

IZA DP No. 6730

Additive Decompositions with Interaction Effects

Martin Biewen

July 2012

Additive Decompositions with Interaction Effects

Martin Biewen

*University of Tübingen,
IZA and DIW Berlin*

Discussion Paper No. 6730
July 2012

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 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

Additive Decompositions with Interaction Effects¹

This paper proposes a comprehensive, path-independent decomposition formula of changes into *ceteris paribus* effects and interaction effects. The formula implies a reassessment of sequential decomposition methods that are widely used in the literature and that are restrictive in how they treat interaction effects. If counterfactual outcomes are correctly specified, it may also be viewed as a description of certain aspects of causality in the situation where more than one causal influence is present.

JEL Classification: J31, J71, C02, D31

Keywords: *ceteris paribus* effects, interaction effects, Oaxaca-Blinder decomposition

Corresponding author:

Martin Biewen
Department of Economics
University of Tübingen
Mohlstr. 36
72074 Tübingen
Germany
E-mail: martin.biewen@uni-tuebingen.de

¹ I am grateful to Andos Juhasz for helpful comments and research assistance. The data used in this paper (SOEP v27, 1998-2010) were made available by the German Socio-Economic Panel Study (SOEP) at the German Institute for Economic Research (DIW), Berlin.

1 Introduction

If an outcome can only be achieved through the joint effort of two persons, it would be nonsensical to ask what percentage each of the two persons has contributed to this outcome. Yet, this is how decomposition analyses in economics are usually carried out. Typically, the aim of such decompositions is to split up a given total effect into disjunct parts that can be attributed to different factors. As indicated above, this may be an unattainable goal in many cases. The interaction effects between different factors do not have to be as strong as in the example given above, but in most real world contexts they are likely to be present. For example, the output produced by two individuals may be higher than the sum of the output that can be produced by each of the two individually (positive interaction). It may also be the case that two factors fully or partly offset each other when they are jointly present as compared to when they are present in isolation (negative interaction). In the case of more than two factors, the situation is even more complex involving possible interaction effects of any subset of the factors considered.

Typical examples for decomposition methods used in economics include the decomposition of distributional change into different factors (Juhn et al. (1993), DiNardo et al. (1996)), and the decomposition of differences between groups into 'characteristics' and 'returns' effects (Blinder (1973), Oaxaca (1973)). For example, in Juhn et al. (1993), the change in US wage inequality is decomposed into the effects due to changes in observable characteristics, changes in observable prices, and changes in unobservable characteristics and/or unobservable prices. In the seminal contribution by DiNardo et al. (1996), changes in the US wage distribution are decomposed into the parts contributed by changes in individual attributes, changes in unionization, changes in the minimum wage, changes in supply and demand, and a contribution due to other changes in the conditional wage structure. In the classic Oaxaca-Blinder decomposition, the differences between two groups, for example black/white or male/female average wage differentials, are decomposed into a part due to differences in characteristics, and a part due to differences in returns to characteristics (also interpreted as 'discrimination').

The common method used in these approaches in order to arrive at an exact decomposition of an overall effect into the parts contributed by different factors is to sequentially add the changes of the different factors until all factors have been accounted for. The incremental changes defined in this way provide an exact additive decomposition of the overall effect into parts contributed by each factor. This widely used method has two drawbacks. The first one is that the result

of the decomposition may depend on the order in which the different factors are added. The second one is that the sequential decomposition provides an answer to the question of splitting up the overall effect into disjunct parts even if there is no such answer. For example, in the case described at the beginning of this section, the individual contribution of each of the two persons is zero, but together they contribute a hundred percent. There is no way of splitting up the overall effect into individual contributions. As another example, consider the case where (among other things) the contributions of changes in unionization and shifts in the industry structure to changes in the wage distribution are of interest. It will not be possible to completely separate the influence of changes in unionization from those in the industry structure because part of the effect of the changes in unionization will only materialize because the industry structure shifts towards or away from industries with high levels of unionization.² Similarly, in the classic Blinder-Oaxaca decomposition which asks why male average wages exceed those of females, part of the effect of higher wage returns for men only materialize because men may also have more favorable characteristics. Again, it will not be possible to separate the effects of 'characteristics' and 'returns' in a strict sense.

It is not that the authors of the studies cited have not recognized the limitations of the kind of sequential decompositions described above that are imposed by the existence of interaction effects (see, e.g., Juhn et al., p. 429). However, to our best knowledge, no attempt has been made to explicitly address these limitations. The aim of this paper is therefore to propose an alternative decomposition scheme that i) takes seriously the existence of interaction effects and therefore does not try to separate the influence of different factors where this is not possible, ii) provides an exact decomposition of an overall effect into different contributions, and iii) is independent of the ordering of the factors under consideration. The proposed alternative decomposition scheme is generally applicable and can be viewed as a formal description of some of the aspects of the notion that an overall effect is additively decomposed into the contributions of different factors and their interactions.

The rest of the paper is organized as follows. Section 2 introduces the alternative decomposition scheme involving interaction effects. Section 3 revisits sequential decomposition schemes and examines their relationship to the decomposition involving interaction effects. In section 4, the

²This example is taken from Fortin et al. (2011) who provide a comprehensive treatment of the various decomposition methods used in the literature including a discussion of many of the issues considered here. The focus of Fortin et al. (2011) is on aspects such as identification which are largely independent of the point made here (see below).

framework is applied to the classic Oaxaca-Blinder decomposition and to an example involving three-way interactions in order to highlight some of implications introduced by the existence of interaction effects. Section 5 presents two empirical illustrations, and section 6 concludes.

2 Additive decompositions with interaction effects

As indicated above, the goal of many analyses in economics (and science in general) is to decompose the difference between two objects into the parts contributed by different factors. For example, it may be investigated to what extent male wages are higher than female wages because men have more favorable characteristics than women, or because returns to given characteristics are higher for men than for women. As another typical example, consider the question of decomposing the change in the distribution of income over time into the contributions of different factors. This section provides a general additive decomposition formula for such changes.

