Population Average Gender Effects*

Tymon Słoczyński†

This version: June 4, 2011

ABSTRACT: In this paper I develop a consistent estimator of the population average treatment effect (PATE) which is based on a nonstandard version of the Oaxaca-Blinder decomposition. What follows, I extend the recent literature which has utilized the treatment effects framework to reinterpret this technique, and propose an alternative solution to the reference group choice problem that is inherent therein. Moreover, I use the new estimator of this paper and its semiparametric extension to decompose gender wage differentials with the Current Population Survey (CPS) data, while providing separate estimates of the average gender effect on men, women, and the whole population.

* I am grateful to Arun Advani, Joshua Angrist, Anna Baranowska, Thomas Crossley, Patrick Kline, and seminar and conference participants in Dublin, Kraków, Odense, and Warsaw for their comments and suggestions at different stages of this project. I would also like to thank Clifford and Mary Corbridge Trust, Cambridge European Trust, and Faculty of Economics, University of Cambridge for financial support which allowed me to undertake graduate studies at the University of Cambridge where this project was started. Any remaining errors are my sole responsibility.

† Warsaw School of Economics, e-mail: t.sloczynski@ci.edu.pl

I. Introduction

Since the seminal contributions of Oaxaca (1973) and Blinder (1973), one of the most extensive strands of literature in empirical labor economics has aimed at decomposing intergroup wage differentials into components attributable to group composition and net effects of group membership, often referred to as the explained component and the unexplained component, respectively. Recently, several researchers (Barsky, Bound, Charles, and Lupton 2002; Melly 2006; Fortin, Lemieux, and Firpo 2011) have noted that the unexplained component in the most basic version of the Oaxaca-Blinder decomposition is also a consistent estimator of the population average treatment effect on the treated (PATT), thus incorporating this technique into the abundant treatment effects literature. Moreover, Kline (2011) has shown that such an estimator of the PATT satisfies a beneficial "double robustness" property (Robins, Rotnitzky, and Zhao 1994), since it constitutes a propensity score reweighting estimator based on a linear model for the treatment odds. At the same time, semiparametric and nonparametric extensions of the Oaxaca-Blinder decomposition have been proposed to account for possible nonlinearity of existing wage structures (Barsky, Bound, Charles, and Lupton 2002; Frölich 2007; Mora 2008; Ñopo 2008).

On the other hand, one of the long-lasting discussions in the Oaxaca-Blinder decomposition literature has concerned the choice of the reference coefficients used to calculate the explained component of the wage differential. Both Oaxaca (1973) and Blinder (1973), whose main goal was to study U.S. gender wage gaps, referred to the problem of choosing either male or female wage structure to compute the explained component as an "index number problem", thus suggesting this choice to be ambiguous. As Oaxaca (1973, p. 695) put it: "On the basis of either of two assumptions, we can estimate the male-female wage ratio that would exist in the absence of discrimination: If there were no discrimination, 1) the wage structure currently faced by females would also apply to males; or 2) the wage structure currently faced by males would also apply to females". Such a tendency to think of the reference coefficients as the "nondiscriminatory" or "competitive" wage structure remained in the literature for long and served as the basis for the most popular alternative solutions to the reference group choice problem (Reimers 1983; Cotton 1988; Neumark 1988; Oaxaca and Ransom 1994; Fortin 2008). Only recently did the analogy between the Oaxaca-Blinder decomposition literature and the treatment effects literature allow Fortin, Lemieux, and Firpo (2011) to distinguish reference wage structures based on the assumption of "simple counterfactual treatment" (the original propositions of Oaxaca 1973 and Blinder 1973) and more general reference wage structures, typically associated with some general equilibrium considerations which retain the notion of "nondiscriminatory" wage coefficients.

In this paper I extend this recent literature which has utilized the treatment effects framework to reinterpret the Oaxaca-Blinder decomposition, and show how this framework can help us better understand the well-known solutions to the reference group choice problem, while I propose an alternative solution to it as well. Namely, I develop a consistent estimator of the population average treatment effect (PATE) which is based on a nonstandard version of the Oaxaca-Blinder decomposition. This alternative estimator uses a linear combination of the regression coefficients for both subpopulations (treated and nontreated, men and women, unionized and nonunionized workers, etc.) as the reference wage structure. While such an approach is similar to the well-known propositions of Reimers (1983) and Cotton (1988), these coefficients are weighted in a nonstandard way, i.e. the sample proportion of group one (two) is used to weight the coefficients for group two (one). Although such a weighting procedure may at first look counterintuitive[1], the treatment effects framework provides a clear

[1] It indeed sometimes does. A similar estimator (not including the explained component) had already been used by Duncan and Leigh (1985) to estimate union wage premiums, but such an

rationale for this proposition. Precisely, the role of each group's wage structure is to serve as counterfactual for the other group (the assumption of "simple counterfactual treatment" in Fortin, Lemieux, and Firpo 2011), so the unexplained component of the Oaxaca-Blinder decomposition can be used as a consistent estimator of the PATE if and only if such a reverse weighting procedure is applied[2].

Similarly, I reinterpret the propositions of Reimers (1983) and Cotton (1988), and show that the former approach estimates the arithmetic mean of the population average treatment effect on the treated (PATT) and the population average treatment effect on the nontreated (PATN), while the latter approach estimates a weighted average of these two effects, but attaches reversed weights to both of them. I provide a similar reinterpretation of the recent proposition of Fortin (2008) and Jann (2008) as well. Consequently, all these approaches as well as the new estimator of this paper can actually be motivated by the assumption of "simple counterfactual treatment" (Fortin, Lemieux, and Firpo 2011), although I believe it is only the new estimator whose estimand should typically be interesting for an applied researcher. In other words, this paper shows that meaningful threefold (or generalized) decompositions of intergroup wage differentials can be carried out without any reference to the "nondiscriminatory" wage structure, although there has been a strong tendency in the literature to claim otherwise. The only extension of the Oaxaca-Blinder decomposition which would then retain its general equilibrium considerations is the approach of Neumark (1988), although this procedure has recently been criticized by Fortin (2008),

---

approach was heavily criticized – as "not a very intuitive procedure" – by Oaxaca and Ransom (1988, p. 143). Of course, neither of these papers made any reference to the treatment effects framework, which was then practically unknown in economics, while showing that such an estimator is consistent for the PATE is the major contribution of the present paper.

[2] Unless, of course, both groups are of equal size.

Jann (2008), and Elder, Goddeeris, and Haider (2010), since its reference coefficients are estimated in a pooled model without the group membership dummy, while such an exclusion can severely bias regression coefficients on other covariates (omitted variables bias).

Moreover, I also contribute to the recent literature which has provided semiparametric and nonparametric extensions of the Oaxaca-Blinder decomposition in order to account for possible nonlinearity of existing wage structures (Barsky, Bound, Charles, and Lupton 2002; Frölich 2007; Mora 2008; Ñopo 2008), and propose a simple combination of stratification (on the propensity score) and the Oaxaca-Blinder decomposition. While a combination of stratification and linear regression performed best in the well-known paper by Dehejia and Wahba (1999) and is suggested by Imbens and Wooldridge (2009, p. 41) as "one of the more attractive estimators [of average treatment effects] in practice", their estimator does not allow for within-strata heterogeneity of treatment effects. This shortcoming is addressed by a combination of stratification and Oaxaca-Blinder which I propose in this paper.

Finally, I provide an empirical example which uses the new estimator of this paper and its semiparametric extension to decompose gender wage differentials with Current Population Survey (CPS) data, for each year from 1983 to 2010. First, I replicate the results of Elder, Goddeeris, and Haider (2010) who have recently compared – both theoretically and empirically – linear regression, both simple estimators of Oaxaca (1973) and Blinder (1973), and the extension of the Oaxaca-Blinder decomposition proposed by Neumark (1988). Second, I utilize their sample and variable selections as well as the new estimator of this paper to provide estimates of an estimand which I refer to as the population average gender effect (PAGE). I also use a combination of stratification and different versions of the Oaxaca-Blinder decomposition to provide separate estimates of the (semiparametrically estimated) average gender effect on men, women, and the whole population. To the best of my knowledge this is the first paper to clarify the distinction between these three estimands (which I call PAGM, PAGW, and PAGE, respectively), and provide separate estimates for each of them.

The remainder of this paper is organized as follows. In the next section, I review the treatment effects framework and the Oaxaca-Blinder decomposition, and present the major equivalence result established in the recent literature. In Section III, I develop a consistent estimator of the population average treatment effect (PATE) which is based on the Oaxaca-Blinder decomposition, reinterpret some of the well-known solutions to the reference group choice problem which is inherent in this technique, and propose a semiparametric extension of the Oaxaca-Blinder decomposition which incorporates stratification on the propensity score. In Section IV, I provide an empirical application to U.S. gender wage differentials using CPS data. Finally, I conclude and review my findings in Section V.

II. Background

A. The Treatment Effects Framework

In this subsection I provide a standard description of the treatment effects framework with its key component, the so-called Rubin Causal Model (RCM). My description borrows notation from a recent survey by Imbens and Wooldridge (2009).

Consider a population of individuals who belong to either of two mutually exclusive groups indexed by $W_i \in \{0,1\}$. The division of individuals into these two groups of interest is based on whether an individual has been exposed to regime referred to as *treatment*. If unit i has been exposed to treatment, he or she is called a treated individual and $W_i = 1$. If this unit has not been exposed to treatment, he or she is called a nontreated (or control) individual and $W_i = 0$. The population consists of N individuals, indexed by $i = 1, \dots, N$, and $\sum_{i=1}^{N} W_i = N_1$,

$N_0 + N_1 = N$, i.e. there are exactly N_1 treated individuals and exactly N_0 nontreated individuals in the population. For each individual i, a column vector of covariates (control variables) X_i is also observed.

The treatment effects framework is based, however, to a large extent on an important notion of *potential outcomes*. Let there be a variable, the outcome, whose codetermination by the treatment is of key interest. For each individual i, the realized (observable) outcome is denoted by Y_i. It is assumed, however, there exist two potential outcomes which could have been observed, $Y_i(0)$ and $Y_i(1)$, dependent on whether the individual has been exposed to treatment. If $W_i = 1$, then it is only $Y_i(1)$, *the treated outcome*, which is realized (and observed). On the other hand, if $W_i = 0$, then we can only observe $Y_i(0)$, *the nontreated outcome*. Consequently, it is the group membership of each individual i which causes one of the potential outcomes to become observable (realized) and the other potential outcome to become counterfactual, while the realized outcome, Y_i, can be written as $Y_i = Y_i(W_i) = Y_i(1) \cdot W_i + Y_i(0) \cdot (1 - W_i)$.

