

IZA DP No. 9909

An Oaxaca Decomposition for Nonlinear Models

Stephen Bazen
Xavier Joutard
Brice Magdalou

April 2016

An Oaxaca Decomposition for Nonlinear Models

Stephen Bazen

*Aix-Marseille University,
CNRS, EHESS and IZA*

Xavier Joutard

LEST, Aix-Marseille University

Brice Magdalou

LAMETA, University of Montpellier I

Discussion Paper No. 9909

April 2016

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 Oaxaca Decomposition for Nonlinear Models*

The widely used Oaxaca decomposition applies to linear models. Extending it to commonly used nonlinear models such as duration models is not straightforward. This paper shows that the original decomposition that uses a linear model can also be obtained by an application of the mean value theorem. By extension, this basis provides a means of obtaining a decomposition formula which applies to nonlinear models which are continuous functions. The detailed decomposition of the explained component is expressed in terms of what are usually referred to as marginal effects. Explicit formulae are provided for the decomposition of some nonlinear models commonly used in applied econometrics including binary choice, duration and Box-Cox models.

JEL Classification: C10, C18, C21

Keywords: Oaxaca decomposition, nonlinear models, duration models, binary choice, Box-Cox transformation

Corresponding author:

Stephen Bazen
Aix-Marseille School of Economics
Aix-Marseille University
2 Rue de la Charité
13002 Marseille
France
E-mail: stephen.bazen@univ-amu.fr

* We are grateful to Habiba Djebbari and seminar participants at GREQAM, Marseille, IZA and the University of Dijon for comments on an earlier version of this paper.

Introduction

Much applied work in economics is devoted to analyzing the sources of differences between individuals and groups. The Oaxaca decomposition (Oaxaca, 1973) is a method of expressing the difference between the mean values of a variable – usually the logarithm of earnings – for two groups based on the coefficients obtained from two group-specific linear regressions¹. The difference is expressed in terms of two components that contribute to the divergence in group means: an explained part or ‘composition effect’ due to differences in the mean characteristics of the two groups, and an unexplained component or ‘structure effect’ due to differences in the estimated coefficients in the group equations. A very similar decomposition was proposed by Blinder (1973), in the same year but after the publication of Oaxaca’s article². The technique was originally developed in order to establish the existence and extent of wage and other forms of discrimination and is widely used in labour economics and to some extent other areas. It can also be applied to analyze group differences, in general. Surveys of this and other decomposition methods are provided by Beblo, Beninger, Heinze and Laisney (2003) and Fortin, Lemieux and Firpo (2011).

Attempts have been made to use the Oaxaca approach to decompose group differences using specific nonlinear models, such as the logit and probit models (Nielsen, 1998; Yun, 2000; Fairlie, 2005; Powers and Pullum, 2006), hazard or duration models (Wagstaff and Nguyen, 2001; Powers and Yun, 2009) and Tobit-type models (Neumann and Oaxaca; 2004, Yun, 2007; Wolff, 2012). More recently, Bauer and Sinning (2008) have proposed a generalization of the Oaxaca approach based on the sample means of estimated functions for nonlinear specifications. This method will be shown to be problematic for the identification of certain components of interest defined in the Oaxaca-linear approach. In particular, in the original version, the existence of discrimination is based on assuming that two groups have the same mean characteristics. The approaches mentioned above are not formulated on this kind of counterfactual basis.

This paper proposes an Oaxaca-type decomposition for any continuous nonlinear model. It uses as a basis the difference between two fitted values, which is decomposed into a composition and a structure effect. It is obtained through an application of the mean value theorem and the resulting decomposition is exact in the sense that there is no remainder even though the model is nonlinear. The paper begins in section 1 with a brief examination of the basis of the Oaxaca decomposition and then explores some of the difficulties encountered when seeking to generalize this approach to nonlinear relations in section 2. In the following section, it is shown that the Oaxaca decomposition can be obtained by the application of the

¹ It is possible to obtain the same estimates in a pooled regression with group specific coefficients and dummy variables.

² In private correspondence with these authors, it emerges that the two papers were prepared independently but the authors had met and discussed their research beforehand.

mean value theorem to the estimated relation for one of the groups being compared. The application of theorem is then used as means of obtaining a decomposition technique, which can be used with any continuous nonlinear function. In section 4, explicit forms for the decomposition of some widely used nonlinear models for binary choice and duration analysis along with a model using the Box-Cox transformation. Empirical examples of each of these are presented.

I Interpretations of the Oaxaca decomposition

The original Oaxaca decomposition has a certain number of features which are inextricably linked to the linear regression model, and which limit the extent to which the method can be directly generalized. It applies to an explicitly linear framework in which the dependent variable for member i of group g is y_{gi} (often this is the logarithm of earnings), the explanatory variables are represented in vector form, x_{gi} , (which contains k elements and x_{gi}^T is its transpose) and the error term is ε_{gi} . The decomposition applies to two groups $g = M, F$. The group-specific parameters are β_g and the linear relationship used is

$$y_{gi} = x_{gi}^T \beta_g + \varepsilon_{gi} \quad g = M, F \quad (1)$$

The Oaxaca decomposition is obtained by first estimating the parameters using ordinary least squares (OLS) to obtain $\hat{\beta}_g$ for each group, and then by defining a counterfactual fitted value of the dependent variable as $\bar{x}_F^T \hat{\beta}_M$ (or $\bar{x}_M^T \hat{\beta}_F$) where \bar{x}_M and \bar{x}_F , are vectors of the respective means of the right hand side variables for the two groups. Defining the difference

$$\Delta = \bar{x}_M^T \hat{\beta}_M - \bar{x}_F^T \hat{\beta}_F$$

and adding and subtracting this counterfactual term, results in the following additive decomposition :

$$\Delta = \bar{x}_F^T (\hat{\beta}_M - \hat{\beta}_F) + \hat{\beta}_M^T (\bar{x}_M - \bar{x}_F) \quad (2)$$

The first term on the right hand side is the unexplained component or structure effect – that is, what the person with mean characteristics in group F would have obtained if they were a member of group M relative to what they actually have. The second term is the explained component or composition effect – the difference due to differences in mean characteristics.

There is discrimination when $\bar{x}_M = \bar{x}_F$, (the situation in which the two groups on average have the same characteristics) and the structure effect is non-zero. This is the original form of the decomposition presented by Oaxaca (1973, p. 697, equation 13). It has the following properties :

(i) The decomposition is *model-based*. A model is specified to determine the value of y that one group would have if it had the same mean value of x of the other group, $\bar{x}_g^T \hat{\beta}_h$ ($g \neq h$). In other words, a model is used to construct a counterfactual situation. Furthermore, the decomposition is only meaningful if the “factual” mean, \bar{y}_g , is equal to the model’s prediction $\bar{x}_g^T \hat{\beta}_g$. The treatment of the factual and counterfactual means is not symmetric.

