



Working Paper Series

**Segregation of women into low-paying
occupations in the US**

Carlos Gradín

ECINEQ WP 2017 - 426

Segregation of women into low-paying occupations in the US*

Carlos Gradín[†]

Universidade de Vigo and EQUALITAS, Spain

Abstract

We present an approach to measure the stratification of occupations by sex. For that, we extend the conventional framework for measuring gender segregation to take into account the quality of jobs (e.g. average earnings) predominantly held by each sex. We complement segregation curves and measures derived from them, with their associated concentration curves and indices, to determine whether women are segregated into low-paying jobs. We investigate with this approach the long-term trends of gender segregation and stratification of occupations by sex in the US using census data. Our results show that de-stratification of occupations by sex was more intense than their desegregation, and lasted longer, even after segregation had stagnated. Neither segregation nor stratification levels can be explained by the different characteristics of male and female workforces, although the profound changes in the composition of workers over time (e.g. education, marital status) did help to substantially explain their trends. Changes in the earnings structure favoring occupations held by women since 1980 additionally contributed to reduce stratification over time. Finally, changes in the conditional occupational distribution by sex only reduced segregation and stratification before 1990.

Keywords: occupational segregation, stratification, low-paying occupations, gender.

JEL Classification: J16, J42, J71, J82, 051.

*I acknowledge financial support from the Spanish *Ministerio de Economía y Competitividad* (ECO2016-76506-C4-2-R) and *Xunta de Galicia* (GRC 2015/014).

[†]**Contact details:** Facultade de CC. Económicas e Empresariais, Universidade de Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Galicia, Spain. E-mail: cgradin@uvigo.es; Phone: +34 986 813527.

1. Introduction

Among the many gender inequalities across labor market outcomes, occupational segregation by sex is one that has received considerable attention so far, both methodologically and empirically. The conventional framework to measure occupational segregation by sex (or by another attribute) involves the analysis of inequality in the distribution of workers from each group across different occupations. This may imply drawing segregation curves by plotting the cumulative proportion of workers from each group with occupations sorted by their sex ratio. In addition, one can quantify the level of segregation using measures, like the most popular dissimilarity index, or the Gini coefficient, among others. One important feature of this framework is to assume symmetry across occupations. The only relevant information of an occupation is the proportion of workers from each population group. However, workers may regard some occupations to be better because of the required skills, social prestige, pay, social benefits, physical risks, instability, or any other labor conditions attached to jobs.

Any type of segregation has potentially negative implications in terms of social cohesion. Nevertheless, one can expect more negative effects when segregation implies stratification among groups, systematically confining one particular group (like women or nonwhites) into less valued jobs. These effects are further aggravated whenever this segregation cannot be justified on the grounds of differences in the accumulation of human capital. The segregation of women, immigrants, or ethnic minorities into low-paying occupations has historically been the main source of social discrimination for these groups all over the world, constraining their economic and social opportunities. In this line, job segregation is one of the main sources that explain the gender wage differentials within countries (e.g. Groshen, 1991; Bayard et al., 2003; Amuedo-Dorantes and De la Rica, 2006; Brynin and Perales, 2015).¹ Occupational segregation by sex strongly declined in the US since 1970 (e.g. Blau, Brummund, and Liu, 2013), but the slower pace at which it decreased since the 1990s (stalling after the 2000s) helps to explain some of the slowing convergence also in the gender pay gap (e.g. Blau and Kahn, 2006). At the same time, gender wage gaps within occupations have increased its relevance in explaining the persistence of the wage gap (e.g. Goldin, 2014).

The main contribution of this paper is to extend the conventional framework for measuring gender segregation to take into account the quality of the jobs held by each population group.

¹ However, the evidence in favor of segregation explaining inter-country differences in the wage gap is weaker (e.g. Blau and Khan, 2003; Dolado et al., 2002; Pissarides et al., 2003).

We aim at measuring the extent to which segregation implies stratification in occupations, with one group systematically holding less-valued jobs. Building on Gradín's (2013) approach to low-paying segregation, we construct a concentration curve that plots the cumulative proportion of workers of each group when occupations are ordered by earnings. We quantify the phenomenon using the dissimilarity and Gini concentration indices derived from this curve. We integrate, with a different rationalization, Blackburn and Jarman's (1997) index of vertical segregation, and deal with labor market stratification like other similar inequities are treated in the welfare literature (like tax redistribution or socioeconomic inequalities of health or education). We additionally use a counterfactual distribution in which men and women are compared using the same distribution of characteristics (e.g. education, age, etc.) to analyze the extent to which segregation or stratification can be explained by gender differences in the attributes that workers bring to the labor market. Using this approach with census data, we investigate the nature of long-term trends in segregation of women in the US labor market.

The next Section summarizes the conventional framework for measuring segregation, extended in Section 3 to measure stratification by sex. Section 4 introduces the conditional analysis. Section 5 reviews the literature about occupational gender segregation in the US, Section 6 describes the data, and Section 7 presents the empirical analysis. The last section summarizes the results.

2. Measuring segregation

2.1 Notation and basic definitions

Let us consider a two-group population of N workers distributed across $J > 1$ occupations. In what follows, subscript $j = 1, \dots, J$ refers to occupation. Superscript $i = c, r$ identifies each of the two groups, which for convenience we label the comparison group (i.e. women) and the reference group (i.e. men), each with N^i workers, $n_j^i \geq 0$ in the j th occupation. Also for convenience, occupations are indexed² in ascending values of the groups' ratio $\rho_j = n_j^r/n_j^c$, with $f^i = (f_1^i, \dots, f_J^i)$ being the vector of relative frequencies ($f_j^i = n_j^i/N^i$). The corresponding cumulative values are $F_j^i = \sum_{s=1}^j f_s^i$, which we use represent the step-function cdf $F^i(\rho)$.³

² The fact that occupations are ordered is relevant for the graphical representation of the segregation curve and some formulations of indices derived directly from it (e.g. Gini and Dissimilarity). However, the ranking of distributions according to the segregation curve is unaffected, and the indices can also be obtained from the distribution of employment with occupations sorted with any other criterion.

³ Thus, $F^i(\rho) = f_j^i$ if $\rho_j \leq \rho < \rho_{j+1}$, for $j = 0, \dots, J - 1$, and $F^i(\rho) = 1$ if $\rho \geq \rho_j$; with $\rho_0 = F_0^i = 0$.

A numerical measure of segregation $S(f^c, f^r)$ is a continuous function, for which a higher value indicates higher segregation. Occupational gender segregation is often seen as inequality in the distribution of workers of each sex across occupations, the extent to which their employment distributions differ from each other (they work in a different subset of occupations).⁴ A segregation index can be expressed as a measure $I_c(\rho)$ of inequality of the groups' ratio ρ among members of the comparison group – with $F^c(\rho)$ being the cdf (e.g. Silber, 1989).⁵ We thus may want S to verify equivalent properties to those usually required for any inequality index I_c . Therefore, the Lorenz curve that allows producing partial orderings of inequality income distributions based on a minimal set of value judgements has its correspondence in the segregation curve, while inequality indices can be used as segregation measures. We now summarize this approach, which we later adapt to introduce stratification of occupations.

2.2 Desirable properties of a measure of segregation and the segregation curve

Hutchens (1991, 2001, 2004) discussed the minimum set of properties or value judgments that are necessary to measure segregation in parallel with income inequality measurement:⁶ Segregation does not change after multiplying each n_j^i by the same positive scalar α^i (*Homogeneity*), after a permutation of people between occupations (*Symmetry in occupations*), or after a *proportional division* of an occupation.⁷ Segregation, however, increases after a *disequalizing movement* (*Principle of movements between occupations*).⁸ According to the first two properties, segregation only depends on the groups' proportions across occupations (f_j^i), not on their population sizes, N^i , or on other characteristics of occupations. The last two properties state that segregation is invariant to merging occupations with the same groups' ratio, and increases after moving workers to occupations with a higher proportion of their own group. There is an equivalence between the orderings obtained using indices verifying these

⁴ We refer here to the evenness approach to segregation (see Massey and Denton, 1988, for this and other alternative approaches) and to the case of two-group occupational segregation. The approach, though, can be extended to the general multigroup case (Silber, 1992; Chakravarty, D'Ambrosio, and Silber, 2009; Frankel and Volij, 2011; Reardon and Firebaugh, 2002; Alonso-Villar and Del Río, 2010).

⁵ For the relationship between inequality and segregation, see also James and Taeuber (1985); Butler (1987); Hutchens (1991); or Deutsch, Fluckiger and Silber (1994).

⁶ See also James and Taeuber (1985), Lasso de la Vega and Volij (2014), and Volij (2016). We present the set of properties according to the formulation in Hutchens (2004). In Hutchens (1991) the Insensitivity to Proportional Divisions was replaced by the assumption that the comparison group had equal shares across occupations ($f_j^c = f_t^c \forall j, t$).

⁷ A *proportional division* of an occupation j occurs when it is divided into M smaller occupations, each with the same ratio ρ_j .

⁸ A *disequalizing movement* of the reference group between occupations j and s ($j < s$) in (f^c, f^r) occurs if we obtain a new distribution $(f^{c'}, f^{r'})$, such that $\rho_j' = \rho_j - d$ and $\rho_s' = \rho_s + d$, for $0 < d \leq \rho_j$ and $\rho_t' = \rho_t, \forall t \neq j, s$.

properties and using the segregation curve (Hutchens, 1991, 2004; Lasso de la Vega and Volij, 2014).⁹

The segregation curve $F^r(p), p \in [0,1]$ is the continuous piece-wise function that connects, with linear segments, the cumulative proportions of workers for the comparison (F_j^c) and reference (F_j^r) groups across occupations (ordered by the sex ratio ρ_j).¹⁰ It can be seen as just a (first order) interdistributional Lorenz or discrimination curve in the interdistributional inequality framework (Butler and McDonald, 1987; Le Breton et al., 2012).¹¹ The segregation curve is non-decreasing and convex (with slopes f_j^r/f_j^c), and takes values between the 45° line (no segregation) and the abscissa (in the case of full segregation, jumping to 1 at $F^c = 1$). Like in the Lorenz case, if the segregation curves for two distributions do not intersect, we can unanimously rank them in terms of segregation by simply agreeing that a measure of segregation should exhibit just the four properties above. A distribution given by (f^c, f^r) dominates (has less segregation than) another one $(f^{c'}, f^{r'})$ if its segregation curve lies at no point below and at some point above the other: $F^r(p) \geq F^{r'}(p) \forall p \in [0,1]$, with strict inequality holding for some p . Segregation curves generate a partial ordering of distributions because whenever they intersect, we cannot judge which one exhibits more segregation without agreeing on additional properties. In this case, measures consistent with the segregation curve can produce different rankings of distributions because they implicitly incorporate additional properties needed to obtain a complete ranking (especially the degree of sensitivity of the index to disequalizing movements at different points of the distribution).

A segregation index might also verify other interesting properties, like the *range property* saying that it should take values between 0 (no segregation) and 1 (full segregation), or *symmetry in types* (Hutchens, 2004), requiring segregation not to change after exchanging the comparison and the reference groups. The latter, in particular, means that it is equivalent to speak of segregation of one group (women or men) with respect to the other or just sex segregation.¹²

⁹ This is similar to the correspondence between income inequality orderings obtained by Lorenz curves and by inequality indices verifying Scale invariance, Population Principle, Symmetry, and Pigou-Dalton transfer principle (e.g. Foster, 1985).

¹⁰ The cdfs are step functions, however. The segregation curve is given by $F^r(p) = F_j^r + \frac{p-F_j^c}{f_{j+1}^c} f_{j+1}^r$ if $F_j^c \leq p < F_{j+1}^c$, for $j = 0, \dots, J-1$, $F_0^c = 0$, and $F^r(1) = 1$. The origin of the segregation curve is not clear, but Duncan and Duncan (1955) provided the first known application.

¹¹ In which the (income) distribution of a comparison group is compared against the distribution of a reference group.

¹² The ranking between distributions produced by the segregation curve is symmetric in types (although the values of the curve change after the permutation of groups).

