

IZA DP No. 2427

## Revisiting Inter-Industry Wage Differentials and the Gender Wage Gap: An Identification Problem

Myeong-Su Yun

November 2006

# Revisiting Inter-Industry Wage Differentials and the Gender Wage Gap: An Identification Problem

**Myeong-Su Yun**

*Tulane University  
and IZA Bonn*

Discussion Paper No. 2427  
November 2006

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

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 company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. 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

### **Revisiting Inter-Industry Wage Differentials and the Gender Wage Gap: An Identification Problem\***

We propose a measure of the industrial gender wage gap which is free from an identification problem by using inter-industry wage differentials, or industrial wage premia. We draw on a recent literature showing that a normalized regression equation can be used to resolve the identification problem in detailed Oaxaca decompositions of wage differentials. By identifying the constant and the coefficients of dummy variables, including the reference category, the normalized equation can resolve the two key identification problems that arise in studying wage gaps: one in detailed Oaxaca decompositions; the other measuring industrial gender wage gaps.

JEL Classification: C20, J70

Keywords: industrial wage differentials, invariance, identification, normalized regression

Corresponding author:

Myeong-Su Yun  
Department of Economics  
Tulane University  
206 Tilton Hall  
New Orleans, 70118 LA  
USA  
E-mail: [msyun@tulane.edu](mailto:msyun@tulane.edu)

---

\* The author wishes to thank Ira Gang for comments.

## I. Introduction

Fields and Wolff (1995) study whether there might be correlation between inter-industry wage differentials and the gender wage gap. Inter-industry wage differentials are defined as the deviations from the (weighted) mean of coefficients of industry dummy variables, following Krueger and Summers (1988). Fields and Wolff (1995) define a gender wage gap in industry  $j$  intuitively as differences between men and women in intercept terms and coefficients of dummy variables for industry  $j$ .

Horrace and Oaxaca (2001) point out that there is an identification problem since the definition of the gender wage gap by Fields and Wolff (1995) is not invariant to the choice of reference groups for other sets of dummy variables. This is because “intercept terms capture not only the omitted industry but also the omitted category for any other binary variables in the specification” (Horrace and Oaxaca, 2001, p. 612). To fix the identification problem, Horrace and Oaxaca (2001) modify the definition of the gender wage gap in industry  $j$  by adding wage differentials due to differences in coefficients of other than industry dummy variables.<sup>1</sup> Though the modified wage gap resolves the identification problem, it does not capture what is usually thought of as the gender wage gap in industry  $j$ , since it includes wage differences due to differences in the coefficients of other variables. Hence, it does not isolate the wage gap due solely to the coefficients of industry  $j$ , which is what Fields and Wolff (1995) measure.

To derive a new measure of the gender wage gap in industry  $j$ , we take one step back and

---

<sup>1</sup> They also suggest another definition of the gender gap based on an order statistic where wage gaps are expressed as a percentage of the largest wage gap. This ranking of wage gaps is further developed in Horrace (2005). We do not discuss this order statistic in detail because the issue is automatically resolved once we are able to compute an industrial gender wage gap which does not suffer from an identification problem.

re-establish the relationship between inter-industry wage differentials and the gender wage gap which motivated Fields and Wolff (1995). We show that the inter-industry wage differentials, or industrial wage premia, are actually coefficients in the normalized regression equation which identifies the intercept and all coefficients of the dummy variables including the reference category. Therefore, the gender gap due to differences in industry wage premia only can be calculated by the differences in the coefficients of industry dummy variables in the normalized equation. It is interesting to see that the normalized regression, a solution to the identification problem in the detailed Oaxaca decomposition, also provides a way to calculate the gender wage gap in each industry which is free from the identification problem.

## II. Calculating Industrial Gender Wage Gap Using a Normalized Equation

We examine the following regression model containing dummy variables for industry ( $d$ 's).

$$y_g = \alpha_g + \left[ \sum_{j=2}^J d_{gj} \gamma_{gj} + \sum_{k=2}^K q_{gk} \theta_{gk} \right] + \sum_{l=1}^L z_{gl} \delta_{gl} + e_g, \quad (1)$$

where there are two sets of categorical variables ( $d$ 's and  $q$ 's) and  $L$  continuous variables ( $z$ 's); 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 subscript,  $g = A$  or  $B$ ; without loss of generality, the reference group is the first category for each set of dummy variables. We refer to equation (1) as the usual regression equation.