Having described the notion of potential outcomes, it is useful to introduce standard estimands which are of interest in the treatment effects literature. For each individual i, the difference between his or her treated outcome and the according nontreated outcome, $Y_i(1) - Y_i(0)$, is referred to as the individual-specific *treatment effect*. In standard settings, these effects are intended to be averaged over certain (sub)populations which are of interest in a given study. For the whole population (which includes both treated and nontreated individuals), the so-called population average treatment effect (PATE) is the appropriate estimand:

$$\tau_{PATE} = \mathbb{E}[Y_i(1) - Y_i(0)]$$

Alternatively, individual treatment effects can be averaged over the subpopulation of treated individuals (the population average treatment effect on the treated, PATT, is then obtained) or the subpopulation of nontreated individuals (to obtain the population average treatment effect on the nontreated, PATN):

$$\tau_{PATT} = \mathbb{E}[Y_i(1) - Y_i(0)|W_i = 1]$$

$$\tau_{PATN} = \mathbb{E}[Y_i(1) - Y_i(0)|W_i = 0]$$

Importantly, identification and estimation of various average treatment effects (the PATE, the PATT, and the PATN) is not a straightforward task. Suppose we focus on τ_{PATT}. Consequently, we are only concerned with the treated subpopulation, and wish to average individual-specific treatment effects over the subpopulation of treated individuals. We do therefore observe the treated outcomes (= the realized outcomes) for all the individuals of interest, but we do not observe the nontreated outcome for any of them. Some identifying assumptions are needed. Suppose, for example, we wish to assume that the treated individuals would have received, on average, what the nontreated individuals do actually receive (on average) now, had they not been exposed to treatment. Such a naïve solution is biased, however, if selection to treatment is present (see, e.g., Cobb-Clark and Crossley 2003):

$$\begin{aligned}
\mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0] &= \mathbb{E}[Y_i(1)|W_i = 1] - \mathbb{E}[Y_i(0)|W_i = 0] \\
&= \mathbb{E}[Y_i(1)|W_i = 1] - \mathbb{E}[Y_i(0)|W_i = 0] \\
&\quad + \{\mathbb{E}[Y_i(0)|W_i = 1] - \mathbb{E}[Y_i(0)|W_i = 1]\} \\
&= \mathbb{E}[Y_i(1) - Y_i(0)|W_i = 1] + \{\mathbb{E}[Y_i(0)|W_i = 1] - \mathbb{E}[Y_i(0)|W_i = 0]\} \\
&= \tau_{PATT} + \{\mathbb{E}[Y_i(0)|W_i = 1] - \mathbb{E}[Y_i(0)|W_i = 0]\}
\end{aligned} \tag{1}$$

Indeed, such a naïve estimator, $\mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0]$, does not separate the estimand of interest, τ_{PATT}, from selection bias, $\mathbb{E}[Y_i(0)|W_i = 1] - \mathbb{E}[Y_i(0)|W_i = 0]$. Precisely, this second component will not be equal to zero if the nontreated outcomes of treated and nontreated individuals are, on average, different. In observational studies, there is typically little reason to believe they are not.

Consequently, identification of various average treatment effects must proceed differently. There are two main strands in the treatment effects literature, typically referred to as selection on observables and selection on unobservables, and such a division is predicated on the identifying assumptions they employ. The present paper is only concerned with selection on observables (aka unconfoundedness), a strand whose key assumptions, unconfoundedness and overlap, were jointly referred to by Rosenbaum and Rubin (1983) as strong ignorability. Under unconfoundedness, it is assumed there are no unobserved individual characteristics which are associated both with the group membership (treatment) and the potential outcomes. In other words:

$$W_i \perp\!\!\!\perp \big(Y_i(0), Y_i(1)\big) | X_i$$

Moreover, the overlap assumption ensures there are both treated and nontreated individuals for all possible values of the control variables:

$$0 < \mathrm{pr}(W_i = 1 | X_i = x) < 1, \text{ for all } x$$

Importantly, strong ignorability allows for identification of various average treatment effects (see Imbens and Wooldridge 2009, p. 26-27), and these estimands of interest are typically estimated using regression methods (see Angrist and Pischke 2009), methods based on the propensity score (see Rosenbaum and Rubin 1983 for the seminal contribution and Hirano, Imbens, and Ridder 2003 for an important theoretical study), matching on X_i (see Abadie and Imbens 2006 for a recent theoretical contribution), and various combinations of these methods. A good survey is provided by Imbens and Wooldridge (2009). Recent contributions of Barsky, Bound, Charles, and Lupton (2002), Melly (2006), and Fortin, Lemieux, and Firpo (2011) have noted that a consistent estimator of the population average treatment effect on the treated (PATT) is also provided by the Oaxaca-Blinder decomposition.

B. The Oaxaca-Blinder Decomposition

The econometric method typically referred to as the Oaxaca-Blinder decomposition was first used in labor economics by Oaxaca (1973) and Blinder (1973), although its history goes back to the work of Kitagawa (1955) in demography. The method aims at decomposing intergroup differentials in outcomes into components attributable to group composition and net effects of group membership, often referred to as the explained component and the unexplained component, respectively. My description of this method continues to use notation from the previous subsection and borrows a lot from a survey by Fortin, Lemieux, and Firpo (2011).

Let the model for outcomes be linear and separable in observable and unobservable characteristics, and let the regression coefficients be different for both groups of interest:

$$Y_i = X_i\beta_1 + \upsilon_{1i} \text{ if } W_i = 1;\ Y_i = X_i\beta_0 + \upsilon_{0i} \text{ if } W_i = 0$$

where $\mathbb{E}[\upsilon_{1i}|X_i] = \mathbb{E}[\upsilon_{0i}|X_i] = 0$. Suppose we are concerned with decomposing gender wage differentials. Then, let $W_i = 1$ for males and $W_i = 0$ for females, and let Y_i be the observed (log) wages. The gender wage differential, $\mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0]$, can then be decomposed into two components:

$$\begin{aligned}
\mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0] &= \mathbb{E}[\mathbb{E}(Y_i|X_i, W_i = 1)|W_i = 1] - \mathbb{E}[\mathbb{E}(Y_i|X_i, W_i = 0)|W_i = 0] \\
&= (\mathbb{E}[X_i|W_i = 1]\beta_1 + \mathbb{E}[\upsilon_{1i}|W_i = 1]) - (\mathbb{E}[X_i|W_i = 0]\beta_0 + \mathbb{E}[\upsilon_{0i}|W_i = 0]) \\
&= \mathbb{E}[X_i|W_i = 1]\beta_1 - \mathbb{E}[X_i|W_i = 0]\beta_0 \\
&= \mathbb{E}[X_i|W_i = 1]\beta_1 - \mathbb{E}[X_i|W_i = 0]\beta_0 + \mathbb{E}[X_i|W_i = 1]\beta_0 - \mathbb{E}[X_i|W_i = 1]\beta_0 \\
&= \mathbb{E}[X_i|W_i = 1](\beta_1 - \beta_0) + (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])\beta_0
\end{aligned} \tag{2}$$

where the first element, $\mathbb{E}[X_i|W_i = 1](\beta_1 - \beta_0)$, reflects gender differences in regression coefficients or net effects of group membership, and is often referred to as *the unexplained component*, while the second element, $(\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])\beta_0$, reflects gender differences in mean covariate values or group composition, and is often referred to as *the explained component*. Similarly, an alternative decomposition can be performed:

$$\begin{aligned}\mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0] &= \mathbb{E}[X_i|W_i = 1]\beta_1 - \mathbb{E}[X_i|W_i = 0]\beta_0 \\ &= \mathbb{E}[X_i|W_i = 1]\beta_1 - \mathbb{E}[X_i|W_i = 0]\beta_0 + \mathbb{E}[X_i|W_i = 0]\beta_1 - \mathbb{E}[X_i|W_i = 0]\beta_1 \\ &= \mathbb{E}[X_i|W_i = 0](\beta_1 - \beta_0) + (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])\beta_1\end{aligned} \tag{3}$$

Clearly, the difference between Equations 2 and 3 rests upon using alternate reference coefficients (reference group) to calculate the explained component of the decomposition as well as measuring the distance between the two regression functions, $\beta_1 - \beta_0$, for a different set of covariate values.

There has been a long-lasting tendency in the Oaxaca-Blinder decomposition literature, originating from the seminal papers of Oaxaca (1973) and Blinder (1973), to claim that the choice of the reference group in this context (choosing between Equations 2 and 3) is necessarily ambiguous (see, e.g., Oaxaca 1973 and Elder, Goddeeris, and Haider 2010). Although I shall soon show that such a claim is actually a misunderstanding, the standard response in the literature has been to suggest alternative, noticeably less simple sets of reference coefficients to solve this reference group choice problem. Such an approach has sometimes been referred to as "generalized Oaxaca-Blinder", while it involves performing an alternative decomposition of the gender wage differential:

$$\begin{aligned}\mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0] &= \mathbb{E}[X_i|W_i = 1]\beta_1 - \mathbb{E}[X_i|W_i = 0]\beta_0 \\ &= \mathbb{E}[X_i|W_i = 1]\beta_1 - \mathbb{E}[X_i|W_i = 0]\beta_0 + \mathbb{E}[X_i|W_i = 1]\beta^* - \mathbb{E}[X_i|W_i = 1]\beta^* \\ &\quad + \mathbb{E}[X_i|W_i = 0]\beta^* - \mathbb{E}[X_i|W_i = 0]\beta^* \\ &= \{\mathbb{E}[X_i|W_i = 1](\beta_1 - \beta^*) + \mathbb{E}[X_i|W_i = 0](\beta^* - \beta_0)\} \\ &\quad + (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])\beta^*\end{aligned} \tag{4}$$

where β^* is the set of reference coefficients, typically referred to as the "nondiscriminatory" or "competitive" wage structure.