2.1 The two factor case

First, consider the case where the change in an object f is thought to be caused by the change of two factors. The overall change in the object can be written as $f_{11} - f_{00}$ where f_{11} is the outcome that results if both factors are changed, while f_{00} denotes the outcome that results if both factors remain in their original state. The object f may be any outcome of interest, for example, a wage, a distribution of wages, a functional of a distribution such as an inequality index, or any other object of interest. In order to decompose the overall change into changes contributed by individual factors, one has to introduce counterfactual outcomes f_{10} , f_{01} which describe what the outcome would be if only one of the factors was changed in isolation.

For example, in the case of the Oaxaca-Blinder decomposition, f is an average wage, f_{11} is the average wage of men, f_{10} the average wage for individuals who have mens' characteristics but womens' returns to characteristics, f_{01} the average wage for individuals with womens' characteristics but with mens' returns to characteristics, and f_{00} the average wage of women (i.e. individuals with womens' characteristics and womens' returns to characteristics). In the case of decomposing changes of the income distribution over time, f_{11} and f_{00} denote the income distributions of periods 1 and 0, f_{10} is the hypothetical income distribution that would result if factor one was changed to its period 1 level but factor two was kept at its period 0 level, and f_{01}

is the hypothetical distribution that would result if factor one was kept at its period 0 level but factor two was changed to its period 1 level.

In such a situation, the difference $f_{11} - f_{00}$ may be decomposed as

$$f_{11} - f_{00} = (f_{10} - f_{00}) \quad (1)$$

$$+ (f_{01} - f_{00}) \quad (2)$$

$$+ [(f_{11} - f_{00}) - (f_{10} - f_{00}) - (f_{01} - f_{00})]. \quad (3)$$

Here, (1) represents the ceteris paribus effect of factor one, (2) the ceteris paribus effect of factor two, and (3) the interaction effect between the two factors. The ceteris paribus effects describe the effects that occur if each of the factors is changed separately. If the two separate changes do not add up to the overall change, this necessarily implies that there is an interaction effect. The interaction effect is the part of the overall change that cannot be explained by changing both of the factors in isolation, i.e. what remains if both ceteris paribus effects are subtracted from the overall change.

The interaction effect has other, more intuitive interpretations:

$$[(f_{11} - f_{00}) - (f_{10} - f_{00}) - (f_{01} - f_{00})] \quad (4)$$

$$= (f_{11} - f_{01}) - (f_{10} - f_{00}) \quad (5)$$

$$= (f_{11} - f_{10}) - (f_{01} - f_{00}) \quad (6)$$

It can be seen that the interaction effect is equal to both the effect of factor one varied by whether or not factor two is present (5), and to the effect of factor two varied by whether or not factor one is present (6).

2.2 The three factor case

In the three factor case, the decomposition is given by

$$f_{111} - f_{000} = (f_{100} - f_{000}) \quad (7)$$

$$+ (f_{010} - f_{000}) \quad (8)$$

$$+ (f_{001} - f_{000}) \quad (9)$$

$$+ [(f_{110} - f_{000}) - (f_{100} - f_{000}) - (f_{010} - f_{000})] \quad (10)$$

$$+ [(f_{101} - f_{000}) - (f_{100} - f_{000}) - (f_{001} - f_{000})] \quad (11)$$

$$+ [(f_{011} - f_{000}) - (f_{010} - f_{000}) - (f_{001} - f_{000})] \quad (12)$$

$$+ [(f_{111} - f_{000}) - (7) - (8) - (9) - (10) - (11) - (12)]. \quad (13)$$

Again, (7) - (9) represent the ceteris paribus effects. Contribution (10) is the two-way interaction effect between factor one and factor two. Contributions (11), (12) are the corresponding two-way interaction effects between factors one and three, and between factors two and three, respectively. Everything that cannot be accounted for by the ceteris paribus and the two-way interaction effects has to be due to a three-way interaction effect between all the three factors. The three-way interaction effect is therefore given by (13).

Note that the three-way interaction effect int^3 is defined as

$$int^3 = total_3 - \sum_{i=1}^3 cp_i - \sum_{k \in P_2} int_k^2, \quad (14)$$

where $total_3$ denotes the total change when all the three factors are changed, cp_i the ceteris paribus effect of factor i , and int_k^2 the two-way interaction effects (over the set P_2 of all possible combinations of two factors out of the three).

2.3 The m factor case

The decomposition presented in the previous section easily generalizes to an arbitrary number of factors m . For this, note that in the case of $m = 4$ factors, the four-way interaction effect results as

$$int^4 = total_4 - \sum_{i=1}^4 cp_i - \sum_{k \in P_2} int_k^2 - \sum_{k \in P_3} int_k^3, \quad (15)$$

where the last term stands for all possible three-way interaction effects between three factors chosen out of the four.

In the general case,

$$int^j = total_j - \sum_{i=1}^j cp_i - \sum_{k \in P_2} int_k^2 - \sum_{k \in P_3} int_k^3 - \dots - \sum_{k \in P_{j-1}} int_k^{j-1}, \quad (16)$$

which implies for $j = m$

$$int^m = total_m - \sum_{i=1}^m cp_i - \sum_{k \in P_2} int_k^2 - \sum_{k \in P_3} int_k^3 - \dots - \sum_{k \in P_{m-1}} int_k^{m-1}. \quad (17)$$

This yields the general decomposition for m factors

$$total_m = \sum_{i=1}^m cp_i + \sum_{k \in P_2} int_k^2 + \sum_{k \in P_3} int_k^3 + \dots + \sum_{k \in P_{m-1}} int_k^{m-1} + int^m. \quad (18)$$

