

IZA DP No. 9381

**An Alternative Estimator for Industrial
Gender Wage Gaps:
A Normalized Regression Approach**

Myeong-Su Yun
Eric S. Lin

September 2015

An Alternative Estimator for Industrial Gender Wage Gaps: A Normalized Regression Approach

Myeong-Su Yun

*Tulane University
and IZA*

Eric S. Lin

*National Tsing Hua University
and IZA*

Discussion Paper No. 9381
September 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

An Alternative Estimator for Industrial Gender Wage Gaps: A Normalized Regression Approach*

Using normalized regression equations, we propose an alternative estimator of industrial gender wage gaps which is identified in the sense that it is invariant to the choice of an unobserved non-discriminatory wage structure, and to the choice of the reference groups of any categorical variables. The proposed estimator measures the pure impact of industry on gender wage gaps after netting out wage differentials due to differences in characteristics and their coefficients between men and women. Furthermore, the proposed estimator is easy to implement, including hypothesis tests. We compare the proposed estimator with existing estimators using samples from 1998 Current Population Survey of US.

JEL Classification: C12, J31, J71

Keywords: industrial gender wage gaps, identification, gender wage discrimination, normalized regression, Oaxaca decomposition

Corresponding author:

Eric S. Lin
Department of Economics
National Tsing Hua University
Hsin-Chu 30013
Taiwan
E-mail: slin@mx.nthu.edu.tw

* Authors wish to thank the Editor, two anonymous reviewers, Hwei-Lin Chuang, Ira Gang, Dan Hamermesh, Bill Horrace, Ching-Kang Ing, Jenn-Hong Tang, and Malathi Velamuri for their valuable comments and suggestions on the earlier version of this article. Special thanks to Bill Horrace for kindly providing the March 1998 CPS data. Ta-Sheng Chou has provided excellent research assistance.

1 Introduction

Researchers have studied to identify the sources of the gender wage gap in order to quantify the role of compositional differences and differential effect of individual characteristics in the labor market, ever since the pioneering work of Blinder (1973) and Oaxaca (1973). The contribution of the Oaxaca decomposition to the observational studies of counterfactual and treatment effect has been recently recognized (e.g., Fortin, Lemieux and Firpo, 2011; Kline, 2011). Some technical difficulties of this decomposition method have been well-known. The decomposition results vary depending on the choice of base group, or non-discriminatory wage structure, known as the index problem (hereafter referred as **IP1**). Another problem, known as the identification problem (hereafter referred as **IP2**), is that the coefficients effect of set of dummy variables, that is, wage gap due to difference in the coefficients of set of dummy variables, is not invariant to the choice of the reference group (see Oaxaca and Ransom, 1999). These problems are well-recognized in the literature, and some resolutions and practical guides are proposed and widely implemented (for example, Neumark, 1988, Oaxaca and Ransom, 1988 and 1994 for the **IP1** and Gardeazabal and Ugidos, 2004, and Yun, 2005, for the **IP2**).

Gender wage gaps by subgroup such as industry and occupation may be able to provide rich and detailed information on disparities in wages between men and women, enabling us to devise appropriate policies for enhancing gender equity.¹ When it comes to study gender wage gap by subgroup, the existing literature provides two approaches. One is doing the standard decompositions after estimating earnings equations of men and women by subgroup, and later aggregating the decomposition results using the share of each subgroup as the weight. With some manipulation, the contributions of disparity of subgroup choice, and those of compositional differences and differential effect of individual characteristics in the earnings equation are computed (e.g., Brown, Moon and Zoloth, 1980). The alternative is estimating the earnings equation of men and women including dummy variables of each subgroup as explaining variables. As an attempt to estimate the pure impact of industry on gender wage gaps, Fields and Wolff (below FW, 1995) using the second approach define gender wage gaps in industry j intuitively as differences in intercepts and coefficients of

¹It is worth noting that the gender wage gap within sub-groups partitioned by other attributes also draws attention to researchers. For instance, Lin (2010) inspects the gender wage gap by college major in Taiwan.

dummy variable for industry j between men and women.²

The same technical difficulties as occurring in the standard Oaxaca decomposition arise for the first approach, and the same resolutions and practical guides may be applied. Hence, there is no additional technical issue for the first approach. However, the second approach, particularly the measure by FW (1995) suffers from an additional identification issue as pointed out by Horrace and Oaxaca (below HO, 2001). Several resolutions are proposed by researchers (e.g., HO, 2001, Horrace, 2005, and Ural, Horrace and Jung (below UHJ, 2009)). One problem is that several estimators suggested in the literature for the second approach may provide us with a different picture of gender wage gaps by subgroup, potentially leading to confusing policy recommendations.

Since most of the existing studies cannot get rid of the two identification problems (except for UHJ, 2009) or capture the pure impact of industry on gender wage gaps, this paper proposes an alternative gender wage gap by subgroup (industry) estimator using a normalized regression approach (e.g., Suits, 1984; Gardeazabal and Ugidos, 2004; Yun, 2005). The new estimator has several nice features. First, there is no need to choose the non-discriminatory wage structure so that **IP1** will be solved. Second, the new estimator is invariant to any choice of the left-out reference group (including industry dummies) and hence **IP2** is overcome. Third, the identified new gender wage gap estimator is not interfered with by the mean individual characteristics, so that the true wage gap from the coefficients will not be masked by the covariates. Furthermore, we do not need to worry about the stochastic nature of the mean characteristics which may change the asymptotic variance estimates of the industry gender wage gap measure (Lin, 2007). Finally, the standard error for the new estimator is straightforward, making it quite easy to implement hypothesis testing.

The remainder of this paper is organized as follows. First we summarize two identification problems in Oaxaca decomposition as a preliminary for our study on gender wage gaps by industry. In Section 3, we critically examine the existing industrial gender wage gap estimators from the perspectives of the two identification problems. We then develop an alternative industrial gender wage gap estimator based on a normalized regression approach in Section 4. Section 5 illustrates the implementation of our new estimator to the industrial gender wage gaps using the March 1998 Current Population Survey (CPS) adopted in HO

²The pure impact of industry on gender wage gaps using the second approach is measured as the gender wage gaps in industry j after eliminating wage gaps due to differences in characteristics and their coefficients between men and women.

(2001). The final section concludes.

2 Identification Problems in Oaxaca Decomposition: Preliminary

For illustration purposes, we examine the following regression model containing dummy variables for industries (d 's) suppressing individual subscript.³

$$y^c = \alpha^c + \sum_{j=2}^J \beta_j^c d_j^c + \sum_{k=2}^K \pi_k^c q_k^c + x^c \theta^c + \varepsilon^c, \quad (1)$$

where there are two sets of categorical variables (d 's and q 's) and x is a vector of continuous variables with dimension L ; the first and second sets of dummy variables (d 's and q 's) have J and K categories and $J - 1$ and $K - 1$ dummy variables in the equation, respectively, and a group superscript, $c = f$ (female) or m (male); without loss of generality, the reference group is the first category for each set of dummy variables ($\beta_1 = \pi_1 = 0$); α , β , π , and θ are parameters to be estimated; ε is a stochastic term. We refer to Equation (1) as the usual regression equation.

The overall wage gaps between men and women can be decomposed into a portion due to differences in characteristics (characteristics effect, explained component), a portion due to differences in coefficients (coefficients effect, unexplained component) as follows:⁴

$$\begin{aligned} \bar{y}^f - \bar{y}^m &= (\hat{\alpha}^f - \hat{\alpha}^m) + \sum_{j=2}^J (\hat{\beta}_j^f - \hat{\beta}_j^m) \bar{d}_j^f + \sum_{k=2}^K (\hat{\pi}_k^f - \hat{\pi}_k^m) \bar{q}_k^f + \bar{x}^f (\hat{\theta}^f - \hat{\theta}^m) \\ &+ \sum_{j=2}^J \hat{\beta}_j^m (\bar{d}_j^f - \bar{d}_j^m) + \sum_{k=2}^K \hat{\pi}_k^m (\bar{q}_k^f - \bar{q}_k^m) + (\bar{x}^f - \bar{x}^m) \hat{\theta}^m, \end{aligned}$$

³It is trivial to extend the model to incorporate more sets of dummy variables. From our private communication with Bill Horrace, we are reminded of the potential problems in regression specification, including the possibility that the choice of industry is endogenous. On the other hand, if we exclude industry dummy variables, then the wage regression equations may suffer from an omitted variable bias.