Several important papers have been devoted to suggesting alternative sets of reference coefficients for Equation 4 (although only one paper, Neumark 1988, provided any theoretical rationale for its suggestion), and these coefficients have often been formulated as $\beta^* = \lambda \cdot \beta_1 + (1-\lambda) \cdot \beta_0$ where $\lambda \in [0,1]$ is a weighting factor. If $\lambda = 0$, then women are used as the reference group, $\beta^* = \beta_0$, and Equation 4 simplifies to Equation 2. Similarly, if $\lambda = 1$, then men are used as the reference group, $\beta^* = \beta_1$, and Equation 4 simplifies to Equation 3. Reimers (1983, p. 573) claimed that "neither group's observed wage-offer function would be likely to exist in a non-discriminatory world. Instead, the no-discrimination wage function lies somewhere between them", and proposed therefore to use $\lambda = \frac{1}{2}$ (or $\beta^* = \frac{1}{2}\beta_1 + \frac{1}{2}\beta_0$). In the context of the black-white wage differential, Cotton (1988, p. 239) suggested that "the nondiscriminatory wage structure will be closer to the current white wage structure than to the current black wage structure", and operationalized this assumption by using $\lambda = \frac{N_1}{N_1+N_0}$, the population proportion of group one. Alternatively, Neumark (1988) developed a simple model of Beckerian discrimination, and showed that identification of the nondiscriminatory wage structure is ensured, for example, if the utility function of the representative producer is homogeneous of degree zero with respect to male and female labor inputs (within each type of labor). This wage structure can be approximated by $\beta^* = \beta_p$, the set of regression coefficients in a pooled model without the group membership dummy (Neumark 1988). Although this solution to the reference group choice problem was later popularized by Oaxaca and Ransom (1994) and has been the most popular alternative to the original Oaxaca-Blinder decomposition (see Weichselbaumer and Winter-Ebmer 2005, Table 2), it has recently been criticized by Fortin (2008), Jann (2008), and Elder, Goddeeris, and Haider (2010), since exclusion of the group membership dummy can bias coefficients on other covariates (omitted variables bias). Consequently, Fortin (2008) and Jann (2008) have proposed to use a pooled model with the group membership dummy as reference. Of course, as noted by Fortin (2008) and Fortin, Lemieux, and Firpo (2011), the unexplained component in such a decomposition is equal to the coefficient on the group membership dummy in a simple linear regression.

C. The Oaxaca-Blinder Unexplained Component as an Estimator of the PATT

As already noted, recent papers of Barsky, Bound, Charles, and Lupton (2002), Melly (2006), and Fortin, Lemieux, and Firpo (2011) have shown that the Oaxaca-Blinder decomposition can also be used to consistently estimate the population average treatment effect on the treated (PATT), while Kline (2011) has provided an interpretation of Oaxaca-Blinder as a propensity score reweighting estimator which enjoys therefore an important property of "double robustness" (Robins, Rotnitzky, and Zhao 1994). Precisely, Equations 1 and 2 are easily shown to be equivalent if the model for outcomes is linear:

$$\begin{aligned}
&\mathbb{E}[X_i|W_i = 1](\beta_1 - \beta_0) + (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])\beta_0 \\
&\qquad = \mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0] = \mathbb{E}[Y_i(1)|W_i = 1] - \mathbb{E}[Y_i(0)|W_i = 0] \\
&\qquad = \tau_{PATT} + \{\mathbb{E}[Y_i(0)|W_i = 1] - \mathbb{E}[Y_i(0)|W_i = 0]\}
\end{aligned}$$

Consequently, the population average treatment effect on the treated (PATT) is measured as the distance between the two regression functions, $\beta_1 - \beta_0$, at the mean covariate values in the treated subpopulation, $\tau_{PATT} = \mathbb{E}[X_i|W_i = 1](\beta_1 - \beta_0)$.

In the next section I will develop a consistent estimator of the population average treatment effect (PATE) which is based on a similar reasoning, and provides therefore a nonstandard solution to the reference group choice problem inherent in the Oaxaca-Blinder decomposition. I will also show that the underlying estimands of the well-known estimators proposed by Reimers (1983) and Cotton (1988) have been unknown so far, while they are not likely to be interesting for an applied researcher and should therefore be avoided. I will also provide a reinterpretation of the underlying estimand in Fortin (2008) and Jann (2008) as well as an easily estimable semiparametric extension of the Oaxaca-Blinder decomposition.

III. A Reanalysis of the Oaxaca-Blinder Decomposition Methodology

As described in the previous section, it has recently been shown (Barsky, Bound, Charles, and Lupton 2002; Melly 2006; Fortin, Lemieux, and Firpo 2011) that the unexplained component in the most basic version of the Oaxaca-Blinder decomposition (Equation 2) is also a consistent estimator of the population average treatment effect on the treated (PATT). In the context of gender wage differentials, such a result calls for a reinterpretation of the Oaxaca-Blinder decomposition estimands and methodology as well as a search for "gender equivalents" of the PATE, the PATT, and the PATN in the data.

Let me first introduce, however, an important "simple counterfactual treatment" assumption which has recently been proposed by Fortin, Lemieux, and Firpo (2011, p. 16), and provides the starting point for this section of the present paper. Most basically, under this assumption no general counterfactual wage structures are sought; instead, the male wage structure is used as counterfactual for females and the female wage structure is used as counterfactual for males. In other words, we attempt to compare current wages of men (women) and current wages of women (men) with similar observable characteristics, and not current wages of both genders with an assumed "nondiscriminatory" wage structure. Such an approach is exactly parallel with the treatment effects framework which compares treated and nontreated outcomes of similar individuals. Consequently, Fortin, Lemieux, and Firpo (2011) have distinguished versions of the Oaxaca-Blinder decomposition with reference wage structures based on the simple counterfactual treatment assumption (the original propositions of Oaxaca 1973 and Blinder 1973) and versions of this decomposition with more general reference wage structures, typically associated with some general equilibrium considerations (all the alternative propositions). Such a division has been, however, premature, and I show in this section that the simple counterfactual treatment assumption can be regarded as the basis of the well-known solutions to the reference group choice problem proposed by Reimers (1983), Cotton (1988), Fortin (2008), and Jann (2008), while Fortin, Lemieux, and Firpo (2011, Subsection 3.3) have suggested otherwise. Furthermore, in this section I also reinterpret the Oaxaca-Blinder decomposition estimands under the simple counterfactual treatment assumption, develop a consistent Oaxaca-Blinder estimator of the population average treatment effect (PATE), and propose a semiparametric extension of the Oaxaca-Blinder decomposition which incorporates stratification on the propensity score.

A. Population Average Gender Effects

As already noted, the simple counterfactual treatment assumption allows for identification and estimation of the population average treatment effect on the treated (PATT) using the Oaxaca-Blinder decomposition in Equation 2. In the context of gender wage differentials, it is problematic, however, to call the estimands of interest "average treatment effects" or to refer to being male or being female as treatment.

Consequently, let me introduce new names for various estimands of interest in the gender wage differentials context. Using notation from Section 2, let $W_i = 1$ for males and $W_i = 0$ for females, and let Y_i be the observed (log) wages and X_i be a column vector of control variables. Then, let $\mathbb{E}[Y_i(1) - Y_i(0) | X_i = x]$ be *the average gender effect* conditional on $X = x$, i.e. the difference between the expected (log) wage of a male with observable characteristics $X = x$ and the expected (log) wage of a female with observable characteristics

$X = x$. Dependent on the question we wish to answer, we may want to average these conditional average gender effects over the whole population, over the subpopulation of women or over the subpopulation of men. First, for the whole population (including both men and women), let me define:

$$\tau_{PAGE} = \mathbb{E}[Y_i(1) - Y_i(0)]$$

as the population average gender effect (PAGE). Clearly, this estimand is technically equivalent to the population average treatment effect (PATE). A Oaxaca-Blinder estimator of the PAGE/PATE has been unknown so far, and will be proposed in the next subsection.

Second, if we wish to average conditional average gender effects over the subpopulation of men (women), let me denote the estimands of interest:

$$\tau_{PAGM} = \mathbb{E}[Y_i(1) - Y_i(0)|W_i = 1]$$

$$\tau_{PAGW} = \mathbb{E}[Y_i(1) - Y_i(0)|W_i = 0]$$

as the population average gender effect on men (PAGM) and the population average gender effect on women (PAGW), respectively. Of course, these estimands are technically equivalent to the PATT (PATN) and the PATN (PATT). Moreover, since $W_i = 1$ for males, the PAGM can be consistently estimated with the unexplained component of the Oaxaca-Blinder decomposition in Equation 2, while the PAGW can be consistently estimated with the unexplained component of the decomposition in Equation 3 (Equations 2 and 3 are the original propositions of Oaxaca 1973 and Blinder 1973). In other words, whenever we use the male wage structure as reference in the Oaxaca-Blinder decomposition, the unexplained component provides an estimate of the population average gender effect on women (PAGW). If we use the female wage structure as reference, we estimate the population average gender effect on men (PAGM). Consequently, there is absolutely no ambiguity about choosing either male or female wage structure to compute the explained component of the Oaxaca-Blinder decomposition, in spite of many views to the contrary (see, e.g., a recent claim of Elder, Goddeeris, and Haider 2010, p. 284). Any such choice should be based solely on the empirical question our research attempts to answer, i.e. whether we are interested in the average male gain in comparison with similar women or in the average female loss in comparison with similar men. Neither are any assumptions about (or any reference to) the "nondiscriminatory" wage structure needed to ensure a meaningful decomposition is performed. There has been, however, a strong tendency in the Oaxaca-Blinder decomposition literature to claim otherwise. See, for example, Oaxaca and Ransom (1994, p. 8) and Mora (2008, p. 464).

There are several important reasons, in my view, to prefer referring to the estimands of interest in the gender wage differentials context as "average gender effects" instead of "average treatment effects". First, an important goal in the treatment effects framework is to establish causal relationships between some treatment and an outcome of interest. As discussed by Fortin, Lemieux, and Firpo (2011), there are clear reasons not to interpret the unexplained components of Oaxaca-Blinder decompositions as "causal" effects (in the gender wage differentials context). Ceasing to use the word "treatment" should ensure no claims to causality are transferred from the treatment effects literature. Second, it might be preferred not to refer to the estimand of interest as a "discrimination effect"[3], although it has often been done in the literature. Third, the treatment effects framework requires referring to one of the two groups in the population as the "treated" group and the other as the "nontreated" group. It is quite problematic to claim that either being male or being female is "treatment" and the other gender is a "nontreated" standard. An introduction of the notion of average gender effects should solve this problem, since we no longer need to refer to any of the groups as "treated" or "nontreated"; we only need to decide whose outcome is subtracted from the other one. If female wages are subtracted from male wages (which is the case in this paper), we will

[3] See Weichselbaumer and Winter-Ebmer (2006) for a good explanation and a discussion of the rhetoric associated with the use of different terms to describe the unexplained component.

typically get positive average gender effects; on the other hand, were male wages subtracted from female wages to compute various average gender effects, these effects would typically be negative. Of course, this is not a problem, since the estimated average gender effects would always be equal in absolute value (when one is negative, the other is positive).

