UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2017/89

Segregation of women into low-paying occupations in the United States

Carlos Gradín*

April 2017

United Nations University World Institute for Development Economics Research
wider.unu.edu

Abstract: We extend the conventional framework for measuring segregation to consider stratification of occupations by gender, i.e. when women or men are predominantly segregated into low-paying jobs. For this, we propose to use concentration curves and indices. Our empirical analysis using this approach shows that the decline over time in occupational gender segregation in the US has been accompanied by a deeper, longer reduction in gender stratification. We further investigate the role of workers' characteristics, showing that gender differences cannot explain the levels of segregation/stratification in any year. However, changes over time for each gender do help to explain their trends.

Keywords: gender inequality, low-paying occupations, occupational segregation, stratification
JEL classification: J16, J42, J71, J82, O51

Acknowledgements: This study has been prepared within the UNU-WIDER project on 'Gender and development' and builds on our previous work which has been circulated under the same title. I acknowledge the support received from AEI/FEDER, UE (ECO2016-76506-C4-2-R).

* UNU-WIDER, Helsinki, Finland, gradin@wider.unu.edu

This study has been prepared within the UNU-WIDER project on 'Gender and development'.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-313-4

Typescript prepared by Luke Finley.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Denmark, Finland, Sweden, and the United Kingdom.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

Occupational segregation of workers by gender has long been a matter of research among social scientists, both empirically and methodologically, at least since the early literature that emerged in the 1940s. There is a persistent lack of overlap between the occupations held by men and by women in many countries, despite the large-scale transformations the labour markets have been through in the last decades (e.g. Anker 1998; Bettio and Verashchagina 2009). The literature has already developed an appropriate measurement framework (reviewed in Silber 2012) in which segregation can be regarded as inequality in the distribution of men and women across occupations. Thus, the segregation curve (SC), plotting the proportions of workers from both groups with occupations indexed by their gender ratio, produces comparisons across time or space with a high level of robustness upon agreement on only a few reasonable value judgements. Because this is a partial ordering of distributions, several well-known inequality measures can be used to account for the level and trend of segregation. Two of them can be directly obtained from the segregation curve: the Gini and Dissimilarity indices, the latter being the most popular for its simplicity and intuitive interpretation.

The literature has been less effective so far, however, in properly integrating the measurement of stratification of occupations, i.e. where there is a tendency for one population group (e.g. women, non-whites) to fill least-valued jobs (because of pay, required skills, prestige, social benefits, risks, etc.). This is an important fact because, whenever stratification occurs, we can expect that any potentially negative effects associated with segregation will be aggravated. In this context, job segregation has been pointed out as the main factor explaining gender wage differentials within countries (e.g. Groshen 1991; Bayard et al. 2003; Amuedo-Dorantes and De la Rica 2006; Brynin and Perales 2015), with similar effects on other disadvantaged groups. In the US—the best-studied case—we already know that occupational segregation by gender has strongly declined since 1970 (e.g. Blau et al. 2013), helping to reduce the gender pay gap. The slower pace at which segregation decreased after the 1990s (stalling after the 2000s) helps also to explain some of the slowing wage convergence (e.g. Blau and Kahn 2006). The persistence of the pay gap is increasingly associated with differentials within occupations (e.g. Goldin 2014), which suggests that the nature of gender segregation is also changing.

For these reasons, we propose here a natural extension of the conventional approach for two-group segregation, to include consideration of the quality of jobs held by each group, which is in line with how the welfare literature deals with other socioeconomic inequalities (e.g. in health, education, or taxation). We aim to measure the extent to which segregation implies stratification in occupations, with one group systematically holding less-valued jobs. We propose to use the concentration curve (CC) that plots the cumulative proportion of workers from both groups with occupations indexed by their quality, here proxied by average earnings. This produces a partial order (first-order stochastic dominance) that may be completed using the Dissimilarity and Gini concentration indices derived from that curve. The combined study of the concentration curve and indices with their segregation counterparts provides us with a well-integrated framework within which to analyse not only the extent of segregation but also its ordinal nature in a way highly consistent with other branches of wellbeing analysis. Using this approach with census data, we investigate the long-term trends in gender segregation and stratification in the US labour market.

The effects of any type of segregation, but especially if it implies stratification of groups, are further aggravated whenever they do not result from differences in the accumulation of human capital or in other relevant characteristics that workers of each group bring to the labour market. This is something commonly assumed in the analysis of the pay gap and other labour market outcomes, but rarely so when it comes to segregation. For this, we use a counterfactual distribution in which men and women are compared using the same distribution of workers' characteristics (e.g. education, age, etc.), obtained by reweighting the female (or male) sample. Using this counterfactual, we analyse the extent to which segregation or stratification can be explained by differences in characteristics between the genders or, alternatively, by differences in the genders' conditional occupational distributions. We also study the role of these factors in explaining the decline in segregation over time, and, alongside the occupational earnings structure, in explaining the trend in stratification too.

The next section summarizes the conventional framework for measuring segregation, which is extended in Section 3 to measure stratification by gender. Section 4 describes the data, Section 5 presents the empirical analysis, and the last section makes some concluding remarks.

2 Measuring segregation

2.1 Notation and basic definitions

Let us consider a two-group population of N individuals (i.e. workers) distributed across $J > 1$ units (i.e. occupations). In what follows, subscript $j = 1, \dots, J$ refers to occupation. Superscript $i = c, r$ identifies groups, which 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 jth occupation. Occupations are indexed in ascending values of the gender ratio $\rho_j = n_j^r / n_j^c$, with $f^i = \left(f_1^i, \dots f_J^i\right)$ 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 to represent the cdf (cumulative distribution function) as a step-function: $F^i(\rho) = F_j^i$ 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, with higher values indicating higher segregation. We follow here what Massey and Denton (1988) called the evenness approach to segregation. Occupational gender segregation can be seen as inequality in the distribution of workers by gender across occupations, i.e. the extent to which their employment distributions differ from each other (they work in a different subset of occupations). Many authors have discussed the close relationship between inequality and segregation (e.g. James and Taeuber 1985; Butler 1987; Hutchens 1991, 2001, 2004; Silber, 1989; Deutsch et al. 1994).

A segregation index can be expressed as a measure $I_c(\rho)$ of inequality of the group ratio ρ among members of the comparison group (with $F^c(\rho)$ being the cdf; e.g. Silber 1989). 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 the production of partial orderings of inequality income distributions based on a minimal set of value judgements (scale invariance, population principle, symmetry, and Pigou-Dalton transfer principle; e.g. Foster 1985) has its correspondence in the segregation curve, while inequality indices can be used as segregation measures (e.g. James and Taeuber 1985; Hutchens 1991, 2001, 2004; Lasso de la Vega and Volij 2014; Volij 2016).

2.2 The segregation curve and properties of a segregation index

The segregation curve $F^r(p), p \in [0,1]$ is the continuous piecewise 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 gender ratio ρ_j). Thus, $F^r(p) = F_j^r + \frac{p-F_j^c}{f_{j+1}^c} f_{j+1}^r$ *if* $F_j^c \le p < F_{j+1}^c$, for $j = 0, \dots J-1$, $F_0^i = 0$, and $F^r(1) = 1$. It is just a (first-order) interdistributional Lorenz or discrimination curve in the interdistributional inequality framework (Butler and McDonald 1987; Le Breton et al. 2012). 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, switching to 1 at $p = 1$).

In Hutchens' (2004) formulation, 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 (when it is divided into M smaller occupations, each with the same gender ratio). Segregation, however, increases after a disequalizing movement (principle of movements between occupations). 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 ($f^{c\prime}, f^{r\prime}$), such that $\rho_j' = \rho_j - d$ and $\rho_s' = \rho_s + d$, for $0 < d \le \rho_j$ and $\rho_t' = \rho_t, \forall t \ne j, s$.

The first two properties imply that segregation depends only 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 group ratio, and increases after moving workers to occupations in which there is a higher proportion of their own group.