2.3 Main segregation indices

Most segregation indices are related to inequality measures. Among the most popular, the dissimilarity and Gini indices were first introduced in segregation analysis by Jahn, Schmid and Schrag (1947) and later popularized by Duncan and Duncan (1955).¹³

The *Dissimilarity* index, D , (Relative Mean Deviation or Pietra index of inequality)¹⁴ is half the sum of discrepancies in groups' shares across occupations, $D(f^c, f^r) = \frac{1}{2} \sum_{j=1}^J |f_j^c - f_j^r|$. In geometrical terms, it is the maximum vertical distance between the diagonal and the segregation curve¹⁵:

$$D(f^c, f^r) = \max_{j \in [1, J]} \{F_j^c - F_j^r\} \tag{1}$$

It is, thus, the difference in the cumulative proportion of both groups in the set of occupations in which one group is overrepresented (e.g. Hornseth, 1947): $D(f^c, f^r) = F_q^c - F_q^r$, where $q = \max_{j \in [1, J]} \{j \mid f_j^c \geq f_j^r\}$ is the critical occupation so that the comparison group is overrepresented below and underrepresented above. Therefore, its interpretation as the proportion of workers of each group that should change occupation to achieve full integration (moving from those in which their group is overrepresented to those in which it is underrepresented). After changing a $D\%$ of one group to remove segregation, the occupational distribution would be different to the original. For that reason, Karmel and MacLachlan (1988) formulated an alternative index: the proportion of people required to change occupation to eliminate segregation while preserving the occupational structure, $KM(f^c, f^r) = \sum_{j=1}^J \left| \frac{N^c n_j^r}{N} - \frac{N^r n_j^c}{N} \right|$. This is a scalar transformation of D :

$$KM(f^c, f^r) = 2 \frac{N^c N^r}{N} D(f^c, f^r) \tag{2}$$

¹³ For a discussion of segregation indices, see James and Taeuber (1985) or more recently, Silber (2012). Other inequality indices, such as the families of Generalized Entropy and Atkinson indices, cannot be written directly in terms of the segregation curve (only in terms of its slopes), an essential feature in our approach to measure stratification.

¹⁴ The paternity of this index of inequality is controversial. With initial contributions from Bortkiewicz, Bresciani-Turroni, Pietra, or Ricci, was later rediscovered several times with different names (see Krämer, 1998).

¹⁵ Alternatively, it also corresponds to the area of the maximum triangle that can be inserted between the diagonal and the segregation curve (divided by its maximum, $\frac{1}{2}$).

The *Gini* index corresponds to the homonym inequality index¹⁶ that, among many other expressions, can be computed as the area between the segregation curve and the diagonal (divided by its maximum, $\frac{1}{2}$), i.e. the weighted sum of the vertical distance between the segregation curve and the diagonal computed at the midpoints between adjacent occupations, \hat{F}_j^c .¹⁷

$$Gini(f^c, f^r) = 2 \sum_{j=1}^J (\hat{F}_j^c - \hat{F}_j^r) f_j^c = 1 - 2 \sum_{j=1}^J \hat{F}_j^r f_j^c \quad (3)$$

Where $\hat{F}_j^i = \frac{1}{2}(F_{j-1}^i + F_j^i)$ with mean equal to $\frac{1}{2}$.

D is the *Gini* between male and female dominated occupations (above and below q) and is insensitive to disequalizing movements within occupations dominated by one sex.¹⁸ *Gini* takes into account segregation between as well as within these two set of occupations, verifying all four basic properties, ranking distributions consistently with non-intersecting segregation curves. Both indices are symmetric in types and satisfy the range property. *KM* does not verify Homogeneity, the Principle of movements between occupations, and range.

3. Occupational stratification or segregation into low-paying occupations

3.1 Previous approaches

The importance of taking into account the information about the quality (e.g. average pay) of occupations held by each group in segregation has been present in several ways in the literature before, and received different names such as ordinal or vertical segregation. Blackburn and Jarman (1997) decomposed the Gini index of (overall) segregation into two orthogonal components in the Euclidean space. On the one hand, vertical segregation refers to the idea of inequality or social advantage and is measured using Somers' (1962) index of statistical association with occupations ordered by the vertical dimension (wage). Horizontal segregation, on the other hand, refers to the extent to which men and women are in different occupations without this giving an occupational advantage to either sex and is obtained indirectly using the

¹⁶ First introduced by Gini (1912, 1914), although the mean deviation in which the index is based had been previously used by German Astronomers in the late 1800s.

¹⁷ Note that \hat{F}^i are used instead of F^i (as in Lerman and Yitzhaki, 1989) because we connect the segregation curve by linear segments while define the cdfs as step functions. If the segregation curve were defined also as a step function, then F^i should be used instead (but for consistency the case of no segregation should also be represented with a step function). Lerman and Yitzhaki (1984) showed that *Gini* can be expressed as a covariance. In our case, the covariance between the groups' ratio and the cdf for group c : $Gini(f^c, f^r) = 2cov\left(\frac{f_j^r}{f_j^c}, \hat{F}_j^c\right) = 2 \sum_{j=1}^J (f_j^r - f_j^c) \hat{F}_j^c = 2 \sum_{j=1}^J f_j^r \hat{F}_j^c - 1$.

¹⁸ D verifies a weak version of the property instead (segregation does not decline after the disequalizing movement), and it will never rank two distributions in the reversed order in the case of dominance.

Euclidean norm.¹⁹ This approach has been used in several empirical studies (e.g. Bettio and Verashchagina's, 2009 report on gender segregation in the EU, or Blackburn, Racko, and Jarman, 2009).

Another line of research has proposed augmented indices of segregation that penalize by the concentration of one group into low-status occupations, such as Hutchens (2009, 2012). The latter characterized a generalization of the squared root index within a more general class of indices verifying a set of properties, and proposed a related dominance criterion. Similarly, Del Río and Alonso-Villar (2012) extended Alonso-Villar and Del Río's (2010) measures of local segregation in the multigroup context. Reardon (2009) used measures of ordinal segregation that can be interpreted as either relative ordinal variation (a measure of the difference in the ordinal variation of the population and the average ordinal variation within each unordered category) or as a weighted average of the binary segregation between those above and below each threshold of the ordered variable. On the other hand, Del Río and Alonso-Villar (2015, 2016) have proposed direct measures of the monetary or wellbeing losses associated with segregation of particular groups in the multigroup context.

In a different approach, Gradín (2013) studied the extent to which a particular group is segregated into low-paying occupations by comparing the proportion of both groups for any possible threshold defining low-pay. This implies the use of first-order stochastic dominance of the employment distribution ordered by occupation's (median) earnings. In what follows, we define our framework that develops this approach and integrates Blackburn and Jarman's (1997) measure of vertical segregation.

3.2 The concept

Let w be a measure of the quality of occupations in one or several relevant dimensions such as pay, prestige, skill level, etc, with w_j being its realization for occupation j . We are interested in knowing whether occupations are stratified with one group systematically holding those with lower quality. In the empirical application, w_j will represent the average earnings of workers in each occupation and thus we call this framework segregation of the comparison group (i.e. women) into low-paying occupations. The distribution of occupations is now given by g^i (with

¹⁹ Defining a right-angle triangle in which overall segregation is the hypotenuse, and vertical and horizontal segregation are the two catheti. Note, however, that the Gini index is based on the 1-norm or "city block distance", instead (e.g. Yitzhaki, 2013).

cdf $G^i(w)$), a permutation of f^i in which occupations have been reranked in increasing values of w (in fact, we do not need to observe w , but just the ranking).

We may say that the comparison group c is segregated into low-paying occupations compared with the reference group r , if for a given threshold z defining low pay, we find a larger proportion of workers from the former group working in low-paying occupations.²⁰ Given that any threshold would be arbitrary, we can extend this definition for a range of thresholds between 0 and a maximum reasonable \bar{z} (that could be the maximum average earnings). Group c is then said to be segregated into low-paying occupations compared with group r , if $G^c(z) \geq G^r(z), \forall z \in [0, \bar{z}]$ (with the strict inequality holding for some z). We can alternatively define the situation in which the comparison group is segregated into high-paying occupations and that in which the distribution of employment is pay-neutral, by just replacing \geq in the previous definition by \leq and $=$ respectively.

Let us consider the following example with $f^c = (.5, .5, 0, 0)$ and the $f^r = (0, 0, .5, .5)$. There is full segregation, because in each occupation we only find individuals of one group. Let $\bar{z} = 5$. If the corresponding earnings are given by $w_a = (1, 1, 1, 1)$ or $w_b = (1, 5, 1, 5)$, this segregation is pay-neutral, because the proportion of workers from both groups is the same below any low-pay threshold. Group c is segregated into low-paying occupations if $w_c = (1, 1, 5, 5)$ or $w_d = (1, 3, 2, 4)$, because it has a proportion of workers larger (or equal) below any low-pay threshold (r is segregated into high-paying occupations).

If segregation of group c into low-paying occupations holds over the entire range of w , there is first-order stochastic dominance (FOSD) of r over c (e.g. Bishop, Zeager, and Zheng, 2011). We can interpret g^i as a lottery for workers of group i entering the labor market; in welfare terms, all workers would always prefer to be of type r , regardless of their risk aversion, provided utility is non-decreasing in w . From previous results in poverty analysis (Foster and Shorrocks, 1988) FOSD also implies that group c would be more segregated into low-paying occupations for any head-count index that measures the proportion of each group below the threshold (and for all possible thresholds).²¹ If low-paying segregation holds only below a certain threshold, we have restricted FOSD, instead.

²⁰ In this line, IWPR (2015) estimated that 5.3 million women in the US in 2014 worked in occupations that had median earnings for full-time work below the federal poverty threshold for a family of four, compared with 3.6 million men.

²¹ If w is cardinal, FOSD implies dominance of higher order and thus the same holds for any other index that weighted each worker in a low-paying occupation by a function of the deficit to the threshold, in line

3.3 The low-paying segregation (or concentration) curve

The concentration curve $G^r(p), p \in [0,1]$ is the continuous piece-wise function that connects the cumulative proportions of workers for the comparison (G_j^c) and reference (G_j^r) groups across occupations. The difference with the segregation curve is that it uses occupations indexed by their earnings instead of by their sex' ratio.²² The properties of the concentration curve are well-known (e.g. Kakwani, 1980b; Lambert, 2002): it is non-decreasing, but not necessarily convex, and may fall above the diagonal. It has been used in other fields in economics, such as to measure horizontal inequality of taxes (e.g. Atkinson, 1980; Plotnick, 1981) or socioeconomic inequalities in the access to health care (e.g. Wagstaff, van Doorslaer, and Paci, 1989) or to education (e.g. Antoninis, Delprato, and Benavot, 2016).²³

Based on previous definition, group c is segregated into low-paying occupations if and only if the concentration curve falls below the diagonal in the corresponding range ($G_j^r \leq G_j^c, \forall w_j \leq \bar{z}$). Similarly, there is segregation into high-paying occupations when the curve falls above the diagonal. The concentration curve is bounded from below by the segregation curve and from above by its mirror image above the diagonal: $F^r(p) \leq G^r(p) \leq 1 - F^r(1 - p)$. These represent the cases in which earnings (w) and sex's ratios (ρ) produce the same and the inverted ranking of occupations respectively. Whenever the distribution of employment is pay-neutral, the concentration curve goes along the diagonal: $G^r(p) = p$.

For these reasons, the concentration curve can be interpreted, in our context, as the low-paying segregation curve of the comparison group. Like the segregation curve, it can be used to produce a partial ordering of distributions. Group c is more segregated into low-paying occupations whenever the corresponding concentration curves falls below. We need to use indices to quantify the phenomenon, and to obtain a complete ordering when the curves intersect.

with the Foster, Greer and Thorbecke's (1984) index of poverty. Indices of the *FGT* family would be in our context $S_\alpha(z; i) = \sum_{j|w_j < z} \left(\frac{z-w_j}{z}\right)^\alpha g_j^i, \alpha \geq 0, z \leq \bar{z}, i = c, r$.

²² A concentration curve is a generalization of the notion of the Lorenz curve in which the variable that is accumulated in the abscissa is not necessarily sorted according to the variable accumulated in the ordinate (a restriction intrinsic to the Lorenz curve that also applies to the segregation curve).

²³ In the former case, the concentration curve of post-tax income using pre-tax rankings was compared with the Lorenz curve of post-tax income. The latter used the concentration curve of the cumulative distribution of health (illness) for the population ranked by income.

3.4 Indices of low-paying segregation

We see segregation into low-paying occupations as a particular form of occupational segregation. It thus seems reasonable that the desirable properties are the same as before, adjusted to take into account w . In particular, we require Homogeneity, Insensitivity to proportional divisions, and the Principle of movements between occupations, after re-defining a proportional division to be pay-preserving, and a disequalizing movement in terms of the re-ranked distributions g^i (instead of f^i). Furthermore, we require low-paying segregation not to change after a permutation of people between occupations with the same w (Symmetry in pay-equivalent occupations). The ordering produced by the concentration curve verifies these four properties.²⁴

Regarding the Range property, it seems reasonable to require the index of low-paying segregation to be bounded in absolute terms by the level of observed segregation. This would be the magnitude of occupational stratification. A positive (negative) sign would then indicate whether the segregation of the comparison group is into low-paying (high-paying) occupations; with 0 representing pay neutral employment distribution. This means that the absolute maximum and minimum (when there is full segregation) are 1 and -1. Symmetry in types should be verified by the magnitude of stratification (the index in absolute value) not changing after exchanging the groups. The sign, however, should change according to the Range property, because if one group is segregated into low-paying occupations, the other one is segregated into high-paying ones.