(ii) The original focus was on the decomposition of differences in sample means, $\bar{y}_M - \bar{y}_F$, using estimated coefficients from a linear model. However, when the parameters of the model are estimated by OLS, the Oaxaca decomposition is exact only if the model contains a constant, i.e. if it is an affine function :

$$y_{gi} = \beta_{0g} + x_{gi}^T \beta_g + \varepsilon_{gi} \quad (3)$$

The presence of a constant ensures that the sum and therefore the mean of the estimated OLS residuals, $\hat{\varepsilon}_{gi}$, are both equal to zero – or equivalently that the mean of the fitted values is equal to the sample mean of the dependent variable :

$$\bar{y}_g = \bar{x}_g^T \hat{\beta}_g + \bar{\hat{\varepsilon}}_g = \bar{x}_g^T \hat{\beta}_g = \bar{\hat{y}}_g$$

It is this equality that permits the decomposition of the difference in means, $\bar{y}_M - \bar{y}_F$, into the characteristics and structure components. This *numerical* property is a consequence of OLS estimation. Oaxaca (1973) assimilates the constant term into the coefficient vector. Blinder (1973, p. 439), for reasons of interpretation separates the constant term from the other components and presents the decomposition as :

$$\begin{aligned} \bar{y}_M - \bar{y}_F &= \hat{\beta}_{0M} + \bar{x}_M^T \hat{\beta}_M - (\hat{\beta}_{0F} + \bar{x}_F^T \hat{\beta}_F) \\ &= \hat{\beta}_{0M} - \hat{\beta}_{0F} + \bar{x}_F^T (\hat{\beta}_M - \hat{\beta}_F) + \hat{\beta}_M^T (\bar{x}_M - \bar{x}_F) \quad (4) \end{aligned}$$

(iii) Although it was not presented in this form originally, it is common nowadays to express the decomposition in terms of the expectations of variables for the population relationships (for example, Fortin et al, 2011, and Rothe, 2012). The decomposition is based on the parameters of a linear specification (1). The Oaxaca decomposition *at the population level* is :

$$\begin{aligned} E(y_{Mi}) - E(y_{Fi}) &= E(x_{Mi}^T)\beta_M - E(x_{Fi}^T)\beta_M + E(x_{Fi}^T)\beta_M - E(x_{Fi}^T)\beta_F \\ &= E(x_{Fi}^T)[\beta_M - \beta_F] + [E(x_{Mi}^T) - E(x_{Fi}^T)]\beta_M \end{aligned} \quad (5)$$

since, by assumption, $E(\varepsilon_{Mi}) = E(\varepsilon_{Fi}) = 0$. In other words, the relation need not contain a constant in order to obtain an exact two component decomposition of the difference in group population means. Note that this form of the decomposition is in terms of population parameters, rather than OLS estimates.

Properties (ii) and (iii) differ since the sample mean of the estimated residual, $\hat{\varepsilon}_{gi}$, in the linear model without a constant (1) will not be equal to zero.

(iv) The Oaxaca decomposition is subject to an index number problem. If the difference is calculated around $\bar{x}_M^T \hat{\beta}_F$, the structure effect is $\bar{x}_M^T (\hat{\beta}_M - \hat{\beta}_F)$, rather than $\bar{x}_F^T (\hat{\beta}_M - \hat{\beta}_F)$ as in equation (2). The choice of reference group characteristics for the decomposition affects the size of the each of components. In general, there is no unique, unambiguous measure of the extent of discrimination in terms of the structure effect.

II Extending the Oaxaca method to nonlinear relations

Extending the Oaxaca (linear) approach to nonlinear relations is not straightforward. First, OLS cannot generally be applied due to the presence of nonlinearities in the relation. The decomposition will not have the original Oaxaca form. Furthermore, as has been pointed out above, the decomposition has certain properties that are related explicitly to the numerical properties of least squares estimation and these will no longer apply. Second, and more importantly, when applied to nonlinear models, an Oaxaca-type decomposition of differences in either sample means or expectations of the left hand side variable will not be exact, and so neither of properties (i) and (ii) carries over to nonlinear functions. This is due to Jensen's inequality, a consequence of which is that, in general, for a nonlinear function $v(x)$:

$$\frac{1}{n_g} \sum_{i=1}^{n_g} v(x_{gi}) \neq v\left(\frac{1}{n_g} \sum_{i=1}^{n_g} x_{gi}\right)$$

Even an exact Oaxaca-type decomposition at the population level in terms of expectations, as in (5), will not be obtained in general³. Due to the (near⁴) impossibility of obtaining an exact decomposition of the group difference in sample means for nonlinear models in terms of the group means of the explanatory variables, the *basis* for a decomposition using a nonlinear model needs to be rigorously specified.

Call the *estimated* functions or fitted values for each group $\tilde{y}_{Mi} = G_M(x_{Mi})$ and $\tilde{y}_{Fi} = G_F(x_{Fi})$, respectively. These functions would normally be the estimated conditional expectations in econometric applications. The original Oaxaca decomposition of difference in the sample means of the left hand side variable, $\bar{y}_M - \bar{y}_F$, is possible because when the functions are affine and the parameters estimated by OLS, and the following equality is obtained :

$$\bar{y}_g = \bar{\tilde{y}}_g = G_g(\bar{x}_g) \quad \text{where} \quad \bar{\tilde{y}}_g = \frac{1}{n_g} \sum_{i=1}^{n_g} \tilde{y}_{gi} \quad g = F, M$$

This implies that the group difference in any of these means can be used as basis for a decomposition in the affine case. When the function is nonlinear these three quantities are not identical. Thus when extending the Oaxaca approach to nonlinear relations, the possible candidates as a *basis* are the decomposition of the difference in :

(a) the sample means of the left hand side variable, $\bar{y}_M - \bar{y}_F$;

(b) the sample means of the fitted values of estimated functional relationship,

$$\bar{\tilde{y}}_M - \bar{\tilde{y}}_F = \frac{1}{n_M} \sum_{i=1}^{n_M} \tilde{y}_{Mi} - \frac{1}{n_F} \sum_{i=1}^{n_F} \tilde{y}_{Fi} = \overline{G_M(x_{Mi})} - \overline{G_F(x_{Fi})};$$

(c) the values of the group estimated functions (or fitted values) evaluated at the means of the right hand side variables for that group, $\tilde{y}^M - \tilde{y}^F = G_M(\bar{x}_M) - G_F(\bar{x}_F)$.

These different bases will not be equal and therefore a choice has to be made. In view of Jensen's inequality, basis (a) is unlikely to prove fruitful for a generalization. Even in the linear regression case, (a) is appropriate only when the relation contains a constant. The earlier approaches of Nielsen (1998) and Yun (2004) and more recently Bauer and Sinning (2008) and Schwiebert (2015), propose using basis (b). This produces a decomposition of the differences in the sample means of the fitted values (or equivalently the sample means of the estimated function) :

³ The equality $E[v(x)] = v(E[x])$ only holds with certainty for affine functions.

⁴ An equality could occur in certain situations since the function here is nonlinear but not necessarily monotonic.

$$\bar{y}_M - \bar{y}_F = \overline{G_M(x_{Mi})} - \overline{G_M(x_{Fi})} + \overline{G_M(x_{Fi})} - \overline{G_F(x_{Fi})} \quad (6)$$

where $\overline{G_g(x_{gi})} = \frac{1}{n_g} \sum_{i=1}^{n_g} G_g(x_{gi})$. The logic of this choice is clear in that in population terms, this corresponds to a decomposition of the following difference $E[G_M(x_{Mi})] - E[G_F(x_{Fi})]$, and by the law of iterated expectations, the expectation of this difference will be equal to the difference in the unconditional population means: $E(y_{Mi}) - E(y_{Fi})$.

However, there are at least two reasons why (6) may be unsatisfactory as a generalization of the Oaxaca method. Firstly, if the functions $G_M(x_{Mi})$ and $G_F(x_{Fi})$ are not affine, then in general from Jensen's inequality:

$$G_M(\bar{x}_M) - G_F(\bar{x}_F) \neq \overline{G_M(x_{Mi})} - \overline{G_F(x_{Fi})} \quad (7)$$

Using (b) as a basis therefore entails disconnecting the decomposition from the mean vectors \bar{x}_M and \bar{x}_F . It does not involve the use of a counterfactual defined in terms of mean characteristics. In other words, when the two groups have identical means, the explained component or composition effect is not equal to zero (as it is in the Oaxaca decomposition) and the decomposition does not reduce to the structure effect. This is serious weakness as the identification of the latter is one of the main reasons for undertaking a decomposition of this kind: this is precisely the component that is associated with discrimination. Secondly, if the means of the *estimated* functions or fitted values are used, there is no guarantee that $\bar{y}_M - \bar{y}_F = \overline{G_M(x_{Mi})} - \overline{G_F(x_{Fi})}$: this equality is not valid in the case of the probit model for example.

