The Taylor Decomposition: A Unified Generalization of the Oaxaca Method to Nonlinear Models

Stephen Bazen, Xavier Joutard

Working Papers / Documents de travail

The Taylor Decomposition: A Unified Generalization of the Oaxaca Method to Nonlinear Models

Stephen Bazen
Xavier Joutard

WP 2013 - Nr 32

The Taylor decomposition:

a unified generalization of the Oaxaca method to nonlinear models

Stephen Bazen* and Xavier Joutard

Aix-Marseille University

(Aix-Marseille School of Economics), CNRS and EHESS

May 2013

Paper presented at the French Econometrics Conference, Rennes, November 2012

We are grateful to Thomas Lemieux and Mareva Sabatier for helpful comments.

Keywords: Oaxaca decomposition; nonlinear models

JEL Codes: C.20

* Corresponding author : stephen.bazen@univ-amu.fr

AMSE-GREQAM, Centre de la Vieille Charité, 2 rue de la Charité 13002 Marseille, France

Abstract

The widely used Oaxaca decomposition applies to linear models. Extending it to commonly used nonlinear models such as binary choice and duration models is not straightforward. This paper shows that the original decomposition using a linear model can be obtained as a first order Taylor expansion. This basis provides a means of obtaining a coherent and unified approach which applies to nonlinear models, which we refer to as a Taylor decomposition. Explicit formulae are provided for the Taylor decomposition for the main nonlinear models used in applied econometrics including the Probit binary choice and Weibull duration models. The detailed decomposition of the explained component is expressed in terms of what are usually referred to as marginal effects and a remainder. Given Jensen's inequality, the latter will always be present in nonlinear models unless an ad hoc or tautological basis for decomposition is used.

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 regressions[1]. The difference is expressed in terms of two components that contribute to the divergence in group means: the 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 article[2]. 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. However, 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.

latter basis is problematic for the identification of certain components of interest in the Oaxaca approach since it is not defined in terms of a counterfactual specified in terms of mean characteristics. A key property of the Oaxaca decomposition is the identification and measurement of discrimination based on assuming that two groups have the same average characteristics.

The current paper has the aim of providing a unified and coherent generalization of the Oaxaca decomposition to nonlinear models by recognizing that the original decomposition can be obtained from a first order Taylor expansion of one group's function around the means of the other group's explanatory variables. The paper begins in section 1 with an examination of the basis of the Oaxaca decomposition and the difficulties encountered when seeking to generalize this approach to nonlinear relations. In the following section, an alternative interpretation of the Oaxaca decomposition is presented as a starting point for the elaboration of a unified generalization applicable to nonlinear relations. This is based on the fact that the original Oaxaca decomposition can be obtained from a first order Taylor expansion and this key result is presented in section 2. The resulting *Taylor decomposition* is proposed as a means of extending the Oaxaca-linear technique to the decomposition of group differences based on nonlinear functions. It differs from the Oaxaca-linear method since the expansions involve a remainder which contains (weighted) polynomial terms that go beyond the first order. Interestingly, the polynomial terms only concern the "explained" component, while the "structure" effect is completely defined. In section 3, explicit forms for the Taylor decomposition of some widely used nonlinear models for binary choice and duration analysis are presented, before illustrating the methodology with two applications in section 4.

I Extending the Oaxaca method to nonlinear relations

It is important to note from the outset that the basic Oaxaca decomposition has a certain number of features which limit the extent to which the method can be directly generalized. First and foremost, it applies to an explicitly linear framework which is specified as follows. 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

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

The Oaxaca decomposition is obtained by first estimating the parameters using ordinary least squares (hereafter 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 $\bar{x}_M^T \beta_F$) 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 :

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

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 $\bar{x}_M = \bar{x}_F$ 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. In other words a model is used to construct a counterfactual situation.

(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 :

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

The presence of a constant ensures that the sum and therefore the mean of the estimated OLS residuals, $\hat{u}_{gi}$, are both equal to zero – or equivalently that the mean of the fitted values is equal to the sample mean of the dependent variable. In this case – which is the form presented explicitly in Blinder (1973, p. 439) – the decomposition can be written as :

$$\bar{y}_M - \bar{y}_F = \hat{\beta}_{0M} + \bar{x}_M^T \hat{\beta}_M - \left(\hat{\beta}_{0F} + \bar{x}_F^T \hat{\beta}_F\right)$$

$$= \hat{\beta}_{0M} - \hat{\beta}_{0F} + \bar{x}_F^T \left(\hat{\beta}_M - \hat{\beta}_F\right) + \hat{\beta}_M^T \left(\bar{x}_M - \bar{x}_F\right) \qquad (4)$$

This property is a consequence of OLS estimation, and the equality in (4) is valid even if the estimates are biased. Oaxaca (1973) assimilates the constant term into the coefficient vector.

(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 :

$$E(y_{Mi}) - E(y_{Fi}) = E(x_{Mi}^T)\beta_M - E(x_{Fi}^T)\beta_M + E(x_{Fi}^T)\beta_M - E(x_{Fi}^T)\beta_F$$

$$= E(x_{Fi}^T)[\beta_M - \beta_F] + [E(x_{Mi}^T) - E(x_{Fi}^T)]\beta_M \qquad (5)$$

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 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 unexplained component is $\bar{x}_M^T \left(\hat{\beta}_M - \hat{\beta}_F\right)$, rather than $\bar{x}_F^T \left(\hat{\beta}_M - \hat{\beta}_F\right)$ as in equation (2). The choice of reference group characteristics for the decomposition affects the size of the each of components, except in the extreme case when $\hat{\beta}_M = \hat{\beta}_F$ in equations (2) or (4). In general, there is no unique, unambiguous measure of the extent of discrimination in terms of an unexplained component.