It also seems natural to consider the Gini and dissimilarity concentration indices that can be derived from the concentration curve as candidates to measure low-paying segregation. These concentration indices can be defined re-writing the corresponding segregation indices in **(1)-(3)** with occupations indexed by earnings (i.e. replacing f^i by g^i).

The concentration $Gini(g^c, g^r)$ index measures twice the area between the diagonal and the concentration curve (summing the area below the diagonal and subtracting the area above). This is also the Somers' (1962) d_{YX} index of statistical association used as a measure of vertical segregation (Blackburn, Jarman, and Siltanen, 1994; Blackburn and Jarman, 1997), for which we

²⁴ In particular, note that a disequalizing movement produces FOSD: either $G^{r'}(w) \leq G^r(w)$ or $G^{c'}(w) \geq G^c(w)$, with strict inequality holding for $w_j \leq w < w_s$. A disequalizing movement defined on g^i implies a shift of the concentration curve to the right, like a disequalizing movement defined in f^i implies a shift in the same direction of the segregation curve.

provide an alternative rationalization within our framework.²⁵ The positive (negative) vertical distance between the diagonal and the concentration curve at occupation j ($G_j^c - G_j^r$) represents the proportion of workers of each group that should change occupation in order to eliminate segregation of group c into low-paying (high-paying) occupations for threshold $z = w_j$. A dissimilarity concentration index $D(g^c, g^r)$ may be then defined to be the maximum proportion of women (men) that should move to a higher (lower) paying occupation so to remove women's segregation into low-paying occupations, for any possible threshold z . Correspondingly, we obtain the concentration version of $KM(g^c, g^r)$ as a scalar transformation of the previous one.²⁶

These concentration indices inherit the properties of their segregation counterparts. $Gini(g^c, g^r)$ verifies the four basic properties defined above, plus range²⁷ and symmetry in types, and will rank distributions consistently with non-intersecting concentration curves. Of these properties, $D(g^c, g^r)$ violates the Principle of movements between occupations. $KM(g^c, g^r)$ violates this property, as well as homogeneity and range.

For each concentration index $S(g^c, g^r)$, we may define a concentration ratio r_S as the proportion of observed segregation of group c that is low-paying (or high-paying), by normalizing it by its maximum value ($S(f^c, f^r)$). The sign still indicates whether segregation of c is into low- or high-paying occupations:²⁸

$$r_S = \frac{S(g^c, g^r)}{S(f^c, f^r)}, \text{ where } -1 \leq r_S \leq 1; \quad S = Gini, D, KM. \quad (4)$$

²⁵ Let Y^s be an outcome variable if $X = s$ ($s = 0,1$), where X is a dummy for group r membership (male sex). If we randomly draw a pair of workers from the population, the index, based on Kedall's tau, indicates the difference between the probability of concordance (a higher value of Y corresponds with the higher value of X) and discordance (the opposite): $d_{YX} = \Pr(Y^1 > Y^0) - \Pr(Y^0 > Y^1)$. We obtain $Gini(f^c, f^r)$ and $Gini(g^c, g^r)$ when Y is ρ and w , respectively.

²⁶ $D(g^c, g^r)$ is in line with Pfähler's (1983) "maximum proportionalization percentage" index computed for the concentration curve of taxes with respect to income.

²⁷ Note that the concentration curve may cross the diagonal, generating positive and negative areas, if c is segregated into both low and high-paying occupations. In these cases, it might be useful to compute the index separately for different intervals of w .

²⁸ Plotnick (1981), in line with Atkinson (1980), constructed a measure of horizontal inequity by normalizing the area between the concentration (ranked by pre-tax income) and Lorenz curves of post-tax income ($A = \frac{1}{2}(Gini(f^c, f^r) - Gini(g^c, g^r))$) by its maximum value: $0 \leq \frac{A}{Gini(f^c, f^r)} = \frac{1-rGini}{2} \leq 1$. Aronson, Johnson and Lambert (1994) regarded twice this area as the reranking component of the decomposition of the redistributive effect of taxes (the difference between Gini in the pre and post-tax income distributions); the other two components are vertical and horizontal redistribution. Similarly, Wagstaff, van Doorslaer, and Paci (1989) used a health/illness concentration index to measure equity in the access to health.

In particular, $r_D = r_{KM}$ and $r_{Gini} = \Gamma_c(\rho, w)$, the *Gini correlation coefficient* between groups' ratio (ρ) and earnings (w) across occupations, computed among members of group c .²⁹ The fact that we measure the extent to which segregation of occupations implies stratification as a correlation coefficient does not come as a surprise, because the difference between both approaches is entirely due to a re-ranking of occupations.

4. Measuring conditional segregation

To some extent, the observed level of sex segregation or stratification of occupations could be the result of the distribution of some relevant characteristics differing across population groups (explained or compositional effect). These characteristics include workers' human capital (i.e. education experience, etc.), demographic factors (e.g. immigration or marital status), geographical location, that potentially affect their opportunities in the local labor market. Alternatively, it could also reflect intrinsic segregation when people of similar characteristics work in different occupations depending on their sex. The identification of this unexplained term with discrimination in the labor market has to be cautious, like in the analysis of wage differentials, because it also may reflect differences in unobserved characteristics (e.g. job preferences, skills). Similarly, the explained part could also reflect anticipated discrimination in the labor market by the discriminated group, or discrimination that occurs prior to entering the labor market.

To disentangle the importance of the explained and unexplained terms, we follow here the approach in Gradín (2013). The aggregate decomposition is obtained comparing observed segregation with that using the counterfactual distribution in which members of the comparison group are reweighted using propensity score to have the reference group's distribution of characteristics (based on DiNardo et al., 1996).³⁰ The detailed contribution of each characteristic to the explained effect is obtained using Shapley decomposition (based on Gradín, 2014).

Let us assume that the probability that workers from group i work in occupation j , f_j^i , is a function of their characteristics X , with domain Ω_X . f_j^i can be thus expressed as the product of the conditional probability of type i workers with a specific combination x of characteristics,

²⁹ This correlation index is based on the Gini covariance, which relates a cardinal variable with the rank of another (a mixture of Pearson's and Spearman's correlations, see Schechtman and Yitzhaki, 1987, 1999, and Yitzhaki and Olkin, 1991).

³⁰ Alternatively, we can give the reference's conditional occupational distribution to the comparison group by reweighting the reference group to reproduce the comparison's characteristics.

$f_j^i(X = x)$, and the marginal probability of occurrence of x in group i , $f^i(x)$, summed up over all possible x :

$$f_j^i(X) = \int_{x \in \Omega_x} f_j^i(X = x) f^i(x) dx \quad (5)$$

Assuming that $f_j^i(X = x)$ does not depend on the distribution of X , we define $f_j^y(X)$ to be the counterfactual share of workers from c in occupation j when they keep their own conditional distribution $f_j^c(X = x)$ but have the marginal distribution of characteristics in r , $f^r(x)$:

$$f_j^y = \int_{x \in \Omega_x} f_j^c(X = x) f^r(x) dx = \int_{x \in \Omega_z} f_j^c(X = x) f^c(x) \Psi_X dx; \text{ where } \Psi_X = \frac{f^r(x)}{f^c(x)}. \quad (6)$$

f_j^y can be obtained by reweighting the f_j^c with the factor Ψ_X , i.e. the relative marginal probability of x in both groups. From Bayes theorem we know that $f^i(x) \equiv \Pr(x|i) = \frac{\Pr(i|x) \Pr(x)}{\Pr(i)}$, where $\Pr(i|x)$ is the probability that a worker with characteristics x belongs to group i , $\Pr(x)$ is the probability of having characteristics x regardless of group membership, and $\Pr(i) = N^i/N$ is the probability of group i membership. Thus:

$$\Psi_X = \frac{f^r(x)}{f^c(x)} = \frac{\Pr(c) \Pr(r|x)}{\Pr(r) \Pr(c|x)} = \frac{N^c \Pr(r|x)}{N^r \Pr(c|x)}. \quad (7)$$

The reweighting factor Ψ_X depends on the unconditional and conditional relative probabilities of group membership. In the pooled sample, we can estimate the former (a constant) using the observed population shares, and the latter with a logit model for the probability of being r conditional on x : $\Pr(r|x) = \frac{\exp(x\beta)}{1+\exp(x\beta)}$, where β are the coefficients, and $\Pr(c|x) = 1 - \Pr(r|x)$.³¹ $S(f^y, f^r)$ is the level of conditional segregation in which both groups are compared with the same distribution of characteristics. The unconditional level $S(f^c, f^r)$ may be then written as:

$$S(f^c, f^r) = [S(f^c, f^r) - S(f^y, f^r)] + S(f^y, f^r). \quad (8)$$

Where the first term (explained compositional effect) is the level of segregation explained by both population groups having different distributions of characteristics (shifting from f^c to f^y); and the second one is the (conditional) segregation that remains unexplained after equalizing

³¹ Characteristics must have a common support (both groups overlap across their different values), avoiding cases with $\Pr(i = r|x)$ close to 1, that would have a disproportional influence on the results.

characteristics in both groups, reflecting only cross-group differences in the conditional distribution across occupations (shifting from f^y to f^r ; note that $S(f^r, f^r) = 0$).

A detailed decomposition of the explained term indicates the contribution of each factor. There is no unique solution given the non-linear nature of the approach.³² Starting with the case in which all estimated coefficients in the logit regression are set to zero, several reweighting factors were obtained sequentially switching the coefficients of each factor to its estimated value. The change in segregation after each set of coefficients were switched on is a measure of the contribution of each factor in that sequence. The final contribution is obtained averaging over all possible sequences (i.e. Shapley decomposition as in Chantreuil and Trannoy, 2013, and Shorrocks, 2013). This approach overcomes two well-known problems in the original DiNardo et al. (1996) approach (omitted-variable bias and path dependence).³³

We use the same approach for the decomposition of $S(g^c, g^r)$ and a similar one to decompose changes over time. Let f^{it} be the unconditional occupational distribution of group i in year t ($t = 0,1$), and f_j^{iy} the counterfactual for group i that uses its marginal distribution of characteristics in year 1, and its conditional occupational distribution in year 0:

$$f_j^{iy} = \int_{x \in \Omega_z} f_j^{i0}(X = x) f^{i0}(x) \varphi_x dx; \text{ where } \varphi_x = \frac{f^{i1}(x)}{f^{i0}(x)}.$$

Then we can decompose the total change in segregation over time as:

$$\Delta S(f) = S(f^{c1}, f^{r1}) - S(f^{c0}, f^{r0}) = \Delta S^E(f) + \Delta S^U(f).$$

Where the first term is the change in segregation associated with a change in characteristics of both population groups over time, evaluated using the conditional occupational distribution in year 0:

$$\Delta S^E(f) = [S(f^{cy}, f^{ry}) - S(f^{c0}, f^{r0})].$$

The second term is the unexplained effect, i.e. the change in conditional occupational distributions of both groups over time, evaluated using each sex's characteristics in year 1:

$$\Delta S^U(f) = [S(f^{c1}, f^{r1}) - S(f^{cy}, f^{ry})].$$

³² The approach does not allow for decomposing the unexplained effect.

³³ This consisted in estimating a series of logit regressions in which independent variables accounting for each factor were added sequentially. The difference in segregation using the reweighting factors obtained from two consecutive regressions would reflect the contribution of the factor included at that stage.