A decomposition using basis (b) is therefore not generally expressed in terms of the means of the variables y and x and can diverge from the Oaxaca approach on both sides of the equation. Using this basis with nonlinear functions will generally involve an approximation (i.e. there will be a remainder). Some approaches, for example, are explicitly based on assuming that (7) is close to being an equality (see, Powers and Pullum, (2006), and Powers and Yun (2009)).

An alternative approach can be derived from the original "Oaxaca-linear" method which is applicable to both linear and nonlinear functions using basis (c), $G_M(\bar{x}_M) - G_F(\bar{x}_F)$. In the next section we show that using this same basis, applying the mean value theorem⁵ to one of

⁵ An earlier version of this paper was entitled 'The MV decomposition'. Since writing that version, an article by Schwiebert (2015) appeared which uses the mean value theorem to obtain a decomposition

these functions ($G_M(\bar{x}_M)$ or $G_F(\bar{x}_F)$) and then subtracting the other function, gives precisely the Oaxaca decomposition presented in equation (2) when the functions are linear. By extension, the same operation enables a decomposition method to be derived which is applicable to any smooth parametric function. When the latter is defined on a single linear index, $x_i^T \beta$, a very straightforward decomposition formula is obtained. The form of the decomposition for some commonly used econometric models is then presented in the subsequent section.

III An Oaxaca decomposition for nonlinear models

Applying the *mean value theorem* to one of the functions over the intervals between \bar{x}_{Fj} and \bar{x}_{Mj} for each of the variables $j=1, 2, \dots, k$ in the vector x , produces an equation that resembles the Oaxaca decomposition. The mean value theorem applied to the function $G_M(\bar{x}_M)$ over the intervals $[\bar{x}_{Mj}, \bar{x}_{Fj}]$ states that there exists a vector $\tilde{x}^T = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k)$ where \tilde{x}_j lies in the interval between \bar{x}_{Mj} and \bar{x}_{Fj} , for $j=1, 2, \dots, k$ such that :

$$G_M(\bar{x}_M) = G_M(\bar{x}_F) + \nabla G_M(\tilde{x})^T [\bar{x}_M - \bar{x}_F]$$

where $\nabla G_M(\tilde{x})$ is the vector of partial derivatives evaluated at the vector \tilde{x} . Subtracting $G_F(\bar{x}_F)$ from both sides gives a decomposition formula for the difference $G_M(\bar{x}_M) - G_F(\bar{x}_F)$:

$$\tilde{y}^M - \tilde{y}^F \equiv G_M(\bar{x}_M) - G_F(\bar{x}_F) = [G_M(\bar{x}_F) - G_F(\bar{x}_F)] + \nabla G_M(\tilde{x})^T [\bar{x}_M - \bar{x}_F] \quad (8)$$

The first component on the right hand side is immediately recognizable as the structure effect – for identical mean characteristics, what is the model’s prediction of the difference in y between the two groups? Given that Oaxaca decompositions are identities, the remaining term must be the composition effect, $G_M(\bar{x}_M) - G_M(\bar{x}_F)$. This term will be zero if the groups have identical means, which is the defining property of the measurement of discrimination in the Oaxaca decomposition.

In the case of one explanatory variable, this form of the decomposition can be presented graphically as in Figure 1. The segment BE is parallel to the tangent representing $G_M'(\tilde{x})$. The graphical representation of the decomposition is identical in form to that presented by

and uses that name. The decomposition here is quite different from that presented in Schwiebert (2015) which uses a different basis and applies only to nonlinear models defined on a linear index.

Oaxaca method when applied to linear models, as shown in Figure 2 except for the difference that corresponding function values in the vertical axis are not the sample means.

The decomposition obtained by applying the mean value theorem permits an *exact* detailed decomposition of the composition effect once the vector \tilde{x} has been determined. In the scalar case there is no difficulty in determining \tilde{x} . However, when the functions are defined on a vector, this is a *theoretical* possibility but it is nearly impossible to implement in practice, since \tilde{x} cannot be determined on an a priori basis. This would require the determination of every element in the vector $\tilde{\alpha}^T = (\alpha_1 \ \alpha_2 \dots \alpha_k)$, which are such that :

$$\tilde{x}_j = \alpha_j \bar{x}_{Mj} + (1 - \alpha_j) \bar{x}_{Fj} \quad \text{where } 0 \leq \alpha_j \leq 1 \text{ for } j = 1, 2, \dots, k$$

An alternative approach is to calculate the vector of derivatives of the function, $\nabla G_M(\tilde{x})$, directly rather than determine the vector \tilde{x} itself. This is shown for the general case in the following proposition.

Proposition 1. Let $x(\lambda) = \lambda \bar{x}_M + (1 - \lambda) \bar{x}_F$ where $\lambda \in [0, 1]$. The composition effect of the decomposition is equal to :

$$\nabla G_M(\tilde{x})^T (\bar{x}_M - \bar{x}_F) = \sum_{j=1}^k \left[(\bar{x}_{Mj} - \bar{x}_{Fj}) \int_0^1 G_{Mj}'(x(\lambda)) d\lambda \right] \quad (9)$$

where $G_{Mj}'(x(\lambda)) = \frac{\partial G_M(x(\lambda))}{\partial x_j(\lambda)}$.

Proof : Define the function: $h(\lambda) = G_M[x(\lambda)]$. Since $h(1) = G_M(\bar{x}_M)$ and $h(0) = G_M(\bar{x}_F)$, the composition effect can be therefore be written as:

$$G_M(\bar{x}_M) - G_M(\bar{x}_F) = h(1) - h(0) = \int_0^1 h'(\lambda) d\lambda$$

The derivative of the function $h(\lambda)$ is the scalar product :

$$h'(\lambda) = \sum_{j=1}^k [(\bar{x}_{Mj} - \bar{x}_{Fj}) G_{Mj}'(x(\lambda))]$$

Integrating this derivative across the range of $\lambda \in [0, 1]$ and using equation 8 gives the result. ■

In this formulation, the individual contribution of each variable (x_j) to the overall composition effect is weighted by an average value of the marginal effect over the interval between the group means of the variable in question, $\int_0^1 G_{Mj}'(x(\lambda))d\lambda$. Cameron and Trivedi (2003) propose this formula as one of the methods of calculating marginal effects (p. 122).

In the linear model, the marginal effect is constant and the Oaxaca approach is a special case of the decomposition presented here.

Proposition 2: The “Oaxaca-linear” decomposition (2) can be obtained by applying the mean value theorem to the function $G_M(\bar{x}_M) = \bar{x}_M^T \hat{\beta}_M$ over the intervals $[\bar{x}_{Mj}, \bar{x}_{Fj}]$, for $j = 1, 2, \dots, k$, and subtracting $G_F(\bar{x}_F) = \bar{x}_F^T \hat{\beta}_F$ (where $\hat{\beta}_g$ are vectors of OLS estimates in the equations $y_{gi} = g(x_{gi}) = x_{gi}^T \beta_g + \varepsilon_{gi}$ for $g = M, F$).