Extending the Oaxaca (linear) approach to nonlinear relations is not straightforward. First, the presence of nonlinearities in the relation means that OLS cannot be applied and the decomposition will not have the original Oaxaca form. The decomposition has certain properties that are related explicitly to the use of least squares. Second, and more importantly, when applied to nonlinear models, an Oaxaca-type decomposition of differences in either sample means or expectations of a variable will not be exact, and so neither of (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 $g(x)$, $E[g(x)] \neq g(E[x])$. This implies that even an exact Oaxaca-type decomposition at the population level in terms of expectations, as in (5), is unlikely to be obtained[3]. Due to the (near) 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 $\tilde{y}_{Mi} = M\left(x_{Mi}\right)$ and $\tilde{y}_{Fi} = F\left(x_{Fi}\right)$, respectively. These functions would normally be the estimated conditional expectations in econometric applications. The original Oaxaca

[3] The equality only holds with certainty for affine functions.

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[4] :

$$\bar{y}_M = \bar{\tilde{y}}_M = M(\bar{x}_M) \qquad \text{where} \quad \bar{\tilde{y}}_M = \frac{1}{n_M}\sum_{i=1}^{n_M}\tilde{y}_{Mi}$$

The same is true for group F. This means that the group difference in any of these means can be used as basis for a decomposition in the affine case. 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 mean of the fitted values of estimated functional relationship, $\bar{\tilde{y}}_M - \bar{\tilde{y}}_F = \frac{1}{n_M}\sum_{i=1}^{n_M}\tilde{y}_{Mi} - \frac{1}{n_F}\sum_{i=1}^{n_F}\tilde{y}_{Fi} = \overline{M(x_{Mi})} - \overline{F(x_{Fi})}$;

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

In view of Jensen's inequality, basis (a) is unlikely to prove fruitful for a generalization. Even in the 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), 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) :

$$\bar{\tilde{y}}_M - \bar{\tilde{y}}_F = \overline{M(x_{Mi})} - \overline{M(x_{Fi})} + \overline{M(x_{Fi})} - \overline{F(x_{Fi})} \qquad (6)$$

[4] This is unsurprising given that $E[g(x)] = g(E[x])$ in the affine (and thus linear) cases.

where $\overline{G(x_{gi})} = \frac{1}{n_g}\sum_{i=1}^{n_g} G(x_{gi})$. In population terms, this corresponds to a decomposition of the following difference : $E[M(x_{Mi})] - E[F(x_{Fi})]$.

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

$$M(\bar{x}_M) - F(\bar{x}_F) \neq \overline{M(x_{Mi})} - \overline{F(x_{Fi})} \tag{7}$$

Using (b) a basis therefore entails disconnecting the decomposition from the mean vectors $\bar{x}_M$ and $\bar{x}_F$. In other words, when the two groups have identical means, the explained component is not equal to zero (as it is in the Oaxaca decomposition) and the decomposition does not reduce to the structure effect. This is a serious weakness since the identification of the latter is one of the main reasons for undertaking a decomposition of this kind because this component is precisely that which is associated with discrimination. Secondly, if *estimated* functions or fitted values are used, there is no guarantee that $\bar{y}_M - \bar{y}_F = \overline{M(x_{Mi})} - \overline{F(x_{Fi})}$: this equality is not valid in the case of the probit model for example. While the equality holds for the logit model when there is a constant in the model and the parameters are estimated by maximum likelihood, this is a special case.

A decomposition using basis (b) is therefore not generally expressed in terms of the means of the variables y and x and thus diverges 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)). Given the inherent characteristics of nonlinear relations, it is preferable to recognize that it involves an approximation and provide a precise estimate of any remainder.

In the next section, an approach derived from the original "Oaxaca-linear" method is developed into a unified framework applicable to both linear and nonlinear functions using basis (c), $M(\bar{x}_M) - F(\bar{x}_F)$. As the results of the next section show, applying a first order Taylor expansion around the mean of x for the other group to one of these functions ($M(\bar{x}_M)$ or $F(\bar{x}_F)$) gives precisely the Oaxaca decomposition, presented in equation (2), when the functions are linear. The same operation enables a decomposition method to be derived which is applicable to any parametric function, be it linear or nonlinear.

II Oaxaca's decomposition as a Taylor expansion

When seeking to generalize the Oaxaca approach to models other than a linear regression, it is important to note that a decomposition of group differences can always be obtained using group-specific estimated functions and by defining a counterfactual case. The decomposition is obtained by adding and subtracting the counterfactual case on the right hand side. Thus aggregate decompositions are "trivial" in the sense that they are identities and can be arbitrarily defined in terms of any number of counterfactual cases. However, the major justification for undertaking a decomposition is precisely to identify what would happen if the two groups being studied had identical characteristics, which is precisely what is required when examining the existence of discrimination and for which the decomposition was originally devised. In the Oaxaca context, this is specified in terms of identical *average* characteristics, so that $\bar{x}_M - \bar{x}_F = 0$. This implies that only two counterfactuals are relevant when decomposing differences in the first moment.

In order to obtain a coherent generalization of the Oaxaca method applicable to nonlinear models, as pointed out above, a choice will therefore have to be made about the basis on which a decomposition is made. Usually the objective is to establish the existence of unjustifiable (and/or discriminatory) components in group differences. In this case, the difference between two estimated functions evaluated at common values of the arguments (usually one group's mean characteristics : $M(\bar{x}_M) - F(\bar{x}_M)$) is the component of interest. The remaining difference (the explained component) should also be evaluated at the same value of the arguments $\bar{x}_M, \bar{x}_F$, which is precisely what the "Oaxaca-linear" approach does. It is clear that the decomposition in using as a basis $M(\bar{x}_M) - F(\bar{x}_F)$ will be more appropriate than that given in (6).