In the case of low-pay segregation, there is a third factor to consider, the change in the ranking of occupations over time. Let h^{it} be the employment distribution of group i at year t across occupations indexed by w^* , a common reference earnings distribution by occupations, and h_j^{iy} be the corresponding counterfactual that uses the marginal distribution of characteristics in year 1 and the conditional occupational distribution in year 0. We decompose the change in the concentration index over time as:

$$\Delta S(g) = S(g^{c1}, g^{r1}) - S(g^{c0}, g^{r0}) = \Delta S^w + \Delta S(h) = \Delta S^w + [\Delta S^E(h) + \Delta S^U(h)].$$

Where ΔS^w is the earnings structure effect, the change in segregation associated with a change in the ranking of occupations (from w^0 and w^1 to w^*). ΔS^E is the effect of the change in characteristics of both groups over time (evaluated using w^* and the conditional occupational distributions in year 0). ΔS^U is the unexplained effect, the result of the change in conditional occupational distributions over time (evaluated using w^* and characteristics at year 1):

$$\Delta S^w = [S(g^{c1}, g^{r1}) - S(h^{c1}, h^{r1})] - [S(g^{c0}, g^{r0}) - S(h^{c0}, h^{r0})].$$

$$\Delta S(h) = S(h^{c1}, h^{r1}) - S(h^{c0}, h^{r0}) = \Delta S^E(h) + \Delta S^U(h)$$

$$\Delta S^E(h) = S(h^{cy}, h^{ry}) - S(h^{c0}, h^{r0}).$$

$$\Delta S^U(h) = S(h^{c1}, h^{r1}) - S(h^{cy}, h^{ry}).$$

We obtain the detailed decomposition of ΔS^E and ΔS^U using the same Shapley decomposition described above. Finally, bootstrapping over the entire process (including the logit regression) will produce standard errors.

5. Occupational segregation in the US

5.1 The gender pay gap and occupational segregation

A growing empirical literature has documented the existence of an important pay gap by gender all over the world. This gap has been traditionally explained as the result of gender differences in productivity and labor market discrimination. The former is the consequence of the different amount and content of human capital (i.e. education, on-the-job training, and experience) accumulated by men and women, and of the particular occupations and establishments in which they work (segregation). Labor market discrimination arises when women and men receive a different treatment in the hiring process, pay, promotion, etc.

Disentangling the relevance of each factor explaining the gender pay gap has proved to be a difficult task. On the one hand, some relevant workers' characteristics might not be observed and remain unexplained. For example, there is a great debate about gender differences in non-cognitive skills, some databases do not provide information about the field of the college major or the establishment, and the occupational classifications tend to be broader than real jobs, thus hiding part of segregation. On the other hand, the main factors are strongly inter-related. Occupational and educational choices, as well as career interruptions might be voluntary if they result of gender differences in preferences. But these decisions might be also made facing or anticipating a discriminatory pressure of any sort from inside or outside the labor market. That is, discrimination could affect the distribution of productivity-related characteristics and job characteristics as well as the unexplained portion of the pay gap, as Gunderson (1989, p 48), among others, noted: "For example, women may have higher turnover and, absenteeism because they are assigned to low-wage, dead-end jobs or because they bear a disproportionate burden of household responsibility. Or women may not enroll in certain education programs, even if they have an aptitude for them, because they perceive that opportunities to use such training will be closed to them in the labor market".

There is, however, a consensus that the portion of the gender pay gap that can be explained by differences in characteristics has largely been reduced if not eliminated as the result of what Goldin (2014) called the converging roles of men and women during the last century, one of the main accomplishments of society. This implied a narrowing between men and women in family roles, in qualifications and in several labor market outcomes, such as participation, paid hours of work, accumulated experience, occupations, and earnings.

The role of occupational segregation in determining the gender pay gap has been deeply analyzed in this context. As pointed by Treiman and Hartmann (1981), given that explanations focusing on the characteristics of individual workers left a substantial portion of the earnings gap unexplained, several studies attempted to explain them focusing on the characteristics of the jobs, because predominantly female jobs tend to pay less. As these authors also noted, the more finely disaggregated an occupational classification is, the larger is the proportion of the total difference in earnings that can be attributed to the segregation of men and women in the labor force. This lead several studies to include detailed job classifications considering also the establishment whenever this information was available. Gunderson (1989) showed in his review of the literature that segregation of men and women across occupations and establishments accounted for a substantial portion of the overall earnings gap, while pay differences for the

same narrowly defined occupation within the same establishment do not account for much of the pay gap. Later, Bayard et al. (2003) using a more representative matched employee-employer dataset showed that about a half of the gender pay gap took the form of wage differences between men and women within narrowly defined occupations within establishments.

Clearly, the upgrading of occupations held by women has substantially decreased the relevance of segregation to explain the pay gap, helping to explain its decline since 1970 in the US (e.g. Blau and Kahn, 2000). Goldin (2014) has recently noted that if the pay gap persists, is due to the fact that the majority of the current earnings gap comes from within occupation differences in earnings rather than from between occupation differences. To prove that, she showed that the inclusion of a full set of three-digit occupation dummies in a log earnings regression decreases the coefficient on female by no more than one-third. Additionally, she also showed that equalizing earnings by gender within each occupation reduced the aggregate gender pay gap far more than equalizing the proportions by each occupation instead. She isolated the factors that make for more equal pay within occupations: the value placed on the hours and job continuity of workers, including the self-employed. The gender gap tends to be lower when earnings are linear with respect to time worked and larger when it is convex. In sum: "The gender gap in pay would be considerably reduced and might vanish altogether if firms did not have an incentive to disproportionately reward individuals who labored long hours and worked particular hours. Such change has taken off in various sectors, such as technology, science, and health, but is less apparent in the corporate, financial, and legal worlds."³⁴ However, Blaun and Khan (2016) have recently noted that the women's segregation by occupation and industry still explains a significant share of the pay gap, more than psychological attributes or non-cognitive skills, factors that have received considerable attention recently.

5.2. Trends in occupational segregation in the US

Regardless of its impact on the pay gap, occupational segregation by sex in the US has been the focus of a large empirical literature (using the dissimilarity index). A growing consensus emerged

³⁴ Compared with business occupations, technology and science have greater time flexibility, fewer client and worker contacts, fewer working relationships with others, more independence in determining tasks, and more specific projects with less discretion over them. Each of these characteristics should produce a more linear relationship between hours and earnings and the greater linearity should produce a lower residual difference in earnings by sex. The role of the overwork effect on trends in the gender gap in wages in professional and managerial occupations has also been emphasized by Mandel and Semyonov (2014).

about the general trend and the main driving forces (e.g. Beller, 1985; Blau, Simpson and Anderson, 1998; Cotter, Hermsen, and Vanneman, 2004; Blau, Brummund, and Liu, 2013; Mandel and Semyonov, 2014). Segregation declined since 1960, this reduction accelerated during the 1970s and continued at a slower pace until it practically stalled during the 2000s. Regarding the way segregation was reduced, most studies point at a higher integration of both sexes within occupations, as opposed to changes in the mix of occupations, although the more rapid growth of integrated occupations also contributed, especially explaining the little improvement after 1990. The higher integration was the result of women entering previously male white-collar and service occupations, especially executive and managerial occupations, rather than women entering blue-collar jobs, or men entering predominantly female occupations. As a result, several occupations shifted from being male dominated to being integrated. The main driving factors were “the entry of new cohorts of women, presumably better prepared than their predecessors and/or encountering less labor market discrimination” (Blau, Brummund, and Liu, 2013), although occupational segregation also decreased within cohorts. The laws on equal employment opportunity may have had an effect, and the larger improvement in women’s education played an important role in explaining these trends because it allowed them to enter high-skilled jobs. The largest decrease in segregation was among college graduates, with very little change in segregation among high school dropouts (e.g. Blau, Brummund, and Liu, 2013).

The downside of this trend in segregation is that it stalled during the 2000s at a still high level, and there is in fact a risk of re-segregation if women continue to enter the same occupations (e.g. Blau, Simpson, and Anderson, 1998). Furthermore, workers are often segregated also at the firm level and clustered at the lower level of hierarchies within occupations, something that is usually ignored in segregation measured using census data (e.g. Blau, Simpson, and Anderson, 1998). Additionally, Mandel and Semyonov (2014) highlighted that while declining gender segregation over time may explain the lesser impact of occupations on the gender pay gap in the private sector, in the public sector, by contrast, gender segregation accounts for a greater portion of the gap in absolute as well as relative terms.

Less consensus exists about the causes of segregation. As Blau and Kahn (2000) noted, economic analyses of occupational segregation, like in the case of the pay gap, have focused on gender differences in preferences and qualifications, as well as in labor market discrimination, including historical institutional barriers, gender differences in promotion rates (helping to explain the glass-ceiling hypothesis) or on-the-job training.

6. Data

The empirical analysis is based on microdata samples extracted from censuses conducted by the US Census Bureau between 1960 and 2000, representing 5% of the country's population (except 1% in 1970), and from the annual American Community Survey (ACS) conducted between 2001 and 2014 (about 0.4% of each year's population in 2001-2004, 1% thereafter). We used the Integrated Public Use Microdata Series (IPUMS-USA) harmonized by the Minnesota Population Center at the University of Minnesota (Ruggles et al., 2015). The use of census data guarantees larger samples to analyze segregation across a more detailed classification of occupations. Our sample consists of all workers employed during the reference week.

We analyze the distribution of employment by gender using the IPUMS-USA modified version of the 2010 Census Bureau occupational classification scheme (453 categories after excluding armed forces) that offers a highly consistent classification of occupations over the 1960-2014 period.³⁵ For the analysis of stratification, we rank occupations according to their average earned income obtained during the year of reference (previous calendar year in censuses, past 12 months in ACS).³⁶

In the analysis of conditional segregation in 2014, we use a detailed set of workers' characteristics that might influence their occupation. The most important is attained education, for which we use 24 census categories, from no schooling completed to doctorate degree; for those with college education we distinguished the field of degree in a detailed format.³⁷ Among the other factors, location includes metropolitan statistical area, with one category for non-metropolitan areas. Demographic variables include marital status, age interval, number of children in the household (under and above age 5), race and ethnicity, and migration profile (place of birth, change of residence, years of residence in the US, citizenship, and English speaking proficiency).³⁸ We have also used a more restricted set of information common across

³⁵ The classification is also valid for 1950, excluded here because the information about earnings refers to a small sample of individuals. The exact number of non-empty occupations varies by year: 216 (1960), 282 (1970), 351 (1980-90), 447-449 (2000-11), 446 (2012-14). Results are highly robust to the use of the 1950 or 1990 classifications instead.

³⁶ Earned income includes wage, and self-employment income (from businesses and farms) calculated from midpoints of intervals before 1990, exact amounts in 1990, and rounded amounts ever since –with capped top incomes.

³⁷ 169 categories, after merging 4 fields with few female observations: Military technologies and mining and mineral, naval architecture, and nuclear engineering.

³⁸ Marital status: married with spouse present; married with spouse absent; separated; divorced; widowed; never married/single. Age: under 24, 25-34, 35-44, 45-54, 55-64, 65 or older. Race: white; black; American Indian or Alaska Native; Chinese; Japanese; other Asian or Pacific Islander; other race). Ethnicity: not Hispanic; Mexican; Puerto Rican; Cuban; etc. Place of birth (State if US-born; country/region if foreign-born). Change of residence: changed residence during the last year within state; between contiguous

samples for the sake of comparability over time.³⁹ In the analysis of trends over time, we omitted location due to difficulties to match different years.

7. Segregation and stratification of occupations by sex

7.1 Unconditional segregation

Occupational segregation of women in the US continuously declined between 1960 and 2014, as shown in Figure 1. Segregation was reduced by about 20% (18% with *Gini*; 23% with *D*), with highest intensity in the 1970-90 period but with little progress ever since. The US labor market thus remains highly segregated by sex, with a *Gini* = 0.660 in 2014; three quarters of it being segregation between occupations dominated by each sex (*D* = 0.495, the most common measure). The labor market was also highly stratified by gender in 1960, with women working in low-paying occupations. This low-pay segregation of women increased during the 1960s (by 9% with *Gini*, 13% with *D*), followed by a sharp reduction since 1970 (67% and 69%), that was more intense during the 1980s and the 2000s, slowing down only after 2010. The current levels are *Gini* = 0.174 and *D* = 0.145. Thus, the decline in stratification started later than segregation, was much more intense, and continued when the latter stagnated after 1990.

The concentration ratios, measuring the proportion of segregation that involves stratification by sex, declined from their peak in 1970 (67% with *Gini*; 72% with *D*) to their lowest levels in the 2010s (26% and 29%), providing a quantification of the outstanding change in the nature of women's segregation, regardless of its level. From the interpretation of the dissimilarity indices, we know that it would be necessary to move half of women (49.5%) from their current female dominated occupations to those predominantly masculine to eliminate segregation by sex. To remove their low-paying segregation instead (for any possible threshold), we would need to move one out of seven women to relatively higher-paying occupations (14.5%, i.e. 29% of 49.5%).