Proof: If the model is linear, $G_{Mj}'(x(\lambda)) = \hat{\beta}_{Mj}$ and independent of x_j and λ , so that

$$\int_0^1 G_{Mj}'(x(\lambda))d\lambda = \hat{\beta}_{Mj} \cdot \blacksquare$$

The method can be simplified when a nonlinear function is defined on a linear index, so that $G_g(x_g) = G_g(x_g^T \beta_g)$. In this case the marginal effects are proportional to the parameters, $\nabla G_g(\tilde{x}_g^T \hat{\beta}_g) = k_{g0} \hat{\beta}_g$. In this special case, the decomposition has the following attractive form:

$$\tilde{y}^M - \tilde{y}^F \equiv G_M(\bar{x}_M) - G_F(\bar{x}_F) = [G_M(\bar{x}_F) - G_F(\bar{x}_F)] + k_{M0} \hat{\beta}_M^T [\bar{x}_M - \bar{x}_F] \quad (10)$$

where $k_{M0} = k(\tilde{x}^T \hat{\beta}_M)$ is a fixed scalar. The term $k_{M0} \hat{\beta}_M$ is the vector of marginal effects evaluated at \tilde{x} . Since many nonlinear econometric models are defined in this way (see the examples presented below), the detailed first order decomposition of the composition effect has a very straightforward interpretation since the marginal effects are proportional to the coefficients. Furthermore the value of the scalar k_{M0} must be equal to :

$$k_{M0} = k(\tilde{x}^T \hat{\beta}_M) = \frac{G_M(\bar{x}_M^T \hat{\beta}_M) - G_F(\bar{x}_F^T \hat{\beta}_M)}{\hat{\beta}_M^T [\bar{x}_M - \bar{x}_F]}$$

Given this tautology, there is no need to determine the elements of the vector \tilde{x} that characterize the vector of marginal effects $\nabla G_M(\tilde{x})$. Written in this way, the detailed decomposition of the composition effect is :

$$\nabla G_M(\tilde{x})^T [\bar{x}_M - \bar{x}_F] = k_{M0} \tilde{\beta}_{M1} [\bar{x}_{1M} - \bar{x}_{1F}] + k_{M0} \tilde{\beta}_{M2} [\bar{x}_{2M} - \bar{x}_{2F}] + \dots + k_{M0} \tilde{\beta}_{Mk} [\bar{x}_{kM} - \bar{x}_{kF}]$$

The weights in the detailed decomposition in fact resemble those in Yun (2004), but are applied to a different basis⁶. This common feature is a consequence of the function being defined on a single linear index.

IV Examples of decompositions for nonlinear models

The proposed decomposition has the advantage of having a coherent basis – it compares a model-based estimate of an actual situation with a model-based estimate of a counterfactual one, where both are specified in terms of a parametrically defined function and the vectors of group means (\bar{x}_g). It means that any parametric function can be decomposed into a structure effect and a composition effect, where the latter is zero when $\bar{x}_M - \bar{x}_F = 0$. This contrasts with Yun (2004), Fairlie (2005) and Bauer and Sinning (2008) who use sums of fitted values divided by the sample size (i.e. the other side of the Jensen inequality).

Decomposing augmented linear models such as the sample selection model has already been addressed by Neumann and Oaxaca (2001) based on mean characteristics along the lines proposed here, although they stress the importance of how one interprets the selectivity term⁷. Other functions of interest in applied work are probability models (in which the population rate is decomposed) and hazard models (which involves either the hazard itself or the average duration of a spell). In this section, we derive explicit formulae the decomposition for these types of model using equation (10). Hereafter, any parameter covered by a hat (for example, $\hat{\theta}$) is assumed to be an appropriate estimate of that parameter.

⁶ In Yun (2004) the formula for the composition effect is $\frac{G_M(x_{Mi}^T \beta_M) - G_F(x_{Mi}^T \beta_M)}{\beta_M^T [\bar{x}_M - \bar{x}_F]} \times \beta_M^T [\bar{x}_M - \bar{x}_F]$.

⁷ Yun (2007) and Wolff (2009) provide alternative decomposition procedures for the sample selection model.

(a) Logit and Probit models

Logit and probit models have the same generic form for each of the groups :

$$E(y_{gi} | x_{gi}) = \text{Prob}(y_{gi} = 1 | x_{gi}) = D(\bar{x}_{gi}^T \beta_g), \quad g = M, F$$

where D is a cumulative distribution function with a common form for both groups. Since the latter is defined on a linear index, the vector of first derivatives has the following, straightforward form :

$$\nabla D(\bar{x}_g^T \beta_g) = d(\bar{x}_g^T \beta_g) \beta_g$$

where d is the associated density function (note that d is a scalar and β a vector of parameters). Using parameters estimated by maximum likelihood ($\hat{\beta}$), the implied decomposition for the probit model⁸ is :

$$\begin{aligned} \tilde{y}^M - \tilde{y}^F &= \Phi(\bar{x}_M^T \hat{\beta}_M) - \Phi(\bar{x}_F^T \hat{\beta}_F) \\ &= \Phi(\bar{x}_F^T \hat{\beta}_M) - \Phi(\bar{x}_F^T \hat{\beta}_F) + k_p \hat{\beta}_M^T [\bar{x}_M - \bar{x}_F] \end{aligned}$$

where $k_p = \phi(\bar{x}_F^T \hat{\beta}_M)$ and Φ and ϕ are the standard normal cumulative distribution and density functions, respectively. It is important to note that k_p is fixed scalar in this decomposition, in the sense that each element in the vector $\tilde{\beta}_M$ is multiplied by the same constant. The value of k_p is trivially given by:

$$k_p = \frac{\Phi(\bar{x}_M^T \hat{\beta}_M) - \Phi(\bar{x}_F^T \hat{\beta}_M)}{\hat{\beta}_M^T [\bar{x}_M - \bar{x}_F]}$$

Unlike the probit model, the function to be decomposed in the logit model has a closed form:

⁸ The variance is normalised equal to one.

$$L(\bar{x}^T \beta) = \frac{\exp(\bar{x}^T \beta)}{1 + \exp(\bar{x}^T \beta)}$$

For maximum likelihood estimates of β_g , the decomposition formula can be written in the same form as for the Probit model :

$$\begin{aligned} \tilde{y}^M - \tilde{y}^F &= L(\bar{x}_M^T \hat{\beta}_M) - L(\bar{x}_F^T \hat{\beta}_F) \\ &= L(\bar{x}_F^T \hat{\beta}_M) - L(\bar{x}_F^T \hat{\beta}_F) + k_L \hat{\beta}_M^T [\bar{x}_M - \bar{x}_F] \end{aligned}$$

where $k_L \hat{\beta}_M$ is the vector of the first derivatives with respect to the vector x evaluated at \tilde{x} . This vector of marginal effects has the following special form:

$$k_L \hat{\beta}_M = L'(\tilde{x}^T \hat{\beta}_M) = L(\tilde{x}^T \hat{\beta}_M) [1 - L(\tilde{x}^T \hat{\beta}_M)] \times \hat{\beta}_M$$

where, tautologically,

$$k_L = \frac{L(\bar{x}_M^T \hat{\beta}_M) - L(\bar{x}_F^T \hat{\beta}_M)}{\hat{\beta}_M^T [\bar{x}_M - \bar{x}_F]} \quad \text{and } 0 < k_L \leq 0.25.$$

Various authors have attempted to decompose the difference in sample means using logit and probit models (Nielsen (1998), Yun (2000, 2004) and Fairlie (2005)). In fact for a logit model *containing a constant term*, when the parameters are estimated by maximum likelihood, the sample mean is related to the estimated function in the following way⁹ :

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n L(x_i^T \hat{\beta})$$

This mean property has been used to obtain a decomposition for the logit model given by :

⁹ This is a consequence of the first order conditions for obtaining a maximum likelihood estimate of the constant term.

$$\bar{y}_M - \bar{y}_F = \frac{1}{n_M} \sum_{i=1}^{n_M} L(x_{Mi} \hat{\beta}_M) - \frac{1}{n_F} \sum_{i=1}^{n_F} L(x_{Fi} \hat{\beta}_M) + \frac{1}{n_F} \sum_{i=1}^{n_F} L(x_{Fi} \hat{\beta}_M) - \frac{1}{n_F} \sum_{i=1}^{n_F} L(x_{Fi} \hat{\beta}_F)$$

Note that this decomposition contains the sample means of the dependent variable but not the means of the right hand side variables, \bar{x}_g . This is a consequence of Jensen's inequality.