Fields and Wolff (1995) define the (average) gender wage gap in industry  $j$  as

$$I_{FW} = (\alpha_A - \alpha_B) + (\gamma_{Aj} - \gamma_{Bj}),$$

where  $A$  and  $B$  are female and male. The differences in the intercept terms are the wage gap for the reference group. If the differences in the intercept terms are dropped, then the wage gap in the reference industry is zero, since  $\gamma_{A1} = \gamma_{B1} = 0$ . This definition of the industry gender wage gap has the nice property that even if the reference group for the industry dummy variables change, the wage gap in industry  $j$  does not change; hence, the ranking of the wage gap of each industry is not changed.

Horrace and Oaxaca (2001) point out that industry gender wage gap of Fields and Wolff (1995) has a serious identification problem, that is, if the reference group of the other set of dummy variables ( $q$ 's) changes, then the wage gap in industry  $j$  is not invariant. This 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  $(\alpha_A - \alpha_B) + (\gamma_{Aj} - \gamma_{Bj}) + (\theta_{AK} - \theta_{BK})$ . This is related to a symptom of the well-known identification problem in detailed decomposition analysis, that the differences in the intercept terms are not invariant to the choice of reference group (Oaxaca and Ransom, 1999).

To obtain invariant estimates of gender wage gaps by industry, Horrace and Oaxaca (2001) suggest modifying the definition of the wage gap in industry  $j$  to

$$I_{HO} = (\alpha_A - \alpha_B) + (\gamma_{Aj} - \gamma_{Bj}) + \sum_{k=2}^K \bar{q}_{Ak}(\theta_{Ak} - \theta_{Bk}) + \sum_{l=1}^L \bar{X}_{Al}(\delta_{Al} - \delta_{Bl}),$$

where the mean values of  $\bar{q}_{Ak}$  and  $\bar{X}_{Al}$  are measured using the whole sample.<sup>2</sup>

---

<sup>2</sup> The first measure of gender wage gap in industry  $j$  by Horrace and Oaxaca (2001) is

$$I_{HO}^j = (\alpha_A - \alpha_B) + (\gamma_{Aj} - \gamma_{Bj}) + \sum_{k=2}^K \bar{q}_{Ak}^j(\theta_{Ak} - \theta_{Bk}) + \sum_{l=1}^L \bar{X}_{Al}^j(\delta_{Al} - \delta_{Bl}),$$

where the mean values of  $\bar{q}_{Ak}^j$  and  $\bar{X}_{Al}^j$  are measured using samples in industry  $j$ . Though this modification resolves the identification problem, it may cause the ranking of wage gaps by industry

This measure is obviously invariant.<sup>3</sup> A shortcoming of this modification is that it does not provide the wage gap due to the industrial wage premium only, that is, the wage gap when the returns to characteristics are identical except for the industrial wage premium by gender. Hence, Horrace and Oaxaca (2001) provide an invariant measure, but what they measure is not exactly what Fields and Wolff (1995) try to measure, the industrial gender wage gap due to the industrial wage premium only. We propose a measure of the industrial gender wage gap which resolves the identification problem while measuring what Fields and Wolff (1995) want to measure.

The proposed gender wage gap is closely related to solving the identification problem in the detailed Oaxaca decomposition for wage differentials. It has been recently recognized that a normalized equation can resolve the identification problem in the detailed Oaxaca decomposition of wage differentials. This paper shows that the normalized equation also provides a foundation for measuring an invariant industrial gender wage gap. We first present an overview of the identification problem in the detailed Oaxaca decomposition of wage differentials and how normalized regression can resolve this problem. Then, we utilize the coefficients of industry dummy variables in the normalized regression equation, that is, inter-industry wage differentials, to calculate the industrial gender wage gap.

---

to change. Therefore, they suggest further modification by replacing mean characteristics of observations in the industry  $j$  with overall mean characteristics, i.e., replacing  $\bar{q}_{Ak}^j$  and  $\bar{X}_{Al}^j$  with  $\bar{q}_{Ak}$  and  $\bar{X}_{Al}$ , respectively, which leads to  $I_{HO}$ .

<sup>3</sup> Suppose that the reference group for  $q$ 's is changed from category 1 to category  $K$ . Then the  $I_{HO}$  is changed as follows;

$$I'_{HO} = (\alpha_A - \alpha_B) + (\gamma_{Aj} - \gamma_{Bj}) + (\theta_{AK} - \theta_{BK}) + \sum_{k=1}^{K-1} \bar{q}_{Ak} ((\theta_{Ak} - \theta_{AK}) - (\theta_{Bk} - \theta_{BK})) + \sum_{l=1}^L \bar{X}_{Al} (\delta_{Al} - \delta_{Bl}).$$

It is easy to show that  $I_{HO} = I'_{HO}$  since  $\theta_{gK} - \sum_{k=1}^{K-1} \bar{q}_{gk} \theta_{gk} = (1 - \sum_{k=1}^{K-1} \bar{q}_{gk}) \theta_{gK} = \bar{q}_{gK} \theta_{gK}$ , and  $\theta_{g1} = 0$ , where  $g = A$  or  $B$ .