An Oaxaca-type decomposition can be straightforwardly derived based on the difference between two estimated functions $M(\bar{x}_M) - F(\bar{x}_F)$, each evaluated at the respective mean vector, $\bar{x}_M$ and $\bar{x}_F$. At the population level, this corresponds to the difference $M(E[x_{Mi}]) - F(E[x_{Fi}])$. Assuming that the first derivatives exist and expanding one of the functions (say, $M(\bar{x}_M)$) around the other group's mean vector (in this case $\bar{x}_F$) to the first order gives :

$$M(\bar{x}_M) = M(\bar{x}_F) + M'(\bar{x}_F)^T [\bar{x}_M - \bar{x}_F] + R$$

where $M'(\bar{x}_F)^T = \left(\frac{\partial M(\bar{x}_F)}{\partial x_1} \quad \frac{\partial M(\bar{x}_F)}{\partial x_2} \cdots\cdots \quad \frac{\partial M(\bar{x}_F)}{\partial x_k} \right)$ is the vector of first order partial derivatives calculated for $x = \bar{x}_F$, and the remainder, R, contains a sequence of higher order polynomials in $[\bar{x}_M - \bar{x}_F]$. Subtracting from $F(\bar{x}_F)$ both sides gives an equation that resembles the Oaxaca decomposition :

$$M(\bar{x}_M) - F(\bar{x}_F) = M(\bar{x}_F) - F(\bar{x}_F) + M'(\bar{x}_F)^T [\bar{x}_M - \bar{x}_F] + R \qquad (8)$$

The expansion around $\bar{x}_F$ does two things in this decomposition. First it automatically defines a counterfactual situation, $M(\bar{x}_F)$, the value of y that the average member of group F would have if they were in group M, according to the estimated function. Secondly, the expansion creates a term involving the difference in mean characteristics $[\bar{x}_M - \bar{x}_F]$. The remainder will also depend on this difference. In the case of a linear function, this approach constitutes a valid representation of the Oaxaca decomposition as the following result makes clear.

Proposition: The "Oaxaca-linear" decomposition (2) can be obtained from a first-order Taylor expansion of one of functions around the mean vector for the other group in the case where $M(\bar{x}_M) = \bar{x}_M^T \hat{\beta}_M$ and $F(\bar{x}_F) = \bar{x}_F^T \hat{\beta}_F$, and $\hat{\beta}_g$ are vectors of OLS estimates in the equations $y_{gi} = g(x_{gi}) = x_{gi}^T \beta_g + \varepsilon_{gi}$ for $g = M, F$ [5].

Proof: A (first order) Taylor expansion of the function $M(\bar{x}_M)$ around $\bar{x}_F$ and subtracting $F(\bar{x}_F)$ gives equation (8). Given the linearity of each function :

(a) $M(\bar{x}_F) - F(\bar{x}_F) = \bar{x}_F^T \hat{\beta}_M - \bar{x}_F^T \hat{\beta}_F = \bar{x}_F^T (\hat{\beta}_M - \hat{\beta}_F)$ – which is the unexplained component.

[5] Yun (2004) mentions a similar property for his decomposition in equation (6) , but the result is only implicit in his paper. He arrives at the property by a different means and he does not appear to attach much importance to it.

(b) $M'(\bar{x}_F)^T[\bar{x}_M - \bar{x}_F] = \hat{\beta}_M^T[\bar{x}_M - \bar{x}_F]$ is the explained component.

(c) Since the function is linear, all derivatives other than the first are equal to zero, thus $R = 0$.

Therefore :

$$M(\bar{x}_F) - F(\bar{x}_F) = \bar{x}_F^T(\hat{\beta}_M - \hat{\beta}_F) + \hat{\beta}_M^T(\bar{x}_M - \bar{x}_F)$$

the right hand side of which is precisely the Oaxaca decomposition in (2). ∎

Remark : The sum of the components of the Oaxaca decomposition is equal to the difference in sample means, $\bar{y}_M - \bar{y}_F$, when $\bar{x}_g^T\hat{\beta}_g = \bar{y}_g$. This however is the case only if $M(x)$ and $F(x)$ are affine functions (that is, one of the elements of $\hat{\beta}_g$ is a constant – and thus one of elements $\bar{x}_g$ of is equal to one). This fact is recognized by Oaxaca (1973) who includes a constant in the coefficient vector $\hat{\beta}_g$, while Blinder (1973) includes a constant separately in the specification of the earnings equation.

A coherent, unified approach to extending the Oaxaca decomposition method to any given function is proposed by extending the equivalence established in the Proposition between the Oaxaca decomposition and a Taylor expansion of one of *the group functions evaluated at the mean values of the right hand side variables* around the mean vector for the other group. This is referred to as a Taylor decomposition.

Definition: *a Taylor decomposition*. Let $\tilde{y}_M = M(\bar{x}_M)$ and $\tilde{y}_F = F(\bar{x}_F)$, where both functions are differentiable. Then the difference between these two functions can be decomposed in the following way[6] :

$$\tilde{y}_M - \tilde{y}_F = [M(\bar{x}_F) - F(\bar{x}_F)] + M'(\bar{x}_F)^T[\bar{x}_M - \bar{x}_F] + R \qquad (8')$$

This formula is very similar to the Oaxaca equation and differs due to the nature of the marginal effects (given by the first derivatives $M'(\bar{x}_F)$) and to the presence of a remainder.