A distribution given by (f^c, f^r) dominates (has less segregation than) another one $(f^{c\prime}, f^{r\prime})$ if its segregation curve lies at no point below and at some point above the other: $F^r(p) \ge F^{r\prime}(p)\ \forall p \in [0,1]$, with strict inequality holding for some p. Whenever they intersect, we need measures consistent with the segregation curve to rank distributions, although these rankings may differ depending on the degree of sensitivity of the index to different points of the distribution.

A segregation index might also verify other interesting properties, such as the range property saying that it should take values between 0 (no segregation) and 1 (full segregation), or symmetry in types, requiring segregation not to change after exchanging the comparison and the reference groups (i.e. men and women are equally segregated from each other).

2.3 Dissimilarity and Gini segregation indices

Among the many possible segregation indices, we focus here on the two that have a straight geometrical interpretation in terms of the segregation curve: the Dissimilarity and Gini indices, first introduced in segregation analysis by Jahn et al. (1947) and later popularized by Duncan and Duncan (1955).

The Dissimilarity index, *D*, (relative mean deviation or Pietra index of inequality) is half the sum of discrepancies in population shares by group across occupations, $D(f^c, f^r) = \frac{1}{2}\sum_{j=1}^{J} \left| f_j^c - f_j^r \right|$. In geometrical terms, it is the maximum vertical distance between the diagonal and the segregation curve or, alternatively, twice the area of the largest triangle that can be inserted between the diagonal and the segregation curve:

$$D(f^c, f^r) = \max_{j\in[1,J]}\left\{F_j^c - F_j^r\right\}. \quad \textbf{(1)}$$

It is, thus, the difference in the cumulative proportion of both groups in the set of occupations in which one group is over-represented (e.g. Hornseth 1947): $D(f^c, f^r) = F_q^c - F_q^r$, where $q = \max_{j\in[1,J]}\left\{j \mid f_j^c \geq f_j^r\right\}$ is the critical occupation so that the comparison group is over-represented below and under-represented above. Therefore, it is interpretated as the proportion of workers of each group that should change occupation to achieve full integration (moving from those in which their group is over-represented to those in which it is under-represented).

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

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

where $\hat{F}_j^i = \frac{1}{2}\left(F_{j-1}^i + F_j^i\right)$ with mean equal to $\frac{1}{2}$ is used instead of F_j^i for consistency between the different formulations of $Gini$ (Lerman and Yitzhaki 1989). $Gini$ can also be written in terms of the covariance between the relative gender ratio and $\hat{F}_j^c$ (Lerman and Yitzhaki 1984):

$$Gini(f^c, f^r) = \frac{2}{N^r/N^c} cov\left(\rho, \hat{F}^c\right) = 2\sum_{j=1}^{J}\left(f_j^r - f_j^c\right)\hat{F}_j^c = 2\sum_{j=1}^{J} f_j^r\, \hat{F}_j^c - 1. \quad \textbf{(3)}$$

It is quite useful to understand the relationship between these two indices. 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 gender. *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. Gini considers segregation between as well as within these two sets 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 as well.

3 Occupational stratification: segregation into low-paying occupations

3.1 Previous approaches

The importance of considering 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 has been given different names, such as stratification, or ordinal, vertical, or status-sensitive 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 gender, and is obtained indirectly using the Euclidean norm. This approach has been used in several empirical studies (e.g. Bettio and Verashchagina 2009; Blackburn et al. 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 their measures of local segregation (Alonso-Villar and Del Río 2010) in the multigroup context. Reardon (2009) used measures of ordinal segregation that can be interpreted either as 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) and Alonso-Villar and Del Río (2017) have proposed direct measures of the monetary or wellbeing losses associated with each group's segregation in the multigroup context.

In a different approach, Gradín (2013) studied the extent to which an ethnic minority is segregated into low-paying occupations by comparing the proportions of the minority and the reference group below any possible threshold defining low earnings. This implies the use of first-order stochastic dominance (FOSD) of the employment distributions of both groups ordered by (median) earnings of occupations. In the next section, we extend this partial ordering approach by proposing the use of the concentration curve, a variant of the segregation curve, to test for FOSD. We also propose to obtain a complete ordering by deriving concentration indices consistent with this approach. In this way, segregation and segregation into low-paying occupations can be jointly analysed in an integrated framework which is strongly consistent with the inequality literature.

3.2 The concept of low-pay segregation (or stratification)

Let w be an ordinal or cardinal measure of the quality of occupations in one or several dimensions, such as pay, prestige, skills, etc., with w_j being its realization for occupation j. In the empirical application, w will be workers' average earnings by occupation. All the relevant information is given by relative frequencies g^i (with cdf $G^i(w)$), a permutation of f^i in which occupations have been re-ranked in increasing values of w.

Based on Gradín (2013), we consider that there is segregation of the comparison group (i.e. women) into low-paying occupations (low-pay segregation, for short) if we find a larger proportion of workers from group c below any threshold z defining low pay: $G^c(z) \geq G^r(z), \forall z \in [0, \overline{z}]$ (with strict inequality holding for some z), where $\overline{z}$ defines the range of reasonable thresholds (possibly, the maximum average earnings). In the same spirit, we can define the situation in which c is segregated into high-paying occupations, and that in which the distribution of employment is pay-neutral, by simply replacing $\geq$ in the previous definition with $\leq$ and $=$ respectively. Whenever one group is segregated into low-paying occupations, we can say that occupations are stratified by gender. Let us consider an example in order to understand these definitions, and then discuss some interesting implications of this approach.

There is full segregation in a four-occupation society if $f^c = (.5, .5, 0, 0)$ and $f^r = (0, 0, .5, .5)$, because in each occupation we only find individuals of one group. 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 (with $\overline{z} = 5$). Alternative wage structures may imply low-pay segregation—for example, with $w_c = (1,1,5,5)$ or $w_d = (1,3,2,4)$, because c has a proportion of workers larger than (or equal to) r below any low-pay threshold (with r segregated into high-paying occupations).

If low-pay segregation holds over the entire range of w, this implies FOSD (e.g. Bishop et al. 2011). If we interpret g^i as a lottery for workers of group i entering the labour 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. Following previous results from poverty analysis (Foster and Shorrocks 1988), FOSD also implies that group c would exhibit higher low-pay segregation than r for any head-count index that measures the proportion of each group below a threshold. And this is true for any possible threshold. If w is cardinal, FOSD implies dominance of higher order and, thus, the same would hold for any other index that weighted each worker in a low-paying occupation by a function of the deficit to the threshold, in line with the Foster et al.'s (1984) index of poverty. Indices of the Foster-Greer-Thorbecke (FGT) family would be, in our context, $S_\alpha(z;i) = \sum_{j|w_j<z} \left(\frac{z-w_j}{z}\right)^\alpha g_j^i,\ \alpha \geq 0,\ z \leq \overline{z}, i = c, r$. If low-pay segregation holds only below a certain threshold, we have restricted FOSD instead.

3.3 The low-pay segregation (or concentration) curve

FOSD can be tested using the concentration curve. This is a generalization of the notion of the Lorenz (or segregation) curve in which the variable that is accumulated in the abscissa is not necessarily sorted by the variable accumulated in the ordinate. The concentration curve $G^r(p), p \in [0,1]$ is the continuous piecewise function that connects the cumulative proportions of workers for the comparison (G_j^c) and reference (G_j^r) groups across occupations (indexed by w instead of by gender ratio ρ). The properties of the concentration curve are well-known (e.g. Kakwani 1980; 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—accumulating post-tax income using pre-tax rankings (e.g. Atkinson 1980; Plotnick 1981). Other examples are the study of socioeconomic inequalities of a population ranked by income or wealth in relation to access to health (illness) (e.g. Wagstaff et al. 1989), or to education (e.g. Antoninis et al. 2016).

Based on the previous definitions, low-pay segregation holds for any threshold if and only if the concentration curve always falls below the diagonal: $G_j^r \leq G_j^c\ \forall j \leftrightarrow\ G^r(p) \leq p, \forall p$. Similarly, the curve falls above the diagonal with high-pay segregation, and overlaps with it if employment distributions are pay-neutral. 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 extreme situations in which earnings (w) and the gender ratio (ρ) produce the same and the inverted ranking of occupations, respectively.