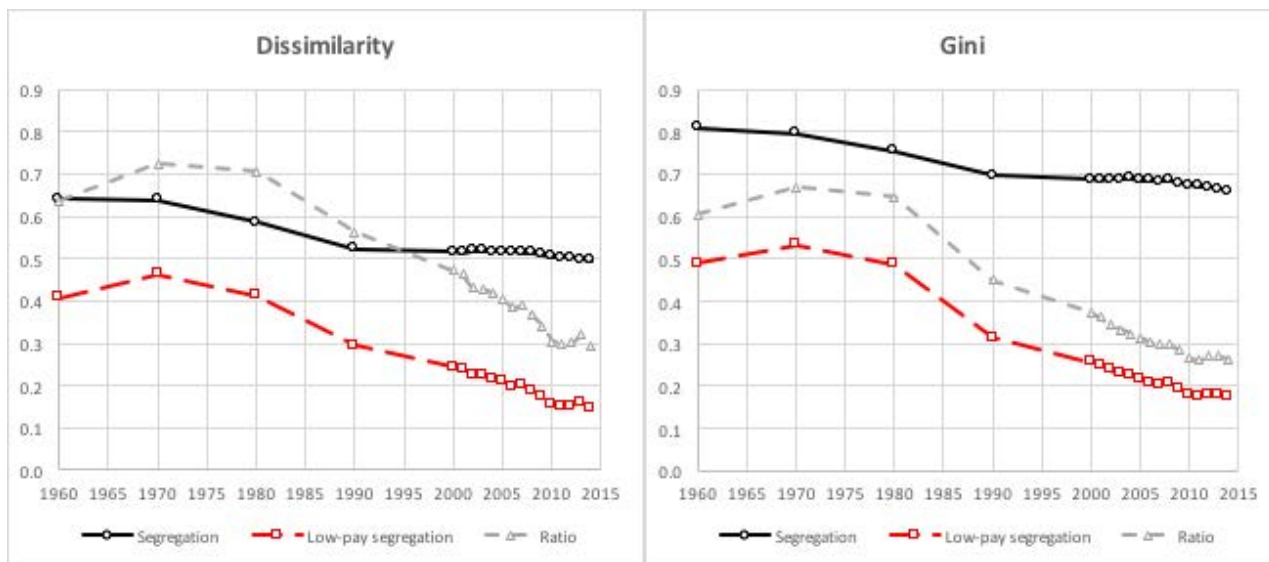
The concentration curves entirely falling below the diagonal in 1970 and 2014 (Figure 2) indicate that women were unambiguously concentrated into low-paying jobs (first-order stochastic dominance) in both years. This is so regardless of the threshold used to define low paying (and of the FGT index used to aggregate the earnings gaps of women for each of those thresholds).

states; between non-contiguous states; abroad; or remained in the same house. Years of residence (up to 5; 6-10; 11-15; 16-20; more than 20; native-born). Citizenship (national or foreign). English speaking proficiency (speaks only English; speaks English very well; well; not well; does not speak English)..

³⁹ This omits the field of college degree (only available after 2010), English speaking proficiency, and migration status. In 1960, we replaced the years of residence in the US for the change of residence during the previous 5 years. Hispanic origin was imputed by IPUMS before 1980.

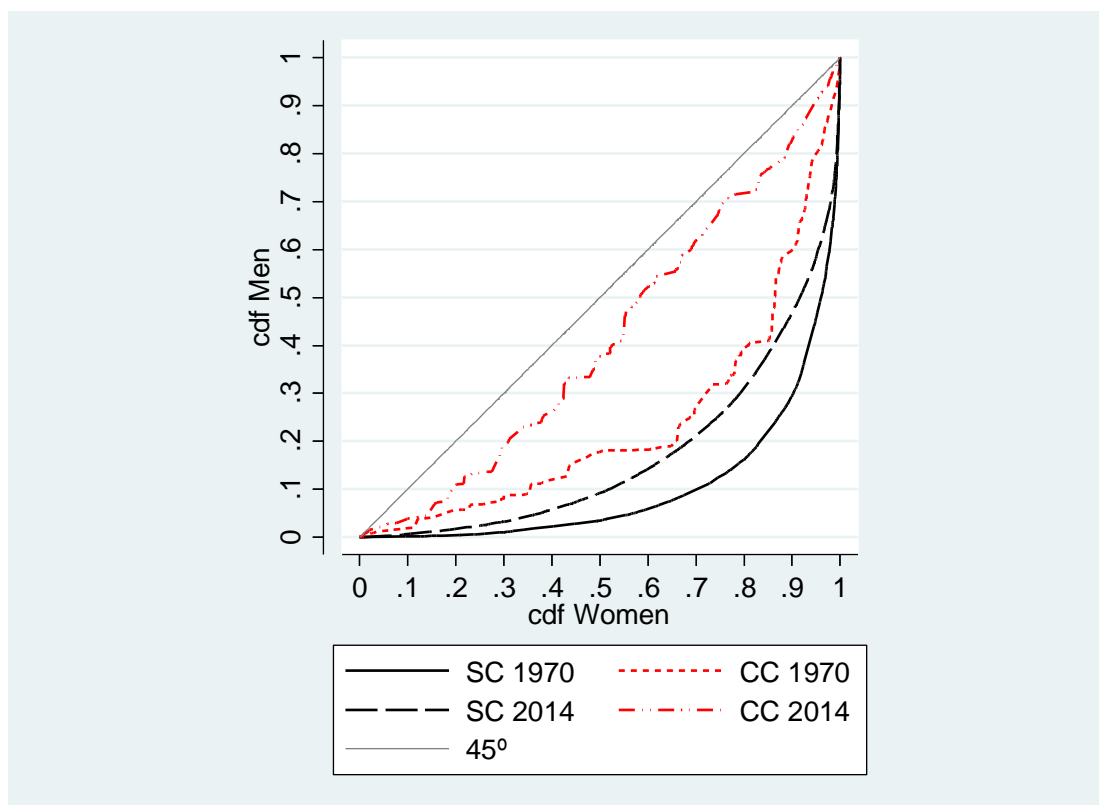
The reductions in both segregation and stratification of occupations by sex over time are robust to all indices consistent with the segregation and concentration curves (of which Gini is a particular case), because the corresponding curves for the latest year dominate those of the earliest.

Figure 1. Women’s (low pay) segregation in the US



Source: Own construction based on microdata from Census and ACS (IPUMS-USA).

Figure 2. Segregation (SC) and Low-pay Segregation (CC) curves by sex in the US



Source: Own construction based on microdata from Census and ACS (IPUMS-USA).

7.2 Explaining segregation levels

Occupational sex segregation in 2014 was the consequence of men and women of similar characteristics working in different occupations (conditional or unexplained segregation in Table 1) rather than the result of a compositional effect by workers' characteristics. Only 10% or less of segregation vanishes after giving women men's distribution of characteristics: 7% with *Gini* (from 0.660 to 0.614), 10% with *D* (from 0.495 to 0.444). This small explained effect is almost entirely driven by gender differences in education (explaining 6% with *Gini* and 9% with *D*), the main determinant that affects the set of occupations available for workers of each sex. Women have more college graduates than men, and different fields of degree.⁴⁰ The latter explains more of sex segregation because the contribution of education to segregation would be only 1% and 2% respectively had we omitted the field of degree in the estimation. The contribution of location and demographic characteristics to explain segregation is almost negligible.⁴¹

Gender differences in observable characteristics altogether do not explain why women are segregated into low-paying occupations, either. The overall explained effect is in fact negative, indicating that based on their current characteristics women should work in higher-paying occupations. These explained effects (-1.2% with *Gini*, -14% with *D*) are, however, the net result of large counterbalancing forces. On the one hand, women's educational mix largely reduce their segregation into low-paying jobs (-15% and -26%). This is due to the higher proportion of women with a college degree, although curbed by their lower rate of doctorates and higher specialization into disciplines with lower average earnings. Indeed, the contribution of education would be larger in this case (-31/32%) had we omitted the field of degree. Women's lower immigration rates also reduce (-3%) their low-paying segregation. On the other hand, women's age profile (9% and 8%) and higher rate of unmarried workers help to explain their segregation into low-paying occupations to some extent (7% and 6%).⁴²

The use of the alternative counterfactual in which we give women men's conditional occupational distribution (reweighting men according to women's characteristics), would raise similar results (Table A3 in the appendix). The explained effects would be just a little higher in the case of segregation (10.5% for *Gini* and 13% for *D*) due to a higher contribution of education

⁴⁰ For attained education, see Table A1 in the appendix. Regarding the field of college degree, female workers are strongly overrepresented in health and education services and underrepresented in engineering or business (see Table A2 in the appendix).

⁴¹ Compared with men, female workers are less likely to be white, Mexican or recent immigrants, and have children under age 5 (but are more likely to have them above that age).

⁴² Women are underrepresented amongst 25-44 year-old workers and among those above 65, and overrepresented among divorced/widowed/separated workers.

(10% and 13%). It would be a little smaller, however, in the case of in low-pay segregation (-0.6% for *Gini* and -9% for *D*) due to a smaller negative contribution of education (-10% and -17%).

Even if workers' characteristics do not explain much of today's segregation or stratification by sex, they could have played a more significant role in the past. Nevertheless, the proportion of segregation explained by characteristics was small in any studied year (Table A4 in the appendix; using information that, among other things, excludes the field of degree for the sake of comparability). The contribution of characteristics to explain women's low-paying segregation was negative in 1960, small in the 1970s and 1980s, and became negative again, with an increasing magnitude, since 2000. The large change in characteristics by sex over time, however, played a more significant role when it comes to explain the trends (instead of the levels) in segregation and stratification, as discussed in the next subsection.⁴³

⁴³ Note that Table A4 is not useful for this, because the contribution of a characteristic over time depends on the contemporary distribution of the characteristics by sex, but also on the contemporary conditional employment distribution by occupations, as well as the corresponding ranking of occupations (in the case of low-paying segregation).

Table 1. Decomposition of (low-paying) segregation of women by characteristics, 2014

		Segregation				Low-Paying Segregation			
		Gini	%	D	%	Gini	%	D	%
Observed		0.6604	100	0.4947	100	0.1737	100	0.1448	100
		(0.0009)		(0.0009)		(0.0012)		(0.0010)	
Unexplained		0.6136	92.9	0.4442	89.8	0.1757	101.2	0.1652	114.1
		(0.0010)		(0.0011)		(0.0011)		(0.0010)	
Explained		0.0468	7.1	0.0505	10.2	-0.0020	-1.2	-0.0204	-14.1
		(0.0005)		(0.0006)		(0.0009)		(0.0007)	
	Location	-0.0001	0.0	-0.0002	0.0	-0.0011	-0.7	-0.0007	-0.5
		(0.0001)		(0.0001)		(0.0002)		(0.0001)	
	Marital Status	0.0003	0.0	0.0005	0.1	0.0127	7.3	0.0086	6.0
		(0.0002)		(0.0002)		(0.0004)		(0.0003)	
	N Children	0.0007	0.1	-0.0005	-0.1	-0.0009	-0.5	-0.0011	-0.8
		(0.0001)		(0.0002)		(0.0002)		(0.0002)	
	Age	0.0023	0.4	0.0034	0.7	0.0161	9.3	0.0120	8.3
		(0.0001)		(0.0002)		(0.0005)		(0.0003)	
	Race	-0.0002	0.0	-0.0004	-0.1	0.0021	1.2	0.0017	1.2
		(0.0001)		(0.0001)		(0.0003)		(0.0002)	
	Hispanic ethnicity	0.0001	0.0	0.0001	0.0	0.0005	0.3	0.0004	0.3
		(0.0000)		(0.0000)		(0.0005)		(0.0004)	
	Migration profile	0.0012	0.2	0.0014	0.3	-0.0052	-3.0	-0.0039	-2.7
		(0.0002)		(0.0002)		(0.0005)		(0.0004)	
	Education	0.0425	6.4	0.0462	9.3	-0.0262	-15.1	-0.0374	-25.8
		(0.0004)		(0.0005)		(0.0010)		(0.0007)	

Notes: Counterfactual: women's 2014 distribution reweighted to reproduce men's 2014 characteristics. Bootstrap standard errors in parentheses (200 replications). See data section for details about the variables used in each category.

Source: Own construction based on ACS 2014 (IPUMS-USA).