Decomposition (18) has number of advantages over other decomposition schemes used in the literature such as the sequential decompositions described in the next section. The first advantage is that the decomposition is *path-independent* because all factors are treated symmetrically. In general, a decomposition is called path-independent if the contribution of a factor does not depend on the order in which the different factors are introduced. A second advantage is that the decomposition not only contains the ceteris paribus effects of changing one factor in isolation but also the ceteris paribus effects of changing any subset of factors at the same time. For example, the ceteris paribus effect of changing factors 1 and 2 at the same time is given by $cp_1 + cp_2 + int_{\{1,2\}}^2$. In general, the ceteris paribus effect of changing a subset S of factors at the same time is given by the sum of all individual ceteris paribus effects and all possible interaction effects between the factors in S . This also means that the decompositions of the joint influence of a subset of factors appear as a part of the larger decomposition involving all factors (this property may be called *subdecomposability*). A third advantage of decomposition (18) is that it is *comprehensive* in the sense that it involves all counterfactual states of the world. This is not true, for example, of the sequential decompositions discussed below which are insensitive to changes in counterfactuals that do not appear in the decomposition.³

3 Sequential decompositions

This section reviews the sequential decomposition schemes that are widely used in the literature and establishes the relationship between sequential decompositions and the interaction effects defined above.

³The property of comprehensiveness is proved as follows. A decomposition with two factors involves all possible counterfactuals for these two factors (see equations (1) to (3)). The decomposition involving three factors implicitly contains all possible two factor decompositions (i.e. all possible counterfactuals for any pair of factors) and adds as a last step the state of the world in which all the three factors are present (see equation (13)). The decomposition involving four factors implicitly contains all lower order decompositions and adds as a last step the state of the world in which all four factors are present, and so on.

3.1 The two factor case

In the two factor case, a possible sequential decomposition of the total change is

$$f_{11} - f_{00} = (f_{10} - f_{00}) \quad (19)$$

$$+ (f_{11} - f_{10}). \quad (20)$$

Here, (19) measures the contribution of factor one, while (20) measures that of factor two. A severe drawback of sequential decompositions is that they are path-dependent. The result of the decomposition depends on the order in which the two factors are introduced. The sequential decomposition in which the two factors are treated in the reverse order is given by

$$f_{11} - f_{00} = (f_{01} - f_{00}) \quad (21)$$

$$+ (f_{11} - f_{01}). \quad (22)$$

Here, the contribution of factor two (21) is measured first, while that of factor one (22) is measured second.

3.2 The three factor case

In the three factor case, a possible sequential decomposition is

$$f_{111} - f_{000} = (f_{100} - f_{000}) \quad (23)$$

$$+ (f_{110} - f_{100}) \quad (24)$$

$$+ (f_{111} - f_{110}). \quad (25)$$

An alternative sequential decomposition is given by

$$f_{111} - f_{000} = (f_{001} - f_{000}) \quad (26)$$

$$+ (f_{101} - f_{001}) \quad (27)$$

$$+ (f_{111} - f_{101}). \quad (28)$$

The latter decomposition first measures the contribution of factor three, then that of factor one, and finally, that of factor two.

In the three factor case, there are $3 \cdot 2 \cdot 1 = 6$ possible sequential decompositions.

3.3 The m factor case

In the m factor case, the simplest sequential decomposition is defined by

$$f_{11111\dots 1} - f_{00000\dots 0} = (f_{10000\dots 0} - f_{00000\dots 0}) \quad (29)$$

$$+ (f_{11000\dots 0} - f_{10000\dots 0}) \quad (30)$$

$$+ (f_{11100\dots 0} - f_{11000\dots 0}) \quad (31)$$

$$+ (f_{11110\dots 0} - f_{11100\dots 0}) \quad (32)$$

$$+ \dots \quad (33)$$

$$+ (f_{11111\dots 1} - f_{11111\dots 0}). \quad (34)$$

In the m factor case, there are $m \cdot (m - 1) \cdot \dots \cdot 2 \cdot 1 = m!$ possible sequential decompositions.

3.4 Sequential decompositions and interaction effects

The following two propositions show that sequential decomposition schemes are very restrictive in how they treat interaction effects.

Proposition 1. *The contributions of a sequential decomposition scheme are path-independent if and only if they are equal to the ceteris paribus effects for all possible decomposition orders.*

Proof. If the contributions of the sequential decomposition scheme are path-independent, then they are equal to the ceteris paribus effects because there is always a sequential decomposition in which a given factor appears first, implying that its contribution is equal to the ceteris paribus effect. On the other hand, if the contributions of a sequential decomposition scheme are equal to the ceteris paribus effects for all possible decomposition orders, then they are independent of the decomposition order and thus path-independent. \square

Proposition 2. *The contributions of a sequential decomposition scheme are path-independent if and only if all interaction effects are zero.*

Proof. If the contributions of a sequential decomposition scheme are path-independent, then they are equal to the ceteris paribus effects (see Proposition 1). Then, all two-way interactions have to be zero because, if the decomposition is path-independent, it does not make a difference

whether a given factor appears in the first or the second position of the corresponding sequential decomposition, i.e.

$$f_{1,1,(0,\dots,0)} - f_{0,1,(0,\dots,0)} = f_{1,0,(0,\dots,0)} - f_{0,0,(0,\dots,0)} \quad (35)$$

(here the first position of the subscript refers to the given factor, the second position to another factor, and the rest to all remaining factors). Equation (35) means that the two-way interactions between the given factor and any other factor are equal to zero (see (5)). This holds for all factors and all two-way interactions. Then, for any three-way interaction

$$int^3 = total_3 - \sum_{i=1}^3 cp_i - \sum_{k \in P_2} int_k^2 = total_3 - \sum_{i=1}^3 cp_i = 0 \quad (36)$$

because, if the sequential decomposition of the m factors is path-independent, also the sequential decomposition involving only the three factors under consideration is path-independent, i.e. $total_3 = \sum_{i=1}^3 cp_i$ (in any path-independent sequential decomposition with three factors, the contributions are equal to the ceteris paribus effects, see Proposition 1). Using this argument recursively,

$$int^j = total_j - \sum_{i=1}^j cp_i - \sum_{k \in P_2} int_k^2 - \sum_{k \in P_3} int_k^3 - \dots - \sum_{k \in P_{j-1}} int_k^{j-1} = 0 \quad (37)$$

for all remaining $j = 4, \dots, m$ (because all preceding $(j-1)$ -way interactions are zero and $total_j = \sum_{i=1}^j cp_i$ because of path-independence). This establishes that if the sequential decomposition scheme is path-independent, all interaction effects have to be zero.