This form of decomposition has been used to obtain a detailed decomposition of both the unexplained and explained components. However, because of different sample sizes, simulation methods have to be used to provide extra data when undertaking detailed decompositions (see Fairlie, 2005). Such an approach cannot be applied in an exact manner to the probit model since $\bar{y} \neq \frac{1}{n} \sum_{i=1}^n \Phi(x_i^T \hat{\beta})$.

(b) Hazard functions and duration models

One of the key differences with duration models is that in most data sets, durations are censored at the time of the survey. This is the case for example with unemployment durations in the Labour Force Surveys used to estimate the unemployment rate according to the ILO definition. In order to analyse differences in unemployment duration or hazard rates between groups, using the difference in sample means as the basis for an Oaxaca-type decomposition is not appropriate because censoring. Most econometric analyses take account of censoring in the estimation of models, but there is an issue of which quantity is to be decomposed. It is in this context that the approach proposed here is particularly relevant. By using the fitted value corresponding mean characteristics as the basis, the decomposition can be straightforwardly obtained.

Using the same notation as above, where D is the cumulative distribution function and d the associated density function for durations or spell lengths, t , the hazard rate is defined as:

$$\lambda(t; x) = \frac{d(t; x)}{1 - D(t; x)} = \frac{d(t; x)}{S(t; x)}$$

where S is the survivor function. The difference between this and earlier models is the dependence of the hazard on time as well as on characteristics. The decomposition technique can be straightforwardly applied for differences in the expected duration¹⁰:

¹⁰ Note that the effect of a variable on the duration of a completed spell is of opposite sign to its effect on the hazard rate.

$$E(s | \bar{x}_g) = G_g(\bar{x}_g)$$

where the survivor function, and thus the hazard function, is linked to the average completed spell duration through the following equality :

$$E(s | \bar{x}) = \int_0^{\infty} t d(t | \bar{x}) dt = \int_0^{\infty} S(t | \bar{x}) dt$$

In what follows, the link between a parametric hazard specification and the corresponding formula for the expected duration is used to obtain decompositions for two popular hazard specifications – the Weibull and loglogistic.

One of the more widely used parametric specifications of the function is the Weibull hazard given by :

$$\lambda(t; x_i) = \alpha t^{\alpha-1} \exp(x_i^T \beta) \quad \alpha > 0 \quad (13)$$

In this case, the expected duration of a completed spell (s) for an individual with the mean value of x , is given by a nonlinear function defined on a linear index :

$$\begin{aligned} E(s | \bar{x}) &= \Gamma\left(\frac{1+\alpha}{\alpha}\right) \exp\left(-\bar{x}^T \frac{\beta}{\alpha}\right) \\ &\equiv \bar{\Gamma}(\alpha) \exp(\bar{x}^T \beta^*) \quad \text{where } \beta^* = -\frac{\beta}{\alpha} \end{aligned} \quad (14)$$

The first term on the right hand side is the gamma function and is independent of both x and s . As noted in the definition of the decomposition, the basis is the difference between two estimated functions evaluated at the means of the explanatory variables, x , which is denoted as $\tilde{y}_M - \tilde{y}_F$. In the current case the decomposition is :

$$\tilde{y}_M - \tilde{y}_F = \bar{\Gamma}(\hat{\alpha}_M) \exp(\bar{x}_F^T \hat{\beta}_M^*) - \bar{\Gamma}(\hat{\alpha}_F) \exp(\bar{x}_F^T \hat{\beta}_F^*) + k_w \hat{\beta}_M^{*T} [\bar{x}_M - \bar{x}_F]$$

where the fixed scalar k_w is given by $k_w = \Gamma\left(\frac{1+\hat{\alpha}_M}{\hat{\alpha}_M}\right) \exp(\bar{x}_F^T \hat{\beta}_M^*)$. Tautologically, this is equal to :

$$k_w = \frac{\bar{\Gamma}(\hat{\alpha}_M) \left(\exp(\bar{x}_M^T \hat{\beta}_M^*) - \exp(\bar{x}_F^T \hat{\beta}_M^*) \right)}{\hat{\beta}_M^{*T} [\bar{x}_M - \bar{x}_F]}$$

This specification contains the exponential specification as a special case when $\alpha = 1$, since $\Gamma(2) = 1$.

The Weibull specification applies only to cases where the hazard rate is monotonic – it is either increasing, $\alpha > 1$, or decreasing, $0 < \alpha < 1$, but cannot be one then the other during time spent in a given state. A hazard specification that permits a non-monotonic form is the log-logistic specification:

$$\lambda(t) = \frac{\gamma \exp(x^T \beta) t^{\gamma-1}}{1 + \exp(x^T \beta) t^\gamma} \quad \text{where } \gamma > 0.$$

When $0 < \gamma < 1$, the hazard rate is first increasing and then decreasing. For $\gamma > 1$, it is monotonic. The corresponding expected duration when $0 < \gamma < 1$ is given by :

$$E(s | \bar{x}) = B\left(1 + \frac{1}{\gamma}, 1 - \frac{1}{\gamma}\right) \exp\left(-\bar{x}^T \frac{\beta}{\gamma}\right) \equiv \bar{B}(\gamma) \exp(\bar{x}^T \beta^0) \quad \beta^0 = -\frac{\beta}{\gamma}$$

where $B(\cdot)$ is the beta function. This expectation formula identical in structure to the Weibull case, in that the first derivatives will all be defined in terms of a scalar multiplicative factor, which in the log-logistic case given by

$$k_{LL} = \frac{\bar{B}(\hat{\gamma}_M) \left(\exp(\bar{x}_M^T \hat{\beta}_M^0) - \exp(\bar{x}_F^T \hat{\beta}_M^0) \right)}{\hat{\beta}_M^{0T} [\bar{x}_M - \bar{x}_F]}$$

The decomposition will thus have a similar form to the Weibull specification.

Uncertainty over the nature of duration dependence in the presence of unobserved heterogeneity has led to the use of specifications involving mixtures of distributions. A final example of a widely used specification is the mixture Weibull hazard function with gamma heterogeneity. The hazard specification is :

$$\lambda(t; x_i) = \alpha t^{\alpha-1} \exp(x_i^T \beta) v_i$$

where v_i has a gamma distribution with unit mean and variance equal to σ^2 . The expected duration in this case (Lancaster, 1979, p. 952) is given by :

$$E(s | \bar{x}) = \frac{B\left(\frac{1}{\alpha}, \frac{1}{\sigma^2} - \frac{1}{\alpha}\right) \exp\left(-\bar{x}^T \frac{\beta}{\alpha}\right)}{\sigma^{2/\alpha} - \alpha} \equiv V(\alpha, \sigma) \exp(\bar{x}^T \beta^*)$$

and the decomposition will have the same form the two earlier specifications with

$$k_{WG} = \frac{V(\hat{\alpha}_M, \hat{\sigma}_M) \left(\exp(\bar{x}_M^T \hat{\beta}_M^*) - \exp(\bar{x}_F^T \hat{\beta}_M^*) \right)}{\hat{\beta}_M^T [\bar{x}_M - \bar{x}_F]}.$$

Example 1: Duration to obtaining a permanent employment for young persons

The data used come from the French Generation 2004 survey, which follows a cohort of individuals leaving the education system in 2004. The age of the person in that year is obviously related to the number of years spent in the education system. However in France, the correspondence between educational attainment in terms of the highest diploma obtained and the age at which the person leaves the system is clouded by the widespread phenomenon of spending more than one year in a particular grade. For example, many university students take their first year twice over. The same occurs lower down the education ladder, where a pupil may spend two years in a particular grade (some pupils even skip a grade). When analyzing access to permanent employment, this lag acts as a signal to employers. The average education lag in the sample is more than two years (see Table 1). The duration of until finding a permanent job is modelled as a function of two education variables: educational attainment measured as the theoretical number of years necessary to obtain a given diploma and the education lag. In addition the overall unemployment rate in the geographical locality of the person's domicile in 2004 is used to measure the influence of the state of the labour market. The duration variable used is the number of months following exit from the education system.

