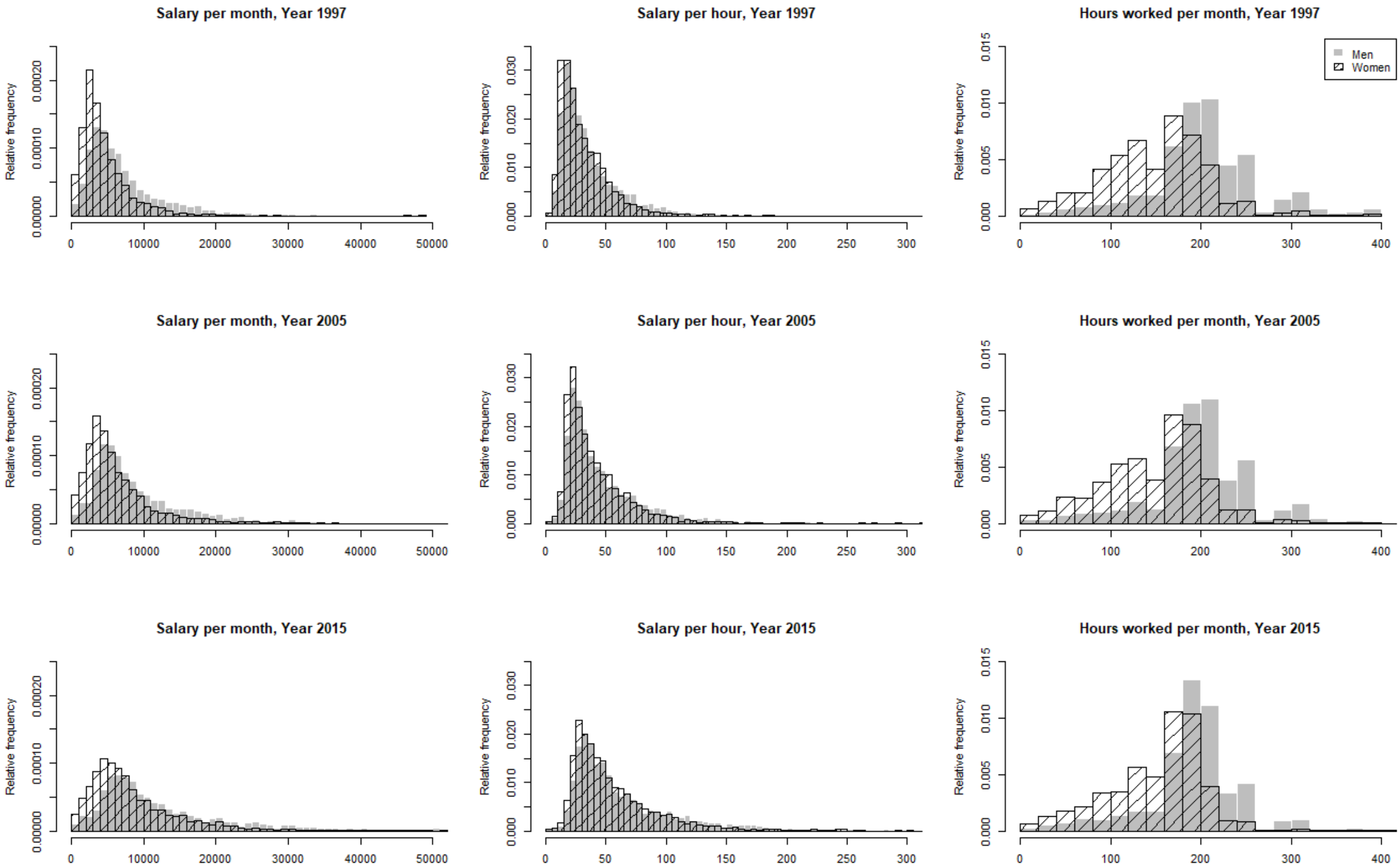


Figure 1. Histograms (relative frequencies) for men (in black) and women (in gray) for each variable and each year. In the columns, variables: monthly salary (left), payment per hour (middle) and hours worked per month (right). In the rows, years: 1997, 2005 and 2015.

Formally, the share of women in the total increases over time. The highest increase is for salary per month: from 33% (0.33 =(4475·2511)/(6204·5406)) in 1997 to 38% and 40% in 2005 and 2015. Similarly, for hours worked per month the share increases from 39% to 42% and 45% over time. The increase in payment per hour is fairly minor (43%, 46% and 46% for 1997, 2005 and 2015, respectively). It is interesting to note that the share of women in the sample of working individuals increases from 46% to 50% and 51%, respectively.