For these reasons, we propose the concentration curve of the re-ranked distributions G^i as a basic tool to analyse low-pay segregation. We can obtain a partial ordering of different distributions if their corresponding concentration curves do not overlap, with those falling below indicating higher low-pay segregation. If they intersect, however, we need to use indices to quantify the phenomenon and obtain a complete ordering.

3.4 Indices of low-pay segregation: desirable properties

We see low-pay segregation as a form of occupational segregation. It thus seems reasonable that the desirable properties and indices are the same as before, but adjusted to consider w. We require homogeneity, insensitivity to proportional divisions, and the principle of movements between occupations, after redefining 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-pay 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. Note that a disequalizing movement produces FOSD: either $G^{r\prime}(w) \leq G^r(w)$ or $G^{c\prime}(w) \geq G^c(w)$, 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, as a disequalizing movement defined in f^i implies a shift in the same direction of the segregation curve.

Regarding the range property, it seems reasonable to require the index of low-pay segregation to be bounded from above by the level of segregation, and from below by its negative magnitude (indicating high-pay segregation instead), with 0 representing a pay-neutral employment distribution. The absolute magnitude of the index gives a measure of the degree of stratification of occupations by gender. When there is full segregation, the index should be either 1 (low-pay) or -1 (high-pay) according to the range property of segregation. 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 per the range property, because low-pay segregation of one group necessarily implies high-pay segregation of the other.

3.5 Indices of low-pay segregation: Gini and Dissimilarity concentration indices

It seems natural in this framework to consider the Gini and Dissimilarity concentration indices that can be derived from the concentration curve as candidates to measure low-pay segregation, as they will be sensitive to the magnitude of the re-ranking of occupations. These concentration indices can be defined by rewriting the corresponding segregation indices in **(1)** and (**2)** with occupations indexed by earnings—that is, by replacing f^i with g^i.

The Gini concentration 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) index of statistical association used as a measure of vertical segregation (Blackburn et al. 1994; Blackburn and Jarman 1997), for which we thus provide an alternative rationalization within our framework:

$$Gini(g^c, g^r) = 2\sum_{j=1}^{J}\left(\hat{G}_j^c - \hat{G}_j^r\right) g_j^c. \qquad \textbf{(4)}$$

The positive (negative) vertical distance $G_j^c - G_j^r$ between the diagonal and the concentration curve represents the proportion of workers of any group that should change occupation to eliminate women's low-pay (high-pay) segregation for threshold $z = w_j$. We define the Dissimilarity concentration index $D(g^c, g^r)$ to be the largest of those distances in absolute terms (while keeping the sign). That is, if s is chosen so that $|G_s^c - G_s^r| = \max_{j\in[1,J]}\left\{\left|G_j^c - G_j^r\right|\right\}$, then:

$$D(g^c, g^r) = G_s^c - G_s^r. \qquad \textbf{(5)}$$

$D(g^c, g^r)$ is in the spirit of Pfähler's (1983) 'maximum proportionalization percentage' index computed for the concentration curve of taxes with respect to income. If the sign of $D(g^c, g^r)$ is positive, the index indicates the maximum proportion of women that should move to a higher-paying occupation in order to remove their low-pay segregation for any possible threshold z. If the index is negative, it indicates instead the proportion of women that should move to a lower-paying occupation to remove their high-pay segregation. A variation of this index would be $KM(g^c, g^r) = \frac{N^r}{N}\frac{N^r}{N}D(g^c, g^r)$, corresponding to Karmel and MacLachlan's (1988) index of segregation.

These indices should be used combined with the concentration curve, as this may cross the diagonal when one group is segregated into both low- and high-paying occupations. This generates positive and negative values for $G_j^c - G_j^r$, affecting the interpretation of the indices. A low value of $Gini$ is compatible with large negative and positive areas cancelling each other out; and a small variation in the distribution may change the sign of D. In these cases, one may want to compute the indices separately for different values of w, or analyse the extent of bipolarization in the employment distribution.

These concentration indices inherit some properties of their segregation counterparts. $Gini(g^c, g^r)$ verifies the four basic properties defined above, plus range and symmetry in types, and will rank distributions consistently with non-intersecting concentration curves. $D(g^c, g^r)$, however, violates the principle of movements between occupations.

It is also interesting to analyse the relationship between segregation and concentration. For each concentration index $S(g^c, g^r)$, we may also define a concentration ratio r_S as the proportion of segregation of group c that is low-pay (or high-pay if the sign is negative), by normalizing it by its maximum value, $S(f^c, f^r)$:

$$r_S = \frac{S(g^c,g^r)}{S(f^c,f^r)}, \text{ where } -1 \leq r_S \leq 1; \qquad S = Gini, D, KM. \qquad \textbf{(6)}$$

In particular, $r_D = r_{KM}$, and from the Gini covariance formulation we also know that $r_{Gini} = \Gamma_c(\rho, w)$, the Gini correlation coefficient between the gender ratio (ρ) and average earnings (w) of occupations, computed among members of group c. This correlation index is based on the Gini covariance, which relates a cardinal variable to the rank of another (a mixture of Pearson's and Spearman's correlations (Schechtman and Yitzhaki 1987, 1999; Yitzhaki and Olkin 1991). Thus, we measure the extent to which segregation of occupations implies stratification based on the correlation between the gender ratio of occupations and their earnings rank.

These ratios are somehow related to other ratios proposed in the related literature. 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 ($A = \frac{1}{2}\left(Gini(f^c, f^r) - Gini(g^c, g^r)\right)$) by its maximum value: $0 \leq \frac{A}{Gini(f^c,f^r)} = \frac{1-r_{Gini}}{2} \leq 1$. Aronson et al. (1994) regarded twice this area as the re-ranking 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.

4 Data

The empirical analysis is based on microdata samples extracted from censuses conducted by the US Census Bureau between 1960 and 2000, representing 5 per cent of the country's population (except in 1970, where the sample is 1 per cent of the population), and from the annual American Community Survey (ACS) conducted between 2001 and 2014 (about 0.4 per cent of each year's population in 2001–2004, 1 per cent thereafter). We used the Integrated Public Use Microdata Series (IPUMS-USA) harmonized by the Minnesota Population Center (Ruggles et al. 2015). Our sample consists of all workers employed during the reference week.

We analyse the distribution of employment by gender using the IPUMS-USA modified version of the 2010 Census Bureau occupational classification scheme (a total of 453 categories after excluding armed forces), which offers a consistent classification of occupations over the 1960–2014 period. 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 in all cases.

In the analysis of conditional segregation in 2014, we use a detailed set of workers' characteristics that might influence their occupation (see summary in Tables A1 and A2, Appendix). Attained education is the most important. 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 (169 categories). 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 the age of five), detailed race and ethnicity, and migration profile (place of birth, change of residence, years of residence in the US, citizenship, and English speaking proficiency). We also use a more restricted set of information common across samples, for the sake of comparability over time, omitting the field of college degree, English speaking proficiency, and migration status. For 1960, we replaced the years of residence in the US with change of residence during the previous five years. Hispanic origin was imputed by IPUMS before 1980. In the analysis of trends over time, we also omitted location.

5 Gender segregation and stratification of occupations

5.1 Unconditional segregation and low-pay segregation

It is a fact well-established in the literature that occupational segregation of women in the US has shown a long-term decline, followed by stagnation in recent years (e.g. Beller 1985; Blau et al. 1998; Cotter et al. 2004; Blau et al. 2013; Mandel and Semyonov 2014). Figures 1a and 1b show that segregation was reduced between 1960 and 2014 by around 20 per cent (18 per cent with $Gini$; 23 per cent with D), with the highest intensity in the 1970–1990 period but with little progress since. The US labour market thus remains highly segregated by gender—$Gini = 0.660$ in 2014— with three quarters of this being segregation between occupations dominated by each gender ($D = 0.495$, the most common measure).