A second phenomenon often associated with difficulty finding a permanent job among young persons is cultural and ethnic origin, and specifically whether the person has parents who are immigrants. In the sample used, 16% have parents who are not of French origin. There are differences in educational attainment and education lag that also suggest that children of immigrants are likely to fare less well in the labour market. In addition to these factors there may also be discrimination in the recruitment of young persons which favours

those whose parents are not immigrants. We therefore use the proposed decomposition to quantify the different components of the difference in durations between the two groups of young persons.

The decomposition uses a model-based estimate of the mean duration for each group and decomposes the difference between these. In the current case, we assume that the hazard function is of the Weibull form - equation (13) – and the corresponding expected duration is given above in equation (14). The parameters are in fact obtained using an accelerated failure time model which is estimated separately for the two groups, and the results are presented in Table 2. The estimated Weibull shape parameters indicate that the hazard function is increasing with duration. The other estimated coefficients suggest that more education, shorter education lag and a smaller unemployment local unemployment all reduce the duration and more so for children of immigrants compared to their French counterparts. There is however a large difference between the estimated constant terms for the two groups which suggests that there is discrimination in access to employment in favour of those of French origin.

The decomposition of the estimated expected duration (in months and not logarithms) is undertaken using the average French origin characteristics in the counterfactual. The difference to be decomposed is the difference between two model-based estimates of the average duration corresponding to the mean characteristics of the respective groups. This is 15.2 months (see Table 3). The structural component of this gap is 13.2 months – or 87%. The composition effect, the part due to differences in characteristics, therefore accounts for only a minor part of the gap (2 months). The detailed decomposition of the composition effect suggests that improving the educational performance (on both fronts) of children whose parents are immigrants will reduce the expected duration to obtaining a permanent job. However, this is limited to the extent that for identical characteristics, those with immigrant parents are at a disadvantage.

(c) Nonlinear models which are not defined on a linear index

The presence of a linear index in a nonlinear function enables the proportional weighting factor $k(\cdot)$ to be obtained directly, independently of each covariate contribution. An example where there is no simple form for the decomposition other than the form of result in Proposition 1 is where a variable is defined using the Box-Cox transformation.

In the case of both dependent and independent variables being transformed, the model to be estimated with only one explanatory variable is:

$$\frac{y_i^\gamma - 1}{\gamma} = \beta_0 + \left(\frac{x_i^\theta - 1}{\theta} \right) \beta_1 + u_i$$

In order to obtain the conditional expectation of y and proceed to decompose the difference between groups, a difficulty arises since the error term, u , will appear in the equation when expressed in terms of y alone:

$$y_i = \left[1 + \gamma \left(\beta_0 + \left(\frac{x_i^\theta - 1}{\theta} \right) \beta_1 + u_i \right) \right]^{\frac{1}{\gamma}}$$

Following Abrevaya and Hausman (2004), a J -th term Taylor expansion around $u = 0$ can be used to avoid this difficulty when the error term is normally distributed. The expected value of y conditional on x is :

$$E(y|x) \approx G(x, u=0) + \sum_{j=1}^J \frac{E(u^j)}{j!} g^j(x, u) \quad (15)$$

where $G(x, u=0) = (1 + \gamma V(x))^{\frac{1}{\gamma}}$, $V(x) = \beta_0 + \left(\frac{x_i^\theta - 1}{\theta} \right) \beta_1$

and $g^j(x, u) = \left. \frac{\partial^j G(x, u)}{\partial u^j} \right|_{u=0}$.

This simplifies in the case when u is normally distributed with a zero mean, since its odd numbered moments are zero. Thus when $J = 1$ the expected value is :

$$E(y|x) \approx (1 + \gamma V(x))^{\frac{1}{\gamma}}$$

and for $J = 2$:

$$E(y|x) \approx (1 + \gamma V(x))^{\frac{1}{\gamma}} + \frac{\sigma^2}{2} (1 - \gamma) (1 + \gamma V(x))^{\frac{1}{\gamma} - 2}$$

In both cases the decomposition will not simplify in the same way as for nonlinear models defined on a linear index and so the integral in Proposition 1 will have to be explicitly evaluated.

Example 2: Male-female differences in earnings in France

Data from the 2005 French Labour Force Survey are used to examine earnings differences between males and females. The sample includes individuals aged 20 to 54 who declare earnings and hours worked enabling an hourly wage to be calculated. The earnings equations are of the form:

$$\frac{y_i^\gamma - 1}{\gamma} = \beta_0 + \beta_1 E_i + \beta_2 Paris_i + \beta_3 \left(\frac{x_i^\theta - 1}{\theta} \right) + u_i \quad (16)$$

where y is earnings, E is education level, $Paris$ is dummy variable equal to one if the person lives in the Paris area and x which is potential labour market experience and enters the model as a Box-Cox transformation. In fact there are only small differences in the means of these variables by gender (see Table 4). Females have more education and almost as much experience as males. There are a slightly higher proportion of females living in the Paris area in this sample (15.3% compared to 13.9%). Nevertheless, males earn 8.9% than females. The parameters of the model are estimated by maximum likelihood, and are presented in Table 5. For both sexes, the Box-Cox transformations are significant (θ and γ are neither zero nor unity), so that there is a nonlinear relation between earnings and experience and more generally between earnings and the other covariates. The coefficients on education and living in Paris are higher for females, but the constant term is higher for males. The decomposition of the difference in earnings (rather than log earnings) is based on the approximation in equation (15) applied to the model given in (16). The difference to be decomposed is 0.838 (less than one euro). Based on the counterfactual, females would earn more than males if their characteristics were rewarded in the same way as those of males. The structure effect is therefore the more important component (+1.27). The composition effect is negative and the key factor indicated by the detailed decomposition is education.

V Conclusion

By recognizing that the Oaxaca technique can be obtained by an application of the mean value theorem, a new decomposition method applicable to nonlinear models is proposed. For continuous nonlinear functions, the decomposition proposed here is based on counterfactuals using the means of the right hand side variables and group differences in fitted values of the estimated functions. In this decomposition the unexplained component or 'structure effect' is completely defined, and a detailed decomposition of the 'composition effect' is possible. There is no remainder. Explicit formulae for decompositions of the binary logit and probit models, duration models based on the Weibull and log-logistic hazard specifications and a Box-Cox model are presented, although the method applies to any a model which is specified as a continuous function of the explanatory variables.

It differs from other approaches in two key respects. First the proposed decomposition is based on the mean characteristics of each group and the fitted value of the model evaluated at the vector of means. It is therefore based on the same clearly defined counterfactual as the original Oaxaca approach. It does not decompose group differences in the observed mean of the dependent variable; but in general this cannot be done in a nonlinear model in an exact manner. Furthermore, in the presence of censoring, as in duration models, the sample mean of the left hand side variable is not a very useful measure. Second, its application is not restricted to the class of nonlinear models defined on a linear index as is the case in most other proposed decomposition techniques. This type of model is used in applied work for

understandable reasons, but there are cases when the nonlinear aspect of the relation between economic variables goes beyond this form of model as in the Box-Cox case.