⁴We are assuming OLS is used for estimation. Therefore, the residuals effect, that is, differences in wages due to different average residuals between men and women is zero. It is well-documented that as long as consistent estimators are obtained, Oaxaca decomposition can be applied easily to decompose the overall wage differentials into characteristics, coefficients and residuals effects (Yun, 2007). All discussions in this paper can be easily extended to decompositions where the residuals effect is not zero.

where the “hat” denotes the estimated counterpart of the true parameter; \bar{y}^c , \bar{d}_j^c , \bar{q}_k^c , and \bar{x}^c , are average values computed using the whole samples of group c ; the first and second lines, respectively, represent coefficients and characteristics effects.

There are two well-known identification problems. The first identification problem, **IP1**, is related to a non-discriminatory wage structure. A non-discriminatory wage structure will pay workers according to their marginal productivity. By taking the unobservable non-discriminatory wage structure into account, the decomposition of the overall gender wage gap can be rewritten as follows:

$$\begin{aligned} \bar{y}^f - \bar{y}^m &= (\hat{\alpha}^f - \hat{\alpha}^m) + \left(\sum_{j=2}^J \tilde{\beta}_j^f \bar{d}_j^f - \sum_{j=2}^J \tilde{\beta}_j^m \bar{d}_j^m \right) + \left(\sum_{k=2}^K \tilde{\pi}_k^f \bar{q}_k^f - \sum_{k=2}^K \tilde{\pi}_k^m \bar{q}_k^m \right) + (\bar{x}^f \tilde{\theta}^f - \bar{x}^m \tilde{\theta}^m) \\ &\quad + \sum_{j=2}^J \beta_j^* (\bar{d}_j^f - \bar{d}_j^m) + \sum_{k=2}^K \pi_k^* (\bar{q}_k^f - \bar{q}_k^m) + (\bar{x}^f - \bar{x}^m) \theta^*, \end{aligned}$$

where $\tilde{\beta}_j^c = \hat{\beta}_j^c - \beta_j^*$, $\tilde{\pi}_k^c = \hat{\pi}_k^c - \pi_k^*$, and $\tilde{\theta}^c = \hat{\theta}^c - \theta^*$, where coefficients with a superscript “*” represent the non-discriminatory wage structure.

The identification of characteristics and coefficients effects at aggregate and detailed levels require identifying the non-discriminatory wage structure. The only exception is identifying coefficients effect of the intercept since $\hat{\alpha}^f - \hat{\alpha}^m = \tilde{\alpha}^f - \tilde{\alpha}^m$ because $\tilde{\alpha}^c = \hat{\alpha}^c - \alpha^*$. That is, the unobservable non-discriminatory wage structure for the constant term, α^* , is swept out.⁵

Since the non-discriminatory wage structure is not identifiable, researchers have to “arbitrarily” choose a non-discriminatory wage structure. Popular choices are pooled sample regression coefficients, or men’s regression coefficients, or women’s regression coefficients, or the average of men’s and women’s regression coefficients (see Neumark, 1988, Oaxaca and Ransom, 1988 and 1994 for details).

The first identification problem above is usually known as an index problem. Typically the second identification problem, **IP2**, is known as the identification problem in the detailed Oaxaca decomposition. This identification problem in the detailed Oaxaca decomposition of wage differentials is that the sum of the detailed coefficients effects attributed to dummy variables is not invariant to the choice of the omitted group(s). That is, $\sum_{j=2}^J (\hat{\beta}_j^f - \hat{\beta}_j^m) \bar{d}_j^f \neq \sum_{j=1}^{J-1} (\hat{\beta}_j^f - \hat{\beta}_j^m) \bar{d}_j^f$, where $\hat{\beta}$ is estimate of coefficient of industry dummy variable when

⁵Still, this does not mean that α^* is not useful and informative to learn the baseline favoritism and pure discrimination. It just says that α^* is not required for identifying the coefficients effect of the intercept.

an omitted industry is switched from the first industry to the last industry J . The same problem exists for coefficients effect of dummy variables of q 's. Since the coefficients of dummy variables shift the intercept, the intercept also changes when the reference groups change (see Oaxaca and Ransom, 1999, for details of this identification problem).

Yun (2005) and Gardeazabal and Ugidos (2004) suggest that the identification problem can be resolved by adopting a normalized equation which can identify the intercept and coefficients of all dummy variables including reference groups. The proto-type normalized regression corresponding to the usual Equation (1) looks like:

$$y = \alpha^+ + \sum_{j=1}^J \beta_j^+ d_j + \sum_{k=1}^K \pi_k^+ q_k + x\theta + \varepsilon.$$

Once consistent estimates of the usual equations are obtained, we follow the approach suggested by Suits (1984) when transforming these estimates by imposing restrictions $\sum_{j=1}^J \beta_j^+ = 0$ and $\sum_{k=1}^K \pi_k^+ = 0$.⁶ When the coefficients of the normalized equation are further specified as $\beta_j^+ = \beta_j + \mu_\beta$ and $\pi_k^+ = \pi_k + \mu_\pi$, then the solution for constraints are $\mu_\beta = -\bar{\beta} = -\sum_{j=1}^J \beta_j/J$ and $\mu_\pi = -\bar{\pi} = -\sum_{k=1}^K \pi_k/K$, where $\beta_1 = \pi_1 = 0$.⁷ The normalized equation is:

$$y = (\alpha + \bar{\beta} + \bar{\pi}) + \sum_{j=1}^J (\beta_j - \bar{\beta}) d_j + \sum_{k=1}^K (\pi_k - \bar{\pi}) q_k + x\theta + \varepsilon, \quad (2)$$

where $\alpha^+ = \alpha + \bar{\beta} + \bar{\pi}$, $\beta_j^+ = \beta_j - \bar{\beta}$, and $\pi_k^+ = \pi_k - \bar{\pi}$.

⁶Since the publication of Suits (1984), there have been several additional approaches for deriving the normalized equation, e.g., Greene and Seaks (1991), Gardeazabal and Ugidos (2004), and Yun (2005, 2008). See Yun (2005) for the details of each method. The restrictions, $\sum_{j=1}^J \beta_j^+ = 0$ and $\sum_{k=1}^K \pi_k^+ = 0$, though somewhat arbitrary, are widely used in ANOVA analysis. Yun (2005) shows that the ‘‘characteristics’’ effect of each individual dummy variable depends on the value of Ω when the restrictions are changed to $\sum_{j=1}^J \beta_j^+ = \Omega$ and $\sum_{k=1}^K \pi_k^+ = \Omega$, where $\Omega \neq 0$.

⁷This paper assumes that the simple average of the coefficients of the dummy variables is used to derive the normalized equation. Though it is easy to derive a normalized equation using the average of the dummy variables’ coefficients weighted by the share of each group as FW (1995) do, this has the implication that the sum of the product of the dummy variables and their coefficients should be zero, which is not attractive for the Oaxaca decomposition (Yun, 2005).

3 Existing Measures for Gender Wage Gaps by Industry

Based on the regression Equation (1), the overall gender wage gap in industry j can be decomposed using the same technique as Oaxaca (1973):

$$\begin{aligned} \bar{y}_j^f - \bar{y}_j^m &= (\hat{\alpha}^f - \hat{\alpha}^m) + (\hat{\beta}_j^f - \hat{\beta}_j^m) + \sum_{k=2}^K (\hat{\pi}_k^f - \hat{\pi}_k^m) \bar{q}_{jk}^f + \bar{x}_j^f (\hat{\theta}^f - \hat{\theta}^m) \\ &+ \sum_{k=2}^K \hat{\pi}_k^m (\bar{q}_{jk}^f - \bar{q}_{jk}^m) + (\bar{x}_j^f - \bar{x}_j^m) \hat{\theta}^m, \end{aligned} \quad (3)$$

where the four terms in the first line of (3) are the coefficients effect (unexplained components), while the two terms in the second line are characteristics effect (explained components) in industry j .⁸

Researchers have defined industrial gender wage gaps more or less related to the coefficients effect in Equation (3).⁹ Some of the proposed measures have suffered from the identification problems similar to those discussed in the previous section. As an attempt to capture the pure effect of industry on gender wage gaps, FW (1995) propose an estimator of gender wage gaps in industry j intuitively as differences in intercepts and coefficients of dummy variables for industry j between men and women as follows:

$$\hat{g}_j = (\hat{\alpha}^f - \hat{\alpha}^m) + (\hat{\beta}_j^f - \hat{\beta}_j^m). \quad (4)$$

The differences in the intercepts are the wage gaps for the reference group. If the differences in the intercepts are dropped, then the wage gap in the reference industry is zero, since $\hat{\beta}_1^f = \hat{\beta}_1^m = 0$.¹⁰

HO (2001) point out that the industrial gender wage gaps of FW (1995) have a serious identification problem, that is, if the reference group of the other set of dummy variables (q 's) changes, then \hat{g}_j is not invariant.¹¹ This is related to a symptom of the well-known

⁸The exact decomposition within a particular industry in Equation (3) can be obtained from estimates using the entire sample in Equation (1). This is due to the fact that the sample mean property of ordinary least squares with an intercept term – the industry sample mean wages (\bar{y}_j) are equivalent to the predicted industry wages (\hat{y}_j).