B. A Consistent Estimator of the PAGE/PATE

In this subsection I develop a consistent estimator of the population average gender effect (PAGE) and, equivalently, of the population average treatment effect (PATE). This alternative estimator is based on a nonstandard version of the Oaxaca-Blinder decomposition, and it uses Equation 4 and a reference wage structure constructed as a linear combination of the regression coefficients for both subpopulations of interest, with the population proportion of group one used to weight the coefficients for group two and the population proportion of group two used to weight the coefficients for group one.

Proposition 1 (***Oaxaca-Blinder as an Estimator of the PAGE/PATE***). *Under the assumptions of unconfoundedness, overlap, and simple counterfactual treatment the population average gender effect (PAGE) can be consistently estimated with Equation 4 if* $\beta^* = \frac{N_0}{N_1+N_0} \cdot \beta_1 + \frac{N_1}{N_1+N_0} \cdot \beta_0$.

Proof. See Appendix 1.

Although using the population proportion of group one (two) to weight the coefficients for group two (one) may at first look counterintuitive, the simple counterfactual treatment assumption provides a clear rationalization for such an approach. Each of the wage structures takes a clearly defined role in such a decomposition: to serve as counterfactual for the other group. This is exactly the reason why more weight should be put on the wage structure of the smaller group: it is used only to provide a counterfactual wage structure for the larger group.

Interestingly, this alternative estimator is equivalent to a flexible OLS estimator of the PATE presented in Wooldridge (2002) and Imbens and Wooldridge (2009), i.e. τ_{PATE} can equivalently be estimated as the coefficient on W_i in the regression of Y_i on 1, W_i, X_i, and $W_i \cdot (X_i - \overline{X})$. Precisely, Imbens and Wooldridge (2009, p. 29) show that such a flexible OLS estimator can alternatively be written as:

$$\hat{\tau}_{PATE} = \overline{Y}_1 - \overline{Y}_0 - \left(\frac{N_0}{N_0 + N_1} \cdot \hat{\beta}_1 + \frac{N_1}{N_0 + N_1} \cdot \hat{\beta}_0\right)\left(\overline{X}_1 - \overline{X}_0\right)$$

which is exactly the sample counterpart of the Oaxaca-Blinder estimator of the PATE presented in Proposition 1.

C. Reinterpreting Reimers (1983) and Cotton (1988)

Now, I reinterpret the well-known solutions to the reference group choice problem provided by Reimers (1983) and Cotton (1988), and show that their underlying estimands associated with the simple counterfactual treatment assumption have remained unknown in the literature so far, while they are unlikely to be interesting for an applied researcher, especially compared with the new estimator of the present paper (developed in Proposition 1).

Proposition 2 (***The Underlying Estimand in Reimers 1983***). *Under the assumptions of unconfoundedness, overlap, and simple counterfactual treatment the unexplained component of the extension of the Oaxaca-Blinder decomposition proposed by Reimers (1983) is a consistent estimator of the arithmetic mean of the population average gender effect on men (PAGM) and the population average gender effect on women (PAGW).*

Proof. See Appendix 2.

Consequently, the underlying estimand of the extension of the Oaxaca-Blinder decomposition proposed by Reimers (1983) has a clear interpretation under the simple counterfactual treatment assumption. This interpretation suggests, however, that this estimator is unlikely to be interesting for an applied researcher, since the arithmetic mean of the population average gender effect on men (PAGM) and the population average gender effect on women (PAGW) is clearly much less natural an estimand than the population average gender effect on men (PAGM) which can be estimated with Equation 2, the population average gender effect on women (PAGW) which can be estimated with Equation 3, and the population average gender effect (PAGE) which can be estimated with Proposition 1.

Interestingly, the underlying estimand of the extension of the Oaxaca-Blinder decomposition proposed by Cotton (1988) has a clear interpretation under the simple counterfactual treatment assumption as well, while it seems that it is even less likely to be interesting in any applied research:

Proposition 3 (***The Underlying Estimand in Cotton 1988***). *Under the assumptions of unconfoundedness, overlap, and simple counterfactual treatment the unexplained component of the extension of the Oaxaca-Blinder decomposition proposed by Cotton (1988) is a consistent estimator of a weighted average of the population average gender effect on men (PAGM) and the population average gender effect on women (PAGW), with reversed weights attached to both these effects, i.e. the population proportion of men is used to weight PAGW and the population proportion of women is used to weight PAGM.*

Proof. See Appendix 3.

Although the reference wage structure proposed by Cotton (1988) is somehow similar to the new structure used in Proposition 1 to provide a consistent estimator of the population average gender effect (PAGE), it is clear that the resulting estimand is unlikely to provide a good answer to any conceivable a research question, since it changes the natural weights attached to the average effects estimated for both subpopulations of interest. Since neither Reimers (1983) nor Cotton (1988) provided any theoretical rationale for their propositions of the reference wage structure in the Oaxaca-Blinder decomposition, their propositions are also unlikely to take any general equilibrium effects into account in a satisfactory way. I therefore strongly recommend these two propositions are no longer used.

D. Reinterpreting Fortin (2008) and Jann (2008)

In this subsection I provide a similar reinterpretation of the underlying estimands in the recent proposition of Fortin (2008) and Jann (2008). These authors have convincingly criticized the Neumark (1988)'s solution to the reference group choice problem which is based on a pooled regression without the group membership dummy, since such an exclusion can bias coefficients on other covariates (omitted variables bias). Instead, they have proposed to use a pooled regression with the group membership dummy as the reference wage structure. Clearly, it is straightforward to show that the unexplained component in such a decomposition is equal by construction to the group membership dummy in a pooled regression. What follows, Propositions 4 and 5 apply to a pooled regression (simple OLS) estimator of various average gender effects as well. Such an estimator has recently been suggested, e.g., by Elder, Goddeeris, and Haider (2010).

Proposition 4 (***The Underlying Estimand in Fortin 2008 and Jann 2008 under Homogeneous Gender Effects***). *Under the assumptions of unconfoundedness, overlap, simple counterfactual treatment, and homogeneous gender effects the unexplained component of the* *extension of the Oaxaca-Blinder decomposition proposed by Fortin (2008) and Jann (2008) is a consistent estimator of the population average gender effect (PAGE), the population average gender effect on men (PAGM), and the population average gender effect on women (PAGW).*

Proof. See Wooldridge (2002, p. 611-612) for a proof that simple OLS provides a consistent estimator of the PATE and the PATT under homogeneous treatment effects.

If gender effects are homogeneous, i.e. $\mathbb{E}[Y_i(1) - Y_i(0)|X_i = x] = \tau$ for all x, the extension of the Oaxaca-Blinder decomposition proposed by Fortin (2008) and Jann (2008) provides a consistent estimator of all the major average gender effects: the PAGE, the PAGW, and the PAGM. This is by no means surprising: it is mere an extrapolation of the established result that simple OLS provides a consistent estimator of various average treatment effects under the assumption of homogeneous treatment effects (see Wooldridge 2002 for a proof and Angrist and Pischke 2009 for a good discussion). On the other hand, it is much less established what exactly is provided by simple OLS if the assumption of homogeneous treatment effects is relaxed and heterogeneity is allowed. Proposition 5 attempts to provide some clarification:

Proposition 5 (***The Underlying Estimand in Fortin 2008 and Jann 2008 under Heterogeneous Gender Effects***). *Under the assumptions of unconfoundedness, overlap, and simple counterfactual treatment the unexplained component of the extension of the Oaxaca-Blinder decomposition proposed by Fortin (2008) and Jann (2008) is a consistent estimator of a weighted average of the population average gender effect on men (PAGM) and the population average gender effect on women (PAGW), with reversed weights attached to both these effects, i.e. the population proportion of men is used to weight PAGW and the* *population proportion of women is used to weight PAGM, provided that X_i is a scalar whose variance is equal in the male and female subpopulations.*

Proof. See Elder, Goddeeris, and Haider (2010, Appendix A) and combine their result with Proposition 3 in the present paper.

In a recent paper, Elder, Goddeeris, and Haider (2010) have provided an interesting attempt to revive the Oaxaca-Blinder decomposition methodology. A serious drawback of their paper lies, however, in a complete lack of communication with the treatment effects literature. For example, Elder, Goddeeris, and Haider (2010, Appendix A) have formally proven that the group membership dummy in a simple linear regression converges to the unexplained component of the extension of the Oaxaca-Blinder decomposition proposed by Cotton (1988) provided that X_i is a scalar whose variance is equal in the male and female subpopulations. Were this last assumption not satisfied, such a relationship could still approximately hold, but there would be no general rule.

Interestingly, this result in Elder, Goddeeris, and Haider (2010, Appendix A) is sufficient to prove Proposition 5, and lack of its recognition by the authors is a consequence of their lack of communication with the treatment effects literature. Precisely, if the assumption of homogeneous gender effects is relaxed, then – under additional assumptions stated in Proposition 5 – simple OLS converges to the underlying estimand in Cotton (1988), while this estimand – under Proposition 3 – is precisely a weighted average of the population average gender effect on men (PAGM) and the population average gender effect on women (PAGW), with reversed weights attached to both these effects. Consequently, both the extension of the Oaxaca-Blinder decomposition proposed by Fortin (2008) and Jann (2008) and simple OLS may tend to overstate the importance of the smaller subpopulation when computing an estimate of the population average gender effect (PAGE) or the population average treatment effect (PATE).

Such a result is indeed intuitive: imagine as an example that returns to schooling are very different for both groups of interest. If these groups are not of approximately equal size, OLS will tend to estimate that average returns to schooling in the whole population are closer to their true underlying values for the larger group. Although this is perfectly fine if our goal is to estimate such an average rate of return, it is not what we need if regression methods are used to construct counterfactuals. In the latter case, we need more weight to be put on the smaller group's rate of return in order to create a useful counterfactual for the larger group.

Consequently, whenever we are concerned with gender (treatment) effects heterogeneity, there is good reason to choose the Oaxaca-Blinder decomposition in Proposition 1 to estimate the population average gender effect (PAGE) or the population average treatment effect (PATE) as well as the most basic versions of the Oaxaca-Blinder decomposition (Equations 2 and 3) to estimate the PAGW, the PAGM, the PATT or the PATN, instead of choosing the extension of the Oaxaca-Blinder decomposition proposed by Fortin (2008) and Jann (2008) or simple OLS. On the other hand, a strong case in favor of simple linear regression is provided by Angrist and Pischke (2009).

E. A Semiparametric Extension

In previous subsections I have shown that the majority of extensions of the Oaxaca-Blinder decomposition can be motivated by the simple counterfactual treatment assumption and that they therefore estimate some function of various average gender (treatment) effects, although the underlying estimands in many of these extensions are unlikely to provide an answer to any feasible research question. Consequently, I have strongly recommended to choose these versions of the Oaxaca-Blinder decomposition which provide consistent estimators of "gender equivalents" of the population average treatment effect (PATE) or the population average treatment effect on the treated (PATT), i.e. either the new estimator of the present paper (Proposition 1) or the original propositions of Oaxaca (1973) and Blinder (1973).