7.3 Explaining trends in segregation

There was a profound change in the distribution of workers' characteristics by sex over time, the combined result of upgrading education of women relative to men, increasing female labor market participation, or trends affecting fertility, marriage, or immigration, among other things.⁴⁴ We first quantify in Table 2 the extent to which this changing composition of workers by sex helps to explain the reduction in segregation and stratification. For that, we use two counterfactuals. The first one gives workers in the initial year (1960 or 1970) the characteristics of their own sex in 2014; the other one gives workers in 2014 the characteristics of their sex in the initial year. In both cases, workers keep their original conditional occupational distributions. Thus, the explained effect is the reduction of segregation driven by the change in characteristics using each sex's initial (or final) conditional occupational distributions. The unexplained

⁴⁴ See a summary in Table A1 in the appendix.

component is the effect of the change in the conditional occupational distributions evaluated using either initial or final distribution of characteristics by sex. The choice of the reference year turned out to be crucial as a natural result of the important structural changes in both characteristics and conditional employment distributions. We focus on the *Gini* index for simplicity, although the results are not much different for *D*.

The magnitude of the extent to which the reduction in segregation between 1960 and 2014 can be explained by the change in worker's characteristics is only 20% using the 1960 conditional occupational distribution (the reference year is 2014), but reaches 58% using that in 2014 instead (the reference is 1960). In the first case, the largest contributions come from changes in marital status (7%), education (5%), and the number of children (4%). In the second case, education outstands with a 45% contribution, followed by marital status (20%), race (8%), and the number of children (4%), partially offset by negative contributions of changes in age (-12%), ethnicity, and migration (about-3% each). The rest of the reduction in segregation is the unexplained term, i.e. the effect of the change in the conditional occupational distributions by sex, which accounts for 80% or 42% (evaluated using 2014 or 1960 characteristics respectively).

In the case of segregation into low-paying occupations, the same exercise is done in two stages in Table 3. We first estimate that about 35% of the reduction between 1970 and 2014 is due to the change in the ranking of occupations by average earnings (favoring those held by women) when we use the 2010 earnings ranking of occupations in both years (but keeping contemporary marginal distribution of characteristics and conditional occupational distributions).

In a second stage, using the 2010 ranking of occupations we decompose the remaining reduction into its explained and unexplained components. We estimate that another 36% or 46% of the overall reduction in low-pay segregation can be explained by changes in characteristics (respectively evaluated using the 1970 or 2014 conditional occupational distributions). In the first case, the most important characteristic was marital status (28%), although the rest of characteristics also contributed: migration and number of children (about 5% each), education and ethnicity (about 4% each), or race (3%). The change in the age composition had a negative impact (-12%), indicating that it curbed the upgrading of occupations held by women. In the second case, there is a much larger relevance of education (27%) again, and smaller of marital status (13%) and the other characteristics. Consequently, the unexplained terms, associated with changes in conditional occupational distributions, account for 29% or 19% of the reduction (evaluated using 2014 or 1970 workers' characteristics by sex).

For a more detailed analysis of segregation over time, Figures 3.a-b draw the trends in segregation and stratification with the explained and unexplained terms evaluated every year using the 2014 conditional occupational distribution or marginal distribution of characteristics respectively. In the case of stratification, these are evaluated using the 2010 earnings structure (see also Table A5 in the appendix).

It becomes clear in Figure 3.a that the change in conditional occupational distributions pushed segregation down only until 1990. The change in the distribution of characteristics additionally helped to reduce segregation before 1990, but continued to do so afterwards at a slower pace. The persistence in the different conditional employment distributions by sex is thus responsible for the stagnation of segregation in the last decades, despite the positive effect of the continuing change in women's characteristics.

The evolution of the gap between the observed trend and that using the 2010 earnings structure reveals that changes in the earnings structure over time have favored male-dominated occupations until 1980 (the gap increases) and female-dominated occupations ever since (the gap decreases). In fact, these sex-biased changes in the earnings structure entirely explain the increase in stratification in the 1960s as well as its reduction in the 1990s, because stratification would remain constant in both decades using the 2010 indexation of better-paying occupations. The trend for the explained term shows that the change in characteristics helped to reduce stratification since 1960 at a nearly constant pace. The change in the occupational distributions by sex, on the contrary, only helped to reduce women's segregation into low-paying jobs between 1970 and 1990, going in the opposite direction ever since. Thus, the fact that stratification continued to be reduced after 1990 is driven by changes in the relative characteristics of women and men, and in the earnings structure of occupations, not to changes in conditional employment distributions by sex.

Table 2. Decomposition of segregation (Gini) trends, 1960/70-2014

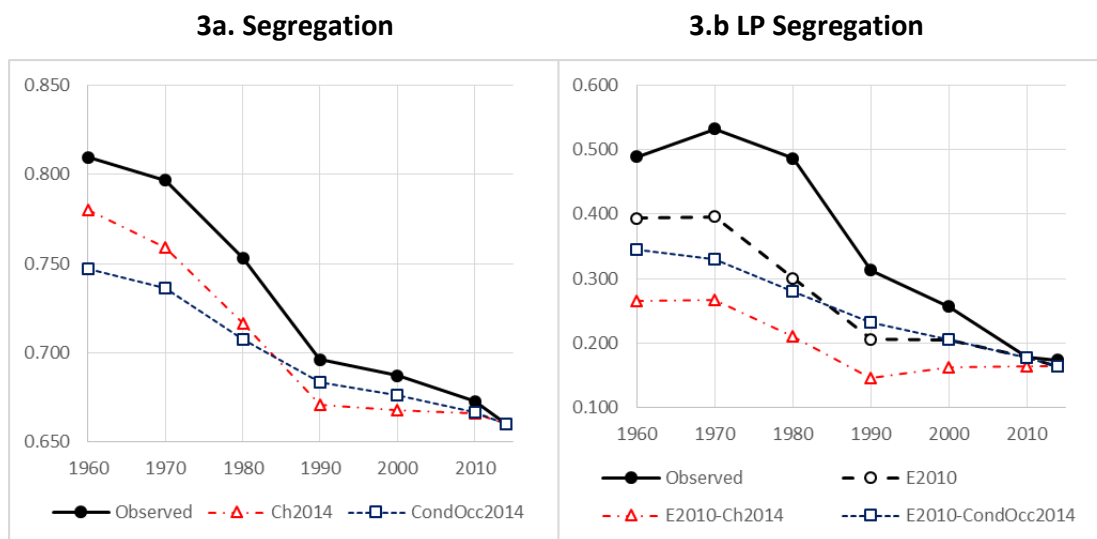
Difference between (low pay) segregation in final and initial years.

	Segregation 1960-2014				LP Segregation 1970-2014			
	Ref. 2014	%	Ref. 1960	%	Ref. 2014	%	Ref. 1970	%
Change	-0.149 (0.001)	100	-0.149 (0.001)	100	-0.358 (0.002)		-0.358 (0.002)	
Earnings structure	--		--		-0.126 (0.001)	35.1	-0.126 (0.001)	35.1
Unexplained	-0.120 (0.001)	80.3	-0.062 (0.001)	41.9	-0.102 (0.004)	28.5	-0.067 (0.004)	18.7
Explained by characteristics	-0.029 (0.001)	19.7	-0.087 (0.003)	58.1	-0.130 (0.004)	36.3	-0.165 (0.004)	46.2
Marital Status	-0.011 (0.000)	7.4	-0.029 (0.000)	19.7	-0.099 (0.002)	27.6	-0.046 (0.002)	13.0
Children	-0.006 (0.000)	3.9	-0.006 (0.000)	3.9	-0.017 (0.001)	4.6	-0.018 (0.001)	5.0
Age	0.001 (0.000)	-0.6	0.018 (0.000)	-12.4	0.043 (0.001)	-12.1	0.016 (0.001)	-4.4
Race	-0.003 (0.000)	2.2	-0.012 (0.000)	8.2	-0.011 (0.002)	3.0	-0.006 (0.001)	1.6
Hispanic	-0.001 (0.000)	0.7	0.005 (0.000)	-3.2	-0.014 (0.002)	3.9	-0.006 (0.001)	1.6
Migration	-0.001 (0.000)	0.9	0.004 (0.000)	-2.6	-0.019 (0.002)	5.2	-0.009 (0.001)	2.5
Education	-0.008 (0.000)	5.2	-0.066 (0.000)	44.5	-0.015 (0.004)	4.1	-0.096 (0.003)	26.9

Notes: Unexplained effect evaluated using reference year' characteristics; Explained effect evaluated using the other year's conditional occupational distribution. In Low-Paying Segregation, Earnings structure uses 2010 ranking of occupations by earnings, also used to estimate explained and unexplained effects. Bootstrap standard errors in parentheses (175-200 replications).

Source: Own construction based on census and ACS (IPUMS-USA).

Figure 3. (LP) Segregation trends, Gini 1960-2014



Note: E2010 = 2010 ranking of occupations indexed by earnings; Ch2014 = each sex’s 2014 characteristics; CondOcc2014 = each sex’s 2014 conditional occupational distribution.

8. Concluding remarks

Occupational segregation of workers by sex has long been a matter of research among social scientists since the early literature that emerged in the 1940s. There is a clear lack of overlapping between the occupations held by men and women despite the large transformations the labor markets went through in the last decades. The literature has developed a corpus that allows for such measurement issues. Segregation can be regarded as inequality in the distribution of men and women across occupations. Several inequality measures can be used to account for the level of segregation, even if a particular one, the Dissimilarity index, has attracted most attention. The use of segregation curves is a powerful tool that allows obtaining comparisons across time or space with a high level of robustness as it depends on the agreement over only a few reasonable value judgements. The literature has been less effective so far, however, in appropriately integrating a measure of stratification of occupations by sex when there is a tendency for one group, women, to fill low-paying jobs. In that direction, we proposed here an extension of the conventional approach for two-group segregation (that can be easily adjusted to include the multigroup case). We apply tools that are well rooted in other fields of the economic literature, like the concentration curve and indices derived from them already used in the analyses of health or educational socioeconomic inequalities, and tax redistribution.

Based on this approach, we have analyzed the long-term trends in the level and nature of women’s segregation in the US. We have shown that in parallel to the long-term reduction in segregation there was a change in its nature that started later, after 1970, but was much deeper

and lasted longer. We already know that this trend is compatible with the persistence of the gender wage gap, as long as within-occupational wage gaps are becoming more relevant.

Occupational segregation by sex is still large and women unambiguously continue to be overrepresented in low-paying occupations in the US. For that, we further investigated the extent to which these are the results of worker's characteristics differing by sex. Using reweighted counterfactuals, we confirm that at most around 10% of women's segregation can be associated with a compositional effect that mostly comes from the different field of college degree, the rest being the result of men and women with similar characteristics working in different occupations. Furthermore, based on their characteristics, especially attained education, women should be rather overrepresented in best-paying occupations.

The weak compositional effect was not much different in the past, but the profound changes since 1960 in the composition of each sex's workforce (e.g. by education, marital status, age, etc.) have played an important role in explaining the long-term trends in the level and nature of segregation. We obtained that 58% of the reduction of segregation and 46% of the much larger decline in stratification can be explained by the changing characteristics of workers if they are evaluated using the current conditional employment distributions by sex. The role of education was important, but the impact of changes in marital status were also noteworthy. Another 35% of the decline in stratification was the result of changes in average earnings favoring occupations mostly held by women after 1980. Finally, the conditional distributions of occupations by sex provide an idea of how segregative the labor markets are against men and women of similar characteristics (given the occupational and earnings structure). Changes on these distributions over time also help to explain a significant proportion of the trends, larger for segregation and smaller for stratification (respectively 42-80% and 19-29%, depending on the characteristics used), but only effectively before 1990.

References

- Alonso-Villar, O. and C. Del Río (2010), "Local versus Overall Segregation Measures", *Mathematical Social Sciences*, 60(1), 30-38.
- Alonso-Villar, O. and C. Del Río (2016), "Local Segregation and Wellbeing", *The Review of Income and Wealth* (forthcoming). DOI: 10.1111/roiw.12224.
- Amuedo-Dorantes, C. and S. de la Rica (2006), "The Role of Segregation and Pay Structure on the Gender Wage Gap: Evidence from Matched Employer-Employee Data for Spain", *The B.E. Journal of Economic Analysis & Policy*, 5(1), 1-34.
- Antoninis, M., M. Delprato, and A. Benavot (2016), "Inequality in Education: The Challenge of Measurement", in ISSC-UNESCO, *Challenging Inequalities: Pathways to a Just World, World Social Science Report*, Paris: UNESCO, 63-67.