Remark: In applied work, the *estimated* conditional expectation of y evaluated at the mean of x (or the fitted value of y for the mean value of x) for each group would be used :

$$\tilde{y}_g = \tilde{E}(y_{gi} \mid \bar{x}_g) = G(\bar{x}_g)$$

In the case of a single explanatory variable, the different components can be presented in graphical form. The first term on the right hand side, which is the unexplained component, is the vertical distance between the two functions evaluated at the mean of group F (the distance AB in Figure 1). If y is earnings, this difference represents the additional earnings an individual with mean characteristics in group F would receive if they were paid on the same basis as someone in group M. This component is usually viewed as an estimate of earnings discrimination. The second term is the marginal effect of changing an individual with mean characteristics in group F into one with those of someone in group M. As before, in the case of a

[6] Clearly the expansion could also have been undertaken around $\bar{x}_M$.

single variable it is measured by the horizontal difference between the means of the variables multiplied by the slope of the function for group M (the distance CD). In the case of earnings differences, this component indicates (partly) the component of earnings differences which is justifiable. R is the remainder and captures both the extent of differences between the mean characteristics of the two groups and the degree of nonlinearity in the function for group M (distance DE). Obviously, the smaller the remainder the easier it is to interpret the Taylor decomposition. It is interesting that the remainder term will contain polynomial terms relating only to the explained part of group differences, $[\bar{x}_M - \bar{x}_F]^j$ i.e. it is part of the explained component or 'composition effect'[7].

It is worth emphasizing a number of properties of the Taylor decomposition.

(i) The basis of the decomposition is two estimated functions evaluated at the vector of means for each group. At the aggregate level, the Taylor decomposition is "trivial" in the sense that

$$M(\bar{x}_M) - F(\bar{x}_F) = M(\bar{x}_M) - M(\bar{x}_F) + M(\bar{x}_F) - F(\bar{x}_F)$$

The sum of the second and third terms on the right hand side of (8′) is therefore equal to :

$$M'(\bar{x}_F)^T [\bar{x}_M - \bar{x}_F] + R = [M(\bar{x}_H) - M(\bar{x}_F)]$$

In other words, the structure effect is exactly defined in spite of the presence of a remainder.

[7] Rothe (2012) provides a full treatment of the detailed decomposition of the composition effect.

(ii) Unlike the approaches proposed by Yun (2004) and Bauer and Sinning (2008), the Taylor decomposition is expressed in terms of sample means for the right hand side variables ($\bar{x}_g$). If the two groups have identical mean characteristics, the explained component or composition effect is equal to zero, because the remainder is zero (and the first order component is clearly equal to zero). Thus while the aggregate decomposition is in a certain sense trivial, the Taylor decomposition shows that the conditions under which the composition effect is zero are expressed in terms of the difference in means $\bar{x}_M - \bar{x}_F$.

The remainder will be equal to zero when $\bar{x}_M = \bar{x}_F$, as can be seen from the Lagrange form of the remainder (see Bartle and Sherbert, 2011) which in this case is the quadratic form :

$$R = \frac{[\bar{x}_M - \bar{x}_F]^T}{2!} M''(\tilde{x})[\bar{x}_M - \bar{x}_F]$$

where $M''(\tilde{x})$ is the matrix of second derivatives for the vector $\tilde{x}^T = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_k)$ where each $\tilde{x}_j$ lies in the interval between $\bar{x}_{Mj}$ and $\bar{x}_{Fj}$, for $j = 1, 2, ..., k$. When $\bar{x}_M = \bar{x}_F$, this remainder is clearly zero, and the sole source of divergence is the structure effect.

(iii) The explained component or composition effect of the Taylor decomposition provides an additive detailed decomposition as :

$$M'(\bar{x}_F)^T[\bar{x}_M - \bar{x}_F] + R = M_1(\bar{x}_F)[\bar{x}_{1M} - \bar{x}_{1F}] + M_2(\bar{x}_F)[\bar{x}_{2M} - \bar{x}_{2F}] + \ldots + M_k(\bar{x}_F)[\bar{x}_{kM} - \bar{x}_{kF}] + R$$

The remainder is there as a consequence of the nonlinearity of the estimated functions used. The first derivative terms M_j are just the usual *marginal effects* for group M, but evaluated at $\bar{x}_F$. In a linear model, the marginal effects are constants and the remainder is zero.

(iv) The Taylor decomposition dissociates the method by which the parameters are estimated from the nature of the decomposition. It does not require the estimated residuals or fitted values to satisfy certain restrictions. For example, if the model is linear and the data are censored in such a way that OLS is biased and inconsistent, a reliable estimator of the function parameters can be used and the resulting estimates plugged in to the initial linear specification. Equality of the mean of the fitted values and the mean of the dependent variable is not required, and this would not occur in general in any case[8].

(v) Like the Oaxaca technique, the Taylor decomposition is subject to the index number problem. The decomposition could have been obtained from an expansion of the function $F(\bar{x}_F)$ around $\bar{x}_M$. The value of the unexplained component will be different if the two functions are different.

The principal advantage of the Taylor decomposition is that is *coherent* in the sense that it is expressed in terms of the means of the right hand side variables and applies in the same way to any parametric function that has first or higher derivatives. If groups have identical means, the structure effect is the only source of divergence between groups. It is a valid extension of the Oaxaca procedure because both the latter and the Taylor decomposition can be obtained from the same first order expansion, that is by expanding $M(\bar{x}_M)$ (respectively, $F(\bar{x}_F)$) around $\bar{x}_F$ (respectively, $\bar{x}_M$). The first term is always the unexplained component (or structure effect) and the second provides the first order composition effect in a detailed manner. The presence of a remainder term is an unavoidable consequence of the nonlinearity of the relation *when a first order expansion is used*, but is only of interest

[8] The logit model is a rare case where this is true – see below.

when undertaking a detailed decomposition of the composition effect. In any case, its size can be quantified. In many situations the sample mean ($\bar{y}$) is not relevant in nonlinear models due to truncation or censoring. When it is relevant, a decomposition of the difference in sample means of the dependent variable ($\bar{y}_M - \bar{y}_F$) the method adopted would have to be dissociated from $\bar{x}_F$ and $\bar{x}_M$, due to Jensen's inequality.