On the other hand, if all the interaction effects are zero, then for any number of factors $j \leq m$

$$total_j = \sum_{i=1}^j cp_i + interactions = \sum_{i=1}^j cp_i. \quad (38)$$

Consider the case of m factors and take any sequential decomposition with order O . We will show that the contributions in this decomposition are equal to the ceteris paribus effects of the corresponding factors. The contribution of the first factor in O is equal to the ceteris paribus effect of this factor by definition. Now consider the second factor. Its contribution in the sequential decomposition is $f_{1,1,(0,\dots,0)} - f_{1,0,(0,\dots,0)}$ (the first subscript refers to the first factor in O , the second to the second factor, and the rest to all other factors). But this contribution is equal to $f_{0,1,(0,\dots,0)} - f_{0,0,(0,\dots,0)} = cp_2$, i.e. the ceteris paribus effect of factor two because there are no two-way interactions (see (6)). Now consider again sequential decomposition O but only the sequential sub-decomposition in O that involves the first three steps. Because all

interaction effects are zero, in this three factor sequential decomposition it also holds that $total_3$ is equal to the sum of the ceteris paribus effects of the the three factors, i.e. $total_3 = \sum_{i=1}^3 cp_i$ (see (38)). This means that in sequential decomposition O (which shares the first three steps with the sub-decomposition), the contribution of the third factor is equal to cp_3 (because the contributions of the first two factors were cp_1 and cp_2). The same argument applies recursively to the contributions of factors 4, 5, \dots , m in O . This means that if all interaction effects are zero, in any sequential decomposition O the contribution of each factor is equal to its ceteris paribus effect. This is equivalent to the sequential decomposition scheme being path-independent according to Proposition 1. \square

Note that Proposition 2 is a rather strong statement as it requires that *all* interaction effects (i.e. also the ones of a higher order) have to be zero for path independence. A possible conclusion from Proposition 2 is that the larger interaction effects are, the more path-dependent the result of a sequential decomposition will be. This justifies the practice of trying out different factor orderings when computing sequential decompositions, see, for example, DiNardo et al. (1996) and Hyslop/Mare (2005). Examples where the contribution of individual factors may be quite dependent on the order in which the different factors are introduced in a sequential decomposition can be found in Biewen/Juhasz (2012) and Biewen (2001). Propositions 1 and 2 also make clear that computing a sequential decomposition is not better than computing ceteris paribus effects. If the results of the sequential decomposition are not identical (or similar to) the ceteris paribus effects, the sequential decomposition is necessarily path-dependent and therefore potentially questionable.

4 Two examples

The purpose of this section is to apply the above framework to two examples in order to highlight some of the issues discussed above.

4.1 The Oaxaca-Blinder decomposition

Starting with the seminal papers by Oaxaca (1973) and Blinder (1973), economists have been asking the question of how to decompose differences between groups or over time into a 'char-

acteristics' and a 'returns' effects.⁴ Oaxaca and Blinder's decomposition has been applied and generalized to a variety of different settings (see Gomoulka/Stern (1990), Fairlie (2005), Yun (2004), Machado/Mata (2005), Biewen/Jenkins (2005), Bauer/Sinnig (2007), among many others). The following remarks also apply to these variants of their idea.

In the Oaxaca-Blinder setup,

$$f_{11} = \bar{x}_M \beta_M, \quad f_{10} = \bar{x}_M \beta_F, \quad f_{01} = \bar{x}_F \beta_M, \quad f_{00} = \bar{x}_F \beta_F, \quad (39)$$

where \bar{x}_M, \bar{x}_F denote the vector of average characteristics of men and women, and β_M, β_F the regression coefficients of male and female wage regressions.

There are two standard variants of the Oaxaca-Blinder decomposition which correspond to the two possible sequential decompositions shown in (19) to (22):

$$\begin{aligned} f_{11} - f_{00} &= (f_{10} - f_{00}) + (f_{11} - f_{10}) \\ &= \bar{x}_M \beta_M - \bar{x}_F \beta_F = (\bar{x}_M - \bar{x}_F) \beta_F + \bar{x}_M (\beta_M - \beta_F) \end{aligned} \quad (40)$$

$$\begin{aligned} f_{11} - f_{00} &= (f_{01} - f_{00}) + (f_{11} - f_{01}) \\ &= \bar{x}_M \beta_M - \bar{x}_F \beta_F = \bar{x}_F (\beta_M - \beta_F) + (\bar{x}_M - \bar{x}_F) \beta_M \end{aligned} \quad (41)$$

In the original context considered by Oaxaca and Blinder, the term involving the differences in coefficients was attributed to 'discrimination'.

The corresponding decomposition with interaction effect (equations (1) to (3)) is given by

$$\begin{aligned} f_{11} - f_{00} &= (f_{10} - f_{00}) + (f_{01} - f_{00}) + [(f_{11} - f_{00}) - (f_{10} - f_{00}) - (f_{01} - f_{00})] \\ &= (\bar{x}_M - \bar{x}_F) \beta_F + \bar{x}_F (\beta_M - \beta_F) + (\bar{x}_M - \bar{x}_F) (\beta_M - \beta_F) \end{aligned} \quad (42)$$

$$= \Delta \bar{x} \beta_F + \bar{x}_F \Delta \beta + \Delta \bar{x} \Delta \beta. \quad (43)$$

Decomposition (42) was introduced at around the same time as the Oaxaca-Blinder decomposition by Winsborough/Dickenson (1971). Surprisingly, this decomposition involving the interaction effect of differences in characteristics and differences in coefficients is rarely used in economics.⁵ The validity of this decomposition is obvious from figure 1.

— Figure 1 about here —

⁴See Fortin et al. (2011) for a comprehensive overview.

⁵An exception is Daymont/Andrisani (1984).