Next, we focus on Fij because it is an easy and intuitive index to interpret. Looking at Fij, the values tend towards 0.5 over time for salary per month and for hours worked per month, but do not change for payment per hour. Recall that Fii (that is, ranking a group within itself) is 0.5. In addition, note that theoretically Fij+Fji=1. However, for discrete variables (such as hours worked per month), the sum may be slightly different than 1 due to ranking in the presence of ties.

For hours worked per month, Fij for men within women slightly decreases from 0.79 in 1997, to 0.78 in 2005 and to 0.77 in 2015 (for salary per month the value decreases from 0.69 in 1997, to 0.68 in 2005 and to 0.65 in 2015). This indicates that women become more similar to men in both variables. The average rank stays unchanged (and relatively low, 0.56) for payment per hour and it is the smallest gap of the three variables under study.

Another parameter of interest is the difference G_b-G_{bp} (Table 1, Line 13), which is referred to as an indication of the overlapping. The results of this difference show that for salary per month and hours worked per month the gap tends to get smaller (the values are smallest in absolute value for 2015), while for payment per hour the trend is not clear.

We now turn to more detailed analyses. We start with the effect of marital status on the overlapping between men and women. The marital status was divided into 3 categories: married, bachelor (never married) and single (widower, separated, divorced), as, for example, in Christofides et al. (2013). Figure 2 illustrates the values of Fij over the 3 decades by marital status. As can be seen from Figure 2, for the three years under study the best separator is hours worked per month – its line lies above the other two for all cases – and the least is salary per hour. Furthermore, while there is almost no change over time for the groups of bachelors and married subjects, an increase in overlapping can be seen for the singles in hours worked and monthly pay, but not in hourly salary. The most interesting finding (but not surprising) is that the group denoted by "bachelor" shows a different pattern than the other two groups. The overlapping between men and women within this group is much higher than in the other two for the three variables under study and the three decades (the values of Fij are closer to 0.5). We note in passing that most of the bachelors are relatively young (72% to 78% lie within ages 25 and 34).

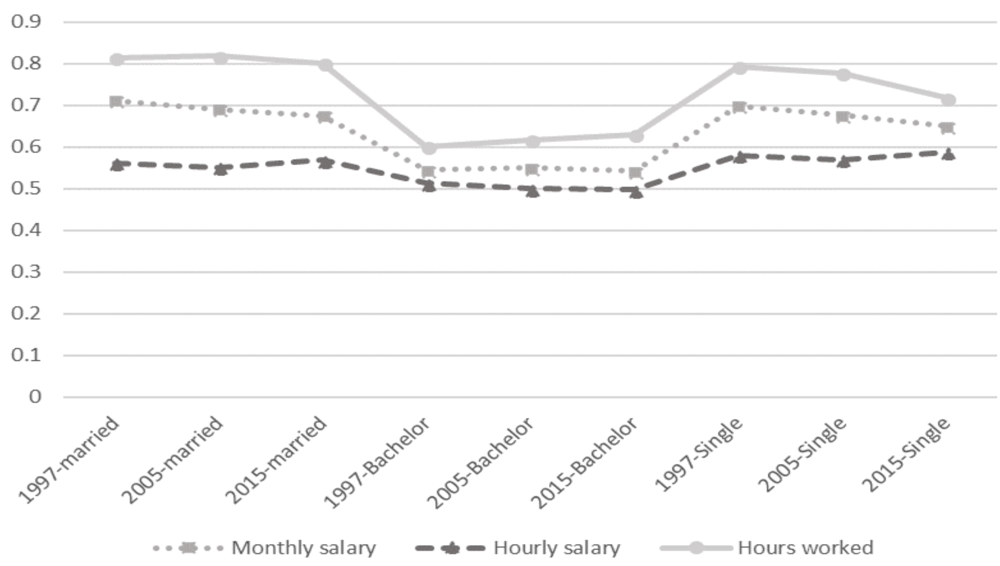


Figure 2. Fij by marital status and year.

Analysis by age groups could shed some light. Unfortunately, when dividing the group denoted by "singles" into subgroups by age, the subsamples are too small and very unbalanced in terms of men and women so could not be analyzed. The only valid comparison was between married and bachelors among young subjects (25-34). The results of Fij values are shown in Figure 3, and are very similar to the previous ones.