IV Examples of Taylor decompositions for nonlinear models

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

Decomposing augmented linear models such as the sample selection model has already been addressed by Neumann and Oaxaca (2001), Yun (2007) and Wolff (2012). Since the estimated equations are linear in structure, they can be treated in principle using the Taylor decomposition in the same way as for a linear or affine model (the estimation method corrects for the sample selection bias)[9]. 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 for these

[9] Although, as stressed by Neumann and Oaxaca (2004), this depends on how one interprets the selectivity term.

types of model using the Taylor decomposition. Hereafter, any parameter covered by a tilde (for example, $\tilde{\theta}$) is assumed to be an appropriate estimate of that parameter.

(a) Logit and Probit models

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

$$\mathrm{E}\left(y_{gi} \mid x_{gi}\right) = \mathrm{Prob}\left(y_{gi} = 1 \mid x_{gi}\right) = H\left(\bar{x}_{gi}^{T}\beta_{g}\right),\ g = M, F$$

where H is a cumulative distribution function. Since the latter is defined on a linear index, the vector of first derivatives has the following, straightforward form :

$$H'\left(\bar{x}_{g}^{T}\beta_{g}\right) = h\left(\bar{x}_{g}^{T}\beta_{g}\right)\beta_{g}$$

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

$$\begin{aligned}\tilde{y}_{M} - \tilde{y}_{F} &= \Phi\left(\bar{x}_{M}^{T}\tilde{\beta}_{M}\right) - \Phi\left(\bar{x}_{F}^{T}\tilde{\beta}_{F}\right) \\ &= \Phi\left(\bar{x}_{F}^{T}\tilde{\beta}_{M}\right) - \Phi\left(\bar{x}_{F}^{T}\tilde{\beta}_{F}\right) + k_{P}\tilde{\beta}_{M}{}^{T}[\bar{x}_{M} - \bar{x}_{F}] + R_{P}\end{aligned}$$

where $k_{P} = \phi\left(\bar{x}_{F}^{T}\tilde{\beta}_{M}\right)$ where $0 < k_{P} < 0.4$ and Φ and ϕ are the standard normal cumulative distribution and density functions, respectively. Note that k_{P} is fixed scalar in this decomposition, in the sense that each element in the vector $\widetilde{\beta}_M$ is multiplied by the same constant.

The function to be decomposed in the logit model has a closed form :

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

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

$$\begin{aligned} \widetilde{y}_M - \widetilde{y}_F &= L\left(\bar{x}_M^T \widetilde{\beta}_M\right) - L\left(\bar{x}_F^T \widetilde{\beta}_F\right) \\ &= L\left(\bar{x}_F^T \widetilde{\beta}_M\right) - L\left(\bar{x}_F^T \widetilde{\beta}_F\right) + k_L \widetilde{\beta}_M{}^T [\bar{x}_M - \bar{x}_F] + R_L \end{aligned}$$

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

$$k_L \widetilde{\beta}_M = L'\left(\bar{x}_F \widetilde{\beta}_M\right) = L\left(\bar{x}_F \widetilde{\beta}_M\right)[1 - L\left(\bar{x}_F \widetilde{\beta}_M\right)] \times \widetilde{\beta}_M \quad \text{where } 0 < k_L \leq 0.25 .$$

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 way[10] :

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

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

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

Note that this decomposition contains the sample means of the dependent variable but not the means of the right hand side variables, $\bar{x}_g$, and is thus subject to the consequences 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 $\bar{y} \neq \frac{1}{n}\sum_{i=1}^{n} \Phi\left(x_i^T \tilde{\beta}\right)$.

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

(b) Hazard functions and duration models

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

$$\lambda(t\,;x)=\frac{h(t\,;x)}{1-H(t\,;x)}=\frac{h(t\,;x)}{S(t\,;x)}$$

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. In order to obtain a two group Taylor decomposition of the difference in hazard functions, it would be necessary to specify the means of both the vector x and the spell length t. This will pose a problem in the majority of economic applications, since at least some spells are incomplete or censored and the mean spell (completed) length is not observed or easily calculated (see for example, Baker and Trivedi, 1993 and Bazen, Joutard and Niang, 2012).

A more straightforward, and arguably more interesting, approach would be to decompose differences in the expected duration[11] :

$$E\left(T\,|\,\bar{x}_g\right)=G\left(\bar{x}_g\right)$$

The survivor function, and thus the hazard function, is linked to the average completed spell duration through the following equality (see the Appendix for the derivation) :

[11] The effect of a variable on the duration of a completed spell is of opposite sign to its effect on the hazard rate.

$$E\left(T \mid \bar{x}\right) = \int_0^{\infty} t \, f\left(t \mid \bar{x}\right) dt \; = \int_0^{\infty} S\left(t \mid \bar{x}\right) dt$$

In what follows, the link between a parametric hazard specification and the corresponding formula for the expected duration is used to obtain Taylor 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 :

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

In this case, the expected duration of a completed spell for an individual with the mean value of x, is given by :

$$E\left(T \mid \bar{x}\right) = \Gamma\left(\frac{1+\alpha}{\alpha}\right) \exp\left(-\bar{x}^T \frac{\beta}{\alpha}\right)$$

$$\equiv \overline{\Gamma}\left(\alpha\right) \exp\left(\bar{x}^T \beta^*\right) \quad \text{where } \beta^* = -\frac{\beta}{\alpha}$$