⁹See Rycx and Tojerow (2002), Adamchik and Bedi (2003), Gannon, *et al.* (2007) and UHJ (2009) for empirical applications of these measures.

¹⁰ \hat{g}_j has the nice property that even if the reference group for the industry dummy variables changes, the wage gap in industry j does not change; hence, the ranking of the wage gap of each industry is not changed.

¹¹This can be easily shown by changing the reference group from category 1 to category K for the other set of dummy variables (q 's). Then the wage gap in industry j is changed to $(\hat{\alpha}^f - \hat{\alpha}^m) + (\hat{\beta}_j^f - \hat{\beta}_j^m) + (\hat{\pi}_K^f - \hat{\pi}_K^m)$.

identification problem in the detailed Oaxaca decomposition analysis (**IP2**), that the differences in the intercept terms are not invariant to the choice of the reference group (Oaxaca and Ransom, 1999).

To obtain invariant estimates of industrial gender wage gaps, HO (2001) propose three alternatives:

$$\begin{aligned}\hat{\phi}_j &= (\hat{\alpha}^f - \hat{\alpha}^m) + (\hat{\beta}_j^f - \hat{\beta}_j^m) + \sum_{k=2}^K (\hat{\pi}_k^f - \hat{\pi}_k^m) \bar{q}_{jk}^f + \bar{x}_j^f (\hat{\theta}^f - \hat{\theta}^m) \\ \hat{\delta}_j &= (\hat{\alpha}^f - \hat{\alpha}^m) + (\hat{\beta}_j^f - \hat{\beta}_j^m) + \sum_{k=2}^K (\hat{\pi}_k^f - \hat{\pi}_k^m) \bar{q}_k^f + \bar{x}^f (\hat{\theta}^f - \hat{\theta}^m) \\ \hat{\gamma}_j &= \max_{n=1\dots J} \hat{g}_n - \hat{g}_j = \max_{n=1\dots J} \hat{\delta}_n - \hat{\delta}_j\end{aligned}$$

Note that \bar{x}_j^f and \bar{q}_{jk}^f in $\hat{\phi}_j$ are measured using female samples in industry j , while \bar{x}^f and \bar{q}_k^f in $\hat{\delta}_j$ are estimated using whole female samples. $\hat{\phi}_j$ is simply the overall coefficients effect when Oaxaca decomposition is applied to gender wage differentials in industry j (see Equation (3)). HO (2001) are concerned that this measure ($\hat{\phi}_j$) may cause the ranking of wage gaps by industry to change from the order of the measure of FW (1995). They suggest further modification by replacing mean characteristics of observations in the industry j with overall mean characteristics, i.e., replacing \bar{x}_j^f and \bar{q}_{jk}^f with \bar{x}^f and \bar{q}_k^f , respectively, which leads to $\hat{\delta}_j$.

$\hat{\gamma}_j$ is different from the two estimators we have just discussed since it is based on order statistic where wage gaps are expressed as differences from the largest wage gap, i.e., the algebraically maximum value of \hat{g}_j (see also Horrace, 2005). UHJ (2009) further develop a relative measure of gender wage gaps in industry j when more than one cross-section of data are available. Their ranking based measure is defined as:

$$\hat{\gamma}_{tj} = \max_{t,j} \hat{\delta}_{tj} - \hat{\delta}_{tj},$$

where t denotes the year, and $\hat{\delta}_{tj} = (\hat{\alpha}_t^f - \hat{\alpha}_t^m) + (\hat{\beta}_{tj}^f - \hat{\beta}_{tj}^m) + \sum_{k=2}^K (\tilde{\pi}_{tk}^f \bar{\mathbf{q}}_k^f - \tilde{\pi}_{tk}^m \bar{\mathbf{q}}_k^m) + (\bar{\mathbf{x}}_t^f \tilde{\theta}_t^f - \bar{\mathbf{x}}_t^m \tilde{\theta}_t^m)$ where $\bar{\mathbf{q}}_k$ is the mean over years in industry k and $\bar{\mathbf{x}}$ is the grand mean over years and industries; $\tilde{\pi}_{tk}^c = \hat{\pi}_{tk}^c - \pi_k^*$ and $\tilde{\theta}_t = \hat{\theta}_t^c - \theta^*$. $\hat{\gamma}_{tj}$ coincides with $\hat{\gamma}_j$ when there is only one year.

The measures of gender wage gaps by industry – except for FW’s (\hat{g}_j) – do not suffer from the second identification problem (**IP2**) when dummy variables are used in the regression.

UHJ (2009), however, point out that although $\hat{\phi}_j$ and $\hat{\delta}_j$ resolve the identification problem related to the coefficients effect of dummy variables (**IP2**), they do not escape from the identification problem related to a non-discriminatory wage structure (**IP1**). The measures $\hat{\phi}_j$ and $\hat{\delta}_j$ in HO (2001) based on an unobservable non-discriminatory wage structure, (θ^*, π_k^*) , are:

$$\begin{aligned}\hat{\phi}_j^* &= (\hat{\alpha}^f - \hat{\alpha}^m) + (\hat{\beta}_j^f - \hat{\beta}_j^m) + \left(\sum_{k=2}^K \tilde{\pi}_k^f \bar{q}_{jk}^f - \sum_{k=2}^K \tilde{\pi}_k^m \bar{q}_{jk}^m \right) + (\bar{x}_j^f \tilde{\theta}^f - \bar{x}_j^m \tilde{\theta}^m) \\ \hat{\delta}_j^* &= (\hat{\alpha}^f - \hat{\alpha}^m) + (\hat{\beta}_j^f - \hat{\beta}_j^m) + \left(\sum_{k=2}^K \tilde{\pi}_k^f \bar{q}_k^f - \sum_{k=2}^K \tilde{\pi}_k^m \bar{q}_k^m \right) + (\bar{x}^f \tilde{\theta}^f - \bar{x}^m \tilde{\theta}^m),\end{aligned}$$

where $\tilde{\theta}^c = \hat{\theta}^c - \theta^*$, and $\tilde{\pi}_k^c = \hat{\pi}_k^c - \pi_k^*$. Obviously, the measures $\hat{\phi}_j^*$ and $\hat{\delta}_j^*$ are not identified since (θ^*, π_k^*) is not identified.¹² The measures $\hat{\gamma}_j$ in HO (2001) and $\hat{\gamma}_{tj}$ in UHJ (2009) do not suffer from **IP1** since the unknown non-discriminatory wage structure is removed by using relative discrimination.

Table 1 summarizes the identification problems these existing industrial gender wage gap measures suffer from. Though it suffers from the second identification problem and motivates other measures, \hat{g}_j does not have the first identification problem. Interestingly, two measures developed to avoid the second identification problem, $\hat{\phi}_j$ and $\hat{\delta}_j$, are not free from the first identification problem. The ranking-based industrial gender wage gap measures, $\hat{\gamma}_j$ and $\hat{\gamma}_{tj}$, are currently the only estimators which overcome both identification problems, **IP1** and **IP2**, simultaneously.