- Aronson, R. and P.J. Lambert (1994), "Decomposing the Gini Coefficient to Reveal the Vertical, Horizontal, and Reranking Effects of Income Taxation", *National Tax Journal*, 47(2), 273-294.
- Aronson, J. R., P. Johnson, and P.J. Lambert (1994), "Redistributive Effect and Unequal Income Tax Treatment", *The Economic Journal*, 104(423), 262-270.
- Atkinson, A., (1980), "Horizontal Equity and the Distribution of the Tax Burden", in H. Aaron and M. Boskin, *The Economics of Taxation*, Washington D.C.: Brookings Institution, 3-18.
- Bayard, K., J. Hellerstein, D. Neumark, and K. Troske (2003), "New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data", *Journal of Labor Economics*, 21, 886-922.
- Beller, 1985, "Changes in the Sex Composition of U.S. Occupations, 1960-1981", *The Journal of Human Resources*, 20(2), 235-250.
- Bettio, F.D. and A. Verashchagina (2009), *Gender Segregation in the Labour Market: Root Causes, Implications and Policy Responses in the EU*, EU Expert Group on Gender and Employment (EGGE), Luxembourg: European Commission.
- Bishop, J.A., L.A. Zeager, and B. Zheng (2011), "Interdistributional Inequality, Stochastic Dominance, and Poverty", mimeo.
- Blackburn R.M. and J. Jarman (1997), "Occupational Gender Segregation", *Social Research Update*, 16 (<http://www.soc.surrey.ac.uk/sru/SRU16/SRU16.html>).
- Blackburn R.M., J. Jarman, and J. Siltanen (1994), "A Reply to Lampard", *Work, Employment and Society*, 8(3), 413-419.
- Blackburn R.M., G. Racko, and J. Jarman (2009), "Gender Inequality at Work in Industrial Countries", *Cambridge Studies in Social Research*, 11.
- Blau, F.D., P. Brummund, and A. Y-H. Liu (2013), "Trends in Occupational Segregation by Gender 1970–2009: Adjusting for the Impact of Changes in the Occupational Coding System." *Demography*, 50(2), 471–92.
- Blau, F.D. and L.M. Kahn (2000), "Gender Differences in Pay", *Journal of Economic Perspectives*, 14(4), 75-99
- Blau, F.D. and L.M. Kahn (2003), "Understanding International Differences in the Gender Pay Gap", *Journal of Labor Economics*, 21(1), 106–144.
- Blau, F. D. and L.M. Kahn (2006), "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence", *Industrial and Labor Relations Review*, 60, 45–66.
- Blau, F. D. and L.M. Kahn (2016), "The Gender Wage Gap: Extent, Trends, and Explanations", IZA Discussion Paper, 9656, Bonn.
- Blau, F.D., P. Simpson, and D. Anderson (1998), "Continuing Progress? Trends in Occupational Segregation in the United States over the 1970s and 1980s", *Feminist Economics*, 4(3), 29-71.
- Brynin, M. and F. Perales (2015), "Gender Wage Inequality: The De-gendering of the Occupational Structure", *European Sociological Review*, 1–13.
- Butler, R.J. (1987), "New Indices of Segregation", *Economics Letters*, 24, 359-362.
- Butler, R.J. and J. B. McDonald, (1987) "Interdistributional Income Inequality", *Journal of Business and Economic Statistics*, 5, 13-18.
- Chakravarty, S.R. (1988), "Extended Gini Indices of Inequality", *International Economic Review*, 29(1), 147-156.
- Chakravarty, S., C. D'Ambrosio, and J. Silber (2009), "Generalized Gini Occupational Segregation Indices", *Research on Economic Inequality*, 17, 71–95.
- Chantreuil, F. and A. Trannoy (2013), "Inequality Decomposition Values: the Trade-Off between Marginality and Consistency", *Journal of Economic Inequality*, 11(1), 83-98.
- Cotter, D.A., J.M. Hermsen, and R. Vanneman (2004), *Gender Inequality at Work, The American People: Census 2000*, New York, NY: Russell Sage Foundation and Population Reference Bureau.

- Del Río, C. and O. Alonso-Villar (2012), "Occupational Segregation Measures: A Role for Status", *Research on Economic Inequality*, 20, 37-62.
- Del Río, C. and O. Alonso-Villar (2015), "The Evolution of Occupational Segregation in the United States, 1940-2010: The Gains and Losses of Gender-Race/Ethnicity Groups", *Demography*, 52(3), 967-988.
- Deutsch, J, Y. Fluckiger, and J. Silber (1994), "Measuring Occupational Segregation: Summary statistics and the impact of classification errors and aggregation", *Journal of Econometrics*, 61,133-146.
- DiNardo, J., N.M. Fortin, and T. Lemieux (1996), "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach", *Econometrica*, 64, 1001-1044.
- Dolado, J. J., F. Felgueroso, and F. Jimeno (2002), "Recent Trends in Occupational Segregation by Gender: A Look across the Atlantic", in A. Argandoña and J. Gual (eds), *The social dimensions of employment: Institutional reforms in labour markets*, Cheltenham, UK and Northampton MA: Elgar; Williston, VT: American International Distribution Corporation.
- Donaldson, D. and J. A. Weymark (1980), "A Single Parameter Generalization of the Gini Indices of Inequality", *Journal of Economic Theory*, 22, 67–86.
- Duclos, J.-Y. (1993) "Progressivity, Redistribution and Equity, with Application to the British Tax and Benefit System", *Public Finance/Finances Publiques*, 48(3), 350–365.
- Duncan, O.D. and B. Duncan, B. (1955), "A Methodological Analysis of Segregation Indexes", *American Sociological Review*, 20(2), 210-217.
- Foster, J.E., (1985), "Inequality Measurement", in Young, W.P. (Ed.), *Fair Allocation, Proceedings of Symposia in Applied Mathematics*, American Mathematical Society, Providence, RI, 31–68.
- Foster, J.E., J. Greer, and E. Thorbecke (1984), "A Class of Decomposable Poverty Measures", *Econometrica*, 52(3), 761–766.
- Foster, J.E. and A.F. Shorrocks, (1988), "Poverty Orderings," *Econometrica*, 56(1), 173-177.
- Frankel, D.M., O. Volij, (2011), "Measuring School Segregation", *Journal of Economic Theory*, 146(1), 1–38.
- Gini, C. (1912), "Variabilità e Mutabilità: Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche", Facoltà di Giurisprudenza della R. Università dei Cagliari, anno III, parte II.
- Gini, C. (1914), "Sulla Misura della Concentrazione e della Variabilità dei Caratteri", *Atti del R. Istituto Veneto di Scienze, Lettere ed Arti*, tomo LXXIII, parte II, 1203-1258.
- Goldin, C. (2014) "A Grand Gender Convergence: Its Last Chapter", *American Economic Review*, 104(4), 1091–1119.
- Gradín, C. (2013), "Conditional Occupational Segregation of Minorities in the U.S.", *Journal of Economic Inequality*, 11(4), 473-493.
- Gradín, C. (2014), "Race and income Distribution: Evidence from the USA, Brazil and South Africa", *Review of Development Economics*, 18(1), 73-92.
- Groshen, E.L. (1991), "The Structure of the Female/Male Wage Differential: Is It Who You Are, What You Do, or Where You Work?", *The Journal of Human Resources*, 26(3), 457-472.
- Gunderson, M. (1989), "Male-Female Wage Differentials and Policy Responses", *Journal of Economic Literature*, 27:1, 46-72.
- Hornseth, R.A. (1947), "A Note on 'The measurement of ecological segregation' by Julius Jahn, Calvin F. Schmid, and Clarence Schrag", *American Sociological Review*, 12(5), 601-606.
- Hutchens, R.M. (1991), "Segregation Curves, Lorenz Curves, and Inequality in the Distribution of People across Occupations", *Mathematical Social Sciences*, 21, 31- 51.
- Hutchens, R.M (2001) "Numerical Measures of Segregation: Desirable Properties and their Implications", *Mathematical Social Sciences*, 42, 13–29.
- Hutchens, R.M (2004), "One Measure of Segregation," *International Economic Review*, 45(2), 555-578.

- Hutchens, R.M (2012), "Measuring Segregation When Hierarchy Matters," IZA Discussion Paper, 6667.
- Hutchens, R.M. (2009), "Occupational Segregation with Economic Disadvantage: an Investigation of Decomposable Indexes", *Research on Economic Inequality*, 17, 99–120.
- Jahn, J., C.F. Schmid, and C. Schrag (1947), "The Measurement of Ecological Segregation", *American Sociological Review*, 12(3), 293-303.
- James, D.R. and K.E. Taeuber (1985), "Measures of Segregation", *Sociological Methodology*, 15, 1-32.
- Jargowsky, P.A. and J. Kim (2009), "The information theory of segregation: uniting segregation and inequality in a common framework", *Research on Economic Inequality*, 17, 3-31.
- Kakwani, N.C. (1980a), "On a Class of Poverty Measures", *Econometrica*, 48(2), 437-446.
- Kakwani, N.C. (1980b), *Income Inequality and Poverty: Methods of Estimation and Policy Applications*, Report 10092, Oxford University Press, Oxford.
- Karmel, T. and M. MacLachlan (1988), "Occupational Sex Segregation - Increasing or Decreasing?", *The Economic Record*, 64(3), 187-1995.
- Kim, J. and P.A. Jargowsky (2009), "The GINI Coefficient and Segregation on a Continuous Variable", *Research on Economic Inequality*, 17, 57-70.
- Krämer, W. (1998), "Measurement of Inequality", in Ullah, A. and D. Guiles (eds.), *Handbook of Applied Economic Statistics*, Marcel Dekker, New York, 39-61.
- IWPR (2014), *The Gender Wage Gap by Occupation 2014 and by Race and Ethnicity*, Institute for Women's Policy Research, Washington DC.
- Lambert, P. (2002), *The Distribution and Redistribution of Income*, 3rd edition, Manchester University Press: Manchester.
- Lasso de la Vega, C. and O. Volij (2014), "Segregation, Informativeness and Lorenz Dominance", *Social Choice and Welfare*, 43(3), 547-564.
- Le Breton, M., A. Michelangeli, and E. Peluso (2012), "A Stochastic Dominance Approach to the Measurement of Discrimination", *Journal of Economic Theory*, 147, 1342–1350.
- Lerman, R.I. and S. Yitzhaki (1984), "Improving the Accuracy of Estimates of Gini Coefficients", *Journal of Econometrics*, 42, 43-47.
- Lerman, R.I. and S. Yitzhaki (1989), "A Note on the Calculation and Interpretation of the Gini Index", *Economic Letters*, 15, 363-368.
- Lerman, R.I. and S. Yitzhaki (1995), "Changing Ranks and the Inequality Impacts of Taxes and Transfers", *National Tax Journal*, 48(1), 45-59.
- Lerman, R., and S. Yitzhaki (1994), "Effect of Marginal Changes in Income Sources on U.S. Income Inequality", *Public Finance Review*, 22, 403-417.
- Mandel, H. and M. Semyonov (2014), "Gender Pay Gap and Employment Sector: Sources of Earnings Disparities in the United States, 1970–2010", *Demography*, 51, 1597-1618.
- Massey, D.S., and N.A. Denton (1988), "The Dimensions of Residential Segregation", *Social Forces*, 67(2), 281-315.
- Oaxaca, R.L. and M-R. Ransom (1994), "On discrimination and the decomposition of wage differentials", *Journal of Econometrics*, 61(1), 5-21.
- Pfähler, W. (1983), "Measuring Redistributive Effects of Tax Progressivity by Lorenz Curves", *Journal of Economics and Statistics*, 198(3), 237-249.
- Pietra, G. (1915), "Delle Relazioni tra gli Indici di Variabilità", *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, LXXIV, II (1914-15), 775-792.
- Pissarides, C., P. Garibaldi, C. Olivetti, B. Petrongolo, and E. Wasmer (2003), *Women in the Labour Force: How well is Europe Doing?*, Annual European Conference of the Fondazione Rudolfo De Benedetti, Milan: Fondazione Rudolfo De Benedetti.
- Plotnick, R. (1981), "A Measure of Horizontal Inequity", *Review of Economics and Statistics*, 63(2), 283-288.
- Reardon, S.F. and G. Firebaugh (2002), "Measures of multigroup segregation", *Sociological Methodology*, 32, 33-76.