If we indeed decide to aim at estimating the population average gender effect (PAGE), the population average gender effect on women (PAGW), and the population average gender effect on men (PAGM), i.e. estimands which are technically equivalent to various average treatment effects, there is no reason to construct semiparametric and nonparametric estimators which try to closely mimic the unexplained component of the parametric Oaxaca-Blinder decomposition (although such estimators have been proposed, e.g., by Frölich 2007, Mora 2008, and Ñopo 2008). Instead, any of the standard estimators of the PATE or the PATT can be used to estimate the population average gender effect (PAGE) or the PAGM (PAGW), and we can safely assume that the better is an estimator for various average treatment effects, the better it is for various average gender effects as well.

In a recent paper, Kline (2011) has convincingly suggested that such an advisable estimator of the PATT is indeed provided by the parametric Oaxaca-Blinder decomposition, since it enjoys a beneficial "double robustness" property (Robins, Rotnitzky, and Zhao 1994), while being a propensity score reweighting estimator itself. On the other hand, a recent survey by Imbens and Wooldridge (2009) provides a strong case for various combinations of linear regression and either matching or methods based on the propensity score. For example, a combination of stratification (on the propensity score) and linear regression is suggested by the authors as "one of the more attractive estimators in practice" (Imbens and Wooldridge 2009, p. 41), while this estimator is also known for having performed best in an important study by Dehejia and Wahba (1999). Such an estimator does not allow, however, for within-strata heterogeneity of treatment effects. Since the Oaxaca-Blinder decomposition can be thought of as a version of flexible OLS (in the sense that it allows for gender/treatment effects heterogeneity), such a shortcoming of the estimator suggested by Imbens and Wooldridge (2009, p. 41) can be addressed by a combination of stratification and the Oaxaca-Blinder decomposition. Indeed, let me propose below such semiparametric extensions of the parametric Oaxaca-Blinder estimators of the PAGE, the PAGW, and the PAGM.

A semiparametric estimator of the population average gender effect (PAGE) requires a first-step estimation of the propensity score (with logit or probit), i.e. the conditional probability that a sample member is female (male) given his or her covariate values. The estimated propensity score is then used to divide the whole sample in J strata. In each stratum, a separate estimation of the average gender effect for the resulting subpopulation is performed using the Oaxaca-Blinder decomposition in Proposition 1. For each stratum j, an estimate $\hat{\tau}_{PAGE,j}$ (and the estimated variance $\hat{V}_{PAGE,j}$) is obtained, and these estimates are averaged using a simple procedure in Imbens and Wooldridge (2009, p. 41):

$$\hat{\tau}_{PAGE} = \sum_{j=1}^{J} \left(\frac{N_{j0}+N_{j1}}{N}\right) \cdot \hat{\tau}_{PAGE,j} \text{ and } \hat{V}_{PAGE} = \sum_{j=1}^{J} \left(\frac{N_{j0}+N_{j1}}{N}\right)^2 \cdot \hat{V}_{PAGE,j}$$

Similarly, an analogous semiparametric estimator of the population average gender effect on women (PAGW) involves running the Oaxaca-Blinder decomposition presented in Equation 3 in each of the strata, while the population average gender effect on men (PAGM) can be estimated on the basis of the Oaxaca-Blinder decomposition in Equation 2 being performed in each stratum. The resulting estimates are then averaged to obtain:

$$\hat{\tau}_{PAGW} = \sum_{j=1}^{J} \left(\frac{N_{j0}}{N}\right) \cdot \hat{\tau}_{PAGW,j} \text{ and } \hat{V}_{PAGW} = \sum_{j=1}^{J} \left(\frac{N_{j0}}{N}\right)^2 \cdot \hat{V}_{PAGW,j}$$

$$\hat{\tau}_{PAGM} = \sum_{j=1}^{J} \left(\frac{N_{j1}}{N}\right) \cdot \hat{\tau}_{PAGM,j} \text{ and } \hat{V}_{PAGM} = \sum_{j=1}^{J} \left(\frac{N_{j1}}{N}\right)^2 \cdot \hat{V}_{PAGM,j}$$

Such an estimator requires, of course, fewer functional form assumptions compared with the traditional approach. It can also provide an alternative estimator outside the gender wage differentials context, and can be used to semiparametrically estimate the population average treatment effect (PATE) or the population average treatment effect on the treated (PATT). This estimator will be tested together with parametric Oaxaca-Blinder decompositions and a combination of stratification and linear regression in the next section's empirical application.

IV. An Empirical Application

In this section the new estimators of this paper are compared empirically in an application to U.S. gender wage differentials[4]. The data source is the pooled 1983-2010 Integrated Public Use Microdata Series (IPUMS) file of the March Current Population Survey (CPS). This dataset (albeit restricted to 1983-2007) has recently been used by Elder, Goddeeris, and Haider (2010) to compare empirically linear regression, both simple estimators of Oaxaca (1973) and Blinder (1973), and the extension of the Oaxaca-Blinder decomposition proposed by Neumark (1988). Although my theoretical conclusions on the use of various estimators based on the Oaxaca-Blinder decomposition are remarkably different from those of Elder, Goddeeris, and Haider (2010), I utilize their sample and variable selections in my analysis to make my results directly comparable to their recent paper.

Consequently, the pooled 1983-2010 CPS dataset is restricted to full-time, full-year workers. This category has been defined by Elder, Goddeeris, and Haider (2010) as those observations who are at least 18 years old, have earned nonzero wage or salary income, and have worked strictly more than 40 weeks a year and 30 hours in a typical week. The outcome variable of interest is the log hourly wage, and the hourly wage is measured as annual earnings divided by annual hours. The set of control variables used both by Elder, Goddeeris, and Haider (2010) and in the present paper is relatively sparse, and includes a quartic in age,

[4] All the applications of the Oaxaca-Blinder decomposition presented in this paper use the oaxaca command in Stata (Jann 2008).

four education categories (no high school diploma, high school diploma either obtained or unclear, 3 years of college or less, and 4 years of college or more), all the "major occupation" categories listed in the CPS (14 categories between 1983 and 2002; 11 categories between 2003 and 2010), and a dummy for whether an individual is black.

Table 1 presents descriptive statistics for outcome and selected control variables and all the yearly samples used in the analysis. To a considerable extent, it is a replication of the corresponding table in Elder, Goddeeris, and Haider (2010, p. 289-290), although their paper has not included the 2008-2010 data and has not reported standard errors.

Table 1. Sample Means of Outcome and Selected Control Variables for

Yearly IPUMS-CPS Datasets

Year	*N*	Wage	Black	Female	Age	HS+
1983	45,637	8.79	0.08	0.40	38.72	0.84
		(5.27)	(0.27)	(0.49)	(12.42)	(0.37)
1984	46,196	9.10	0.08	0.40	38.61	0.85
		(5.43)	(0.28)	(0.49)	(12.27)	(0.36)
1985	48,499	9.62	0.09	0.40	38.52	0.85
		(5.96)	(0.28)	(0.49)	(12.15)	(0.35)
1986	48,365	10.09	0.09	0.40	38.44	0.86
		(6.26)	(0.28)	(0.49)	(12.00)	(0.34)
1987	48,402	10.44	0.09	0.41	38.47	0.86
		(6.52)	(0.29)	(0.49)	(11.93)	(0.34)
1988	49,495	10.80	0.09	0.41	38.58	0.87
		(6.66)	(0.28)	(0.49)	(11.88)	(0.34)
1989	46,741	11.12	0.09	0.41	38.68	0.87
		(6.91)	(0.28)	(0.49)	(11.79)	(0.33)
1990	52,015	11.71	0.09	0.41	38.68	0.87
		(7.28)	(0.28)	(0.49)	(11.70)	(0.33)
1991	51,402	11.99	0.09	0.41	38.89	0.88
		(7.34)	(0.28)	(0.49)	(11.62)	(0.33)
1992	50,018	12.40	0.09	0.43	39.08	0.89
		(7.61)	(0.28)	(0.49)	(11.47)	(0.32)
1993	49,405	12.91	0.09	0.43	39.35	0.89
		(7.87)	(0.28)	(0.49)	(11.38)	(0.31)
1994	47,948	13.19	0.09	0.43	39.54	0.90
		(8.16)	(0.28)	(0.49)	(11.44)	(0.30)
1995	48,839	13.67	0.09	0.42	39.66	0.90
		(8.56)	(0.28)	(0.49)	(11.35)	(0.30)
1996	43,719	14.55	0.09	0.42	39.86	0.89
		(13.00)	(0.28)	(0.49)	(11.39)	(0.31)
1997	44,727	15.14	0.09	0.42	40.06	0.89

		(14.31)	(0.29)	(0.49)	(11.47)	(0.31)
1998	44,941	15.82	0.09	0.43	40.15	0.90
		(14.73)	(0.29)	(0.50)	(11.41)	(0.31)
1999	46,314	16.41	0.09	0.43	40.27	0.89
		(15.24)	(0.29)	(0.49)	(11.45)	(0.31)
2000	47,551	16.61	0.09	0.43	40.45	0.89
		(13.13)	(0.29)	(0.49)	(11.58)	(0.31)
2001	76,647	18.12	0.11	0.43	40.25	0.90
		(16.89)	(0.32)	(0.50)	(11.14)	(0.30)
2002	75,429	19.02	0.11	0.43	40.63	0.90
		(18.02)	(0.32)	(0.50)	(11.23)	(0.30)
2003	73,809	19.48	0.11	0.43	40.98	0.90
		(18.86)	(0.31)	(0.50)	(11.33)	(0.30)
2004	72,351	19.88	0.11	0.43	41.34	0.91
		(18.44)	(0.31)	(0.50)	(11.45)	(0.29)
2005	71,711	20.39	0.11	0.43	41.40	0.91
		(19.81)	(0.31)	(0.50)	(11.54)	(0.29)
2006	72,170	20.97	0.10	0.43	41.48	0.90
		(19.80)	(0.31)	(0.50)	(11.64)	(0.29)
2007	72,500	21.93	0.11	0.43	41.69	0.91
		(21.69)	(0.31)	(0.50)	(11.76)	(0.29)
2008	72,884	22.36	0.11	0.44	41.94	0.91
		(20.68)	(0.31)	(0.50)	(11.94)	(0.28)
2009	71,359	23.30	0.11	0.44	42.21	0.92
		(21.10)	(0.31)	(0.50)	(11.94)	(0.27)
2010	68,399	23.80	0.11	0.45	42.44	0.92
		(22.20)	(0.31)	(0.50)	(12.01)	(0.27)

NOTE: Standard errors are in parentheses. Wages are in nominal dollars. HS+ is equal to 1 for observations who belong to the "high school diploma either obtained or unclear" education category or a higher one, 0 otherwise.