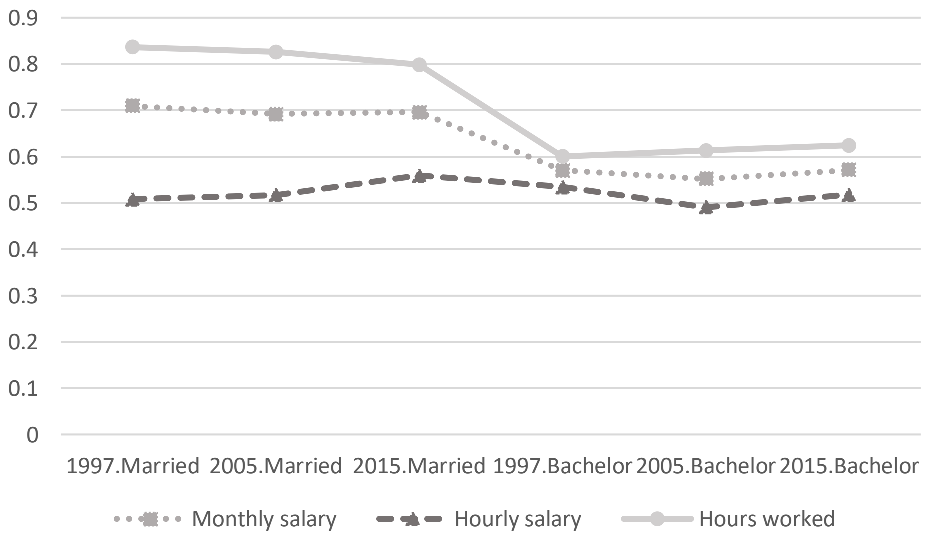
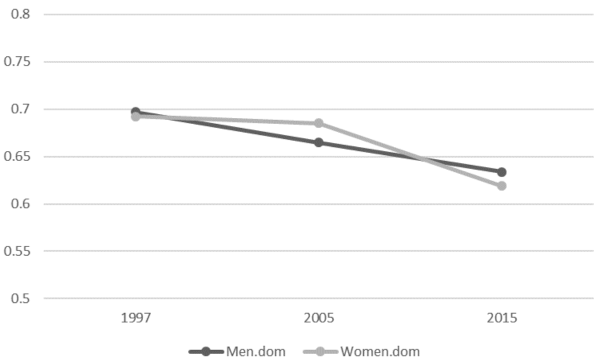
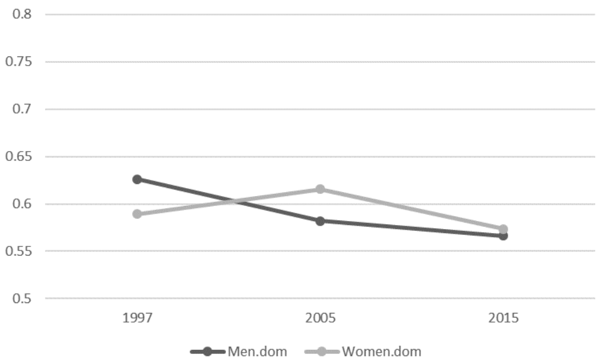


Figure 3. Fij for ages 25-34 by marital status (married and bachelor) and year.

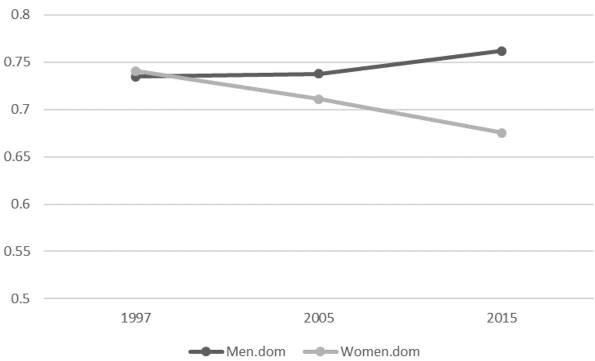
Next, we focus on trends over time for men-dominated and women-dominated branches. We identified branches which had at least 70 percent men (or women) and refer to them as men-dominated (or women-dominated) branches. Figures 4a, 4b and 4c present Fij values for monthly salary, hourly salary and hours worked per month, respectively, over the 3 decades for men-dominated (in black) and women-dominated (in gray) branches. In men-dominated branches, the overlapping decreases for hours worked (average rank Fij increases), but increases for the other two variables. In women-dominated branches, the overlapping increases for all variables and especially for hours worked. These patterns indicate that in women-dominated branches, men's salaries (per month and per hour) are adapted to fit the branch's customary incomes and in terms of hours worked, the distributions are becoming more similar. However, in men-dominated branches, women do not tend to adapt their hours worked to fit those of men and therefore, the overlapping decreases while for the other two variables women tend to adapt over time.