Even though the estimators in HO (2001) and UHJ (2009) are partially or completely able to resolve the two identification problems, they have several shortcomings. We start with the measures $\hat{\phi}_j$ and $\hat{\delta}_j$ by HO (2001). The main drawback of $(\hat{\phi}_j, \hat{\delta}_j)$ and $(\hat{\phi}_j^*, \hat{\delta}_j^*)$ is that they depend on the mean characteristics in each industry and the whole sample, hence,

¹²Note that the decomposition equation for the overall gender wage gap in industry j can be decomposed with an unobserved non-discriminatory wage structure as follows:

$$\begin{aligned}\bar{y}_j^f - \bar{y}_j^m &= (\hat{\alpha}^f - \hat{\alpha}^m) + (\hat{\beta}_j^f - \hat{\beta}_j^m) + \left(\sum_{k=2}^K \tilde{\pi}_k^f \bar{q}_{jk}^f - \sum_{k=2}^K \tilde{\pi}_k^m \bar{q}_{jk}^m \right) + (\bar{x}_j^f \tilde{\theta}^f - \bar{x}_j^m \tilde{\theta}^m) \\ &\quad + \sum_{k=2}^K \pi_k^* (\bar{q}_{jk}^f - \bar{q}_{jk}^m) + (\bar{x}_j^f - \bar{x}_j^m) \theta^*.\end{aligned}$$

the pure industry effects on the gender wage gap are not measured.¹³ Though $\hat{\phi}_j$ and $\hat{\delta}_j$ in HO (2001) overcome the second identification problem (**IP2**), what they measure is not what researchers strive to measure: pure impact of industry on gender wage gaps. In addition, the standard errors for $\hat{\phi}_j$ and $\hat{\delta}_j$ are obtained according to the fixed regressors assumption on \bar{x}_j^f and \bar{x}^f . As Lin (2007) discusses, the standard errors of these measures can be varying depending on whether the regressors are fixed or stochastic.

Now we turn to the measures $\hat{\gamma}_j$ and $\hat{\gamma}_{tj}$ which are free from the two identification problems. Obviously, we cannot use them to test whether the industrial wage gaps in each industry are significantly different from zero. The usual statistical inference may suffer from the finite sample bias due to the sampling variability of the covariance matrix derived in HO (2001) and UHJ (2009). Horrace (2005) employs a sophisticated multiple comparison with the best (MCB) method to construct the confidence interval for $\hat{\gamma}_j$ via simulation. However, the MCB method to make inferences with $\hat{\gamma}_{tj}$ has not been yet fully developed.

Finally, we can find out that the ranking of $\hat{\delta}_j$ is identical to that of \hat{g}_j ; the ranking of $\hat{\gamma}_j$ is opposite to that of \hat{g}_j by construction, that is, the industry with the highest gender gap measured by \hat{g}_j has a value of zero and the industry with lower gender gap will have a positive value of $\hat{\gamma}_j$. $\hat{\gamma}_{tj}$ will coincide with $\hat{\gamma}_j$ in the year where the standard industry, i.e, the maximum $\hat{\delta}_{tj}$ over years and industries, is located. Note that the ranking of $\hat{\gamma}_{tj}$ is also opposite to that of \hat{g}_j by construction within each year. Though the values of two measures ($\hat{\gamma}_{tj}$ and $\hat{\gamma}_j$) do not coincide in years when the standard industry is not located, but the ranking of these two measures are identical within each year. Since $\hat{\phi}_j$ depends upon the mean individual characteristics of the females in each industry, the ranking of the measure based on $\hat{\phi}_j$ will be different from that based on other measures.

4 Estimating Gender Wage Gap by Industry Using Normalized Equations

We have seen pros and cons of existing measures of industrial gender wage gaps in the previous section. The ideal measure should capture pure impact of industry on gender wage gaps, while overcoming both identification problems and enable us to perform a hypothesis test whether the industrial gender wage gaps are significantly different from zero. We propose

¹³HO (2001, p. 616) note that “these results show that gender differences in personal characteristics by industry can mask the pure industry effects on the gender wage gap.”

new measures of industrial gender wage gaps based on a “normalized equation” introduced to resolve the second identification problem in detailed Oaxaca decomposition. The normalized equation is defined in Equation (2). The proposed measures based on normalized equations are free from the two identification problems. Furthermore, we can implement a simple hypothesis test whether industrial gender wage gaps by the proposed measures are significantly different from zero.

Though the potential usefulness of this approach to resolving the identification problems in measuring industrial gender wage gaps is recognized in Yun (2005, p. 771, footnote 14), new measures based on the normalized equations have not been proposed. This paper shows the usefulness of the normalized equations in resolving identification problems in measuring industrial gender wage gaps. Let us first look at a decomposition equation for gender wage gaps in industry j when normalized equations are used. The overall gender wage gaps in industry j can be decomposed using normalized equations as follows:

$$\begin{aligned} \bar{y}_j^f - \bar{y}_j^m &= (\hat{\alpha}^{f+} - \hat{\alpha}^{m+}) + (\hat{\beta}_j^{f+} - \hat{\beta}_j^{m+}) + \sum_{k=1}^K (\hat{\pi}_k^{f+} - \hat{\pi}_k^{m+}) \bar{q}_{jk}^f + \bar{x}_j^f (\hat{\theta}^f - \hat{\theta}^m) \\ &+ \sum_{k=1}^K \hat{\pi}_k^{m+} (\bar{q}_{jk}^f - \bar{q}_{jk}^m) + (\bar{x}_j^f - \bar{x}_j^m) \hat{\theta}^m, \end{aligned} \quad (5)$$

where $\hat{\alpha}^{c+} = \hat{\alpha}^c + \hat{\beta}^c + \hat{\pi}^c$, $\hat{\beta}_j^{c+} = \hat{\beta}_j^c - \hat{\beta}^c$, and $\hat{\pi}_k^{c+} = \hat{\pi}_k^c - \hat{\pi}^c$; the first and second lines represent coefficients and characteristics effects, respectively.

It is natural to devise a new estimator of industrial gender wage gaps using the coefficients of industry dummy variables in the normalized equation (5), i.e., inter-industry wage differentials ($\hat{\beta}_j^{f+}$ and $\hat{\beta}_j^{m+}$). Obviously, the normalized estimates ($\hat{\beta}_j^{f+}$ and $\hat{\beta}_j^{m+}$) do not suffer from the invariance problem unlike $\hat{\beta}_j^f$ and $\hat{\beta}_j^m$ in Equation (3) which are not invariant to the choice of the reference group. Following the spirit of FW (1995) shown in Equation (4), we propose a new measure that captures the pure impact of industry on gender wage gaps using the normalized equation as differences in coefficients of dummy variable for industry j

between men and women:¹⁴

$$\hat{h}_j^+ = \beta_j^{f+} - \beta_j^{m+} = (\hat{\beta}_j^f - \hat{\beta}_j^f) - (\hat{\beta}_j^m - \hat{\beta}_j^m).$$

The proposed estimator of industrial gender wage gaps represents gender wage gaps in industry j after eliminating all differences in characteristics and their coefficients except for inter-industry wage differentials in industry j between men and women.¹⁵ In addition to the fact that \hat{h}_j^+ captures the pure effect of industry on gender wage gaps which researchers have strived to measure, the proposed estimator has several additional nice properties as measures of the gender wage gap in industry j . First, \hat{h}_j^+ is free from the two identification problems noted in Section 2. It is not affected by the choice of the unobservable non-discriminatory wage structure, (θ^*, π_k^*) , hence, the first identification problem (**IP1**) disappears in estimating the proposed measure of industrial gender wage gaps.¹⁶ It is obvious that the proposed measure overcomes the second identification problem (**IP2**), since it is based on the estimates of the normalized equations.