The labour market was also highly stratified by gender in 1960, with women predominantly working in low-paying occupations, which is also consistent with the well-known historical role of segregation in explaining the pay gap (e.g. Treiman and Hartmann 1981; Gunderson 1989; Bayard et al. 2003). This low-pay segregation of women increased during the 1960s (by 9 per cent with

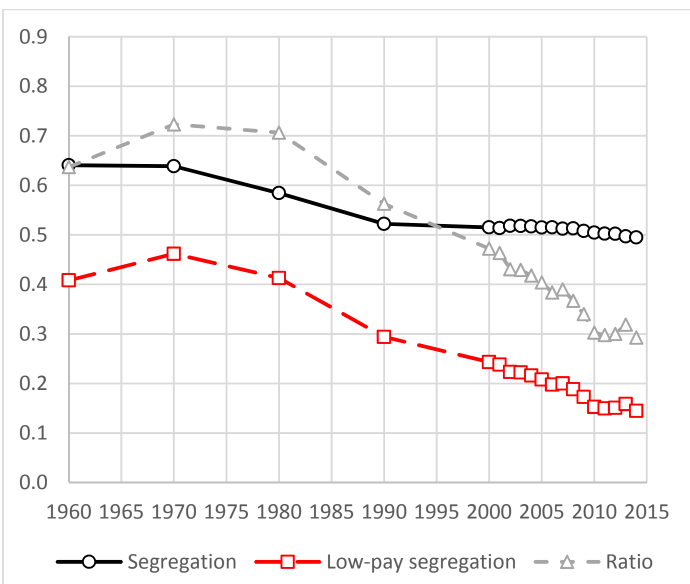
$Gini$, 13 per cent with D), followed by a sharp reduction since 1970 (67 per cent and 69 per cent). This reduction was more intense during the 1980s and the 2000s, and slowed down only after 2010. In 2014 the values were $Gini = 0.174$ and $D = 0.145$. Thus, the decline in stratification started later than in segregation, was much more intense, and continued when the latter stagnated after 1990 for reasons that will become clear later. Not surprisingly, the upgrading of occupations held by women in the US has substantially decreased the relevance of segregation in explaining the pay gap, contributing to its decline since 1970 (e.g. Blau and Kahn 2000). Goldin (2014) has recently noted that if the pay gap persists, this is because most of the current earnings gap comes from differences in earnings within occupations.

The concentration ratios, measuring the proportion of segregation that involves stratification by gender, declined from their peak in 1970 (67 per cent with $Gini$; 72 per cent with D) to their lowest levels in the 2010s (26 per cent and 29 per cent), providing a quantification of the outstanding change in the nature of women's segregation, regardless of its level.

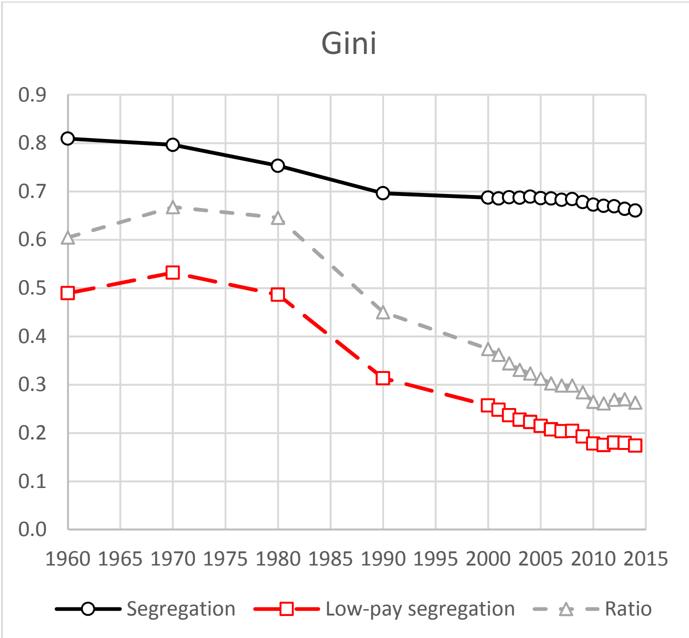
From the interpretation of the Dissimilarity indices, we know that it would be necessary to move 49.5 per cent of women from their current female-dominated occupations to those which are predominantly male-dominated in order to eliminate segregation by gender. To at least remove their low-pay segregation (for any possible threshold), we would need to move one in seven women to relatively higher-paying occupations (14.5 per cent, that is, 29 per cent of 49.5 per cent).

Figure 1: Women's (low-pay) segregation in the US

a: Dissimilarity



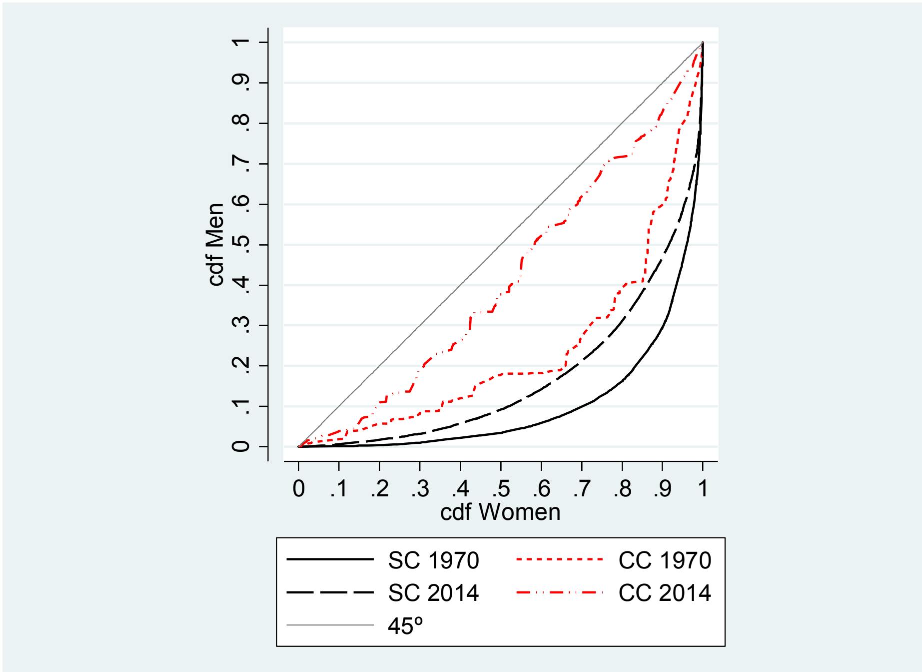
b: Gini



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

The concentration (or low-pay segregation) curves entirely falling below the diagonal in 1970 and 2014 in Figure 2 (implying FOSD) indicate that women were unambiguously concentrated into low-paying occupations in both years. This is the case regardless of the threshold used to define low-pay (and regardless of the FGT index used to aggregate the earnings gaps of women across occupations for each of those thresholds). The reductions in both segregation and stratification of occupations by gender over time are robust to other indices consistent with the segregation and concentration curves, because the corresponding curves for the latest year dominate those of the earliest.

Figure 2: Segregation (SC) and low-pay segregation (CC) curves by gender in the US



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

5.2 Explaining (low-pay) segregation levels

To identify the determinants of (low-pay) segregation, we follow Gradín's (2013) propensity score semi-parametric technique (adapting DiNardo et al. 1996), which is explained in detail in the Appendix. Based on the set of characteristics described above, we reweight the female sample f^c in 2014 to obtain a counterfactual distribution f^γ, indicating the proportion of women across occupations that we would expect to observe if they had the same marginal distribution of characteristics as men, but kept their own conditional occupational distribution. That is, we measure segregation in a hypothetical situation with the same proportion of women and men by workers' type (e.g. unmarried, black, native-born, aged 25–34, holding a degree in Economics, from the New York area), but with the original distribution by occupation for each type and gender. Due to the difficulties in obtaining these reweighting factors non-parametrically, we use for this a logit regression of the probability of being male on workers' characteristics.