(a)



(b)



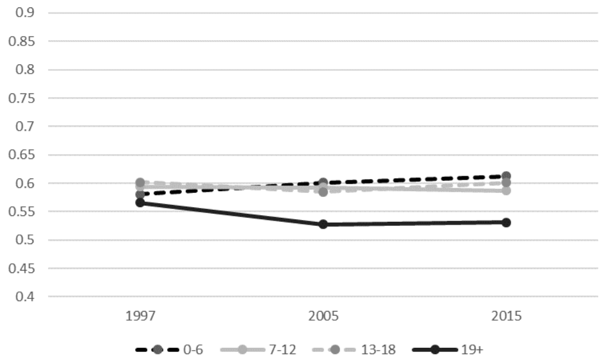
(c)

Figure 4. Fij of monthly salary (a); hourly salary (b) and hours worked per month (c) for men-dominated vs. women-dominated branches by year.

Next, we examine the effect of education on trends over time. We divided the range of years of study to 4 categories (0-6 years of study, 7-12, 13-18 and at least 19). Figures 5a, 5b and 5c present Fij values for monthly salary, hourly salary and hours worked per month, respectively, over the 3 decades, for 0-6 years of study (in dashed black), 7-12 (in solid gray), 13-18 (in dashed gray) and at least 19 (denoted 19+ in solid black). In Figures

5b and 5c, only slight changes occur over time. In Figure 5a (monthly salary) changes occur for two groups: with the lowest and the highest levels of education. For 0-6 years of study (elementary school) the overlapping decreases along time, however for 19+ years of study the overlapping increases.

(a)



(b)

(c)

Figure 5. Fij of monthly salary (a); hourly salary (b) and hours worked per month (c) for years of study by year.

4. Conclusions

The objective of this paper is to investigate the trends in the gaps between men and women in the labor market by looking at three variables: monthly salary, hourly pay and hours worked per month. We used Israeli data over three decades as a case study. While the common practice is to use regressions, we suggest to use a methodology that is based on decomposition of the Gini coefficient of a population according to population sub-groups. The decomposition offers parameters that describe the degree of overlapping between the distributions of the sub-groups. The lower the overlapping between the distributions of women and men and the further from 0.5 the average rank is, the higher the degree of stratification and bigger gap. The advantage of this methodology is that it gives a wide perspective because it takes the entire distribution into account, rather than looking at a choice of representative parameters.

Results showed that the shares of working women in the total sample increase over the three decades under study. In addition, it was found that women become more similar to men for both salary per month and hours worked per month. For payment per hour, however, there appears to be a relatively small gap, which stays the same during the three decades under study. We note in passing that the best classifier (out of the three variables under study) was working hours per month. When focusing on the marital status, the overlapping between men and women within bachelor group is much higher than in the other two for the three variables and the three decades under study. It is interesting to note that for singles the overlapping increases over the three decades for monthly salary and hours worked, but not for salary per hour, for which the gap is relatively small.

Another interesting finding is that the overlapping between men and women increases over the three decades in branches which are dominated by women, meaning that men adapt their monthly salary, hourly salary and hours worked to the one that is common in the branch. However, in branches which are dominated by men this is not the case and the overlapping decreases for hours worked per month (the best classifier). Predominantly female occupations were examined by Macpherson and Hirsch (1995) and Addison et al. (2018), who found that wage levels were lower in those occupations. However, this finding by itself is not concerned with the gap between men and women. It does not indicate whether men are occupying the higher range of wages for those occupations or not. Our contribution is by showing that within branches which are dominated by women, the overlapping between men and women increases along time. That is, the distributions become more alike. An additional insight - the influence of education on monthly salary. It was observed that the overlapping decreases along time for the lowest level of education, while for the highest level of education it increases. To summarize, the good news is that in our case study the gap is closing in most of the analyses shown.