It is well-known that the detailed Oaxaca decomposition of wage differentials is not invariant to the choice of reference group when sets of dummy variables are used (Oaxaca and Ransom, 1999). That is, if we use dummy variable(s), then the sum of the detailed coefficients effects attributed to dummy variables is not invariant to the choice of the omitted group(s).<sup>4</sup> Since the “identification problem in the decomposition equation is a disguised identification problem of the constant and dummy variables in a regression equation” (Yun, 2005, p. 766), we need to construct a normalized equation which can identify the intercept and coefficients of all dummy variables including the reference category.

The proto-type normalized regression corresponding to the usual equation (1) looks like

$$y_g = \alpha_g^* + \left[ \sum_{j=1}^J d_{gj} \gamma_{gj}^* + \sum_{k=1}^K q_{gk} \theta_{gk}^* \right] + \sum_{l=1}^L z_{gl} \delta_{gl} + e_g.$$

Once consistent estimates of the usual equation are obtained, we follow the approach by Suits (1984)

when transforming these estimates by imposing restrictions  $\sum_{j=1}^J \gamma_{gj}^* = 0$  and  $\sum_{k=1}^K \theta_{gk}^* = 0$ .<sup>5</sup> When

the coefficients of normalized equation are further specified as  $\gamma_{gj}^* = \gamma_{gj} + m_{g\gamma}$  and  $\theta_{gk}^* = \theta_{gk} + m_{g\theta}$ ,

then the solution for constraints are  $m_{g\gamma} = \bar{\gamma}_g = -\sum_{j=1}^J \gamma_{gj}/J$  and  $m_{g\theta} = \bar{\theta}_g = -\sum_{k=1}^K \theta_{gk}/K$ , where

---

<sup>4</sup> Oaxaca decomposition explains wage differentials in terms of differences in individual characteristics (characteristics effect) and differences in the regression coefficients of wage equations (coefficients effect).

<sup>5</sup> Since the publication of Suits (1984), there have been several additional approaches for deriving the normalized equation, e.g., Greene and Seaks (1991), Gardeazaba and Ugidos (2004), and Yun (2005). See Yun (2005) for details of each method.

$\gamma_{g1} = \theta_{g1} = 0$ .<sup>6</sup> Then the normalized equation is:

$$y_g = (\alpha_g + \bar{\gamma}_g + \bar{\theta}_g) + \left[ \sum_{j=1}^J d_{gj}(\gamma_{gj} - \bar{\gamma}_g) + \sum_{k=1}^K q_{gk}(\theta_{gk} - \bar{\theta}_g) \right] + \sum_{l=1}^L z_{gl}\delta_{gl} + e_g. \quad (1')$$

Using the normalized equation (1'), we can easily calculate an industrial gender wage gap which can satisfy both Fields and Wolff's (1995) intention of measuring the gender wage gap when returns to other factors are identical between the two comparison groups, and Horrace and Oaxaca's (2001) concern on the identification problem. We propose a measure of the gender wage gap in industry  $j$ , which is a modified measure of Fields and Wolff using the coefficients of industry dummy variables in the normalized equation, i.e., inter-industry wage differentials, as follows;

$$I_{FW}^* = \gamma_{Aj}^* - \gamma_{Bj}^* = (\gamma_{Aj} - \bar{\gamma}_A) - (\gamma_{Bj} - \bar{\gamma}_B).^7$$

Note that all three measures of industrial wage gender gap,  $I_{FW}$ ,  $I_{HO}$  and  $I_{FW}^*$ , have the same ranking. Both  $I_{HO}$  and  $I_{FW}^*$  are invariant to the choice of reference group of  $q$ 's, but  $I_{HO}$  includes the effect of variables and their coefficients other than industry variable  $j$ . Therefore,  $I_{FW}^*$  is the measure which is not only free from the identification problem but also capturing what Fields and

---

<sup>6</sup> This paper assumes 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 Fields and Wolff (1995) do, it 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).

<sup>7</sup> Standard errors of  $I_{FW}$ ,  $I_{HO}$  and  $I_{FW}^*$  can be calculated using the covariance matrix from the usual regression. It is trivial to compute the standard error of  $I_{FW}^*$  once asymptotic covariance matrix of estimates of the normalized equation is computed. See Appendix for computation of the asymptotic covariance matrix of estimates of the normalized equation.