Second, hypothesis testing for existence of gender wage gaps in each industry is easy to conduct using the new gender wage gap estimator, since the standard errors for the proposed measure is quite easy to derive by simply transforming estimates of the usual regression

¹⁴Alternatively, one might define the measure of FW (\hat{g}_j) using coefficients of the normalized equations as a measure of industrial gender wage gaps, that is, devising \hat{g}_j^+ as:

$$\hat{g}_j^+ = (\hat{\alpha}^{f+} - \hat{\alpha}^{m+}) + (\hat{\beta}_j^{f+} - \hat{\beta}_j^{m+}),$$

or equivalently,

$$\hat{g}_j^+ = (\hat{\alpha}^f - \hat{\alpha}^m) + (\hat{\pi}^f - \hat{\pi}^m) + (\hat{\beta}_j^f - \hat{\beta}_j^m).$$

\hat{g}_j^+ represents gender wage gaps in industry j after eliminating all differences in characteristics and their coefficients except for intercept and inter-industry wage differentials in industry j . This means that it captures gender wage gaps due to differences in baseline compensation in addition to differences in inter-industry wage differentials of industry j between men and women. That is, though \hat{g}_j^+ is free from both **IP1** and **IP2**, it fails to measure the pure industry effect on gender wage gaps.

¹⁵This means that \hat{h}_j^+ is absent from the differences in baseline compensation unlike \hat{g}_j^+ , therefore capturing the pure effect of industry on gender wage gaps.

¹⁶Note that, by taking the unobservable non-discriminatory wage structure into account, the decomposition of the overall gender wage gap in industry j could be rewritten as:

$$\begin{aligned} \bar{y}_j^f - \bar{y}_j^m &= (\hat{\alpha}^{f+} - \hat{\alpha}^{m+}) + (\hat{\beta}_j^{f+} - \hat{\beta}_j^{m+}) + \left(\sum_{k=1}^K (\hat{\pi}_k^{f+} - \pi_k^*) \bar{q}_{jk}^f - \sum_{k=1}^K (\hat{\pi}_k^{m+} - \pi_k^*) \bar{q}_{jk}^m \right) + (\bar{x}_j^f \bar{\theta}^f - \bar{x}_j^m \bar{\theta}^m) \\ &+ \sum_{k=1}^K \pi_k^* (\bar{q}_{jk}^f - \bar{q}_{jk}^m) + (\bar{x}_j^f - \bar{x}_j^m) \theta^*. \end{aligned}$$

equations readily available from standard statistical packages. This is a nice property of the proposed measure relative to ranking-based measures $\hat{\gamma}_j$ and $\hat{\gamma}_{tj}$ which require complicated simulation to compute the critical values. A simple algorithm to compute the standard errors for our estimator is presented in the Appendix.

Third, unlike $\hat{\gamma}_j$ and $\hat{\gamma}_{tj}$, we can explicitly test the significance of the proposed measure for all j including so-called standard industry to compute $\hat{\gamma}_j$ and $\hat{\gamma}_{tj}$. There is no way of testing the significance of the gap for the standard industry relative to itself, since the standard industry values for the relative measures $\hat{\gamma}_j$ and $\hat{\gamma}_{tj}$ are set to zero.

Finally, the proposed measure is absent from individual mean characteristics such as \bar{x}_j^f , \bar{q}_{jk}^f , \bar{x}^f , and \bar{x}^m , unlike previous measures, e.g., $\hat{\phi}_j$, $\hat{\delta}_j$ and $\hat{\gamma}_{tj}$. Furthermore, we can avoid complicating computation of the standard errors required when the regressors are stochastic.

5 Empirical Illustration

We demonstrate how to implement the proposed estimator using the March 1998 CPS employed in HO (2001), and compare it with previous estimators.¹⁷ The samples from 1998 CPS used in HO (2001) contain 27,426 males and 25,444 females. The log-hourly wages are regressed on various socio-economic variables such as education, potential experience, size of population of residence, urban residence, three dummy variables for region of residence, marital status, race, and twelve dummy variables for occupation, in addition to thirteen dummy variables for industry (i.e., $J = 14$).

Table 2 reports both usual and normalized regression estimates for men and women. The normalized equation is derived by transforming the estimates of the usual regression equation applying the algorithm presented in the Appendix. To check identification issues, we compute industrial gender wage gaps using the two proposed measures in addition to other previous measures while changing the choice of non-discriminatory wage structures (men or women) and the reference group for race (non-whites or whites). The proposed measure (\hat{h}_j^+) is free from the two identification problems, that is, the values of the measure does not change whether the assumed non-discriminatory wage structure is male or female,

¹⁷Since the empirical example is based on single cross sectional data, we cannot compare our suggested estimator with some existing estimators derived for data over multiple years. Another empirical example using Taiwan's Manpower Utilization Survey from 1999 to 2003 is available from the online version of this paper. In the additional example, the suggested estimator is compared with existing estimators designed for multiple year data including $\hat{\gamma}_{tj}$ and $\hat{\delta}_{tj}$.

and whether the reference group for race is white or non-white, as shown in the last column (column (9)) in Table 3; The columns 2-5 in Table 3 show that the values of $\hat{\delta}_j$ and $\hat{\phi}_j$ proposed by HO (2001) are changing depending whether the choice of non-discriminatory wage structure is male or female, suffering from the first identification problem (**IP1**); As reported in columns 6-7 in Table 3, \hat{g}_j , the measure FW (1995) propose, suffers from the second identification problem (**IP2**), that is, the value of \hat{g}_j changes when the reference group of race is changed from non-white to white; $\hat{\gamma}_j$ does not suffer from the two identification problems as the proposed measure in this paper, as shown in the last two columns.

We can easily verify that all absolute measures, \hat{g}_j , $\hat{\delta}_j$, and \hat{h}_j^+ (except for $\hat{\phi}_j$), have identical ordering because they are devised via affine transformations. Table 3 shows that the differences among these absolute measures except for $\hat{\phi}_j$ are constant, that is, -0.256 between \hat{g}_j and $\hat{\delta}_j$, 0.075 between \hat{g}_j and \hat{h}_j^+ , and 0.181 between $\hat{\delta}_j$ and between \hat{h}_j^+ for all j . As for the relative measure $\hat{\gamma}_j$, the ranking is just opposite to those of the absolute measures by construction, which is also confirmed in the table. In addition, it is worth noting that the overall gap $(\bar{y}_j^f - \bar{y}_j^m)$ and measure $\hat{\phi}_j$ – aggregate coefficients effect – in general have different rankings from absolute and relative measures of industrial gender wage gaps.

The relative measure, $\hat{\gamma}_j$, shows that the differences in industrial gender wage gaps between the standard (Agriculture, Forestry and Fisheries) and industry j are not statistically significantly different from zero. We can examine whether there are substantial gender wage gaps in each industry too. The values of $\hat{\delta}_j$ and $\hat{\phi}_j$ are negative for all industries and most of them are significant at the 5% level. This implies that women are receiving lower wages after controlling differences in characteristics.

On the other hand, the value for the proposed measure \hat{h}_j^+ , is mixed, which indicates that women are not receiving lower wages in every industry after controlling differences in characteristics and their coefficients (see e.g. positive signs in Agriculture, Forestry and Fisheries, Wholesale Trade industry, and Durables Manufacturing, and etc.). It is also interesting that gender wage gaps measured by \hat{h}_j^+ are not significant in most industries. This indicates that there may not be substantial industrial gender wage gaps after controlling differences in characteristics and their coefficients, a stark contrast to the finding that there are significant gender wage gaps when $\hat{\delta}_j$ and $\hat{\phi}_j$ are used. This contrast provides us a cautionary note that our view on existence and size of industrial gender wage gaps may depend on what measures we are using, and that different estimators measure industrial

gender wage gaps from different perspectives. In particular, our propose measure (\hat{h}_j^+) focuses on only differences in inter-industry wage differentials for industry j , while $\hat{\phi}_j$ captures the overall coefficients effect for industry j .¹⁸ Thus, it would be more appropriate to adopt our measure (\hat{h}_j^+) if researchers seek to estimate the pure industrial effect on gender wage differentials.

6 Conclusion

The identification problem raised by HO (2001) for the measure of industrial gender wage gaps proposed by FW (1995) can be resolved by using the normalized regression equation used for resolving the well-known identification problem in the detailed Oaxaca decomposition. The proposed measure overcomes another identification problem, the so-called index problem, since it does not rely on the unobserved non-discriminatory wage structure. That is, the proposed measure is free from two types of the identification problems. Though the existing remedies proposed by HO (2001) and others may overcome one or both identification problems, they do not accomplish what researchers including FW (1995) set out to measure, that is, pure impact of industry on gender wage gaps. The measure of industrial gender wage gaps proposed in this paper using the normalized regression equation satisfies all requirements to be a good measure of gender wage gaps by industry — it provides an invariant estimator immune to identification problems while successfully measuring what researchers strive to capture, the pure impact of industry to gender wage gaps. Additional nice feature is that the standard error of the proposed estimator can be easily calculated for hypothesis testing, while existing relative measures based on ranking have some difficulties.