Declarations

Compliance with Ethical Standards:

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Availability of data and material

The data that support the findings of this study are available from Israel Central Bureau of Statistics. Restrictions apply to the availability of these data, which are confidential and were used under license for this study.

References

Abendroth, A. K., S. Melzer, A. Kalev, and D. Tomaskovic-Devey. 2017. "Women at Work: Women's Access to Power and the Gender Earnings Gap." *ILR Review* 70(1):190–222.

Addison, J. T., O. D. Ozturk, and S. Wang. 2018. "The Occupational Feminization of Wages." *ILR Review* 71(1):208–41.

Arvesen, J. N. 1969. "Jackknifing U-Statistics." *The Annals of Mathematical Statistics* 40(6):2076–2100.

Bergmann, B. R. 1974. "Occupational Segregation, Wages and Profits When Employers Discriminate by Race or Sex." *Eastern Economic Journal* 1(2):103–10.

Bertrand, M. 2017. "The Glass Ceiling." *Becker Friedman Institute for Research in Economics Working Paper No. 2018-38. Available at SSRN: Https://Ssrn.Com/Abstract=3191467*.

Blau, F. D., and L. M. Kahn. 1996. "Wage Structure and Gender Earnings Differentials: An International Comparison." *Economica* 63:S29–62.

Blau, F. D., and L. M. Kahn. 2003. "Understanding International Differences in the Gender Pay Gap." *Journal of Labor Economics* 21(1):106–44.

Christofides, L. N., A. Polycarpou, and K. Vrachimis. 2013. "Gender Wage Gaps, 'Sticky Floors' and 'Glass Ceilings' in Europe." *Labour Economics* 21:86–102.

Cowell, F. A. 1980. "On the Structure of Additive Inequality Measures." *The Review of Economic Studies* 47(3):521–31.

Frick, J. R., J. Goebel, E. Schechtman, G. G. Wagner, and S. Yitzhaki. 2006. "Using Analysis of Gini (ANOGI) for Detecting Whether Two Subsamples Represent the Same Universe: The German Socio-Economic Panel Study (SOEP) Experience." *Sociological Methods & Research* 34(4):427–68.

Huffman, M. L., J. King, and M. Reichelt. 2017. "Equality for Whom? Organizational Policies and the Gender Gap across the German Earnings Distribution." *ILR Review* 70(1):16–41.

Killingsworth, M. R. 1990. *The Economics of Comparable Worth*. Michigan: Upjohn Press, Kalamazoo.

Kleven, H., C. Landais, and J. E. Søgaard. 2018. "Children and Gender Inequality: Evidence from Denmark." *NBER Working Paper No. 24219*.

Lerman, R., and S. Yitzhaki. 1984. "A Note on the Calculation and Interpretation of the Gini Index." *Economics Letters* 15(3–4):363–68.

Lerman, R., and S. Yitzhaki. 1989. "Improving the Accuracy of Estimates of Gini Coefficients." *Journal of*

Econometrics 42(1):43–47.

Macpherson, D. A., and B. T. Hirsch. 1995. "Wages and Gender Composition: Why Do Women's Jobs Pay Less?" *Journal of Labor Economics* 13(3):426–71.

Mookherjee, D., and A. Shorrocks. 1982. "A Decomposition Analysis of the Trend in UK Income Inequality." *The Economic Journal* 92:886–902.

Mussard, S., and P. Mornet. 2018. "A Note on A-Gini Measures." *Review of Income and Wealth* To appear.

O'Neill, J. 2003. "The Gender Gap in Wages, circa 2000." *American Economic Review* 93(2):309–14.

Polachek, S. W. 1975. "Potential Biases in Measuring Male-Female Discrimination." *Journal of Human Resources* 10(2):205–29.

Pyatt, G. 1976. "On the Interpretation and Disaggregation of Gini Coefficients." *The Economic Journal* 86:243–55.

Schechtman, E., and S. Wang. 2004. "Jackknifing Two-Sample Statistics." *Journal of Statistical Planning and Inference* 119(2):329–40.

Shaw, J., and D. S. Tu. 1995. *The Jackknife and the Bootstrap*. New York: Springer.

Shorrocks, A. F. 1984. "Inequality Decomposition by Population Subgroups." *Econometrica* 52:1369–85.

Silber, J. 1989. "Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality." *The Review of Economics and Statistics* LXXI:107–15.