Table 2 presents a comparison of various parametric estimators of the covariate adjusted gender wage differential as well as yearly measures of the unadjusted wage gap. Between 1983 and the mid-1990's both the unadjusted differential and various measures of the adjusted wage gap were generally falling; such a trend has been considerably less clear since the mid-1990's, although the unadjusted differential seems to have fallen as well.

Table 2. A Comparison of Various Oaxaca-Blinder Estimators of the Adjusted Gender Wage Differential

Year	Unadjusted	OLS	OB (*PAGE*)	OB (*PAGW*)	OB (*PAGM*)	OB (Neumark)
1983	0.392	0.361	0.375	0.345	0.395	0.270
	(0.007)	(0.008)	(0.010)	(0.007)	(0.013)	(0.006)
1984	0.370	0.364	0.384	0.350	0.406	0.275
	(0.007)	(0.008)	(0.012)	(0.008)	(0.016)	(0.006)
1985	0.372	0.361	0.371	0.346	0.388	0.276
	(0.007)	(0.007)	(0.009)	(0.007)	(0.013)	(0.006)
1986	0.353	0.338	0.347	0.332	0.357	0.259
	(0.007)	(0.007)	(0.010)	(0.007)	(0.014)	(0.006)
1987	0.343	0.326	0.324	0.305	0.337	0.249
	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.005)
1988	0.322	0.302	0.296	0.293	0.298	0.231
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.004)
1989	0.319	0.310	0.315	0.305	0.322	0.238
	(0.006)	(0.006)	(0.007)	(0.006)	(0.009)	(0.005)
1990	0.302	0.299	0.299	0.290	0.306	0.232
	(0.006)	(0.006)	(0.007)	(0.006)	(0.009)	(0.004)
1991	0.283	0.289	0.288	0.278	0.295	0.225
	(0.006)	(0.006)	(0.007)	(0.006)	(0.009)	(0.004)
1992	0.267	0.284	0.278	0.275	0.281	0.221
	(0.006)	(0.006)	(0.007)	(0.006)	(0.008)	(0.004)
1993	0.262	0.275	0.273	0.263	0.280	0.215
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.005)
1994	0.255	0.265	0.261	0.254	0.266	0.209
	(0.006)	(0.006)	(0.007)	(0.006)	(0.009)	(0.005)
1995	0.253	0.254	0.245	0.249	0.242	0.201
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.005)
1996	0.263	0.281	0.273	0.268	0.276	0.222
	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.005)
1997	0.248	0.270	0.263	0.265	0.261	0.214
	(0.006)	(0.006)	(0.007)	(0.007)	(0.008)	(0.005)
1998	0.260	0.278	0.280	0.264	0.292	0.221
	(0.007)	(0.006)	(0.007)	(0.007)	(0.008)	(0.005)
1999	0.252	0.273	0.274	0.257	0.286	0.217
	(0.006)	(0.006)	(0.007)	(0.006)	(0.008)	(0.005)
2000	0.265	0.286	0.285	0.275	0.293	0.228
	(0.006)	(0.006)	(0.007)	(0.006)	(0.009)	(0.005)
2001	0.285	0.294	0.290	0.280	0.297	0.233
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.004)
2002	0.274	0.290	0.279	0.276	0.281	0.231
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.004)
2003	0.259	0.276	0.266	0.259	0.271	0.224
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.004)
2004	0.253	0.277	0.263	0.260	0.265	0.225
	(0.005)	(0.005)	(0.006)	(0.005)	(0.008)	(0.004)
2005	0.255	0.283	0.266	0.268	0.264	0.231
	(0.005)	(0.005)	(0.007)	(0.005)	(0.010)	(0.004)
2006	0.249	0.281	0.269	0.263	0.273	0.229
	(0.005)	(0.005)	(0.007)	(0.005)	(0.010)	(0.004)
2007	0.243	0.275	0.265	0.257	0.271	0.225

	(0.005)	(0.005)	(0.006)	(0.005)	(0.008)	(0.004)
2008	0.235	0.268	0.249	0.253	0.245	0.219
	(0.005)	(0.005)	(0.007)	(0.005)	(0.010)	(0.004)
2009	0.237	0.267	0.266	0.250	0.278	0.218
	(0.005)	(0.005)	(0.007)	(0.005)	(0.010)	(0.004)
2010	0.235	0.260	0.260	0.245	0.273	0.215
	(0.005)	(0.005)	(0.006)	(0.005)	(0.009)	(0.004)

NOTE: Robust standard errors are in parentheses.

These examples provide an illustration of several issues discussed in Section 3, but are also interesting on their own. First, the population average gender effect on men (PAGM) is larger than the population average gender effect on women (PAGW) in a relatively robust way. The difference is as large as approximately 5 percentage points in 1983-1985; then, the two average gender effects converged to a considerable extent, but have again diverged very recently. In 2009-2010 this difference has risen again to approximately 2.8 percentage points in both years. Hence, men gain typically more in comparison with similar women than women lose in comparison with similar men.

Second, the Oaxaca-Blinder estimates of the population average gender effect (PAGE) lie always (by construction) in between the corresponding estimates of the PAGW and the PAGM, and this is not necessarily true for the linear regression estimates. The PAGE estimates are, however, not only bounded by the PAGM and the PAGW, but also exactly equal (again by construction) to their weighted average, with weights equal to sample proportions of men and women. As already noted by Elder, Goddeeris, and Haider (2010), the Neumark (1988) estimates of the covariate adjusted gender wage differential are always significantly lower than any other estimates.

Third, the empirical relationship between the Oaxaca-Blinder estimates of the population average gender effect (PAGE) and the corresponding linear regression estimates is less clear than Proposition 5 would suggest. On the other hand, Proposition 5 accurately predicts that the Oaxaca-Blinder estimates of the population average gender effect (PAGE) should typically be closer to the Oaxaca-Blinder estimates of the population average gender effect on men (PAGM) than the corresponding linear regression estimates are.

Furthermore, Appendix Tables 1-3 present a comparison of various parametric and semiparametric estimators of the PAGE (Appendix Table 1), the PAGW (Appendix Table 2), and the PAGM (Appendix Table 3). Each table compares linear regression, the appropriate parametric Oaxaca-Blinder decomposition, simple stratification on the estimated propensity score, a combination of stratification and linear regression, and an appropriate combination of stratification and the Oaxaca-Blinder decomposition (as proposed in Section 3). All the three stratification-based estimators are performed both on yearly full samples and on samples which have been trimmed to improve overlap. Such an improvement has been done according to a rule of thumb recently proposed by Crump, Hotz, Imbens, and Mitnik (2009). These authors have suggested to discard all the observations whose estimated propensity score is either smaller than 0.1 or larger than 0.9, and such an approach to addressing lack of overlap should allow for the most precise estimation of average treatment effects.

Again, several interesting patterns emerge. First, linear regression seems to provide relatively poor estimates of both the population average gender effect on women (PAGW) and the population average gender effect on men (PAGM). In accordance with Proposition 5 and the empirical result that the PAGM is typically larger than the PAGW, simple OLS tends to provide robustly larger estimates of the PAGW compared with estimators which allow for gender effects heterogeneity (including the parametric Oaxaca-Blinder decomposition), while linear regression estimates of the PAGM are typically smaller than the corresponding estimates based on stratification, either adjusted or not.

Second, in contrast with the empirical results in Mora (2008) and Ñopo (2008), my estimates suggest there actually might be a relatively general pattern with respect to whether more functional flexibility provides smaller or larger estimates of various average gender effects. The parametric Oaxaca-Blinder decomposition tends to provide robustly smaller estimates of both the PAGE and the PAGM compared with methods that combine stratification on the estimated propensity score with either Oaxaca-Blinder or linear regression. This result is less clear for the PAGW, but it still seems to be there.

Third, simple stratification used to provide larger estimates of all the three estimands of interest compared with both the combination of stratification and linear regression and the combination of stratification and the Oaxaca-Blinder decomposition, but this pattern has recently switched; for the last few years, simple stratification has been providing relatively smaller estimates of the PAGE, the PAGM, and the PAGW. All these results are, again, less clear for the PAGW.

Fourth, there has rarely been any economically significant a difference between the combination of stratification and linear regression and the combination of stratification and the Oaxaca-Blinder decomposition (for any of the estimands of interest). On the other hand, an interesting pattern still emerges: while the differential is very small and generally does not exceed 0.5 percentage points, the combination of stratification and linear regression provides robustly smaller estimates of the PAGE, the PAGM, and the PAGW compared with the combination of stratification and the Oaxaca-Blinder decomposition.

Fifth, my analysis reconfirms one of the conclusions in Ñopo (2008), i.e. that restricting our attention to the common support region leads to significantly smaller estimates of the unexplained gender wage differential. Precisely, while all the points made above were generally concerned with estimates obtained using full samples, trimming these samples to improve overlap (in accordance with the rule of thumb proposed by Crump, Hotz, Imbens, and Mitnik 2009) leads to robustly smaller estimates of the PAGE, the PAGM, and the PAGW. There is a clear reason for that: while the average gender effect conditional on the estimated propensity score generally (but not monotonously) rises with the propensity that a sample member is male, there are typically quite many observations in the bin which corresponds to the 0.9-1 conditional probability of being male as well as typically very few (or even none at all) observations in the bin which corresponds to the 0.9-1 conditional probability of being female.

V. Summary

In this paper I have introduced a new notion of average gender effects (which refers to the difference between the expected male wage and the expected female wage, conditional on observable characteristics), and have shown how this notion can be used to reshape the Oaxaca-Blinder decomposition estimands and methodology. I have proposed an alternative solution to the reference group choice problem which is inherent in Oaxaca-Blinder, while this new decomposition of the present paper can be used to consistently estimate the population average gender effect (PAGE) or the population average treatment effect (PATE). I have also reinterpreted the estimands which are implicit in the well-known extensions of the Oaxaca-Blinder decomposition proposed by Reimers (1983), Cotton (1988), Fortin (2008), and Jann (2008), while I have also suggested that these estimators are considerably less likely to be interesting for an applied researcher in comparison with the new estimator of the present paper. I have also proposed a semiparametric extension of the Oaxaca-Blinder decomposition which incorporates stratification on the estimated propensity score.