Decomposition (42) asks why area $ACGI = \bar{x}_M\beta_M$ is larger than area $DEGH = \bar{x}_F\beta_F$. It is easily seen that the difference between the two is composed of $ABDE = \Delta\bar{x}\beta_F$, $EFHI = \bar{x}_F\Delta\beta$, and $BCEF = \Delta\bar{x}\Delta\beta$. Why do women have lower average wages than men? One part of the difference is due to their less favorable characteristics ($= \Delta\bar{x}\beta_F$), another one due to their lower returns ($= \bar{x}_F\Delta\beta$), and a third part can only be explained by both of these factors together ($= \Delta\bar{x}\Delta\beta$). This latter part would be zero if either $\Delta\bar{x} = 0$ or $\Delta\beta = 0$. It therefore appears questionable to assign this interaction term to either the 'characteristics' or the 'returns' effect as it is done in sequential decompositions (40) and (41) (see Jones/Kelley (1984)). In (40), the interaction term is assigned to the 'returns' effect (compare (40) with (42)), while in (41), it is assigned to the 'characteristics' effect (compare (41) with (42)).

Jones/Kelley (1984) have pointed out that, in the context of measuring discrimination, using (40) to calculate the 'discrimination' (= 'returns') effect corresponds to the thought experiment in which discrimination is ended by raising women's returns to those of men, while (41) corresponds to the reverse experiment in which discrimination is ended by reducing men's returns to those of women. Both cases are extreme and it appears hard to justify the choice of either of them. There is also an intermediate form of the Oaxaca-Blinder decomposition

$$\bar{x}_M\beta_M - \bar{x}_F\beta_F = (\bar{x}_M - \bar{x}_F)\beta^* + [\bar{x}_M(\beta_M - \beta^*) + \bar{x}_F(\beta^* - \beta_F)] \quad (44)$$

(see, e.g., Jann (2008), eq. (6)) in which men's and women's returns are compared to 'appropriate' returns β^* . This however seems to confuse the question of discrimination with that of appropriate remuneration. Discrimination appears to be exclusively about making a difference, no matter what the appropriate level of remuneration is. For example, it could be that both men and women receive less than their appropriate remuneration (in (44), this would in many cases lead to an implausible negative discrimination term), still there would be discrimination if, for example, men were less underpaid than women. Taken together, it is hard to find reasons to allocate the interaction effect either in whole or in part to either the 'characteristics' or the 'returns' effect. Instead, it seems to make more sense to report it separately as the part of the difference that only arises if both factors change together.

Figure 1 also shows that the complications created by the interaction effect are the smaller, the smaller the differences $\Delta\bar{x}$ and $\Delta\beta$ are. This is the reason why in infinitesimal settings (where $\Delta\bar{x}, \Delta\beta \rightarrow 0$), interaction effects are small of a higher order and therefore vanish (with the consequence that the decomposition is only valid locally). This is true, for example, of decompositions such as those underlying the growth accounting approach (Solow (1957)). In a

more discrete setting in which changes are large in relation to levels, interaction effects may also become relevant for these kind of decompositions.

Note that decomposition (43) can also be seen as a general way to decompose changes in aggregate value (= prices \times quantities) into price changes, quantity changes and an interaction term involving both price and quantity changes. In this context, it also does not make sense to ask what percentage of the overall change is due to price and what percentage due to quantity changes because there is a third component that cannot be exclusively assigned to either price or quantity changes. This argument is relevant to the concept of constructing price indices.

Finally, note that the decision to use one particular form of decompositions (40), (41) and (42) is unrelated to the questions involved in identifying the components in (39). For example, Fortin et al. (2011) provide an in-depth analysis of the conditions needed to identify the components in decomposition (40) but their arguments can easily be used to identify the components of decomposition (42). This can be seen by writing the interaction effect as $(\bar{x}_M - \bar{x}_F)(\beta_M - \beta_F) = \bar{x}_M(\beta_M - \beta_F) - \bar{x}_F(\beta_M - \beta_F)$ (which involves similar terms as in (40) or (41)).

4.2 An example with three-way interactions

The aim of this section is to provide further motivation for the general decomposition formula (18) and to discuss some further issues. As a further motivation for decomposition (18), look at figure 2. Why is block $x_1y_1z_1$ larger than block $x_0y_0z_0$? The difference between the two blocks can be written as

$$x_1y_1z_1 - x_0y_0z_0 = [(x_1y_0z_0 - x_0y_0z_0)] \quad (45)$$

$$+ [(x_0y_1z_0 - x_0y_0z_0)] \quad (46)$$

$$+ [(x_0y_0z_1 - x_0y_0z_0)] \quad (47)$$

$$+ [(x_1y_1z_0 - x_0y_0z_0) - (x_1y_0z_0 - x_0y_0z_0) - (x_0y_1z_0 - x_0y_0z_0)] \quad (48)$$

$$+ [(x_1y_0z_1 - x_0y_0z_0) - (x_1y_0z_0 - x_0y_0z_0) - (x_0y_0z_1 - x_0y_0z_0)] \quad (49)$$

$$+ [(x_0y_1z_1 - x_0y_0z_0) - (x_0y_1z_0 - x_0y_0z_0) - (x_0y_0z_1 - x_0y_0z_0)] \quad (50)$$

$$+ [(x_1y_1z_1 - x_0y_0z_0) - (45) - (46) - (47) - (48) - (49) - (50)], \quad (51)$$

which is just decomposition (7) to (13). Figure 2 nicely illustrates that the difference between blocks $x_1y_1z_1$ and $x_0y_0z_0$ is composed of a number of smaller blocks that represent the different two-way and the three-way interaction effects. It also illustrates that any sequential decomposition

would involve a probably arbitrary assignment of these interaction effects to one of the ceteris paribus effects $(x_1y_0z_0 - x_0y_0z_0)$, $(x_0y_1z_0 - x_0y_0z_0)$, and $(x_0y_0z_1 - x_0y_0z_0)$.