Tyrowicz, J., L. Van der Velde, and I. van Staveren. 2017. "Identifying Age Penalty in Women's Wages: New Method and Evidence from Germany 1984-2014." *IZA Discussion Paper No. 11295. Available at SSRN: Https://Ssrn.Com/Abstract=3117299*.

Yitzhaki, S. 1991. "Calculating Jackknife Variance Estimators for Parameters of the Gini Method." *Journal of Business and Economic Statistics* 9(2):235–39.

Yitzhaki, S. 1994. "Economic Distance and Overlapping of Distributions." *Journal of Econometrics* 61:147–59.

Yitzhaki, S. 1998. "More than a Dozen Alternative Ways of Spelling Gini." *Research on Economic Inequality* 8:13–30.

Yitzhaki, S., and R. I. Lerman. 1991. "Income Stratification and Income Inequality." *Review of Income and Wealth* 37(3):313–29.

Appendix. A detailed review of the ANOGI (ANalysis Of GIni) methodology

The methodology used in this research is called ANOGI (ANalysis Of GIni). Generally, ANOGI is similar to ANOVA (ANalysis Of VAriance), and hence the name. The ANOGI method is based on the decomposition of the Gini coefficient of, say, the entire population's income into two components: a "within groups" component, also called "intra", and a "between groups" component, also called "inter". Intuitively, the larger the between component is, relative to the within component, the better the stratification is. (See Frick et al. (2006), for a full description of the methodology, the derivation and the properties of the parameters and estimators involved). The Gini coefficient can be expressed in more than a dozen ways (Yitzhaki 1998). The formula of the Gini coefficient used in this paper is (Lerman and Yitzhaki 1984, 1989):

$$G = \frac{2\operatorname{cov}\left(Y, F(Y)\right)}{\mu} \ ,$$

which is twice the covariance between the income Y and its cumulative distribution $F(Y)$, standardized by mean income μ.

The Gini coefficient is the most popular measure of inequality. Naturally, one would wish to decompose the Gini of a population into the contributions of the sub populations (males and females in this research). Generally, the Gini is not additively decomposable by population sub-groups. As a result, many economists argue that it is not meaningful to decompose it (Cowell 1980; Shorrocks 1984). However, as demonstrated by Frick et al. (2006), the Gini decomposition according to population sub-groups offers a parameter (called the overlapping index) which quantifies the amount of overlapping between the distributions. Stratification is interpreted as an indicator of the quality of classification into groups.

In this paper we focus on overlapping and gain additional insights regarding the data at hand. We note that while other ways of comparing groups are based on a few parameters of the distributions (i.e., means, medians, variances), the proposed method is based on the entire distributions. The decomposition and the statistical properties of the estimators are described in Frick et al. (2006). Therefore this section will only describe the main properties of the decomposition referred to as ANalysis Of GIni (ANOGI).

Let

$Y_u = Y_1 \cup Y_2 \cup \ldots \cup Y_n$,

where Y_u is the income of the entire population and Y_i is the income of sub population i (i=1,…,n).

The Gini of the entire population, denoted by G_u, can be presented as

$$G_u = \sum_{i=1}^{n} s_i G_i O_i + G_b = \sum_{i=1}^{n} s_i G_i + \sum_{i=1}^{n} s_i G_i (O_i - 1) + G_{bp} + (G_b - G_{bp}), \qquad \text{(A.1)}$$

where s_i denotes the share of group *i* in the overall income, O_i is the overlapping index of subpopulation *i* with the entire population (explained below), G_b measures the between-group inequality and G_{bp} is Pyatt's between-group Gini (Pyatt 1976). Formally G_b is given by

$$G_b = \frac{2\operatorname{cov}(\bar{Y}, \bar{F}_u)}{\mu_u}.$$

G_b is twice the covariance between the mean incomes of the sub-populations and the sub-populations' mean of the ranks in the overall population, divided by overall expected income. That is, each sub-population is represented by its mean income and by the mean rank of its members in the overall distribution.

An alternative between-groups Gini (G_{bp}) was defined by Pyatt (1976). (Mookherjee and Shorrocks (1982), Shorrocks (1984) and Silber (1989) also follow Pyatt). In this definition, the between-groups Gini is based on the covariance between the mean income in each sub-population and its rank among the mean incomes of the sub-populations. This between-group component **is** a pure Gini (of the vector of means). The difference between the two definitions is in the rank that is used to represent the group: under Pyatt's approach it is the rank of the mean income of the sub-population, while under Yitzhaki-Lerman it is the mean rank of all members belonging to the sub-population.