Next, I have provided an empirical illustration of the new estimators of this paper, and have used the 1983-2010 CPS data to estimate the population average gender effect (PAGE), the population average gender effect on men (PAGM), and the population average gender effect on women (PAGW) for the U.S. working population. To the best of my knowledge this is the first paper to provide separate estimates for these newly introduced parameters.

Appendix 1

Proof of Proposition 1 (*Oaxaca-Blinder as an Estimator of the PAGE/PATE*). Beginning with Equation 4, I combine it with $\beta^* = \frac{N_0}{N_1+N_0} \cdot \beta_1 + \frac{N_1}{N_1+N_0} \cdot \beta_0$ and reformulate:

$$\mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0]$$

$$= \{\mathbb{E}[X_i|W_i = 1](\beta_1 - \beta^*) + \mathbb{E}[X_i|W_i = 0](\beta^* - \beta_0)\}$$

$$+ (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])\beta^*$$

$$= \left\{\mathbb{E}[X_i|W_i = 1]\left(\beta_1 - \left(\frac{N_0}{N_1 + N_0} \cdot \beta_1 + \frac{N_1}{N_1 + N_0} \cdot \beta_0\right)\right)\right.$$

$$\left. + \mathbb{E}[X_i|W_i = 0]\left(\left(\frac{N_0}{N_1 + N_0} \cdot \beta_1 + \frac{N_1}{N_1 + N_0} \cdot \beta_0\right) - \beta_0\right)\right\}$$

$$+ (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0]) \cdot \left(\frac{N_0}{N_1 + N_0} \cdot \beta_1 + \frac{N_1}{N_1 + N_0} \cdot \beta_0\right)$$

$$= \left\{\mathbb{E}[X_i|W_i = 1]\left(\frac{N_1}{N_1 + N_0} \cdot \beta_1 - \frac{N_1}{N_1 + N_0} \cdot \beta_0\right)\right.$$

$$\left. + \mathbb{E}[X_i|W_i = 0]\left(\frac{N_0}{N_1 + N_0} \cdot \beta_1 - \frac{N_0}{N_1 + N_0} \cdot \beta_0\right)\right\}$$

$$+ (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0]) \cdot \left(\frac{N_0}{N_1 + N_0} \cdot \beta_1 + \frac{N_1}{N_1 + N_0} \cdot \beta_0\right)$$

$$= \left\{\frac{N_1}{N_1 + N_0} \cdot \mathbb{E}[X_i|W_i = 1](\beta_1 - \beta_0) + \frac{N_0}{N_1 + N_0} \cdot \mathbb{E}[X_i|W_i = 0](\beta_1 - \beta_0)\right\}$$

$$+ (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0]) \cdot \left(\frac{N_0}{N_1 + N_0} \cdot \beta_1 + \frac{N_1}{N_1 + N_0} \cdot \beta_0\right)$$

$$= \left\{\frac{N_1}{N_1 + N_0} \cdot \tau_{PAGM} + \frac{N_0}{N_1 + N_0} \cdot \tau_{PAGW}\right\} + (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])$$

$$\cdot \left(\frac{N_0}{N_1 + N_0} \cdot \beta_1 + \frac{N_1}{N_1 + N_0} \cdot \beta_0\right)$$

$$= \tau_{PAGE} + (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])\beta^*$$

End of proof.

Appendix 2

Proof of Proposition 2 (*The Underlying Estimand in Reimers 1983*). The proof is analogous to Proposition 1. Beginning with Equation 4, I combine it with $\beta^* = \frac{1}{2} \cdot \beta_1 + \frac{1}{2} \cdot \beta_0$ and reformulate, economizing on the intermediate steps:

$$\mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0]$$

$$= \{\mathbb{E}[X_i|W_i = 1](\beta_1 - \beta^*) + \mathbb{E}[X_i|W_i = 0](\beta^* - \beta_0)\}$$

$$+ (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])\beta^*$$

$$= \left(\frac{1}{2} \cdot \tau_{PAGM} + \frac{1}{2} \cdot \tau_{PAGW}\right) + (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])\beta^*$$

End of proof.

Appendix 3

Proof of Proposition 3 (*The Underlying Estimand in Cotton 1988*). The proof is analogous to Propositions 1 and 2. Beginning with Equation 4, I combine it with $\beta^* = \frac{\mathrm{N}_1}{\mathrm{N}_1+\mathrm{N}_0} \cdot \beta_1 + \frac{\mathrm{N}_0}{\mathrm{N}_1+\mathrm{N}_0} \cdot \beta_0$ and reformulate, economizing on the intermediate steps:

$$\mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0]$$

$$= \{\mathbb{E}[X_i|W_i = 1](\beta_1 - \beta^*) + \mathbb{E}[X_i|W_i = 0](\beta^* - \beta_0)\}$$

$$+ (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])\beta^*$$

$$= \left(\frac{\mathrm{N}_0}{\mathrm{N}_1 + \mathrm{N}_0} \cdot \tau_{PAGM} + \frac{\mathrm{N}_1}{\mathrm{N}_1 + \mathrm{N}_0} \cdot \tau_{PAGW}\right)$$

$$+ (\mathbb{E}[X_i|W_i = 1] - \mathbb{E}[X_i|W_i = 0])\beta^*$$

End of proof.

Appendix Table 1. A Comparison of Various Estimators of the Population Average Gender Effect (PAGE)

Year	OLS	OB	Stratification	Stratification and OLS	Stratification and OB	Stratification	Stratification and OLS	Stratification and OB
			Full sample			Trimmed sample		
1983	0.361	0.375	0.398	0.389	0.392	0.352	0.349	0.349
	(0.008)	(0.010)	(0.013)	(0.012)	(0.013)	(0.009)	(0.008)	(0.009)
1984	0.364	0.384	0.393	0.380	0.385	0.340	0.346	0.345
	(0.008)	(0.012)	(0.013)	(0.013)	(0.013)	(0.008)	(0.008)	(0.008)
1985	0.361	0.371	0.400	0.383	0.388	0.341	0.358	0.364
	(0.007)	(0.009)	(0.012)	(0.012)	(0.012)	(0.009)	(0.009)	(0.010)
1986	0.338	0.347	0.364	0.351	0.354	0.312	0.333	0.334
	(0.007)	(0.010)	(0.011)	(0.011)	(0.012)	(0.009)	(0.008)	(0.009)
1987	0.326	0.324	0.345	0.335	0.334	0.312	0.324	0.323
	(0.007)	(0.007)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
1988	0.302	0.296	0.312	0.301	0.303	0.300	0.297	0.298
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
1989	0.310	0.315	0.324	0.314	0.316	0.305	0.305	0.306
	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)
1990	0.299	0.299	0.311	0.301	0.304	0.301	0.294	0.294
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
1991	0.289	0.288	0.313	0.299	0.301	0.292	0.283	0.284
	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
1992	0.284	0.278	0.295	0.284	0.283	0.278	0.277	0.276
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
1993	0.275	0.273	0.286	0.279	0.280	0.268	0.266	0.265
	(0.006)	(0.006)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
1994	0.265	0.261	0.268	0.258	0.259	0.264	0.262	0.264
	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
1995	0.254	0.245	0.259	0.251	0.254	0.265	0.259	0.261
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
1996	0.281	0.273	0.283	0.278	0.280	0.275	0.275	0.275
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)
1997	0.270	0.263	0.271	0.265	0.267	0.259	0.261	0.261
	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)

1998	0.278	0.280	0.284	0.281	0.283	0.265	0.267	0.268
	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
1999	0.273	0.274	0.283	0.278	0.278	0.264	0.265	0.266
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
2000	0.286	0.285	0.289	0.286	0.286	0.281	0.283	0.282
	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
2001	0.294	0.290	0.295	0.291	0.292	0.291	0.287	0.288
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
2002	0.290	0.279	0.280	0.281	0.281	0.285	0.284	0.285
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
2003	0.276	0.266	0.274	0.273	0.277	0.262	0.263	0.265
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
2004	0.277	0.263	0.267	0.272	0.275	0.268	0.267	0.268
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
2005	0.283	0.266	0.269	0.274	0.275	0.269	0.269	0.270
	(0.005)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
2006	0.281	0.269	0.277	0.280	0.282	0.282	0.275	0.276
	(0.005)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
2007	0.275	0.265	0.269	0.271	0.274	0.258	0.263	0.264
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
2008	0.268	0.249	0.250	0.256	0.259	0.274	0.258	0.258
	(0.005)	(0.007)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
2009	0.267	0.266	0.263	0.271	0.274	0.277	0.259	0.260
	(0.005)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
2010	0.260	0.260	0.250	0.261	0.264	0.276	0.254	0.255
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)

NOTE: Robust standard errors are in parentheses. The propensity score has been estimated using a logit model. Trimmed sample refers to discarding all the individuals whose estimated propensity score is less than 0.1 or greater than 0.9 (as in Crump, Hotz, Imbens, and Mitnik 2009). Full samples have been divided into five strata of equal width (0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, and 0.8-1) and trimmed samples have been divided into four strata of equal width (0.1-0.3, 0.3-0.5, 0.5-0.7, and 0.7-0.9).