By adding and subtracting segregation in the counterfactual, we decompose unconditional segregation into its explained and unexplained components:

$$S(f^c, f^r) = S^E(f) + S^U(f).$$

The first component (explained segregation) gives us the compositional effect—how much segregation goes away after equalizing workers' characteristics by gender, while keeping their own conditional occupational distributions (replacing the original female sample with its counterfactual):

$$S^E(f) = S(f^c, f^r) - S(f^\gamma, f^r).$$

The individual contribution of each set of characteristics is later obtained using a Shapley decomposition that overcomes the path dependency problem when the contribution of each factor is obtained sequentially (constructing reweighting factors from the logit regression after switching the coefficients of each factor from 0 to their estimated values).

The second component is conditional or unexplained segregation and indicates how much segregation is left when workers of each gender are compared with the same distribution of characteristics, but differ in their conditional occupational distributions:

$S^U(f) = S(f^\gamma, f^r)$.

The same procedure is applied to low-pay segregation. Results are shown in Table 1.

Occupational gender segregation in 2014 was mostly the consequence of men and women of similar characteristics working in different occupations. Only 10 per cent or less of segregation vanishes after giving women men's distribution of characteristics: 7 per cent with Gini (from 0.660 to 0.614), 10 per cent with D (from 0.495 to 0.444). This small explained or compositional effect is almost entirely driven by gender differences in education (explaining 6 per cent with Gini and 9 per cent with D), the main determinant that affects the set of occupations available to workers of each gender. Women are more likely to be college graduates than men, and have different fields of degree (they are strongly over-represented in health and education services and under-represented in engineering and business). Field of degree is the main contributor to gender segregation: when it is omitted in the estimation, the contribution of education to segregation is only around 1–2 per cent with both indices. Compared with men, female workers are also less likely to be white, Mexican, or recent immigrants, and to have children under the age of five (but are more likely to have them above that age). However, the contribution of location and demographic characteristics in explaining segregation is almost negligible.

Gender differences in observable characteristics altogether do not explain why women are segregated into low-paying occupations in 2014 either. The overall explained effect is in fact negative (given their characteristics, women should work in higher-paying occupations). These explained effects (−1.2 per cent with Gini, −14 per cent with D) are, however, the net result of larger counterbalancing forces. On the one hand, women's educational mix largely reduces their segregation into low-paying jobs (−15 per cent and −26 per cent). This is due to the higher proportion of women with a college degree, although it is curbed by their lower rate of doctorates and higher specialization into disciplines with lower average earnings. Indeed, the (negative) impact of education would be larger in this case (−31–32 per cent) had we omitted the field of degree. Women's lower immigration rates also reduce (−3 per cent) their low-pay segregation. On the other hand, demographic differences in age and marital status are also relevant. Segregation into low-paying occupations can be explained, to some extent, by women's under-representation among 25–44-year-old workers and among those above 65 years old (9 per cent and 8 per cent), and over-representation among unmarried workers (7 per cent and 6 per cent).

The alternative counterfactual, in which we reweight the male sample to reproduce women's characteristics, produces qualitatively similar results (Table A3, Appendix).

There has been a profound change in the distribution of workers' characteristics by gender over time, the combined result of the upgrading of the education of women relative to men, increasing female labour market participation, and trends affecting fertility, marriage, or immigration, among other things. However, these characteristics, evaluated with contemporary conditional occupational distributions, do not explain much of the segregation or low-pay segregation in previous years, either (Table A4, Appendix).

Table 1: Decomposition of (low-pay) occupational segregation of women by characteristics, 2014

		Segregation				Low-pay 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)	
	No. of 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. In parentheses, bootstrap standard errors over the entire process (200 replications). See Section 4 and Appendix for details about the variables used in each category.

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

5.3 Explaining trends in (low-pay) segregation

Characteristics by gender play a more significant role when it comes to explaining segregation and stratification trends, especially when the change in characteristics is valued using men's and women's current conditional occupational distributions (Table 2).

If f^{it} is the unconditional occupational distribution of group i in year t, the counterfactual is now $f^{i\lambda}$, which gives each gender in 2014 its distribution of characteristics in 1960, while keeping its own conditional occupational distribution. By adding and subtracting segregation in the counterfactual, the change in segregation over time is thus decomposed into the corresponding explained and unexplained changes:

$$\Delta S(f) = \Delta S^E(f) + \Delta S^U(f);$$

$$\Delta S^E(f) = S(f^{c2014}, f^{r2014}) - S(f^{c\lambda}, f^{r\lambda});$$

$$\Delta S^U(f) = S(f^{c\lambda}, f^{r\lambda}) - S(f^{c1960}, f^{r1960}).$$

The explained term now indicates how much of the reduction in segregation over time can be explained by the change in characteristics of workers of each gender (valued using 2014 conditional occupational distribution). The unexplained term now indicates how much of the reduction can instead be attributed to changes in their conditional distributions (valued using 1960 marginal distribution of characteristics). Note that, unlike in the previous exercise, here men and women are compared with their own distribution of characteristics; the equalization is made over time for each gender separately.

Our results show that 58 per cent of the reduction in segregation between 1960 and 2014 can be explained by the change in workers' characteristics. The largest contributions come from changes in education (45 per cent), followed by those in marital status (20 per cent), race (8 per cent), and number of children (4 per cent), partially offset by the negative contributions of changes in age (−12 per cent), ethnicity, and migration (about −3 per cent each). The remaining 42 per cent of the reduction in segregation is the unexplained term—that is, it results from changes in the conditional occupational distributions of men and women over time.

In the case of low-pay segregation, there is a third factor to consider: the change in the ranking of occupations by average earnings over time. The decomposition is thus done in two steps: we first change the ranking of occupations in 1970 and 2014 to a common one (for convenience, we use 2010 earnings), while keeping contemporary marginal distribution of characteristics and conditional occupational distributions. We label this distribution by $h = h^c, h^r$. The change in low-pay segregation, ΔS^w, is then attributed to changes in the earnings structure. In a second stage, using this 2010 earnings structure, we measure the components explained ΔS^E and unexplained by characteristics ΔS^U over distribution h, as we did in the case of segregation (reweighting the 2014 sample to reproduce 1970 characteristics):

$$\Delta S(g) = \Delta S^w(g, h) + \Delta S^E(h) + \Delta S^U(h).$$

We estimate that about 35 per cent of the reduction in low-pay segregation between 1970 and 2014 is due to the change in the ranking of occupations by average earnings (favouring those held by women). Another 46 per cent of the overall reduction in low-pay segregation can be explained by changes in characteristics. The most important of them was education (27 per cent), with smaller contributions from marital status (13 per cent) and the other characteristics. Consequently, the unexplained term, associated with changes in conditional occupational distributions, accounts for 19 per cent of the reduction.

The choice of the reference year, however, turned out to be crucial in explaining these trends, as a natural result of the important structural changes in both characteristics and conditional employment distributions. Table A5 (Appendix) shows that if we use the alternative counterfactual with 2014 characteristics and initial year's conditional occupational distribution, characteristics explain 20 per cent and 36 per cent of the reduction in segregation and low-pay segregation respectively—a smaller but still substantial proportion.

Figures 3a and 3b (also Table A6, Appendix, for both indices) draw the trends in *Gini* (low-pay) segregation, unconditional and using the 2014 conditional occupational distribution or the 2014 marginal distribution of characteristics respectively (with the 2010 earnings structure in the case of low-pay segregation). It becomes clear (Ch2014 in Figure 3a) that the change in conditional occupational distributions pushed segregation down only until 1990. The change in the distribution of characteristics (CondOcc2014) 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 gender 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 trend in low-pay segregation estimated using the 2010 earnings structure (E2010) in Figure 3b uncovers the fact that changes in the structure of earnings by occupation favoured male-dominated occupations until 1980 and have favoured female-dominated occupations ever since. These changes in earnings thus entirely explained the increase in low-pay segregation in the 1960s, curbed its reduction during the next decade, and entirely explained the reduction in the 1990s.