Generally, it can be shown (Frick et al. 2006) that $G_b \le G_{bp}$.

The overlapping of subpopulation i with the entire population can be presented as composed of the overlapping between each pair of groups:

$$O_i = \sum_j p_j O_{ji} = p_i O_{ii} + \sum_{j \neq i} p_j O_{ji} = p_i + \sum_{j \neq i} p_j O_{ji} \qquad \text{(A.2)}$$

where O_{ji} is the overlapping of sub population j by sub-population i. Formally, it is given by

$$O_{ji} = \frac{\operatorname{cov}_i(Y, F_j(Y))}{\operatorname{cov}_i(Y, F_i(Y))} \qquad \text{(A.3)}$$

where $F_j(y)$ is the rank of observation *y*, coming from F_i, had it been ranked within the observations from distribution *j*. The overlapping coefficient can tell us how much the distributions are intertwined.

The overlapping coefficient was introduced by Yitzhaki and Lerman (1991) and modified in Yitzhaki (1994). Intuitively, it measures to what extent one group is overlapped by the other. It takes values between 0 and 2, where the extreme lower bound (zero) occurs when there is a complete stratification, i.e., when each group occupies a given range and the ranges do not intersect (no overlapping, perfect stratification), and the extreme upper bound for group A (at most 2, depending on the distribution) occurs when group B is concentrated inside the range of A, around the mean of group A, with no member of group A lying inside the range of group B. In this case, group A cannot be considered to be a group because the members of group B separate the members of A that are below the average of A from those that are above it. An overlapping index that is equal to one means that the distributions of the two sub-populations are similar. The measure is based on ranking the members of one group according to the ranking of the other. O_{ji} is an index that measures the extent to which population *j* is included in the range of population *i*. Note that the indices O_{ji} and O_{ij} are not inter-related by a simple relationship. It is clear that the indices of overlapping are not independent. (A detailed list of properties of the overlapping coefficient can be found in Frick et al. (2006)).

The other components of Equation (A.1) that require an additional interpretation are G_b and G_{bp}. G_b is based on the covariance between the mean value of each sub-group and the average rank of its members in the overall distribution (Yitzhaki and Lerman 1991). On the other hand, G_{bp} is based on the covariance between the mean value of each sub-group and the ranking of the mean value in the distribution of mean-values (Pyatt 1976). By construction, $G_b \leq G_{bp}$. The role of the overlapping in Equation (A.1) can be seen from the second and fourth terms on the right-hand side of the equation. The terms G_u, G_i (i=1, …, n) and G_{bp} are not affected by the degree of overlapping. Therefore, the higher the degree of overlapping between the sub-groups the higher the second term on the right-hand side of Equation (A.1) (intra-group component) and the lower the fourth term (between-group component). The difference between G_{bp} and G_b is the following: G_{bp} represents the maximum value that the between-group component can achieve (assuming that all members of the groups are identical). G_b represents the between group component that takes into account the overlapping. The difference between them represents the deterioration in the classification into groups, due to overlapping.

For the benefit of readers who are interested in a quick comparison with ANOVA, a summary table of ANOGI is given in Table A.1 and is detailed in Frick et al. (2006).

Table A.1. A Summary of ANOGI components in comparison to ANOVA

Components Parallel to ANOVA	Formula	Range
Intra-Group	$IG = \sum_{i=1}^{n} s_i G_i$	$0 \leq IG \leq G_u$
Between-Groups-Pyatt	$BG_p = G_{bp}$	$0 \leq BG_p \leq G_u$
Additional Information		
Overlapping Effect on Intra-Group	$IGO = \sum_{i=1}^{n} s_i G_i (O_i - 1)$	
Overlapping Effect on Between-Groups	$BGO = G_b - G_{bp}$	$-BG_p - IGO - IG \leq BGO \leq 0$

Equation (A.1) ivolves 4 parameters which need to be estimated: G_i, O_i, G_b and G_{bp}. The estimation technique used here is based on U-statistics theory. For each parameter, a kernel of the proper degree is found and then, a U-statistic is constructed. Details can be found in Frick et al. (2006). In addition, the jackknife method for variance estimation works well for U-statistics (see Arvesen 1969; Schechtman and Wang 2004; Shaw and Tu 1995) and was used in order to estimate the standard errors of the estimators.