References

- Abreyeva, Jason. and Jerry Hausman (2004) Response error in a transformation model with an application to earnings equation estimation, *Econometrics Journal*, 7, 366-388.
- Andrews, George, Richard Askey and Ranjan Roy (2001) *Special Functions (Encyclopedia of Mathematics and its Applications)*, Cambridge University Press, Cambridge.
- Baker, George and Pravin Trivedi (1985) Estimation of unemployment duration from grouped data: a comparative study, *Journal of Labor Economics*, 3, 153-174.
- Bartle, Robert and Donald Sherbert (2011) *Introduction to Real Analysis*, Fourth Edition, John Wiley and Sons.
- Bauer, Thomas and Mattias Sinning (2008) An extension of the Blinder-Oaxaca decomposition to nonlinear models, *Advances in Statistical Analysis*, 92, 197-206.
- Bazen, Stephen, Xavier Joutard and Moustapha Niang (2014) The measurement of unemployment using completed durations: evidence on the gender gap in unemployment in France", *Journal of Economic Inequality*, 12, 517-534.
- Beblo, Miriam, Denis Beninger, Anja Heinze, and Francois Laisney (2003) Methodological issues related to the analysis of gender gaps in employment, earnings and career progression, *Final Report for the European Commission Employment and Social Affairs Directorate*, Brussels.
- Blinder, Alan (1973) Wage discrimination: reduced form and structural estimates, *Journal of Human Resources*, 8, 436-465.
- Cameron, Colin and Pravin Trivedi (2003) *Microeconometrics*, Oxford University Press, Oxford.
- Fairlie, Douglas (2005) An extension of the Blinder-Oaxaca decomposition technique to logit and probit models, *Journal of Economic and Social Measurement*, 30, 305-316.
- Nicole Fortin, Thomas Lemieux and Sergio Firpo, (2011) Decomposition methods in economics, *Handbook of Labor Economics Volume 4*, Elsevier, Amsterdam.
- Neumann, Shoshana and Ronald Oaxaca (2004) Wage decompositions with selectivity-corrected wage equations: a methodological note, *Journal of Economic Inequality*, 2, 3-10.
- Nielson, Helena (1998) Discrimination and detailed decomposition in a logit model, *Economics Letters*, 61, 115-120.
- Oaxaca, Ronald, (1973) Male-female differentials in urban labor markets, *International Economic Review*, 14, 673-709.
- Powers, Daniel and George Pullum (2006) Multivariate decomposition for nonlinear models, Paper presented at PPA conference, Princeton University.
- Powers, Daniel and Myong Sue Yun (2009) Decomposition of multivariate hazard rates, *Sociological Methodology*, 39, 233-263.
- Rothe, Christophe (2012) Decomposing the composition effect, *IZA Discussion Paper No 145*, Insititute for the Study of Labor, Bonn.

Schwiebert, Jorge (2015) A detailed decomposition for nonlinear models, *Journal of Economic Inequality*, 13, 53-67.

Wagstaff, Adam and Nga Nguyet Nguyen (2002) Poverty and survival prospects of Vietnamese children under Doi Moi, *World Bank Policy Research Working Paper* No 2832.

Wolff, Francois-Charles (2012) Decomposition of nonlinear models using simulated residuals, *Economics Letters*, 116, 346-348.

Yun, Myong Sue (2000) Decomposition analysis for a binary choice model, *IZA Discussion Paper* No 145, Insititute for the Study of Labor, Bonn.

Yun, Myong Sue (2004) Decomposing differences in the first moment, *Economics Letters*, 82, 275-280.

Yun, Myong Sue (2007) An extension of the Oaxaca decomposition using generalized residuals, *Journal of Economic and Social Measurement*, 32, 15-22.

Figure 1 Graphical representation of the decomposition for a nonlinear function

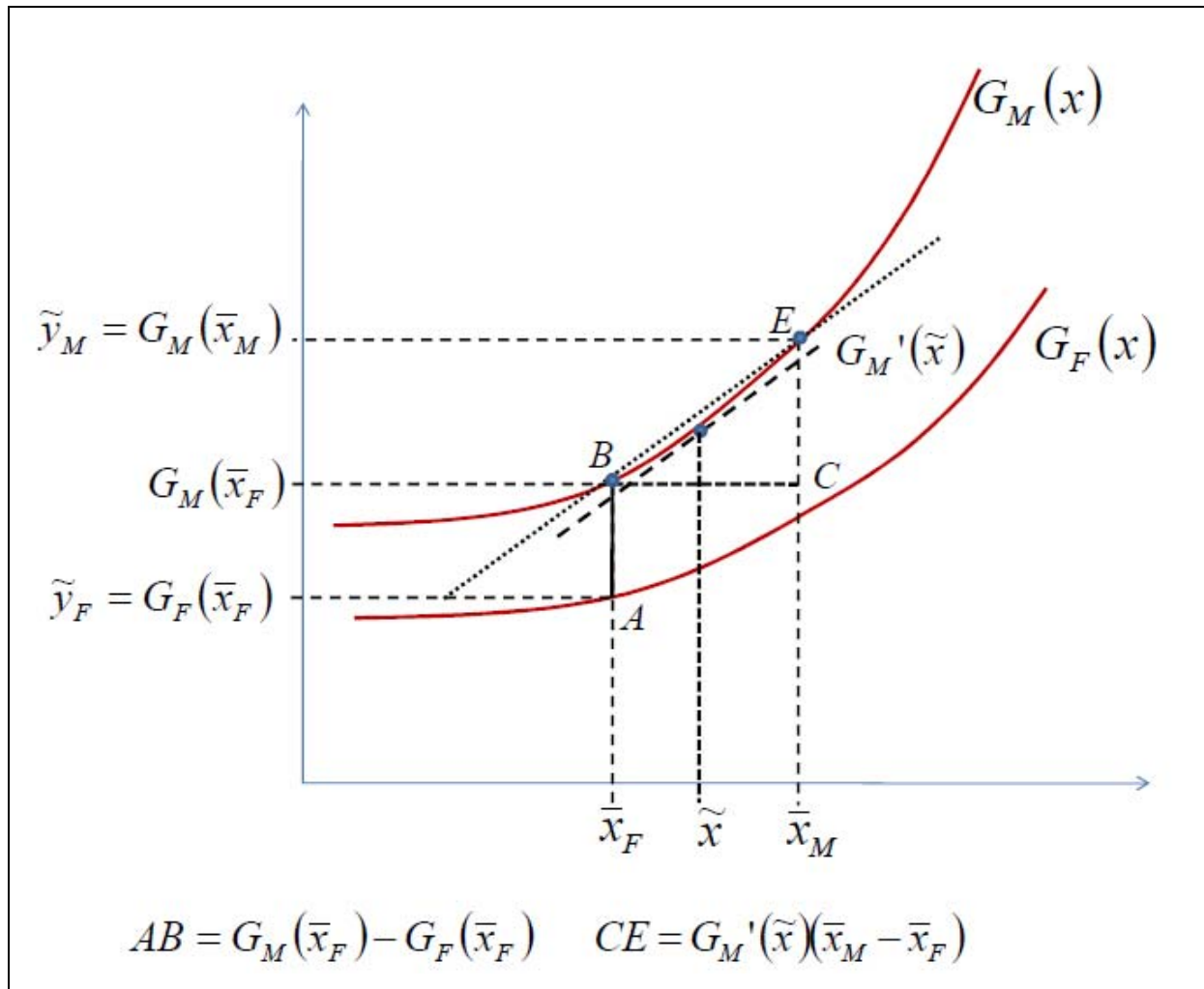
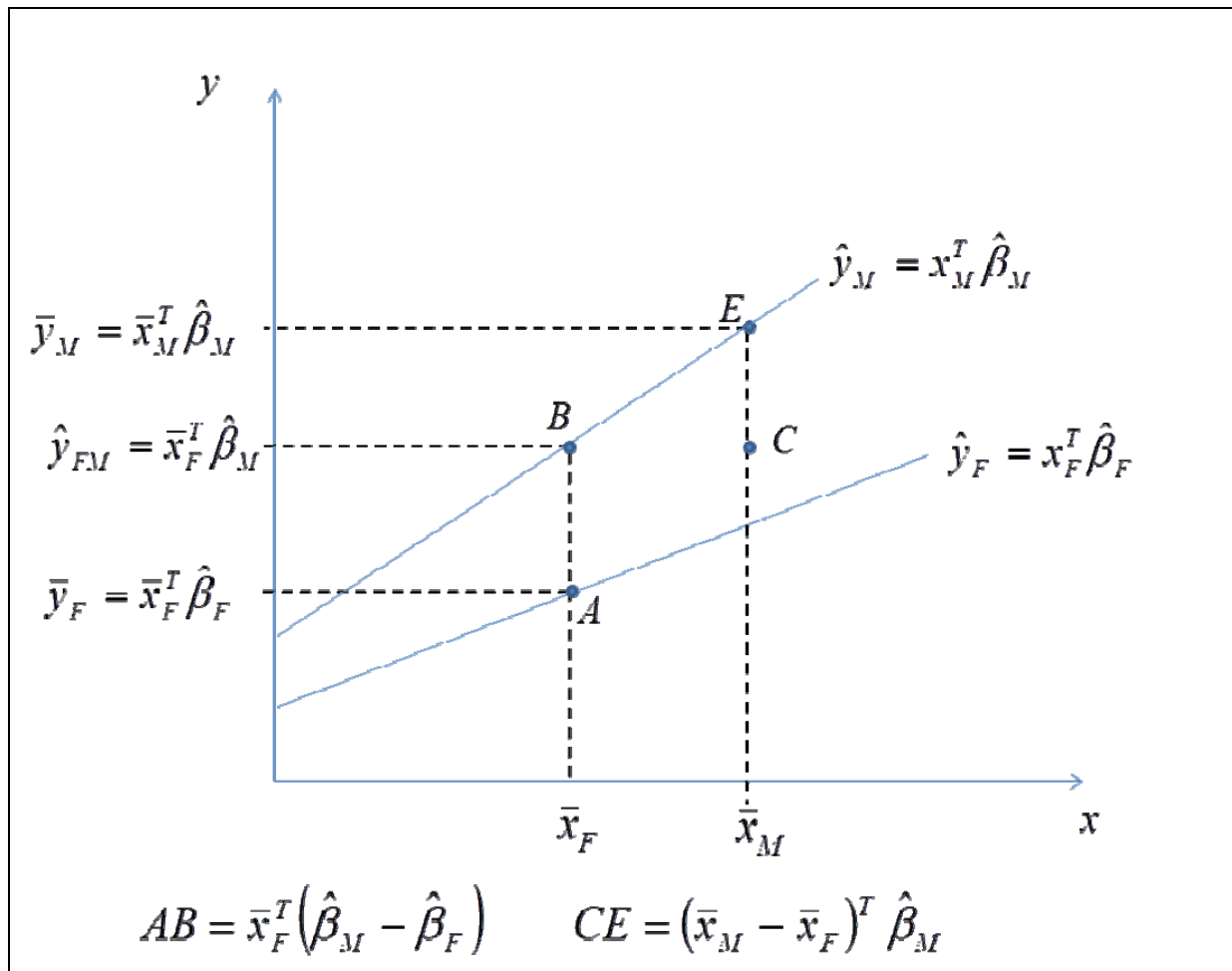