Wolff (1995) intend to measure, the impact of differences in wage premium in industry  $j$ . Using the normalized equation, we can resolve not only the identification problem of the detailed Oaxaca decompositions of wage differentials, but also the identification problem of measuring industrial wage gaps.

### **III. Conclusion**

We have shown that the identification problem raised by Horrace and Oaxaca (2001) for the measure of industrial gender wage gap by Fields and Wolff (1995) can be resolved by using a normalized regression equation. Though the remedy by Horrace and Oaxaca (2001) is invariant, it does not accomplish what Fields and Wolff (1995) set out to measure, differences in industrial wage premia (inter-industry wage differentials). The measure of industrial gender wage gaps suggested in this paper using the normalized regression equation provides an invariant measure while capturing what Fields and Wolff (1995) intended. This remedy becomes natural and obvious when we realize that inter-industry wage differentials are one of the building blocks of a normalized regression equation devised to resolve the identification problem.

## References

- Fields, Judith and Edward N. Wolff. 1995. "Interindustry Wage Differentials and the Gender Wage Gap." *Industrial and Labor Relations Review*, Vol. 49, No. 1 (October), 105-120.
- Gardeazabal, Javier and Arantza Ugidos. 2004. "More on Identification in Detailed Wage Decompositions." *Review of Economics and Statistics*, Vol. 86, No. 4 (November), 1034-1036.
- Greene, William H. and Terry G. Seaks. 1991. "The Restricted Least Squares Estimator: A Pedagogical Note." *Review of Economics and Statistics*, Vol. 73, No. 3 (August), 563-567.
- Haisken-DeNew, John P. and Christoph M. Schmidt. 1997. "Interindustry and Interregion Differentials: Mechanics and Interpretation." *Review of Economics and Statistics*, Vol. 79, No.3 (August), 516-521.
- Horrace, William C. and Ronald L. Oaxaca. 2001. "Inter-Industry Wage Differentials and the Gender Wage Gap: An Identification Problem." *Industrial and Labor Relations Review*, Vol.54, No.3 (April), 611-618.
- Horrace, William C. 2005. "On the Ranking Uncertainty of Labor Market Wage Gaps." *Journal of Population Economics*, Vol. 18, No.1 (March), 181-187.
- Krueger, Alan B. and Lawrence H. Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica*, Vol. 56, No. 2 (March), 259-293.
- Oaxaca, Ronald L. and Michael R. Ransom. 1999. "Identification in Detailed Wage Decompositions." *Review of Economics and Statistics*, Vol. 81, No. 1 (February), 154-157.
- Suits, Daniel B. 1984. "Dummy Variables: Mechanics v. Interpretation." *Review of Economics and Statistics*, Vol. 66, No.1 (February), 177-180.
- Yun, Myeong-Su. 2005. "A Simple Solution to the Identification Problem in Detailed Wage Decompositions." *Economic Inquiry*, Vol. 43, No. 4 (October), 766-772 with 2006, "Erratum." *Economic Inquiry*, Vol. 44, No. 1 (January), 198.

### Appendix: Calculation of Asymptotic Variance of Estimates of the Normalized Equation

A matrix representation of equation (1) is,  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$  while suppressing a group subscript.<sup>8</sup>  $\mathbf{Y}$  is the vector of the dependent variable ( $N \times 1$ );  $\mathbf{X} = (\mathbf{1} : \mathbf{D} : \mathbf{Q} : \mathbf{Z})$ , where  $\mathbf{D} = (\mathbf{d}_2 : \dots : \mathbf{d}_J)$ ,  $\mathbf{Q} = (\mathbf{q}_2 : \dots : \mathbf{q}_K)$ , and  $\mathbf{Z} = (\mathbf{z}_1 : \dots : \mathbf{z}_L)$ ;  $\mathbf{X}$ ,  $\mathbf{D}$ ,  $\mathbf{Q}$  and  $\mathbf{Z}$  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  $\mathbf{D}$  and  $\mathbf{Q}$  are matrices of two sets of dummy variables, and  $\mathbf{Z}$  is a matrix of continuous variables;  $\mathbf{1}$  is a vector of ones ( $N \times 1$ );  $\boldsymbol{\beta} = (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\delta})$  is a coefficient vector ( $T \times 1$ ), where  $\boldsymbol{\gamma} = (\gamma_2, \dots, \gamma_J)$ ,  $\boldsymbol{\theta} = (\theta_2, \dots, \theta_K)$ , and  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_L)$ .