The trend in low-pay segregation with the 2014 conditional occupational distribution (E2010–Ch2014) reveals that the change in characteristics has helped to reduce stratification since 1960 at a nearly constant pace. The trend in low-pay segregation with the 2014 distribution of characteristics (E2010–Ch2014), shows that the change in the conditional occupational distributions, on the contrary, only helped the reduction between 1970 and 1990, going in the opposite direction ever since. Thus, the fact that stratification continued to be reduced after 1990 was driven by changes in the relative characteristics of women and men, and in the earnings structure of occupations, with changes in conditional employment distributions by gender now operating in the opposite direction.

Table 2: Decomposition of (low-pay) occupational segregation trends; Gini index, 1960/70–2014. Difference between (low-pay) segregation in final and initial years.

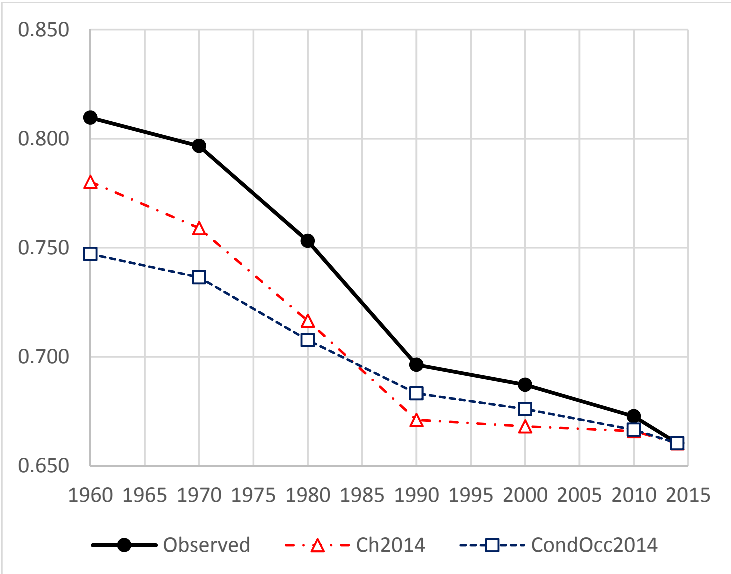
		Segregation 1960–2014	%	Low-pay segregation 1970–2014	%
Change		−0.149	100	−0.358	100
		(0.001)		(0.002)	
Explained by Earnings structure		-		−0.126	35.1
				(0.001)	
Unexplained		−0.062	41.9	−0.067	18.7
		(0.001)		(0.004)	
Explained by characteristics		−0.087	58.1	−0.165	46.2
		(0.003)		(0.004)	
	Marital status	−0.029	19.7	−0.046	13.0
		(0.000)		(0.002)	
	Children	−0.006	3.9	−0.018	5.0
		(0.000)		(0.001)	
	Age	0.018	−12.4	0.016	−4.4
		(0.000)		(0.001)	
	Race	−0.012	8.2	−0.006	1.6
		(0.000)		(0.001)	
	Hispanic	0.005	−3.2	−0.006	1.6
		(0.000)		(0.001)	
	Migration	0.004	−2.6	−0.009	2.5
		(0.000)		(0.001)	
	Education	−0.066	44.5	−0.096	26.9
		(0.000)		(0.003)	

Notes: Unexplained effect evaluated using initial characteristics; explained effect evaluated using the final conditional occupational distribution. In low-pay segregation, earnings structure uses 2010 ranking of occupations by earnings, also used to estimate explained and unexplained effects. In parentheses, bootstrap standard errors over the entire process (200 replications). See Section 4 and Appendix for details about the variables used in each category.

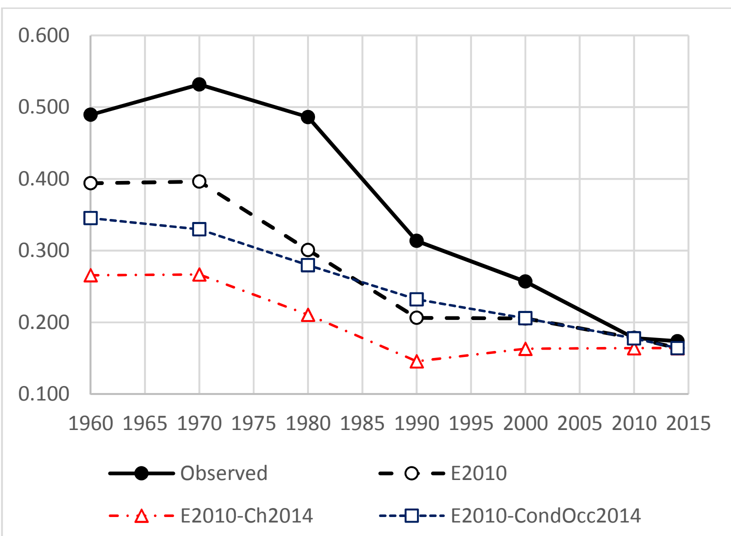
Source: Own construction based on Census and ACS (IPUMS-USA).

Figure 3: (Low-pay) occupational segregation trends, Gini 1960–2014

a: Segregation



b: Low-pay segregation



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

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

6 Concluding remarks

We have discussed a natural extension of the conventional framework used to measure segregation, in line with other fields of welfare analysis, that provides a more complete understanding of its ordinal or vertical nature. Using this integrated framework, we have shown that in parallel to the long-term reduction in segregation in the US during the last decades there was a deeper change in stratification that started later, in 1970, but lasted longer. This trend is consistent with previous results underlining the declining role of segregation in explaining the wage gap, with pay differentials within occupations becoming more relevant over time.

Occupational segregation by gender is still significant in the US, however, and women unambiguously continue to be over-represented in low-paying occupations. This has little or nothing to do with differences in the characteristics of the genders. At most, around 10 per cent of women's segregation can be associated with a compositional effect that mostly comes from differences in field of college degree. Furthermore, based on their attained education, women should be over-represented in the best-paying occupations (even after accounting for their different fields of specialization). Most of the segregation and all of the stratification found occurs because women work in different and lower-paying occupations compared with men of similar characteristics, and this was not much different in the past. The profound changes in the gender composition of the workforce by education or marital status after 1960, however, played an important role in explaining the long-term trends in the level and nature of segregation. Using the current conditional occupational distributions by gender, changes in characteristics account for 58 per cent of the reduction in segregation and 46 per cent of the much larger decline in stratification. Another 35 per cent of the decline in stratification was the result of changes in average earnings, after 1980, favouring occupations mostly held by women. Finally, changes over time in the conditional distributions of occupations by gender also help to explain a significant proportion of the trends, but only effectively before 1990, with no progress since then in either segregation or stratification.

The approach presented here may be easily used to analyse the extent to which other forms of segregation (e.g. residential, educational, etc.) also imply stratification of two (or more) population groups defined by any given characteristic. Combined with the decomposition approach, it helps to disentangle the role of compositional effects. However, as in most counterfactual analyses, we must assume here that there are no general equilibrium effects and that there is no selection of individuals based on unobservables (e.g. Fortin et al. 2011). This approach thus provides a good complement to the vast literature aimed at identifying the true mechanisms by which segregation and stratification operate (which can more easily deal with those two well-known problems).

References

Alonso-Villar, O., and C. Del Río (2010). 'Local versus Overall Segregation Measures'. *Mathematical Social Sciences*, 60: 30–38.

Alonso-Villar, O., and C. Del Río (2017). 'Local Segregation and Wellbeing'. *The Review of Income and Wealth*, forthcoming, DOI: 10.1111/roiw.12224.

Anker, R. (1998). 'Gender and Jobs: Sex Segregation of Occupations in the World'. Geneva: International Labour Organization.

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 and Policy*, 5: 1–34.

Antoninis, M., M. Delprato, and A. Benavot (2016). 'Inequality in Education: The Challenge of Measurement'. In ISSC, IDS, and UNESCO, *World Social Science Report 2016, Challenging Inequalities: Pathways to a Just World*. Paris: UNESCO.