The first term on the right hand side is the gamma function and is independent of both x and t. As noted in the definition of a Taylor decomposition, the basis is the difference between two estimated functions evaluated at the means of the explanatory variables, x. In the current case it is :

$$\widetilde{E}(T|\,\overline{x}_M) - \widetilde{E}(T|\,\overline{x}_F) = \overline{\Gamma}(\widetilde{\alpha}_M)\exp\left(\overline{x}_F^T\widetilde{\beta}_M^*\right) - \overline{\Gamma}(\widetilde{\alpha}_F)\exp\left(\overline{x}_F^T\widetilde{\beta}_F^*\right) + k_W\widetilde{\beta}_M^{*T}[\overline{x}_M - \overline{x}_F] + R_W$$

where the fixed scalar k_W is given by $k_W = \Gamma\left(\frac{1+\widetilde{\alpha}_M}{\widetilde{\alpha}_M}\right)\exp\left(\overline{x}_F^T\widetilde{\beta}_M^*\right) > 0$. 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 – duration dependence is either positive, $\alpha > 1$, or negative, $0 < \alpha < 1$, but cannot be one then the other during time spent in a given state. A hazard specification that permits increasing, followed by decreasing duration dependence, as well as the negative monotonic form is the log-logistic specification:

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

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

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

where $B(.)$ is the beta function. This expectation formula has a similar structure to the Weibull case, in that the first derivatives will all be defined in terms of a scalar multiplicative factor $k_{LL} = -\overline{B}(\widetilde{\gamma}_M)\exp\left(-\overline{x}_F^T\widetilde{\beta}_M\right) < 0$ which is a constant. The Taylor decomposition thus has a similar form :

$$\widetilde{E}\left(T|\,\overline{x}_M\right)-\widetilde{E}\left(T|\,\overline{x}_F\right)=\overline{B}\left(\widetilde{\gamma}_M\right)\exp\left(-\overline{x}_F^T\widetilde{\beta}_M\right)-\overline{B}\left(\widetilde{\gamma}_F\right)\exp\left(-\overline{x}_F^T\widetilde{\beta}_F\right)+k_{LL}\widetilde{\beta}_M^T[\overline{x}_M-\overline{x}_F]+R_{LL}$$

In all cases, the choice of method of estimation is independent of the decomposition. Thus for example the presence of censoring does not alter the form of the decomposition. Clearly the most appropriate procedure would be to use unbiased or consistent estimates of the parameters that define the conditional expectation.

V Two practical applications of the Taylor decomposition

Example 1: Male-female differences in unemployment in France

In contrast to most OECD countries, female unemployment in France has been historically higher than that of males, both in terms of the numbers and the proportion of the labour force (see for example, Azmat, Guell and Manning, 2006 and Bazen, Joutard and Niang, 2012). In 2005, the French Labour Force survey indicates that the respective unemployment rates of males and females aged between 20 and 54 were 10.2% and 12.7%. Using the same data source, the individual probability of being unemployed is modelled using a logit specification. The right hand side variables are: living in a couple, theoretical school leaving age corresponding to highest diploma (see example 2 for further information), current age, number of children and whether the person lives in large town or not. The data are for persons aged 20 to 54 and who are members of the labour force as defined by the ILO criteria. There are in fact few differences between the means of the variables for the two groups : females have slightly more education, are a little older, have fewer children, less likely to live in a couple and slightly more likely to live in a large town. These negligible differences mean that most of the gap in the unemployment rate will be due to the structural component rather than the composition effect in the Taylor decomposition.

The estimated parameters of the logit model are proportional to the marginal effects. Thus, more education, living in a couple and being older are factors which reduce the probability of being unemployed in France for both males and females. On the other hand, the presence of children and living in a large town tend to increase an individual's probability of being unemployed. As seen above, the Taylor decomposition for the logit model is given by :

$$L\left(\bar{x}_F^T \tilde{\beta}_F\right) - L\left(\bar{x}_M^T \tilde{\beta}_M\right) = L\left(\bar{x}_M^T \tilde{\beta}_F\right) - L\left(\bar{x}_M^T \tilde{\beta}_M\right) + k_L \tilde{\beta}_F^{\,T}[\bar{x}_F - \bar{x}_M] + R_L$$

where $k_L = L\left(\bar{x}_M^T \tilde{\beta}_F\right)\left(1 - L\left(\bar{x}_M^T \tilde{\beta}_F\right)\right)$ and $L(z) = \exp(z)/(1+\exp(z))$. The counterfactual is defined in terms of average male characteristics. The difference to be decomposed is 2.16% compared to a gross difference of 2.47% – a consequence of the nonlinearity of the logit function (and evidence of Jensen's inequality). The structural component (3.31%) is greater than the gap to be decomposed. This means that were a male with average characteristics to be treated in the same way as a female in the labour market, he would have a higher probability of being unemployed than at present and a higher probability than average female. As a consequence the composition effect is negative (-1.15): a female with average characteristics treated in the same way (the same logit function) as a male has a lower unemployment probability than that of an average male. The detailed first-order analysis of the composition effect indicates that education differences are a key factor. The remainder is a third of the composition effect.

Example 2: Duration of unemployment of young persons in France

As in many countries, youth unemployment in France is a major concern. The rate of unemployment among the under 25s is typically more than twice the average rate. In this example, the determinants of unemployment durations among young persons are examined. The data used come from the 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 employment, this lag acts as a signal to employers. The average education lag among those experiencing unemployment is two years (see Table 4). The duration of unemployment is therefore 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 spent unemployed in the four years following exit from the education system. For modelling purposes, we assume that this duration corresponds to a single spell rather than a series of shorter spells.

A second phenomenon often associated with unemployment among young persons is cultural and ethnic origin, and specifically whether the person has parents who are immigrants. Among those experiencing unemployment, 18% have parents who are not of French origin. The average duration of unemployment for the children of immigrants is two and half months longer than those with parents who are not immigrants – see Table 4. 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 and retention of young persons which favours those whose parents are not immigrants. We therefore use a Taylor decomposition to quantify the different components of the difference in unemployment durations between the two groups of young persons.