Figure 2 Graphical representation of the Oaxaca decomposition for a linear model



	Parents French	Parents Not French	Difference
Censored duration (months)	8.64 (4.63)	9.1 (4.43)	0.37
Censoring Indicator	0.4 (0.49)	0.34 (0.47)	
Education level (years)	13.23 (2.76)	12.64 (2.84)	-0.59
Education lag (years)	2.43 (1.96)	2.92 (2.02)	0.59
Local unemployment rate	8.64 (2.2)	9.13 (2.18)	0.49
Sample size	13,728	2,639	

	Parents French	Parents Not French	Difference
Constant	3.314 (0.087)	3.815 (0.217)	-0.53
Education level	-0.077 (0.004)	-0.106 (0.01)	0.029
Education lag	-0.0433 (0.006)	-0.012 (0.016)	-0.0213
Local unemployment rate	0.0267 (0.006)	0.018 (0.015)	0.0087
Weibull parameter	0.714 (0.008)	0.7 (0.02)	0.014
Sample size	13,728	2,639	
Estimated standard errors in parentheses			

Table 3 A decomposition of group difference in duration (in months)		
	Our decomposition	Yun's method
$N(\bar{x}_N) = \bar{\Gamma}(\hat{\alpha}_N) \exp(\bar{x}_N^T \hat{\beta}_N^*)$	51.97	$\overline{N(x_{Ni})} = 56.59$
$F(\bar{x}_F) = \bar{\Gamma}(\hat{\alpha}_F) \exp(\bar{x}_F^T \hat{\beta}_F^*)$	36.81	$\overline{F(x_{Fi})} = 38.55$
Difference to be decomposed:	15.16	18.04
$N(\bar{x}_F) = \bar{\Gamma}(\hat{\alpha}_N) \exp(\bar{x}_F^T \hat{\beta}_N^*)$	38.79	$\overline{N(x_{Fi})} = 40.72$
Structure effect	13.18	15.82
Composition effect	1.98	2.22
Detailed decomposition:		
education level	$0.108 \times (13.23 - 12.64) \times (37.8) = 2.4$	2.7
education lag	$0.060 \times (2.43 - 2.92) \times (37.8) = -1.25$	-1.26
local unemployment rate	$-0.037 \times (8.64 - 9.13) \times (37.8) = 0.69$	0.78

Table 4 Means and standard deviations of variables used in earnings equations			
	Males	Females	Difference
Hourly earnings	10.04 (3.87)	9.22 (3.56)	0.78
Education level (age at end of full-time education)	18.01 (1.99)	18.40 (2.05)	-0.39
Labour Market Experience	19.41 (10.04)	19.36 (10.24)	0.05
Live in Paris	0.139 (0.34)	0.153 (0.36)	0.014
Sample size	8,006	7,363	

Table 5 Male and female earnings equations with Box-Cox transformations			
	Males	Females	Difference
Constant	0.621 (0.007)	0.541 (0.002)	0.049
Education level	0.039 (0.0008)	0.044 (0.001)	-0.005
Paris region	0.041 (0.004)	0.054 (0.004)	-0.013
Experience	0.041 (0.001)	0.03 (0.001)	0.011
Box-Cox parameter (earnings)	-0.381 (0.025)	-0.353 (0.025)	-0.028
Box-Cox parameter (experience)	0.227 (0.052)	0.293 (0.069)	-0.065
Sigma	0.119	0.131	-0.012
Sample size	8,006	7,363	
Estimated standard errors in parentheses			

Table 6 Decomposition of gender differences in hourly earnings (2 nd order Taylor expansion)	
$G(\bar{x}_F^T, \hat{\beta}_F, \hat{\lambda}_F)$	9.961
$G(\bar{x}_M^T, \hat{\beta}_M, \hat{\lambda}_M)$	9.123
Difference to be decomposed :	0.838
$G(\bar{x}_F^T, \hat{\beta}_M, \hat{\lambda}_M)$	10.347
Structure effect	1.224
Composition effect	-0.386
Detailed decomposition	
Education level	-0.367
Labour Market Experience	0.005
Lives in Paris	-0.014