Appendix Table 2. A Comparison of Various Estimators of the Population Average Gender Effect on Women (PAGW)

Year	OLS	OB	Stratification	Stratification and OLS	Stratification and OB	Stratification	Stratification and OLS	Stratification and OB
			Full sample			Trimmed sample		
1983	0.361	0.345	0.349	0.351	0.352	0.349	0.344	0.338
	(0.008)	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)
1984	0.364	0.350	0.351	0.346	0.351	0.345	0.344	0.339
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
1985	0.361	0.346	0.357	0.352	0.357	0.342	0.348	0.346
	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)
1986	0.338	0.332	0.336	0.332	0.335	0.317	0.326	0.324
	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)
1987	0.326	0.305	0.321	0.317	0.318	0.311	0.311	0.311
	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)
1988	0.302	0.293	0.299	0.300	0.302	0.294	0.290	0.294
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
1989	0.310	0.305	0.311	0.311	0.312	0.301	0.301	0.305
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)
1990	0.299	0.290	0.292	0.291	0.298	0.293	0.287	0.288
	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)
1991	0.289	0.278	0.297	0.288	0.290	0.285	0.279	0.279
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
1992	0.284	0.275	0.293	0.284	0.282	0.277	0.276	0.275
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
1993	0.275	0.263	0.270	0.268	0.270	0.260	0.259	0.259
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
1994	0.265	0.254	0.260	0.249	0.253	0.257	0.257	0.259
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
1995	0.254	0.249	0.250	0.248	0.253	0.258	0.256	0.259
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
1996	0.281	0.268	0.271	0.265	0.268	0.262	0.265	0.268
	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
1997	0.270	0.265	0.274	0.268	0.273	0.260	0.263	0.265
	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)

1998	0.278	0.264	0.266	0.262	0.266	0.255	0.262	0.264
	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)
1999	0.273	0.257	0.268	0.260	0.261	0.251	0.256	0.257
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
2000	0.286	0.275	0.280	0.276	0.279	0.271	0.273	0.273
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
2001	0.294	0.280	0.280	0.278	0.278	0.280	0.281	0.280
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
2002	0.290	0.276	0.273	0.273	0.272	0.277	0.280	0.280
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
2003	0.276	0.259	0.261	0.259	0.261	0.243	0.251	0.253
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
2004	0.277	0.260	0.255	0.260	0.260	0.245	0.251	0.253
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
2005	0.283	0.268	0.262	0.265	0.268	0.252	0.259	0.260
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
2006	0.281	0.263	0.257	0.259	0.261	0.264	0.262	0.261
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
2007	0.275	0.257	0.251	0.253	0.252	0.236	0.248	0.249
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
2008	0.268	0.253	0.252	0.254	0.254	0.261	0.247	0.247
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
2009	0.267	0.250	0.244	0.251	0.251	0.262	0.248	0.249
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
2010	0.260	0.245	0.230	0.238	0.237	0.259	0.244	0.245
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)

NOTE: Robust standard errors are in parentheses. The propensity score has been estimated using a logit model. Trimmed sample refers to discarding all the individuals whose estimated propensity score is less than 0.1 or greater than 0.9 (as in Crump, Hotz, Imbens, and Mitnik 2009). Full samples have been divided into five strata of equal width (0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, and 0.8-1) and trimmed samples have been divided into four strata of equal width (0.1-0.3, 0.3-0.5, 0.5-0.7, and 0.7-0.9).

Appendix Table 3. A Comparison of Various Estimators of the Population Average Gender Effect on Men (PAGM)

Year	OLS	OB	Stratification	Stratification and OLS	Stratification and OB	Stratification	Stratification and OLS	Stratification and OB
			Full sample			Trimmed sample		
1983	0.361	0.395	0.430	0.414	0.418	0.354	0.353	0.358
	(0.008)	(0.013)	(0.018)	(0.018)	(0.019)	(0.011)	(0.010)	(0.011)
1984	0.364	0.406	0.421	0.402	0.407	0.336	0.349	0.350
	(0.008)	(0.016)	(0.018)	(0.018)	(0.019)	(0.010)	(0.009)	(0.009)
1985	0.361	0.388	0.430	0.404	0.409	0.339	0.367	0.380
	(0.007)	(0.013)	(0.016)	(0.016)	(0.017)	(0.011)	(0.011)	(0.014)
1986	0.338	0.357	0.383	0.363	0.367	0.309	0.340	0.343
	(0.007)	(0.014)	(0.016)	(0.016)	(0.016)	(0.011)	(0.010)	(0.012)
1987	0.326	0.337	0.361	0.347	0.345	0.313	0.336	0.334
	(0.007)	(0.009)	(0.011)	(0.011)	(0.011)	(0.010)	(0.009)	(0.010)
1988	0.302	0.298	0.322	0.302	0.304	0.306	0.303	0.303
	(0.006)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.007)	(0.008)
1989	0.310	0.322	0.333	0.317	0.319	0.309	0.309	0.308
	(0.006)	(0.009)	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
1990	0.299	0.306	0.325	0.307	0.309	0.308	0.300	0.299
	(0.006)	(0.009)	(0.010)	(0.009)	(0.010)	(0.008)	(0.008)	(0.008)
1991	0.289	0.295	0.324	0.307	0.308	0.298	0.288	0.288
	(0.006)	(0.009)	(0.010)	(0.009)	(0.010)	(0.008)	(0.008)	(0.008)
1992	0.284	0.281	0.297	0.284	0.283	0.280	0.278	0.277
	(0.006)	(0.008)	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
1993	0.275	0.280	0.297	0.287	0.287	0.276	0.273	0.271
	(0.006)	(0.008)	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
1994	0.265	0.266	0.275	0.264	0.263	0.271	0.266	0.269
	(0.006)	(0.009)	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
1995	0.254	0.242	0.265	0.254	0.255	0.271	0.261	0.263
	(0.006)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
1996	0.281	0.276	0.293	0.288	0.289	0.287	0.285	0.282
	(0.007)	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)	(0.008)	(0.009)
1997	0.270	0.261	0.269	0.263	0.263	0.258	0.258	0.257
	(0.006)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)

1998	0.278	0.292	0.298	0.295	0.296	0.273	0.273	0.272
	(0.006)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
1999	0.273	0.286	0.294	0.291	0.291	0.276	0.275	0.275
	(0.006)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
2000	0.286	0.293	0.297	0.293	0.292	0.291	0.293	0.290
	(0.006)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
2001	0.294	0.297	0.305	0.301	0.302	0.302	0.293	0.295
	(0.005)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
2002	0.290	0.281	0.286	0.287	0.287	0.293	0.287	0.289
	(0.005)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
2003	0.276	0.271	0.284	0.283	0.289	0.281	0.275	0.276
	(0.005)	(0.007)	(0.008)	(0.007)	(0.007)	(0.006)	(0.005)	(0.005)
2004	0.277	0.265	0.277	0.282	0.287	0.289	0.281	0.283
	(0.005)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.005)	(0.006)
2005	0.283	0.264	0.275	0.281	0.281	0.286	0.279	0.279
	(0.005)	(0.010)	(0.007)	(0.007)	(0.007)	(0.006)	(0.005)	(0.005)
2006	0.281	0.273	0.292	0.296	0.298	0.299	0.287	0.290
	(0.005)	(0.010)	(0.008)	(0.007)	(0.007)	(0.006)	(0.005)	(0.006)
2007	0.275	0.271	0.283	0.285	0.291	0.279	0.277	0.278
	(0.005)	(0.008)	(0.008)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)
2008	0.268	0.245	0.249	0.258	0.263	0.287	0.269	0.269
	(0.005)	(0.010)	(0.008)	(0.007)	(0.007)	(0.006)	(0.005)	(0.005)
2009	0.267	0.278	0.278	0.286	0.293	0.292	0.270	0.270
	(0.005)	(0.010)	(0.008)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)
2010	0.260	0.273	0.267	0.279	0.286	0.292	0.265	0.265
	(0.005)	(0.009)	(0.008)	(0.007)	(0.008)	(0.006)	(0.006)	(0.006)

NOTE: Robust standard errors are in parentheses. The propensity score has been estimated using a logit model. Trimmed sample refers to discarding all the individuals whose estimated propensity score is less than 0.1 or greater than 0.9 (as in Crump, Hotz, Imbens, and Mitnik 2009). Full samples have been divided into five strata of equal width (0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, and 0.8-1) and trimmed samples have been divided into four strata of equal width (0.1-0.3, 0.3-0.5, 0.5-0.7, and 0.7-0.9).

References

Abadie A, Imbens GW. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica* 74: 235-267.

Angrist JD, Pischke J-S. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press: Princeton, NJ; Oxford.

Barsky R, Bound J, Charles KK, Lupton JP. 2002. Accounting for the black-white wealth gap: a nonparametric approach. *Journal of the American Statistical Association* 97: 663-673.

Blinder AS. 1973. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources* 8: 436-455.

Cobb-Clark DA, Crossley T. 2003. Econometrics for evaluations: an introduction to recent developments. *Economic Record* 79: 491-511.

Cotton J. 1988. On the decomposition of wage differentials. *Review of Economics and Statistics* 70: 236-243.

Crump RK, Hotz VJ, Imbens GW, Mitnik OA. 2009. Dealing with limited overlap in estimation of average treatment effects. *Biometrika* 96: 187-199.

Dehejia RH, Wahba S. 1999. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 94: 1053-1062.

Duncan GM, Leigh DE. 1985. The endogeneity of union status: an empirical test. *Journal of Labor Economics* 3: 385-402.

Elder TE, Goddeeris JH, Haider SJ. 2010. Unexplained gaps and Oaxaca-Blinder decompositions. *Labour Economics* 17: 284-290.

Fortin NM. 2008. The gender wage gap among young adults in the United States: the importance of money versus people. *Journal of Human Resources* 43: 884-918.

Fortin N, Lemieux T, Firpo S. 2011. Decomposition methods in economics. In *Handbook of Labor Economics, Volume 4A*, Ashenfelter O, Card D (eds). Elsevier: San Diego, CA; Amsterdam.

Frölich M. 2007. Propensity score matching without conditional independence assumption – with an application to the gender wage gap in the United Kingdom. *Econometrics Journal* 10: 359-407.

Hirano K, Imbens GW, Ridder G. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71: 1161-1189.

Imbens GW, Wooldridge JM. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47: 5-86.

Jann B. 2008. The Blinder-Oaxaca decomposition for linear regression models. *Stata Journal* 8: 453-479.

Kitagawa EM. 1955. Components of a difference between two rates. *Journal of the American Statistical Association* 50: 1168-1194.

Kline P. 2011. Oaxaca-Blinder as a reweighting estimator. *American Economic Review* 101: 532-537.

Melly B. 2006. Applied quantile regression. PhD dissertation. University of St. Gallen.

Mora R. 2008. A nonparametric decomposition of the Mexican American average wage gap. *Journal of Applied Econometrics* 23: 463-485.

Neumark D. 1988. Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources* 23: 279-295.

Ñopo H. 2008. Matching as a tool to decompose wage gaps. *Review of Economics and Statistics* 90: 290-299.

Oaxaca R. 1973. Male-female wage differentials in urban labor markets. *International Economic Review* 14: 693-709.

Oaxaca RL, Ransom MR. 1988. Searching for the effect of unionism on the wages of union and nonunion workers. *Journal of Labor Research* 9: 139-148.

Oaxaca RL, Ransom MR. 1994. On discrimination and the decomposition of wage differentials. *Journal of Econometrics* 61: 5-21.

Reimers CW. 1983. Labor market discrimination against Hispanic and black men. *Review of Economics and Statistics* 65: 570-579.

Robins JM, Rotnitzky A, Zhao LP. 1994. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 89: 846-866.

Rosenbaum PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41-55.

Weichselbaumer D, Winter-Ebmer R. 2005. A meta-analysis of the international gender wage gap. *Journal of Economic Surveys* 19: 479-511.

Weichselbaumer D, Winter-Ebmer R. 2006. Rhetoric in economic research: the case of gender wage differentials. *Industrial Relations* 45: 416-436.

Wooldridge JM. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA; London.