- Reardon, S.F. (2009), "Measures of Ordinal Segregation", *Research on Economic Inequality*, 17, 129-155.
- Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2015), *Integrated Public Use Microdata Series: Version 6.0* [Machine-readable database], Minneapolis: University of Minnesota.
- Salas, R. J. Bishop, and L. Zeager (2015), "Second-Order Discrimination and Stochastic Dominance", 6th ECINEQ Meeting, Luxembourg.
- Schechtman, E. and S. Yitzhaki (1987), "A Measure of Association Based on Gini's Mean Difference", *Communications in Statistics*, 16(1), 207-231.
- Schechtman, E. and S. Yitzhaki (1999), "On the Proper Bounds of the Gini Correlation", *Economics Letters*, 63, 133-138.
- Shorrocks, A.F. (1980), "The Class of Additively Decomposable Inequality Measures", *Econometrica*, 48, 613-625.
- Shorrocks, A.F. (2013), "Decomposition Procedures for Distributional Analysis: A Unified Framework based on the Shapley Value", *Journal of Economic Inequality*, 11(1), 99-126.
- Silber, J. (1989), "On the Measurement of Employment Segregation", *Economics Letters*, 30, 237-243.
- Silber, J. (1992), "Occupational Segregation Indices in the Multidimensional Case: A Note", *Economic Record*, 68, 276-277.
- Silber, J. (2012), "Measuring Segregation: Basic Concepts and Extensions to other Domains", *Research on Economic Inequality*, 20, 1-35.
- Somers, R.H. (1962), "A New Asymmetric Measure of Association for Ordinal Variables", *American Sociological Review*, 27(6), 799-811.
- Sommers, D. (1974), "Occupational rankings for men and women by earnings", *Monthly Labor Review*, 97(8), 34-51.
- Volij, O. (2016), "Segregation: Theoretical Approaches", forthcoming in C. D'Ambrosio (Ed.), *Handbook of Research on Economic and Social Well-Being*, Edward Elgar.
- Wagstaff, A., E. van Doorslaer, and P. Paci (1989), "Equity in the finance and delivery of health care: some tentative cross-country comparisons", *Oxford Review of Economic Policy*, 5 (1), 89-112.
- Wolfson, M. C. (1994), "When Inequality Diverge", *American Economic Review*, 84, 353-358.
- Yitzhaki, S. and I. Olkin (1991), "Concentration indices and concentration curves", in K. Mosler and M. Scarsini (eds.), *Stochastic Orders and Decisions under Risk*, Institute of Mathematical Statistics, Lecture-Notes Monograph Series, 19, 380-392.

	Women							Men						
	1960	1970	1980	1990	2000	2010	2014	1960	1970	1980	1990	2000	2010	2014
Married	58.0	58.8	56.4	56.2	54.8	50.9	48.9	80.2	77.1	69.0	64.1	61.4	58.1	55.8
Divorced/Widowed	18.9	18.4	18.9	19.3	19.7	19.9	19.3	4.5	5.4	8.2	10.2	11.5	12.1	11.8
1 child	17.4	15.7	17.2	19.6	19.5	19.3	18.9	18.6	15.9	16.1	16.7	16.1	15.4	14.8
2+ children	20.5	23.9	22.7	21.2	20.9	19.5	18.8	28.2	29.5	23.6	20.3	19.4	18.3	17.8
1 child (<6)	8.1	8.8	9.6	10.8	10.1	9.0	8.4	15.7	14.3	12.3	11.9	10.6	9.2	8.6
2+ children (<6)	3.4	2.5	2.2	2.8	2.7	2.6	2.4	10.4	6.3	4.3	4.2	3.8	3.4	3.1
Aged <24	18.7	23.8	23.6	16.7	15.2	13.4	13.8	13.6	17.6	19.5	14.9	14.1	12.1	12.6
Aged 25-34	18.4	18.6	27.3	28.3	22.2	21.1	21.7	22.3	22.1	27.9	28.9	22.9	21.5	22.2
Aged 35-44	24.0	19.4	19.3	26.1	26.9	21.8	20.8	24.5	21.0	19.9	25.7	27.0	22.7	21.7
Aged 45-54	22.1	20.8	15.9	16.9	22.5	24.0	22.2	20.9	20.7	16.9	17.0	21.8	23.7	22.0
Aged 55-64	12.8	13.7	11.1	9.2	10.2	15.8	16.7	13.8	14.3	12.5	10.2	10.6	15.4	16.2
Aged 65+	4.0	3.8	2.8	2.8	3.1	4.0	4.9	4.9	4.3	3.4	3.4	3.6	4.6	5.3
White	87.9	87.4	86.1	82.1	77.9	75.7	74.2	91.2	90.4	89.2	84.1	79.7	77.9	76.7
Black	11.3	11.4	11.2	11.4	11.6	12.5	12.9	8.0	8.4	8.3	8.6	8.6	9.2	9.9
Mexican	1.4	1.6	2.9	3.9	4.9	8.0	8.8	1.9	2.0	3.5	5.2	6.9	10.6	11.2
Other Hispanic	1.0	1.3	2.3	3.1	4.2	5.4	6.0	0.9	1.3	2.3	3.2	4.3	5.7	6.3
Foreign-born	5.8	5.0	6.5	8.8	11.9	15.7	16.5	6.5	5.3	6.8	10.4	14.5	19.2	19.6
Grade 11 or less	46.0	35.1	21.7	11.7	9.5	7.1	6.6	55.3	41.6	25.9	15.5	12.8	10.5	9.5
High school diploma	34.1	40.5	42.2	34.1	37.9	32.0	30.2	24.7	31.3	35.0	31.9	37.6	34.9	35.1
College (1-2 years)	11.5	13.5	20.3	32.3	25.5	28.2	28.2	9.7	13.0	18.5	27.9	22.2	24.0	24.2
College (3+ years)	8.4	10.9	15.9	21.9	27.1	32.7	35.0	10.3	14.0	20.6	24.7	27.5	30.6	31.3

Source: Own construction based on Census and ACS (IPUMS-USA).

Table A2. College workers' field of degree by sex 2014

Fields with largest over/underrepresentation of women

	%Women	%Men		%Women	%Men
Nursing	7.3	0.8	Mechanical Engineering	0.3	2.8
Elementary Education	5.1	0.6	Computer Science	1.0	3.4
Psychology	6.1	2.8	General Business	3.6	5.8
General Education	5.0	1.8	General Engineering	0.3	2.2
English Language & Literature	3.7	2.0	Economics	1.2	2.9
Social Work	1.6	0.3	Business Management & Administration	5.6	7.1
Family and Consumer Sciences	1.4	0.1	Finance	1.3	2.9
Sociology	2.0	1.1	Civil Engineering	0.3	1.6
Special Needs Education	0.9	0.1	Political Science & Government	1.8	3.1
Communication Disorders Sciences & Services	0.8	0.1	History	1.4	2.6

Source: Own construction based on Census and ACS, 2014 (IPUMS-USA).

	Segregation				Low-paying Segregation			
	Gini	%	D	%	Gini	%	D	%
Observed	0.6604		0.4947		0.1737		0.1448	
Unexplained	0.5913	89.5	0.4296	86.8	0.1748	100.6	0.1581	109.1
Explained	0.0690	10.5	0.0651	13.2	-0.0011	-0.6	-0.0132	-9.1
Location	0.0002	0.0	0.0001	0.0	-0.0007	-0.4	-0.0004	-0.3
Marital Status	-0.0017	-0.3	-0.0020	-0.4	0.0113	6.5	0.0073	5.0
N Children	-0.0061	-0.9	-0.0058	-1.2	-0.0144	-8.3	-0.0112	-7.8
Age	0.0053	0.8	0.0047	0.9	0.0196	11.3	0.0149	10.3
Race	0.0029	0.4	0.0021	0.4	0.0050	2.9	0.0040	2.8
Hispanic ethnicity	0.0002	0.0	0.0001	0.0	0.0004	0.2	0.0003	0.2
Migration profile	0.0023	0.4	0.0020	0.4	-0.0045	-2.6	-0.0039	-2.7
Education	0.0660	10.0	0.0637	12.9	-0.0177	-10.2	-0.0241	-16.6

Note: Counterfactual: men’s 2014 distribution reweighted to reproduce women’s 2014 characteristics. See data section for details about the variables used in each category.
 Source: Own construction based on ACS 2014 (IPUMS-USA).

Table A4. Decomposition of (low-paying) occupational segregation of women, 1960-2014

Gini	Segregation							Segregation into low-paying occupations						
	1960	1970	1980	1990	2000	2010	2014	1960	1970	1980	1990	2000	2010	2014
Observed	0.809	0.796	0.753	0.696	0.687	0.673	0.660	0.489	0.532	0.486	0.313	0.257	0.178	0.174
Unexplained	0.795	0.790	0.745	0.685	0.676	0.660	0.649	0.494	0.521	0.465	0.304	0.261	0.198	0.201
Explained	0.014	0.007	0.008	0.011	0.011	0.013	0.011	-0.004	0.010	0.022	0.010	-0.004	-0.020	-0.027
Location	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	-0.002	-0.002	-0.001	0.000	0.000
Marital St.	0.002	-0.005	-0.002	-0.001	-0.001	-0.001	0.000	0.011	0.003	0.002	0.003	0.004	0.008	0.012
N Children	-0.001	0.000	-0.001	0.001	0.001	0.001	0.001	-0.011	-0.004	-0.001	0.002	0.002	0.000	-0.001
Age	0.002	0.004	0.002	0.001	0.002	0.003	0.002	0.001	0.005	0.006	0.011	0.015	0.017	0.016
Race	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.005	0.002	0.002	0.003	0.003	0.003
Hispanic	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001
Migration	0.000	0.000	0.001	0.001	0.002	0.003	0.003	0.003	0.002	-0.003	-0.003	-0.005	-0.008	-0.006
Education	0.008	0.008	0.007	0.008	0.007	0.006	0.005	-0.020	0.000	0.016	-0.004	-0.022	-0.042	-0.053
Dissimilarity														
Observed	0.641	0.639	0.584	0.522	0.515	0.505	0.495	0.413	0.462	0.413	0.294	0.243	0.153	0.145
Unexplained	0.617	0.629	0.577	0.511	0.502	0.489	0.481	0.388	0.440	0.388	0.288	0.249	0.173	0.170
Explained	0.024	0.009	0.007	0.011	0.013	0.016	0.014	0.025	0.022	0.025	0.006	-0.006	-0.020	-0.025
Location	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	-0.001	0.000	0.000	0.000
Marital St.	0.003	-0.004	-0.002	-0.001	-0.001	-0.001	0.000	0.003	0.004	0.003	0.003	0.003	0.006	0.008
N Children	0.001	0.000	-0.001	0.001	0.000	0.000	-0.001	0.001	0.001	0.001	0.001	0.001	0.000	-0.001
Age	0.002	0.004	0.002	0.002	0.003	0.004	0.003	0.005	0.005	0.005	0.008	0.011	0.012	0.012
Race	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.001	0.001	0.002	0.002	0.003
Hispanic	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001
Migration	0.000	0.000	0.001	0.001	0.003	0.003	0.003	-0.002	0.001	-0.002	-0.002	-0.004	-0.006	-0.005
Education	0.017	0.009	0.007	0.008	0.009	0.010	0.008	0.018	0.010	0.018	-0.005	-0.019	-0.035	-0.043

Note: Counterfactual: women’s distribution in each year reweighted to reproduce contemporary men’s characteristics. See data section for details about the variables used in each category.

Source: Own construction based on Census and ACS (IPUMS-USA).

Table A5. Decomposition of (low-paying) occupational segregation of women by characteristics, 1960-2014

Gini	Segregation							Segregation into low-paying occupations						
	1960	1970	1980	1990	2000	2010	2014	1960	1970	1980	1990	2000	2010	2014
Observed	0.810	0.797	0.753	0.696	0.687	0.673	0.660	0.489	0.532	0.486	0.313	0.257	0.178	0.174
(E2010)								0.266	0.267	0.211	0.146	0.163	0.164	0.164
+Ch2014	0.780	0.759	0.716	0.671	0.668	0.666	0.660	0.394	0.396	0.301	0.206	0.205	0.178	0.164
+CondOcc2014	0.747	0.736	0.708	0.683	0.676	0.667	0.660	0.345	0.330	0.280	0.232	0.206	0.178	0.164
Dissimilarity														
Observed	0.641	0.639	0.584	0.522	0.515	0.505	0.495	0.408	0.462	0.413	0.294	0.243	0.153	0.145
(E2010)								0.361	0.351	0.266	0.182	0.182	0.153	0.136
+Ch2014	0.625	0.599	0.549	0.503	0.500	0.499	0.495	0.259	0.246	0.188	0.120	0.140	0.140	0.136
+CondOcc2014	0.573	0.565	0.538	0.515	0.508	0.500	0.495	0.295	0.276	0.230	0.196	0.174	0.148	0.136

Notes: E2010 = 2010 ranking of occupations indexed by earnings; Ch2014 = each sex's 2014 characteristics; CondOcc2014 = each sex's 2014 conditional occupational distribution.

Source: Own construction based on Census and ACS (IPUMS-USA).