— Figure 2 about here —

Note that the decomposition formula (18) is more general than the example given in figure 2. Formula (18) applies to any mechanism generating counterfactual outcomes. The specific mechanism xyz is restrictive in the sense that necessarily all two-way and three-way interactions have to be present (provided that $\Delta x, \Delta y, \Delta z \neq 0$). In a general mechanism, any kind of interaction effect could be present or absent. For example, there could be three-way interactions but no two-way interactions, it could be a mechanism without interaction effects at all, or one with negative interaction effects (see the examples given in the introduction).

5 Empirical illustrations

5.1 Oaxaca-Blinder decomposition

In the first illustration, we decompose differences between wages paid in the services sector and those paid in other sectors of the economy into a characteristics, a coefficient, and an interaction effect (see decomposition (42)). The data for this exercise (and also for the one in the next section) are taken from the German Socio-Economic Panel (GSOEP). The Oaxaca-Blinder type decomposition considered is based on a standard wage regression explaining log hourly wages by years of education, experience, experience squared, tenure, and a female dummy. According to the results shown in table 1, the difference of mean log wages between the services sector and other sectors of the economy ($= 0.1454$) is accounted for by differences in the endowment with wage relevant characteristics (resulting in a contribution of 0.0834), by differences in returns to characteristics (contributing 0.0830 of the difference), and a negative interaction effect of characteristics and returns (contributing -0.211 of the difference).

— Tables 1 and 2 about here —

A closer look at the more detailed results in table 2 reveals that the negative interaction effect is driven by the fact that female workers earn less in the services sector but their share there is

higher than in the other sectors, and by the fact that the return to tenure is higher in the services sector but the average tenure there is lower than in other sectors. As argued above it would be wrong to attribute these effects to either the characteristics or the returns effect as they only materialize because *both coefficients and characteristics* differ between the sectors.

5.2 Decomposition of distributional change

The next example considers three factors explaining possible changes in the distribution of equivalized incomes in Germany between 1999/2000 (= period 0) and 2005/2006 (= period 1). The example focuses on changes due to shifts in the labor market returns to household characteristics (= factor 1), changes in the tax system (= factor 2), and other changes (= factor 3). The dependent variable of the analysis is personal equivalized disposable income, i.e. household income from all sources minus taxes and social security contributions, divided by an equivalence scale in order to arrive at a measure of personal income for each household member.⁶

Changes in labor market returns are modeled by regressions of (log) household labor market income y_{lab} on household characteristics z which include information on household employment outcomes and the composition of the household with respect to variables such as age, gender, educational qualifications, disability status, marital status, region, and nationality.⁷ All regressions are carried out separately for six different household types (single and multi-adult pensioner households, single and multi-adult households with or without children). From the perspective of period 0, the expected change in household labor income that results if labor market returns are counterfactually set to their period 1 level but household characteristics are kept at their period 0 level is given by

$$\widehat{\Delta}y_{lab}^{01} = z_0' \hat{\beta}_1 - z_0' \hat{\beta}_0 \quad (52)$$

where $\hat{\beta}_0, \hat{\beta}_1$ are the labor market returns to household characteristics in periods 0 and 1, respectively, and z_0 are the characteristics of the household in period 0. From the perspective of period 1, the expected shift that results if labor market returns are counterfactually set to their period 0 levels is defined by

$$\widehat{\Delta}y_{lab}^{10} = z_1' \hat{\beta}_0 - z_1' \hat{\beta}_1. \quad (53)$$

The tax system $tax(\cdot)$ is modeled as described in Biewen/Juhasz (2012).

⁶The setup for this example is taken from Biewen/Juhasz (2012). See the more detailed descriptions there.

⁷Again, see Biewen/Juhasz (2012) for more details.

As an example, the counterfactual household income y_{110} of period 0 that would result if labor market returns and the tax system were set to their period 1 level but everything else was kept as in period 0 is given by

$$y_{110} = y_{gross,0} + \widehat{\Delta}y_{lab}^{01} + y_{transf,0} - y_{sscontr,0} - tax_1(y_{gross,0} + \widehat{\Delta}y_{lab}^{01}), \quad (54)$$

where $y_{gross,0}$ are period 0 market incomes from all sources, $y_{transf,0}$ period 0 government transfers, $y_{sscontr,0}$ period 0 household social security contributions, and $tax_1(\cdot)$ is the counterfactual tax system of period 1. In the notation y_{110} , the first subscript refers to labor market returns, the second to the tax system, and the third to all other factors.

Using this notation, all remaining factual or counterfactual incomes are given by

$$y_{000} = y_{gross,0} + y_{transf,0} - y_{sscontr,0} - tax_0(y_{gross,0}) \quad (55)$$

$$y_{001} = y_{gross,1} + \widehat{\Delta}y_{lab}^{10} + y_{transf,1} - y_{sscontr,1} - tax_0(y_{gross,1} + \widehat{\Delta}y_{lab}^{10}) \quad (56)$$

$$y_{010} = y_{gross,0} + y_{transf,0} - y_{sscontr,0} - tax_1(y_{gross,0}) \quad (57)$$

$$y_{011} = y_{gross,1} + \widehat{\Delta}y_{lab}^{10} + y_{transf,1} - y_{sscontr,1} - tax_1(y_{gross,1} + \widehat{\Delta}y_{lab}^{10}) \quad (58)$$

$$y_{100} = y_{gross,0} + \widehat{\Delta}y_{lab}^{01} + y_{transf,0} - y_{sscontr,0} - tax_0(y_{gross,0} + \widehat{\Delta}y_{lab}^{01}) \quad (59)$$

$$y_{101} = y_{gross,1} + y_{transf,1} - y_{sscontr,1} - tax_0(y_{gross,1}) \quad (60)$$

$$y_{111} = y_{gross,1} + y_{transf,1} - y_{sscontr,1} - tax_1(y_{gross,1}). \quad (61)$$

We are interested in decomposing the change in inequality in equivalized income between periods 0 and 1, i.e. $I(y_{111}) - I(y_{000})$, into the contributions by the three different factors and their interactions. The results for this exercise are shown in table 3.⁸ According to these results, changing the labor market returns to their period 1 levels but keeping everything else constant accounts for around 35 percent of the overall inequality change. Changing the tax system in isolation accounts for around 25 percent of the overall change. Changing all other factors (but keeping returns and the tax system at their period 0 level) accounts for around 49 percent of the overall change. There is a substantial negative interaction effect between the changes in labor market returns and the changes in the tax system, amounting to some minus 12 percent of the overall inequality change. This means that, although the isolated contribution of changes in returns and changes in the tax system add up to some $35 + 25 = 60$ percent of the overall change, their combined effect is only $35 + 25 - 12 = 48$ percent. All other interaction effects (including the three-way interaction) are economically and statistically insignificant.