We demonstrate the easiness and usefulness of the proposed estimator with the March 1998 CPS adopted in HO (2001). Not surprisingly, each measure shows different levels of and trends in industrial gender wage gaps, reminding us of the importance of developing and employing an appropriate measure. Overall, the proposed measure shows that the gender wage gaps due to differences in inter-industry wage differentials between men and women are not significant in most industries. Of course, this does not mean that there are no substantial gender wage gaps since the overall gender wage gaps by industry due to

¹⁸Note that \hat{g}_j^+ mentioned in Footnote 14 captures baseline gender disparity too by adding differences in intercepts to \hat{h}_j^+ .

differences in regression estimates of all variables are substantial in most industries.¹⁹

We face several identification problems when constructing various measures of gender wage gaps, either overall or industry-wise, which have hampered undertaking of rigorous examination of the gender wage gaps. As shown in devising the alternative measure for industrial gender wage gaps, the normalized equation holds the key to various identification problems when studying wage differentials using Oaxaca decomposition. Thanks to the normalized equation, we are finally turning the corner in overcoming identification problems. As a final note, though the measures are developed to study the industrial gender wage gaps, their application is not limited to studying industrial gender wage gaps – they can be easily applied other areas such as studying racial wage gaps by industry or occupation.

¹⁹One might interpret the results as indicating that significance of industry in studying gender wage gaps is low. However, this does not necessarily mean that choice of industry is not important since the compositional differences in industry between men and women contribute to the gender wage gaps. See FW (1995) for the role of composition of industry in explaining gender wage gaps. Detailed and in-depth studies on the role of industry, in terms of composition and inter-industry wage differentials, in explaining gender wage gaps is required to set appropriate policies. The proposed measure in addition to other various existing measures will provide useful tools for the future studies.

References

- [1] Abowd, J. M., F. Karamariz, and D. N. Margolis (1999): “High Wage Workers and High Wage Firms,” *Econometrica*, 67(2), 251–333.
- [2] Adamchik, V. A. and A. S. Bedi (2003): “Gender Pay Differentials During the Transition in Poland,” *Economics of Transition*, 11(4), 697–726.
- [3] Blinder, A. S. (1973): “Wage Discrimination: Reduced Form and Structural Estimates,” *Journal of Human Resources*, 8(4), 436–455.
- [4] Brown, R. S., M. Marilyn, and S. Z. Barbara (1980): “Incorporating Occupational Attainment in Studies of Male-female Earnings Differentials,” *Journal of Human Resources*, 15(1), 3–28.
- [5] Fang, Z. and C. Sakellariou (2011): “A Case of Sticky Floors: Gender Wage Differentials in Thailand,” *Asian Economic Journal*, 25(1), 35–54.
- [6] Fields, J. and E. M. Wolff (1995): “Interindustry Wage Differentials and the Gender Wage Gap,” *Industrial and Labor Relations Review*, 49(1), 105–120.
- [7] Fortin, N., T. Lemieux, and S. Firpo (2011): “Decomposition Methods in Economics,” in Orley Ashenfelter and David Card, *Handbook of Labor Economics*, Volume 4A, 1–102
- [8] Gannon, B., R. Plasman, F. Rycx, and I. Tojerow (2007): “Interindustry Wage Differentials and the Gender Wage Gap: Evidence from European Countries,” *The Economic and Social Review*, 38(1), 135–155.
- [9] Gardeazabal, J. and A. Ugidos (2004): “More on Identification in Detailed Wage Decompositions,” *Review of Economics and Statistics*, 86(4), 1034–1036.
- [10] Gibbons, F. and L. Katz (1992): “Does Unmeasured Ability Explain Inter-industry Wage Differentials?” *Review of Economic Studies*, 59, 515–535.
- [11] Greene, W. and T. Seaks (1991): “The Restricted Least Squares Estimator: A Pedagogical Note,” *Review of Economics and Statistics*, 73, 563–567.

- [12] Haisken-DeNew, J. P. and C. M. Schmidt (1997): “Interindustry and Interregion Differentials: Mechanics and Interpretation,” *Review of Economics and Statistics*, 79(3), 516–521.
- [13] Helwege, J. (1992): “Sectoral Shifts and Inter-Industry Wage Differentials,” *Journal of Labor Economics*, 10(1), 55–84.
- [14] Horrace, W. C. (2005): “On the Ranking Uncertainty of Labor Market Wage Gaps,” *Journal of Population Economics*, 18, 181–187.
- [15] Horrace, W. C. and R. Oaxaca (2001): “Inter-industry Wage Differentials and the Gender Wage Gap: An Identification Problem,” *Industrial and Labor Relations Review*, 54(3), 611–618.
- [16] Kapsos, S. (2009): “The Gender Wage Gap in Bangladesh,” ILO Asia-Pacific Working Paper Series, International Labour Organization.
- [17] Kline, P. (2011): “Oaxaca-Blinder as a Reweighting Estimator,” *American Economic Review: Papers & Proceedings*, 101(3), 532–537.
- [18] Lin, E. S. (2007): “On the Standard Errors of Oaxaca-type Decompositions for Inter-industry Gender Wage Differentials,” *Economics Bulletin*, 10(6), 1–11.
- [19] Lin, E. S. (2010): “Gender Wage Gaps by College Major in Taiwan: Empirical Evidence from 1978-2003 Manpower Utilization Survey,” *Economics of Education Review*, 29(1), 156–164.
- [20] Nielsen, H. S. (2000): “Wage Discrimination in Zambia: An Extension of the Oaxaca-Blinder Decomposition,” *Applied Economics Letters*, 7, 405–408.
- [21] Neumark, D. (1988): “Employers’ Discriminatory Behavior and the Estimation of Wage Discrimination,” *Journal of Human Resources*, 23(3), 279–295.
- [22] Oaxaca, R. L. (1973): “Male-Female Wage Differentials in Urban Labor Markets,” *International Economic Review*, 14(3), 693–709.
- [23] Oaxaca, R. L. and M. R. Ransom (1988): “Searching for the effect of unionism on the wages of union and nonunion workers,” *Journal of Labor Research*, 9(2), 139–148.

- [24] Oaxaca, R. L. and M. R. Ransom (1994): “On Discrimination and the Decomposition of Wage Differentials,” *Journal of Econometrics*, 61(1), 5–21.
- [25] Oaxaca, R. L. and M. R. Ransom (1999): “Identification in Detailed Wage Decomposition,” *Review of Economics and Statistics*, 81(1), 154–157.
- [26] Reimers, C. W. (1983): “Labor Market Discrimination Against Hispanic and Black Men,” *Review of Economics and Statistics*, 65(4), 570–579.
- [27] Rycx, F. and I. Tojerow (2002): “Inter-Industry Wage Differentials and the Gender Wage Gap in Belgium,” *Brussels Economic Review*, 45(2), 119–141.
- [28] Suits, D. B. (1984): “Dummy Variables: Mechanics v.s. Interpretation,” *Review of Economics and Statistics*, 66(1), 177–180.
- [29] Ural, B. P., W. C. Horrace, and J. H. Jung (2009): “Inter-Industry Gender Wage Gaps by Knowledge Intensity: Discrimination and Technology in Korea,” *Applied Economics*, 41(11), 1437–1452.
- [30] Yun, M.-S. (2005): “A Simple Solution to the Identification Problem in Detailed Wage Decomposition,” *Economic Inquiry*, 43(4), 766–772.
- [31] Yun, M.-S. (2007): “An Extension of the Oaxaca Decomposition using Generalized Residuals,” *Journal of Economic and Social Measurement*, 32(1), 15–22.
- [32] Yun, M.-S. (2008): “Identification Problem and Detailed Oaxaca Decomposition: A General Solution and Inference,” *Journal of Economic and Social Measurement*, 33(1), 27–38.