The Taylor decomposition uses a model-based estimate of the mean unemployment 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 :

$$\lambda(t) = \alpha t^{\alpha-1} \exp\left(x^T \beta\right)$$

and the corresponding expected duration is (see above):

$$E\left(T \mid x\right) = \Gamma\left(\frac{1+\alpha}{\alpha}\right)\exp\left(x^T \beta^*\right) = \overline{\Gamma}(\alpha)\exp\left(x^T \beta^*\right) \qquad \text{where } \beta^* = -\frac{\beta}{\alpha}$$

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 5. The estimated Weibull shape parameters indicate that there is positive duration dependence for both groups. The hazard function is increasing with duration but the probability of leaving unemployment at a given duration is higher for young persons with immigrant parents. A similar conclusion emerges from the other estimated coefficients : more education, shorter education lag and a smaller unemployment local unemployment reduce the duration of unemployment of children of immigrants compared to their French counterparts. There is 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 Taylor decomposition of the estimated expected duration of unemployment (in months and not logarithms) using the average French origin characteristics in the counterfactual is :

$$N(\bar{x}_N) - F(\bar{x}_F) = \bar{\Gamma}(\tilde{\alpha}_N)\exp\left(\bar{x}_F^T \tilde{\beta}_N^*\right) - \bar{\Gamma}(\tilde{\alpha}_F)\exp\left(\bar{x}_F^T \tilde{\beta}_F^*\right) \\ + k\tilde{\beta}_N^{*T}(\bar{x}_N - \bar{x}_F) + R$$

$$k = N(\bar{x}_F) = \bar{\Gamma}(\tilde{\alpha}_N)\exp\left(\bar{x}_F^T \tilde{\beta}_N^*\right)$$

where F (N) stands for individuals whose parents are (not) of French origin. 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 2.36 months (see Table 6) compared to the gross difference of 2.54 months (the divergence is evidence of Jensen's inequality). The structural component of this gap is 0.7 months – or 30%. The composition effect, the part due to differences in characteristics, therefore accounts for the majority of the gap. The first order part of the composition effect is 0.8 months and the remainder, which picks up the higher order parts due to the nonlinearity of the function, is 0.86 months – more half the overall composition effect. The results suggest that improving the educational performance (on both fronts) of children whose parents are immigrants will reduce the expected duration of unemployment and thereby the gap in durations between the two groups.

VI Conclusion

By recognizing that the Oaxaca technique can be obtained from a first order Taylor expansion, a new decomposition method applicable to nonlinear models is proposed. The Taylor decomposition provides a coherent, unified basis for decomposing differences in estimated functions (evaluated at their respective means) rather than differences in sample means. It provides a decomposition based on counterfactuals using the means of the right hand side variables on the basis of group differences in fitted values. In this decomposition the unexplained component or 'structure effect' is completely defined, and a first-order additive, detailed decomposition of the 'composition effect' is provided. When a nonlinear function is defined on a linear index – a common feature of many widely used nonlinear econometric models – the Taylor decomposition has a very attractive form. Explicit formulae for decompositions of the binary logit and probit models and duration models based on the Weibull and log-logistic hazard specifications are presented. The technique can be used with any smooth nonlinear function for a single left hand side variable such as the CES and Box-Cox. The proposed method – being based on a first order Taylor expansion of a nonlinear function – includes a non-zero remainder when a detailed decomposition is sought. This is a necessary feature of a first order Taylor expansion of a nonlinear function and, in terms of decompositions involving the mean, is a consequence of Jensen's inequality.

References

Andrews, George, Richard Askey and Ranjan Roy (2001) *Special Functions (Encylopedia of Mathematics and its Applications)*, Cambridge University Press, Cambridge.

Azmat, G., M. Guell, and A. Manning (2006): "Gender gaps in unemployment rates in OECD countries," *Journal of Labor Economics*, 24(1), 1–38.

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 (2012) The duration-based measurement of unemployment: estimation issues and an application to male-female unemployment differences in France, *Aix-Marseille School of Economics Discussion Paper* No 2012-6, Marseille.

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.

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.

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*, forthcoming.

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.

Table 1 Means and standard deviations of variables used			
	Female	Males	Difference
Unemployed	0.127 (0.33)	0.102 (0.302)	0.025
Education level (age at end of full-time education)	17.80 (2.45)	17.42 (2.41)	0.38
Couple	0.675 (0.47)	0.698 (0.46)	-0.023
Age	37.35 (9.74)	37.16 (9.70)	0.19
Children	0.81 (0.979)	0.84 (1.06)	-0.03
Lives in large town	0.791 (0.41)	0.781 (0.41)	0.01
Sample size	12,746	14,226	

Table 2 Logit model of the probability of being unemployed		
	Females	Males
Constant	3.079 (0.28)	1.148 (0.27)
Education level	-0.201 (0.013)	-0.144 (0.013)
Couple	-0.533 (0.058)	-0.992 (0.068)
Age	-0.044 (0.003)	-0.027 (0.003)
Children	0.154 (0.028)	0.087 (0.031)
Live in large town	0.343 (0.074)	0.769 (0.085)
Sample size	12,746	14,226
Estimated standard errors in parentheses		