⁸The results are shown for the Theil coefficient. Results for other inequality indices are very similar.

— Table 3 about here —

The source of the negative interaction effect between changes in labor market returns and changes in the tax system becomes evident from figures 3 and 4. Figure 3 shows that the *ceteris paribus* effect of changes in labor market returns was inequality increasing because these changes implied a shift of distributional mass from the upper half to the lower half of the (log) income distribution. On the other hand, figure 4 confirms that the *ceteris paribus* effect of the changes in the German tax system between 1999/2000 and 2005/2006 was also inequality increasing as it stretched the middle and the top of the distribution to the right.⁹ The bottom of the distribution was unaffected by changes in the tax system. Individuals in the bottom of the distribution usually do not pay taxes at all because their pre-tax income is below the basic tax allowances for their household. The fact that changes in labor market returns increased the share of individuals in the bottom of the distribution then implies a negative interaction effect: the scope for effects coming from changes in the tax system is smaller after changes in labor market returns have been accounted for because, after these changes, fewer individuals are affected by the tax system. The effect of changing labor market returns and the tax system together is smaller than the sum of the effects from changing them individually. As argued above, it would be wrong to ascribe this interaction effect to either one of the two factors because it is a genuine product of the simultaneous presence of the two of them.

— Figures 3 and 4 about here —

For comparison, table 4 shows all possible sequential decompositions for the given case. First of all, note that the results for any sequential decomposition can be in principle be computed from the entries in table 3, i.e. from the results for the decomposition with interaction effects. This implies that the latter decomposition is always more informative than the first. The reason why the sequential decomposition can be computed from the results of the decomposition with interaction effects is that the latter allows one to compute *ceteris paribus* effects of changing any subset of factors together (see above), and the entries of the sequential decomposition are generated by comparing situations where j factors are present with situations where $j + 1$ factors are present (see equations (29 to (34))). According to the results shown in table 4, the sequential

⁹The changes in the German tax system between 1999/2000 and 2005/2006 consisted of a series of reforms that reduced marginal tax rates across the whole range of pre-tax incomes with reductions being somewhat higher at the top (see Biewen/Juhasz (2012) for more details).

decomposition results are quite path-dependent for the given case. The effect of the changes in labor market returns on increasing inequality vary between 22.24 and 35.09 percent, while the effect of changes in the tax system vary between 13.02 and 28.33 percent. As explained above, this variability comes from the presence of the interaction effect between two of the factors which is arbitrarily assigned to one of the factors depending on the ordering chosen. Moreover, the question of splitting up the overall effect into three separate parts contributed by the three factors is ill-posed because part of the overall effect comes from the simultaneous presence of more than one factor.

— Table 4 about here —

6 Conclusion

This paper has explored the challenges posed to sequential decomposition schemes by the existence of interaction effects. It has been argued that, instead of attributing interaction effects in an arbitrary manner to individual factors considered in the decomposition, they should be considered and reported separately and receive their own contribution in an additive decomposition. The paper has proposed a general additive and path-independent decomposition formula that decomposes any difference between two objects into *ceteris paribus* effects of individual factors and all possible interaction effects. The proposed decomposition has the additional advantage of being comprehensive in the sense that it uses information on all possible counterfactual situations of the problem under consideration. The empirical illustrations have shown that, in many contexts, interaction effects are a feature of reality that should not be ignored.

The proposed decomposition formula may either be used as a descriptive tool to decompose differences between objects into the parts contributed by different factors and their interactions, or as a way to summarize the causal effects of a number of factors that possibly interact with each other. If all counterfactual outcomes involved are correctly identified, the decomposition formula may be viewed as an anatomy of causal effects. Starting with the reference situation in which all factors are unchanged, one first asks how far one can get in explaining the target situation by changing each factor individually. Then it is asked what one can explain in addition by taking into account interaction effects by any subset of two factors. After that, it is asked what can be further explained by taking into account interactions of any subset of three factors, and so on.