Aronson, J.R., P. Johnson, P.J. Lambert (1994). 'Redistributive Effect and Unequal Income Tax Treatment'. *The Economic Journal*, 104: 262–70.

Atkinson, A. (1980). 'Horizontal Equity and the Distribution of the Tax Burden'. In H. Aaron, and M. Boskin (eds), *The Economics of Taxation*. Washington, DC: Brookings Institute.

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, A.H. (1985). 'Changes in the Sex Composition of US Occupations, 1960–1981'. *The Journal of Human Resources*, 20: 235–50.

Bettio, F.D., and A. Verashchagina (2009). 'Gender Segregation in the Labour Market: Root Causes, Implications and Policy Responses in the EU'. Luxembourg: EU Expert Group on Gender and Employment (EGGE), European Commission.

Bishop, J.A., L.A. Zeager, and B. Zheng (2011). 'Interdistributional Inequality, Stochastic Dominance, and Poverty'. Unpublished paper, East Carolina University.

Blackburn, R.M., and J. Jarman (1997). 'Occupational Gender Segregation'. *Social Research Update*, 16 (Spring 1997). Guildford: Department of Sociology, University of Surrey.

Blackburn, R.M., J. Jarman, and J. Siltanen (1994). 'A Reply to Lampard'. *Work, Employment and Society*, 8: 413–19.

Blackburn, R.M., G. Racko, and J. Jarman (2009). 'Gender Inequality at Work in Industrial Countries'. *Cambridge Studies in Social Research*, 11. Cambridge: SSRG Publications.

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: 471–92.

Blau, F.D., and L.M. Kahn (2000). 'Gender Differences in Pay'. *Journal of Economic Perspectives*, 14: 75–99.

Blau, F. D., and L.M. Kahn (2006). 'The US Gender Pay Gap in the 1990s: Slowing Convergence'. *Industrial and Labor Relations Review*, 60: 45–66.

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: 29–71.

Brynin, M., and F. Perales (2015). 'Gender Wage Inequality: The De-gendering of the Occupational Structure'. *European Sociological Review*, 32(1): 1–13.

Butler, R.J. (1987). 'New Indices of Segregation'. *Economics Letters*, 24: 359–62.

Butler, R.J., and J.B. McDonald (1987). 'Interdistributional Income Inequality'. *Journal of Business and Economic Statistics*, 5: 13–18.

Chantreuil, F., and A. Trannoy (2013). 'Inequality Decomposition Values: The Trade-Off between Marginality and Consistency'. *Journal of Economic Inequality*, 11: 83–98.

Cotter, D.A., J.M. Hermsen, and R. Vanneman (2004). 'Gender Inequality at Work, the American People: Census 2000'. New York: 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: 967–88.

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: 1001–44.

Duncan, O.D., and B. Duncan (1955). 'A Methodological Analysis of Segregation Indexes'. *American Sociological Review*, 20: 210–17.

Fortin, N.M., T. Lemieux, and S. Firpo (2011). 'Decomposition Methods in Economics'. In O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, 4. Amsterdam: North Holland.

Foster, J.E (1985). 'Inequality Measurement'. In W.P. Young, *Fair Allocation, Proceedings of Symposia in Applied Mathematics*. Providence, RI: American Mathematical Society.

Foster, J.E., J. Greer, and E. Thorbecke (1084). 'A Class of Decomposable Poverty Measures'. *Econometrica*, 52: 761–66.

Foster, J.E., and A.F. Shorrocks (1988). 'Poverty Orderings'. *Econometrica*, 56: 173–77.

Goldin, C. (2014). 'A Grand Gender Convergence: Its Last Chapter'. *American Economic Review*, 104: 1091–19.

Gradín, C. (2013). 'Conditional Occupational segregation of Minorities in the US'. *Journal of Economic Inequality*, 11: 473–93.

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: 457–72.

Gunderson, M. (1989). 'Male–Female Wage Differentials and Policy Responses'. *Journal of Economic Literature*, 27: 46–72.

Hornseth, R.A. (1947). 'A Note on "The Measurement of Ecological Segregation" by Julius Jahn, Calvin F. Schmid, Clarence Schrag'. *American Sociological Review*, 12: 601–06.

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: 555–78.

Hutchens, R.M. (2009). 'Occupational Segregation with Economic Disadvantage: An Investigation of Decomposable Indexes'. *Research on Economic Inequality*, 17: 99–120.

Hutchens, R.M. (2012). 'Measuring Segregation when Hierarchy Matters'. IZA Discussion Paper 6667. Bonn: IZA.

Jahn, J., C.F. Schmid, and C. Schrag (1947). 'The Measurement of Ecological Segregation'. *American Sociological Review*, 12: 293–303.

James, D.R., and K.E. Taeuber (1985). 'Measures of Segregation'. *Sociological Methodology*, 15, 1–32.

Kakwani, N.C. (1980). *Income Inequality and Poverty: Methods of Estimation and Policy Applications, Report 10092*. Oxford: Oxford University Press.

Karmel, T., and M. MacLachlan (1988). 'Occupational Sex Segregation—Increasing or Decreasing?'. *The Economic Record*, 64: 187–95.

Lambert, P. (2002). 'The Distribution and Redistribution of Income'. 3rd edition. Manchester: Manchester University Press.

Lasso de la Vega, C., and O. Volij (2014). 'Segregation, Informativeness and Lorenz Dominance'. *Social Choice and Welfare*, 43: 547–64.

Le Breton, M., A. Michelangeli, and E. Peluso (2012). 'A Stochastic Dominance Approach to the Measurement of Discrimination'. *Journal of Economic Theory*, 147: 1342–50.

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: 363–68.

Mandel, H., and M. Semyonov (2014). 'Gender Pay Gap and Employment Sector: Sources Of Earnings disparities in the United States, 1970–2010'. *Demography*, 51: 1597–618.

Massey, D.S., and N.A. Denton, N.A. (1988). 'The Dimensions of Residential Segregation'. *Social Forces*, 67: 281–315.

Pfähler, W. (1983). 'Measuring Redistributional Effects of Tax Progressivity by Lorenz Curves'. *Journal of Economics and Statistics*, 198: 237–49.

Plotnick, R. (1981). 'A Measure of Horizontal Inequity'. *Review of Economics and Statistics*, 63: 283–88.

Reardon, S.F. (2009). 'Measures of Ordinal Segregation'. *Research on Economic Inequality*, 17: 129–55.

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2015). 'Integrated Public Use Microdata Series: Version 6.0'. Minneapolis: University of Minnesota.

Schechtman, E., and S. Yitzhaki, S. (1987). 'A Measure of Association based on Gini's Mean Difference'. *Communications in Statistics*, 16: 207–31.

Schechtman, E., and S. Yitzhaki, S. (1999). 'On the Proper Bounds of the Gini Correlation'. *Economics Letters*, 63: 133–38.

Shorrocks, A.F. (2013). 'Decomposition Procedures for Distributional Analysis: A Unified Framework based on the Shapley Value'. *Journal of Economic Inequality*, 11: 99–126.

Silber, J. (1989). 'On the Measurement of Employment Segregation'. *Economics Letters*, 30: 237–43.

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: 799–811.

Treiman, D., and H. Hartmann (1981). *Women, Work and Wages: Equal Pay for Jobs of Equal Value*. Washington DC: National Academy Press.

Volij, O. (2016), 'Segregation: Theoretical Approaches'. In C. d'Ambrosio (ed.), *Handbook of Research on Economic and Social Well-Being*. Cheltenham: 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: 89–112.

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: 380–92.