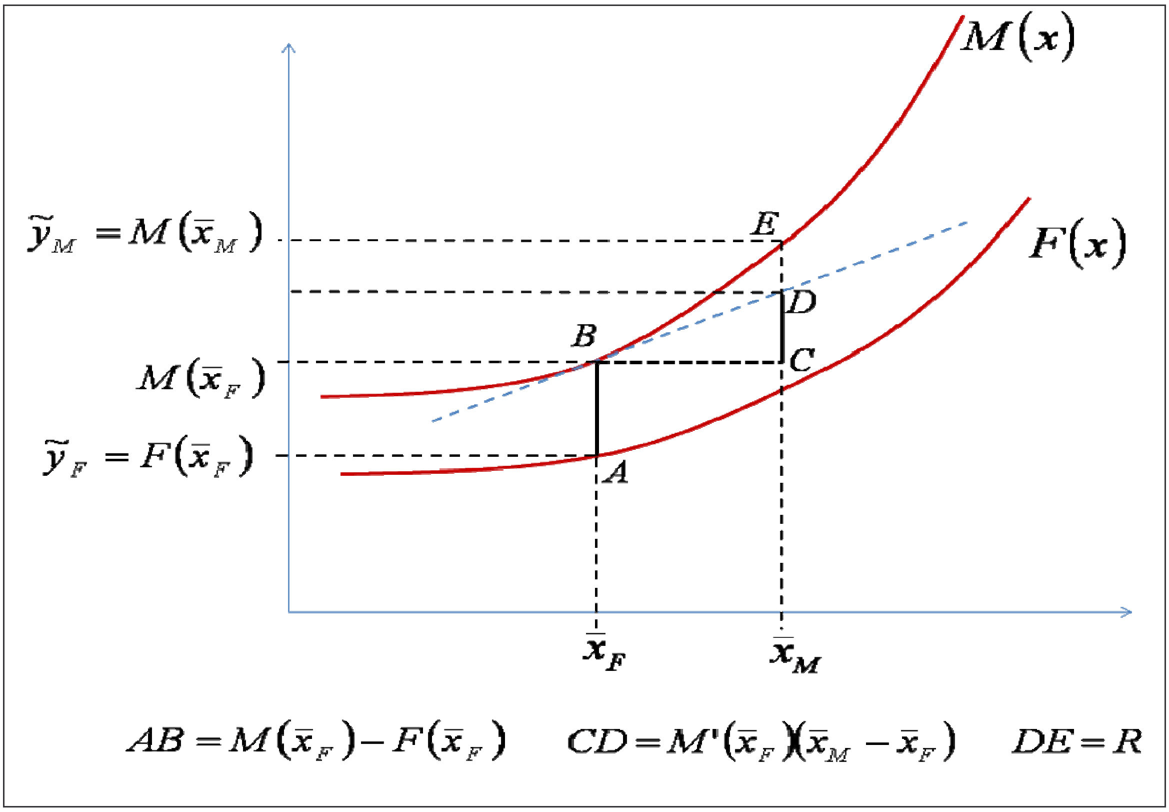
Table 3 Taylor decomposition of the gender difference in unemployment rates (in percentages)			
$L\left(\bar{x}_F^T \widetilde{\beta}_F\right)$	10.61		
$L\left(\bar{x}_M^T \widetilde{\beta}_M\right)$	8.45		
Difference to be decomposed :	2.16		
$L\left(\bar{x}_M^T \widetilde{\beta}_F\right)$	11.76		
Structural component	3.31		
Composition effect	– 1.15		
First order effects		-0.761	
education level			-0.790
lives in a couple			0.128
age			-0.087
children			-0.048
lives in large town			0.036
Remainder		-0.389	

Table 4 Means and standard deviations of variables used			
	Parents French	Parents Not French	Difference
Unemployment duration (months)	9.57 (8.56)	12.11 (9.83)	2.54
Education level (years)	13.05 (2.86)	12.30 (2.87)	-0.75
Education lag (years)	1.97 (1.71)	2.51 (1.72)	0.54
Local unemployment rate	8.78 (2.27)	9.24 (2.15)	0.36
Sample size	13,507	2,998	

Table 5 Accelerated Failure Time model estimates for unemployment durations			
	Parents French	Parents Not French	Difference
Constant	2.449 (0.045)	2.979 (0.092)	0.53
Education level	-0.045 (0.002)	-0.065 (0.005)	-0.02
Education lag	0.032 (0.004)	0.030 (0.009)	-0.002
Local unemployment rate	0.043 (0.003)	0.031 (0.007)	-0.012
Weibull parameter	1.184 (0.008)	1.265 (0.018)	0.081
Sample size	13,507	2,998	
Estimated standard errors in parentheses			

Table 6 Taylor decomposition of group difference in unemployment duration (in months)	
$N(\bar{x}_N) = \bar{\Gamma}(\tilde{\alpha}_N)\exp(\bar{x}_N^T \tilde{\beta}_N^*)$	11.793
$F(\bar{x}_F) = \bar{\Gamma}(\tilde{\alpha}_F)\exp(\bar{x}_F^T \tilde{\beta}_F^*)$	9.437
Difference to be decomposed :	2.356
$N(\bar{x}_F) = \bar{\Gamma}(\tilde{\alpha}_N)\exp(\bar{x}_F^T \tilde{\beta}_N^*) = k$	10.130
Structural component	0.693
Composition effect	1.663
First order effects	0.803
education level	0.496
education lag	0.162
local unemployment rate	0.145
Remainder	0.860

Figure 1 Graphical representation of the Taylor decomposition for a nonlinear function



Appendix The expected duration of completed spell is equal to the integral of the survival function.

The expected value of a (non-negative) random variable over its whole support $[\,0, Z\,]$ is defined as

$$E(T) = \int_0^Z t\ f(t)\,dt$$

Integrating by parts yields :

$$E(T) = Z - \int_0^Z F(t)\,dt$$

The integral of the survivor function is equal to right-hand side of this expression. From the definition of the survivor function

$$S(t) = 1 - \mathrm{F}(t)$$

The integral of the survivor function is then :

$$\int_0^Z S(t)\,dt = \int_0^Z (1 - \mathrm{F}(t))\,dt$$

$$= [\,t\,]_0^Z - \int_0^Z \mathrm{F}(t)\,dt = Z - \int_0^Z \mathrm{F}(t)\,dt = E(T)$$