Appendix

This appendix shows an algorithm to derive the normalized equation by transforming the usual regression equation.²⁰

Algorithm for Computing Normalized Regression Estimates

First, we augment the omitted categories to the matrix of independent variables ($Z \rightarrow Z^*$), and zeros to usual regression estimates ($\eta \rightarrow \eta_0$) and their covariance matrix ($\Sigma_{\hat{\eta}} \rightarrow \Sigma_{\hat{\eta}_0}$). Second, using a weight matrix, \mathbf{W} , estimates of the normalized regression equation can be obtained by transforming η_0 to $\eta^* = \mathbf{W}\eta_0$. This transformation sets the deviation from the mean of estimates of dummy variables as normalized estimates of the dummy variables. A matrix representation of the equation (1) is, $Y = Z\eta + \varepsilon$ while suppressing the gender superscript. Y is a $(N \times 1)$ vector of the dependent variable; $Z = (\iota : d : q : x)$, where $d = (d_2, \dots, d_J)$, $q = (q_2, \dots, q_K)$, and $x = (x_1, \dots, x_L)$; Z , d , q , and x are, respectively, $N \times T$, $N \times (J - 1)$, $N \times (K - 1)$ and $N \times L$ matrices where $T = (J + K + L - 1)$, and d and q are matrices of two sets of dummy variables, and x is a matrix of continuous variables; ι is a vector of ones; $\eta = (\alpha, \beta_2, \dots, \beta_J, \pi_2, \dots, \pi_K, \theta_1, \dots, \theta_L)'$ is a $(T \times 1)$ coefficient vector. We can define $\beta = (\beta_2, \dots, \beta_J)'$, $\pi = (\pi_2, \dots, \pi_K)'$, and $\theta = (\theta_1, \dots, \theta_L)'$. In order to obtain the normalized equations (2), it is useful to rewrite the equation as $Y = Z^*\eta_0 + \varepsilon$, where $Z^* = (\iota : d_1 : d : q_1 : q : x)$ and $\eta_0 = (\alpha, 0, \beta', 0, \pi', \theta)'$. The normalized regression equation, $Y = Z^*\eta^* + \varepsilon$, is obtained by transforming η_0 to η^* using a weight matrix, \mathbf{W} , that is, $\eta^* = \mathbf{W}\eta_0$, which yields a $(T^* \times 1)$ vector of $((\alpha + \bar{\beta} + \bar{\pi}), (\beta_1 - \bar{\beta}), (\beta - \bar{\beta} \cdot \iota)', (\pi_1 - \bar{\pi}), (\pi - \bar{\pi} \cdot \iota)', \theta)'$, where $T^* = T + 2 = J + K + L + 1$. The weight matrix \mathbf{W} is defined as:

$$\mathbf{W} = \begin{bmatrix} 1 & (1/J) \cdot \iota_{(1 \times J)} & (1/K) \cdot \iota_{(1 \times K)} & 0_{(1 \times L)} \\ 0_{(J \times 1)} & R_{(J \times J)} & 0_{(J \times K)} & 0 \\ 0_{(K \times 1)} & 0_{(K \times J)} & R_{(K \times K)} & 0 \\ 0_{(L \times 1)} & 0_{(L \times J)} & 0_{(L \times K)} & \mathbf{I}_{(L \times L)} \end{bmatrix},$$

where $R_{(P \times P)} = \mathbf{I}_{(P \times P)} - (1/P) \cdot \mathbf{I}_{(P \times P)}$, and 0 and \mathbf{I} are a matrix of zeros and an identity matrix.²¹

²⁰The derivation of the normalized regression equation is developed by extending a method employed by Haisken-DeNew and Schmidt (1997).

²¹A normalized regression with weighted average of coefficients of dummy variables can be easily obtained by changing the weight matrix, \mathbf{W} . Let $W_d = (W_{d_1}, \dots, W_{d_J})$ and $W_q = (W_{q_1}, \dots, W_{q_K})$ be vectors of shares of dummy variables, d and q . In order to find a weight matrix for obtaining a normalized regression

Algorithm for Computing Standard Errors

Similarly, covariance matrix of the estimates of the normalized regression equation can be obtained by transforming $\Sigma_{\hat{\eta}_0}$ to $\Sigma_{\hat{\eta}^*} = \mathbf{W}\Sigma_{\hat{\eta}_0}\mathbf{W}'$. The covariance matrix of estimates of the normalized regression equation (η^*) is computed as $\Sigma_{\hat{\eta}^*} = \mathbf{W}\Sigma_{\hat{\eta}_0}\mathbf{W}'$, where the covariance matrix for $\hat{\eta}_0$ ($\Sigma_{\hat{\eta}_0}$) can be obtained by adding zero vectors to the covariance matrix of $\hat{\eta}$ as:

$$\Sigma_{\hat{\eta}_0} = \begin{bmatrix} \sigma_{\hat{\alpha}}^2 & 0 & \Sigma_{\hat{\alpha},\hat{\beta}'} & 0 & \Sigma_{\hat{\alpha},\hat{\pi}'} & \Sigma_{\hat{\alpha},\hat{\theta}'} \\ 0 & 0 & \mathbf{0}_{1 \times (J-1)} & 0 & \mathbf{0}_{1 \times (K-1)} & \mathbf{0}_{1 \times L} \\ \Sigma_{\hat{\beta},\hat{\alpha}} & \mathbf{0}_{(J-1) \times 1} & \Sigma_{\hat{\beta},\hat{\beta}'} & \mathbf{0}_{(J-1) \times 1} & \Sigma_{\hat{\beta},\hat{\pi}'} & \Sigma_{\hat{\beta},\hat{\theta}'} \\ 0 & 0 & \mathbf{0}_{1 \times (J-1)} & 0 & \mathbf{0}_{1 \times (K-1)} & \mathbf{0}_{1 \times L} \\ \Sigma_{\hat{\pi},\hat{\alpha}} & \mathbf{0}_{(K-1) \times 1} & \Sigma_{\hat{\pi},\hat{\beta}'} & \mathbf{0}_{(K-1) \times 1} & \Sigma_{\hat{\pi},\hat{\pi}'} & \Sigma_{\hat{\pi},\hat{\theta}'} \\ \Sigma_{\hat{\theta},\hat{\alpha}} & \mathbf{0}_{L \times 1} & \Sigma_{\hat{\theta},\hat{\beta}'} & \mathbf{0}_{L \times 1} & \Sigma_{\hat{\pi}\hat{\theta}} & \Sigma_{\hat{\theta},\hat{\theta}'} \end{bmatrix}.$$

Once asymptotic covariance matrix for the estimates in the normalized equation is computed, then it is trivial task to calculate the asymptotic variance of \hat{h}_j^+ .

with weighted averages, replace $(1/J) \cdot \iota_{(1 \times J)}$ and $(1/K) \cdot \iota_{(1 \times K)}$ with W_d and W_q , and replace $R_{(P \times P)} = \mathbf{I}_{(P \times P)} - (1/P) \cdot \mathbf{I}_{(P \times P)}$ with $R_{(J \times J)} = \mathbf{I}_{(J \times J)} - W_d^\dagger$ and $R_{(K \times K)} = \mathbf{I}_{(K \times K)} - W_q^\dagger$ when $P = J$ and K , respectively. $W_d^\dagger = (W'_d, \dots, W'_d)$ and $W_q^\dagger = (W'_q, \dots, W'_q)$, which are $J \times J$ and $K \times K$ matrices, respectively.

Table 1: Existing Measures of Industrial Gender Wage Gaps

		IP1	IP2	Absolute or Relative
\hat{g}_j	(FW, 1995)	NO	YES	Absolute
$\hat{\phi}_j$	(HO, 2001)	YES	NO	Absolute
$\hat{\delta}_j$	(HO, 2001)	YES	NO	Absolute
$\hat{\gamma}_j$	(HO, 2001)	NO	NO	Relative
$\hat{\gamma}_{tj}$	(UHJ, 2009)	NO	NO	Relative

* YES means that the measure is not free from the identification problem.