Appendix 1: Description of workers' characteristics (ACS, 2014)

- **Location:** metropolitan statistical area, with one category for non-metropolitan areas.
- **Marital status:** married with spouse present; married with spouse absent; separated; divorced; widowed; never married/single.
- **Number of children** in the household under and above age 5.
- **Age interval:** 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.
- **Hispanic ethnicity:** not Hispanic; Mexican; Puerto Rican; Cuban; Costa Rican; Guatemalan; Honduran; Nicaraguan; Panamanian; Salvadoran; Other Central American; Argentinean; Bolivian; Chilean; Colombian; Ecuadorian; Paraguayan; Peruvian; Uruguayan; Venezuelan; Other South American; Spaniard; Dominican; other.
- **Migration profile**
    - **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 states; between non-contiguous states; from abroad; or remained in the same house.
    - **Years of residence in the US:** 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.
- **Education:**
    - **Attained education:** no schooling completed; nursery school, preschool; kindergarten; grade1; …; grade 11; grade 12, no diploma; regular high school diploma; GED (Genderal Educational Development) or alternative credential; some college, but less than 1 year; 1 or more years of college credit, no diploma; associate's degree, type not specified; bachelor's degree; master's degree; professional degree beyond a bachelor's; doctoral degree.
    - **Field of college degree:** 169 categories, after merging four fields with few female observations (military technologies and mining and mineral, naval architecture, and nuclear engineering).

Appendix 2: Measuring conditional segregation

We explain here in more detail the procedure used to obtain the aggregate decomposition of the level and trend in segregation. The detailed contribution is obtained using a Shapley decomposition described below. The procedure is very similar for segregation and low-pay segregation. Bootstrapping over the entire process (including the logit regression), we obtain the corresponding standard errors.

a Decomposition of (low-pay) segregation in each year

To some extent, the observed level of gender 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), and geographical location, which potentially affect their opportunities in the local labour market. Alternatively, segregation could also reflect that people of similar characteristics work in different occupations depending on their gender. The identification of this unexplained term with discrimination in the labour market has to be cautious, as in the analysis of wage differentials, because it may also reflect gender differences in unobserved characteristics (e.g. job preferences, skills). Similarly, the explained part could also reflect anticipated discrimination in the labour market by the group discriminated against, or discrimination that occurs prior to entering the labour market.

To disentangle the importance of the explained and unexplained terms, we follow here the approach in Gradín (2013). The aggregate decomposition is obtained by 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). Alternatively, we can give the reference's conditional occupational distribution to the comparison group (by reweighting the former to obtain the distribution of characteristics of the latter).

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

$$f_j^i(X) = \int_{x\in\Omega_X} f_j^i(X=x) f^i(x) dx.$$

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

$$f_j^\gamma = \int_{x\in\Omega_X} f_j^c(X=x) f^r(x) dx = \int_{x\in\Omega_z} f_j^c(X=x) f^c(x) \Psi_X dx; \qquad \text{where} \qquad \Psi_X = \frac{f^r(x)}{f^c(x)}.$$

f_j^{γ} can be obtained by reweighting 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:

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

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 $Pr(c|x) = 1 - Pr(r|x)$.[1] $S(f^{\gamma}, f^r)$ 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:

$$S(f^c, f^r) = S^E(f) + S^U(f),$$

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^{γ}):

$$S^E(f) = S(f^c, f^r) - S(f^{\gamma}, f^r).$$

The second term is the (conditional) segregation that remains unexplained after equalizing characteristics in both groups, reflecting only cross-group differences in the conditional distribution across occupations (shifting from f^{γ} to f^r; note that $S(f^r, f^r) = 0$):

$$S^U(f) = S(f^{\gamma}, f^r) - S(f^r, f^r) = S(f^{\gamma}, f^r).$$

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.[2] Starting with the case in which all estimated coefficients in the logit regression are set to 0, several reweighting factors were obtained by 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).[3] We use the same approach for the decomposition of $S(g^c, g^r)$.

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

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

3 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.

b Decomposition of a change in (low-pay) segregation

Let f^{it} be the unconditional occupational distribution of group i in year t ($t = 0{,}1$), and $f^{i\lambda}$ the counterfactual for group i that uses its marginal distribution of characteristics in year 0, and its conditional occupational distribution in year 1:

$f_j^{i\lambda} = \int_{x\in\Omega_z} f_j^{i1}(X = x) f^{i0}(x)dx = \int_{x\in\Omega_z} f_j^{i1}(X = x) f^{i1}(x)\varphi_X dx$; where $\varphi_X = \frac{f^{i0}(x)}{f^{i1}(x)}$.

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

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

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 1:

$$\Delta S^E(f) = S(f^{c1}, f^{r1}) - S\big(f^{c\lambda}, f^{r\lambda}\big).$$

The second term is the unexplained effect, i.e. the change in conditional occupational distributions of both groups over time, evaluated using each gender's characteristics in year 0:

$$\Delta S^U(f) = S\big(f^{c\lambda}, f^{r\lambda}\big) - S(f^{c0}, f^{r0}).$$

In the case of low-pay segregation, there is a third factor to consider: the change in the ranking of occupations by average earnings 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 $h_j^{i\lambda}$ 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 low-pay 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 1). ΔS^U is the unexplained effect, the result of the change in conditional occupational distributions over time (evaluated using w^* and characteristics at year 0):

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

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

$$\Delta S^E(h) = S(h^{c1}, h^{r1}) - S\big(h^{c\lambda}, h^{r\lambda}\big).$$

$$\Delta S^U(h) = S\big(h^{c\lambda}, h^{r\lambda}\big) - S(h^{c0}, h^{r0}).$$

An alternative decomposition can also be obtained with a counterfactual that uses the characteristics of year 1 and the conditional distribution of year 0. A Shapley decomposition produces the detailed decomposition in all cases.

Appendix 3: Tables

Table A1: Selected characteristics by gender, 1960–2014

	% 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 gender, 2014. Fields with largest over-/under-representation 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 and Literature	3.7	2.0	Economics	1.2	2.9
Social Work	1.6	0.3	Business Management and 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 and Government	1.8	3.1
Communication Disorders Sciences and Services	0.8	0.1	History	1.4	2.6

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

Table A3: Decomposition of (low-pay) occupational segregation of women by characteristics, 2014 (alternative counterfactual)

		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
	No. of 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

Notes: Counterfactual: men's 2014 distribution reweighted to reproduce women's 2014 characteristics. See Section 4 and this Appendix for details about the variables.

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

Table A4: Decomposition of (low-pay) occupational segregation of women, 1960–2014

	Segregation							Low-pay segregation						
Gini	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 status	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
No. of 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 Status	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
No. of 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 Section 4 and this Appendix for details about the variables used in each category (among other things, it excludes the field of degree for the sake of comparability).

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

Table A5: Decomposition of (low-pay) occupational segregation (Gini) trends, 1960/70–2014. Difference between (low-pay) segregation in final and initial years. Alternative counterfactual.

		Segregation 1960–2014	%	Low-pay segregation 1970–2014	%
Change		−0.149	100	−0.358	100
		(0.001)		(0.002)	
Explained by Earnings structure		-		−0.126	35.1
				(0.001)	
Unexplained		−0.120	80.3	−0.102	28.5
		(0.001)		(0.004)	
Explained by characteristics		−0.029	19.7	−0.130	36.3
		(0.001)		(0.004)	
	Marital status	−0.011	7.4	−0.099	27.6
		(0.000)		(0.002)	
	Children	−0.006	3.9	−0.017	4.6
		(0.000)		(0.001)	
	Age	0.001	−0.6	0.043	−12.1
		(0.000)		(0.001)	
	Race	−0.003	2.2	−0.011	3.0
		(0.000)		(0.002)	
	Hispanic ethn.	−0.001	0.7	−0.014	3.9
		(0.000)		(0.002)	
	Migration profile	−0.001	0.9	−0.019	5.2
		(0.000)		(0.002)	
	Education	−0.008	5.2	−0.015	4.1
		(0.000)		(0.004)	

Notes: Unexplained effect evaluated using final characteristics; explained effect evaluated using the initial conditional occupational distribution. In low-pay segregation, earnings structure uses 2010 ranking of occupations by earnings, also used to estimate explained and unexplained effects. In parentheses, bootstrap standard errors over the entire procedure (175–200 replications). See Section 4 and this Appendix for details about characteristics included in each category.

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

Table A6: Decomposition of (low-pay) occupational segregation of women by characteristics, 1960–2014

	Segregation							Low-pay segregation						
Gini	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

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

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