In order to obtain the normalized equation (1') it is useful to rewrite the equation as  $\mathbf{Y} = \mathbf{X}^* \boldsymbol{\beta}_0 + \mathbf{e}$ , where  $\mathbf{X}^* = (\mathbf{1} : \mathbf{d}_1 : \mathbf{D} : \mathbf{q}_1 : \mathbf{Q} : \mathbf{Z})$  and  $\boldsymbol{\beta}_0 = (\boldsymbol{\alpha}, 0, \boldsymbol{\gamma}, 0, \boldsymbol{\theta}, \boldsymbol{\delta})$ . The normalized regression equation,  $\mathbf{Y} = \mathbf{X}^* \boldsymbol{\beta}^* + \mathbf{e}$ , is obtained by transforming  $\boldsymbol{\beta}_0$  to  $\boldsymbol{\beta}^*$  using a weight matrix,  $\mathbf{W}$ , that is,  $\boldsymbol{\beta}^* = \mathbf{W} \boldsymbol{\beta}_0$ , which yields ( $T^* \times 1$ ) vector of  $((\boldsymbol{\alpha} + \bar{\boldsymbol{\gamma}} + \bar{\boldsymbol{\theta}}), (\gamma_1 - \bar{\gamma}), (\boldsymbol{\gamma} - \bar{\boldsymbol{\gamma}}), (\boldsymbol{\theta}_1 - \bar{\boldsymbol{\theta}}), (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}), \boldsymbol{\delta})$ , where  $T^* = T+2 = J+K+L+1$ . The weight matrix  $\mathbf{W}$  is defined as

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

---

<sup>8</sup> The derivation of the normalized regression equation is developed by extending a method employed by Haisken-DeNew and Schmidt (1997).

where  $\mathbf{M}_{P \times P} = \mathbf{I}_{P \times P} - (1/P) \cdot \mathbf{1}_{P \times P}$ , and  $\mathbf{0}$ ,  $\mathbf{1}$ , and  $\mathbf{I}$  are a matrix of zeros, a matrix of ones and an identity matrix.<sup>9</sup>

The covariance matrix of estimates of the normalized regression equation ( $\boldsymbol{\beta}^*$ ) is computed as  $\Sigma_{\boldsymbol{\beta}^*} = \mathbf{W} \Sigma_{\boldsymbol{\beta}_0} \mathbf{W}'$  where the covariance matrix for  $\boldsymbol{\beta}_0$  ( $\Sigma_{\boldsymbol{\beta}_0}$ ) can be obtained by adding zero vectors to the covariance matrix of  $\boldsymbol{\beta}$  as

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

Once asymptotic covariance matrix for the estimates in the normalized equation is computed, then

it is a trivial task to calculate the asymptotic variance of  $I_{FW}^*$ , that is,  $\sigma_{I_{FW}^*} = \sqrt{\sigma_{\gamma_{A_j}^*}^2 + \sigma_{\gamma_{B_j}^*}^2}$ .

---

<sup>9</sup> A normalized regression with weighted average of coefficients of dummy variables can be easily obtained by changing the weight matrix,  $\mathbf{W}$ . Define  $\mathbf{W}_D = (\mathbf{W}_{d_1} \dots \mathbf{W}_{d_j})$  and  $\mathbf{W}_Q = (\mathbf{W}_{q_1} \dots \mathbf{W}_{q_K})$  be vectors of shares of dummy variables,  $\mathbf{D}$  and  $\mathbf{Q}$ . In order to find a weight matrix for obtaining a normalized regression with weighted averages, replace  $(1/J) \cdot \mathbf{1}_{1 \times J}$  and  $(1/K) \cdot \mathbf{1}_{1 \times K}$  with  $\mathbf{W}_D$  and  $\mathbf{W}_Q$ , and replace  $\mathbf{M}_{P \times P} = \mathbf{I}_{P \times P} - (1/P) \cdot \mathbf{1}_{P \times P}$  with  $\mathbf{M}_{J \times J} = \mathbf{I}_{J \times J} - \mathbf{W}_D^+$  and  $\mathbf{M}_{K \times K} = \mathbf{I}_{K \times K} - \mathbf{W}_Q^+$  when  $P = J$  and  $K$ , respectively.  $\mathbf{W}_D^+ = (\mathbf{W}_D, \dots, \mathbf{W}_D)$  and  $\mathbf{W}_Q^+ = (\mathbf{W}_Q, \dots, \mathbf{W}_Q)$ , which are  $J \times J$  and  $K \times K$  matrices, respectively.