Table 2: Split Sample Regression Results for 1998 CPS

<i>Variables</i>	Women					Men				
	Mean	Coeff.	S.E.	Coeff.	S.E.	Mean	Coeff.	S.E.	Coeff.	S.E.
Constant	1.000	1.199*	.073	1.080*	.032	1.000	1.055*	.067	1.099*	.030
Education	13.531	.068*	.002	.068*	.002	13.460	.070*	.002	.070*	.002
Experience	19.514	.021*	.001	.021*	.001	19.534	.032*	.001	.032*	.001
Experience ² /1000	574.32	-.321*	.020	-.321*	.020	536.87	-.464*	.020	-.464*	.020
Urban	.792	-.004	.014	-.002	.007	.799	.010	.014	.005	.007
SMSA	3.955	.032*	.002	.032*	.002	4.029	.025*	.002	.025*	.002
Married	.543	.044*	.008	.022*	.004	.621	.152*	.009	.076*	.004
White	.848	.020	.011	.010*	.005	.877	.092*	.011	.046*	.006
<i>Regions:</i>										
Northeast [†]	.211			.034*	.007	.216			.017*	.007
Midwest	.242	-.039*	.012	-.005	.007	.235	-.009	.011	.008	.007
South	.300	-.069*	.011	-.034*	.006	.300	-.051*	.011	-.034*	.006
West	.247	-.030*	.011	.005	.007	.259	-.009	.011	.008	.006
<i>Occupations:</i>										
Exec., Admin. & Managerial [†]	.139			.228*	.017	.147			.248*	.018
Professional Specialty	.177	.019	.014	.247*	.018	.136	-.021	.015	.228*	.019
Technician	.039	-.017	.022	.211*	.022	.031	-.078*	.023	.170*	.025
Sales	.122	-.217*	.017	.011	.018	.110	-.144*	.016	.104*	.019
Admin. Support & Clerical	.258	-.260*	.013	-.032*	.016	.062	-.346*	.018	-.097*	.021
Private Households	.013	-.420*	.041	-.192*	.039	.000	-.490*	.177	-.241	.163
Protective Service	.007	-.239*	.046	-.011	.044	.029	-.243*	.026	.005	.027
Service	.150	-.359*	.016	-.130*	.017	.083	-.430*	.018	-.182*	.020
Precision Products	.018	-.294*	.031	-.065*	.030	.192	-.213*	.014	.035	.018
Machine operator	.051	-.405*	.023	-.177*	.023	.076	-.336*	.018	-.087*	.020
Transportation	.009	-.318*	.041	-.090*	.039	.070	-.336*	.018	-.088*	.020
Handlers & Cleaners	.016	-.348*	.033	-.120*	.031	.060	-.357*	.019	-.109*	.021
Farming, Forestry & Fisheries	.001	-.109	.156	.120	.143	.004	-.234*	.077	.014	.071
<i>Industry:</i>										
Agric., Forestry, & Fisheries [†]	.004			-.114	.061	.007			-.183*	.054
Mining	.002	.363*	.107	.248*	.080	.011	.370*	.068	.187*	.033
Construction	.012	.196*	.073	.082*	.033	.096	.272*	.060	.089*	.014
Durables Manufacturing	.058	.222*	.067	.108*	.018	.141	.266*	.059	.083*	.012
Non-Durables Manufacturing	.055	.122	.067	.008	.019	.078	.233*	.060	.050*	.014
Trans., Comm., & Utilities	.046	.244*	.067	.130*	.019	.099	.299*	.059	.116*	.013
Wholesales Trade	.023	.177*	.069	.063*	.025	.051	.184*	.061	.001	.017
Retail Trade	.181	-.080	.066	-.194*	.014	.162	.013	.059	-.170*	.012
Finance, Insur. & Real Estate	.087	.152*	.066	.038*	.015	.051	.310*	.061	.127*	.017
Business & Repair Service	.045	.038	.067	-.076*	.019	.073	.108	.060	-.075*	.014
Personal Services	.047	-.010	.068	-.124*	.022	.019	.051	.064	-.132*	.026
Entertainment	.017	-.034	.071	-.148*	.029	.016	.091	.065	-.092*	.028
Professional & Related Services	.376	.033	.065	-.081*	.012	.144	.108	.059	-.075*	.012
Public Administration	.048	.178*	.067	.063*	.019	.054	.256*	.061	.073*	.018
R-square			0.256					0.325		
Observations			25,444					27,426		

Notes: (1) † denotes reference groups. (2) Complementary groups for single dummy variables, Urban, SMSA (Standard Metropolitan Statistical Area), Married, and White, are not reported. For the normalized equation, the coefficients of complementary groups are identical to those of the single dummy variables, but in opposite sign; their standard errors are identical to those of single dummy variables. (3) * denotes statistically significant at the 5% significance level.

Table 3: Comparison of Industrial Gender Wage Gap Estimators for 1998 CPS

	Overall		Non-discriminatory wage structure			Reference group		Identified measures	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\hat{y}_j^f - \bar{y}_j^m$	$\hat{\delta}_j$	$\hat{\phi}_j$	$\hat{\delta}_j$	$\hat{\phi}_j$	\hat{g}_j	\hat{g}_j	$\hat{\gamma}_j(\hat{\tau}_j)$	\hat{h}_j^+
Ranked Industries									
Agric., Forestry, & Fisheries	-0.098 (1.139)	-0.112 (.087)	-0.097 (.078)	-0.161 (.087)	-0.129 (.089)	.144 (.099)	.072 (.098)	.000 (.000)	.069 (.081)
Wholesale Trade	-0.187 (.901)	-0.118* (.030)	-0.153* (.030)	-0.168* (.031)	-0.203* (.030)	.138* (.057)	.065 (.055)	.007 (.092)	.062* (.030)
Mining	-0.101 (.827)	-0.119 (.093)	-0.147 (.093)	-0.169* (.093)	-0.244* (.093)	.137 (.104)	.065 (.103)	.007 (.127)	.061 (.086)
Durables Manufacturing	-0.284 (.901)	-0.156* (.021)	-0.214* (.019)	-0.205* (.020)	-0.245* (.019)	.100 (.052)	.028 (.050)	.044 (.089)	.025 (.022)
Transp. & Communications	-0.190 (.909)	-0.167* (.022)	-0.150* (.022)	-0.206* (.022)	-0.223* (.023)	.089 (.053)	.017 (.051)	.055 (.090)	.013 (.023)
Personal Services	-0.304 (1.003)	-0.173* (.034)	-0.136* (.054)	-0.222* (.035)	-0.170* (.033)	.083 (.059)	.011 (.057)	.061 (.094)	.008 (.034)
Business & Repair Services	-0.221 (1.049)	-0.182* (.023)	-0.161* (.023)	-0.232* (.023)	-0.212* (.023)	.074 (.053)	.002 (.050)	.070 (.090)	-.002 (.024)
Prof. & Related Services	-0.339 (1.043)	-0.187* (.013)	-0.178* (.012)	-0.237* (.015)	-0.204* (.012)	.069 (.051)	-.003 (.048)	.075 (.088)	-.007 (.017)
Construction	-0.085 (.961)	-0.188* (.038)	-0.196* (.038)	-0.237* (.037)	-0.285* (.041)	.068 (.060)	-.004 (.057)	.076 (.094)	-.008 (.036)
Public Administration	-0.285 (.834)	-0.190* (.026)	-0.187* (.025)	-0.239* (.026)	-0.243* (.027)	.066 (.055)	-.006 (.053)	.078 (.091)	-.009 (.026)
Retail Trade	-0.203 (.980)	-0.204* (.015)	-0.206* (.013)	-0.254* (.016)	-0.208* (.013)	.052 (.049)	-.021 (.046)	.093 (.098)	-.024 (.018)
Non-Durables Manufacturing	-0.341 (.903)	-0.223* (.023)	-0.282* (.021)	-0.272* (.022)	-0.302* (.021)	.034 (.053)	-.039 (.050)	.111 (.090)	-.042* (.024)
Entertainment	-0.245 (1.075)	-0.237* (.041)	-0.200* (.041)	-0.286* (.042)	-0.217* (.041)	.019 (.062)	-.053 (.060)	.125 (.096)	-.057 (.040)
Finance, Insur. & Real Estate	-0.431 (1.045)	-0.270* (.021)	-0.268* (.021)	-0.320* (.022)	-0.318* (.021)	-.014 (.053)	-.086 (.050)	.158 (.090)	-.090* (.023)

Notes: (1) The measures from columns 2-3 and columns 4-5 are computed based on the male and female wage structures, respectively.

(2) We consider the change of the left-out reference group for the race dummy at columns 6-7. (3) Standard errors are in parentheses.

(4) “***” Denotes statistically significant at the 5% significance level.