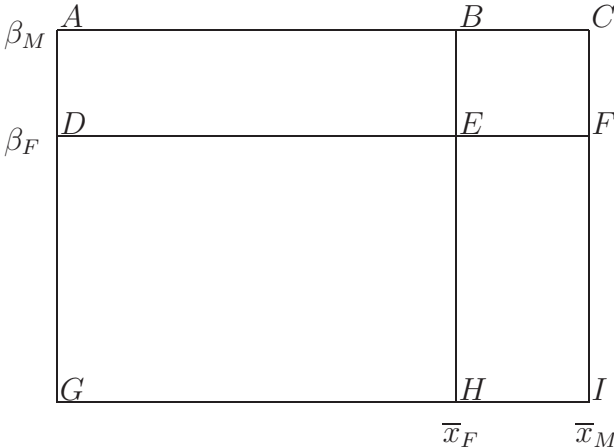
7 References

- Bauer, T., M. Sinnig (2007): Gender Differences in Smoking Behavior, *Health Economics* 16, pp. 895 - 909.
- Biewen, M. (2001): Measuring the Effects of Socio-Economic Variables on the Income Distribution: An Application tot the East German Transition Process, *Review of Economics and Statistics* 83, pp. 185 - 190.
- Biewen, M., S. Jenkins (2005): A framework for the decomposition of poverty differences with an application to poverty differences between countries, *Empirical Economics* 30, pp. 331 - 358.
- Biewen, M., A. Juhasz (2012): Understanding Rising Income Inequality in Germany, 1999/2000 to 2005/2006, forthcoming in: *Review of Income and Wealth*.
- Blinder, A. (1973): Wage Discrimination: Reduced Form and Structural Estimates, *Journal of Human Resources* 8, pp. 436 - 455.
- Daymont, T.N., P.J. Andrisani (1984): Job Preferences, College Major, and the Gender Gap in Earnings, *Journal of Human Resources* 19, pp. 409 - 428.
- DiNardo, J. N.M. Fortin, and T. Lemieux (1996): Labor Market Institutions and the Distribution of Wages, 1973 - 1992: A Semiparametric Approach, *Econometrica* 64, pp. 1001 - 1044.
- Fairlie, R.W. (2005): An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models, *Journal of Economic and Social Measurement* 30, pp. 305 - 316.
- Fortin, N., T. Lemieux, and S. Firpo (2011): Decomposition Methods in Economics, in: Ashenfelter, O., and D. Card (eds.), *Handbook of Labor Economics*, Volume 4A, North-Holland, Amsterdam, pp. 1 - 102.
- Gomoulka, J., N. Stern (1990): The Employment of Married Women in the United Kingdom, 1970 - 83, *Economica* 57, pp. 171 - 199.
- Jann, B. (2008): A Stata implementation of the Blinder-Oaxaca decomposition, *Stata Journal* 8, pp. 453 - 479.
- Jones, F.L., J. Kelley (1984): Decomposing Differences Between Groups. A Cautionary Note on Measuring Discrimination, *Sociological Methods and Research* 12, pp. 323 - 343.

- Juhn, C., K.M. Murphy, and B. Pierce (1993): Wage Inequality and the Rise in Returns to Skill, *Journal of Political Economy* 101, pp. 410 - 442.
- Hyslop, D.R., D.C. Mare (2005): Understanding New Zealand's Changing Income Distribution, 1983-1998: A Semi-parametric Analysis, *Economica* 72, pp. 469 - 495.
- Machado, J.F., J. Mata (2005): Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression, *Journal of Applied Econometrics* 20, pp. 445 - 465.
- Oaxaca, R.L. (1973): Male-Female Wage Differentials in Urban Labor Markets, *International Economic Review* 14, pp. 693 - 709.
- Solow, R. (1957): Technical change and the aggregate production function, *Review of Economics and Statistics* 3, pp. 312 - 320.
- Winsborough, H.H., P. Dickenson (1971): Components of negro-white income differences, *Proceedings of the American Statistical Association, Social Statistics Section*, pp. 6 - 8.
- Yun, M.S. (2004): Decomposing differences in the first moment, *Economics Letters* 82, pp. 275 - 280.

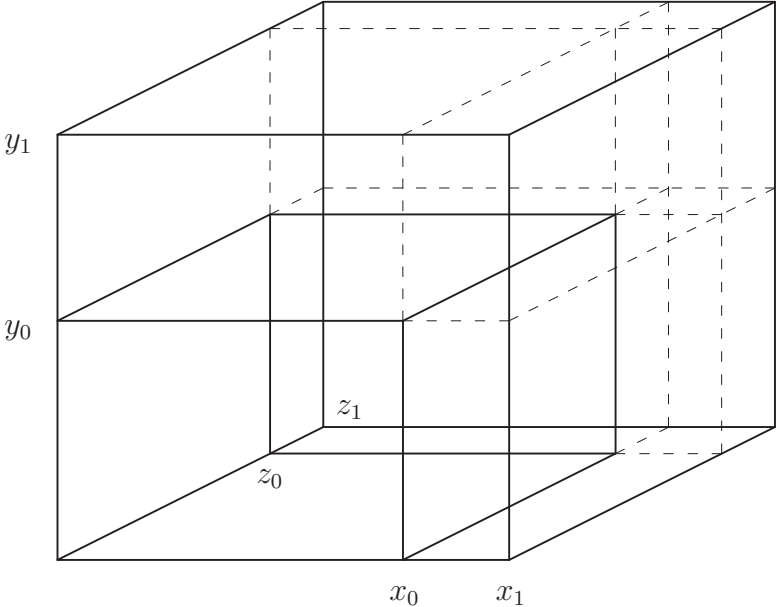
8 Figures

Figure 1 – Illustration of Winsborough/Dickenson (1971) decomposition



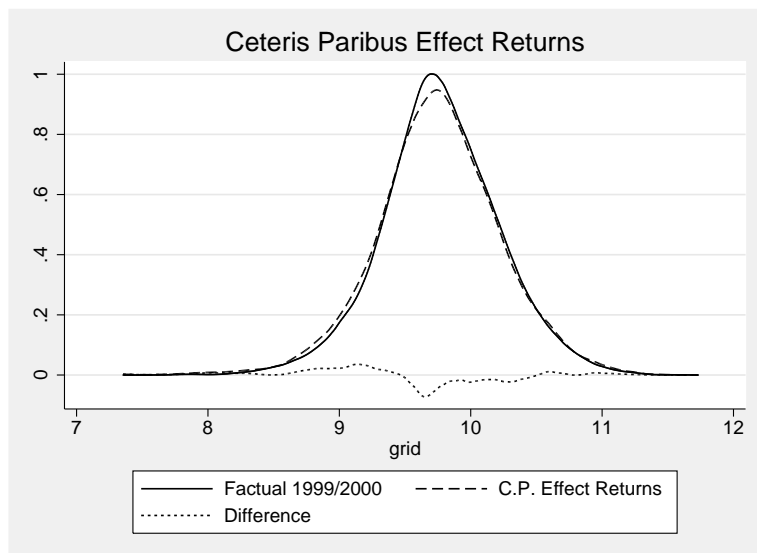
Notation: \bar{x}_M, \bar{x}_F represent mens' and womens' average characteristics,
 β_M, β_F represent their regression coefficients

Figure 2 – Illustration of general decomposition with three factors



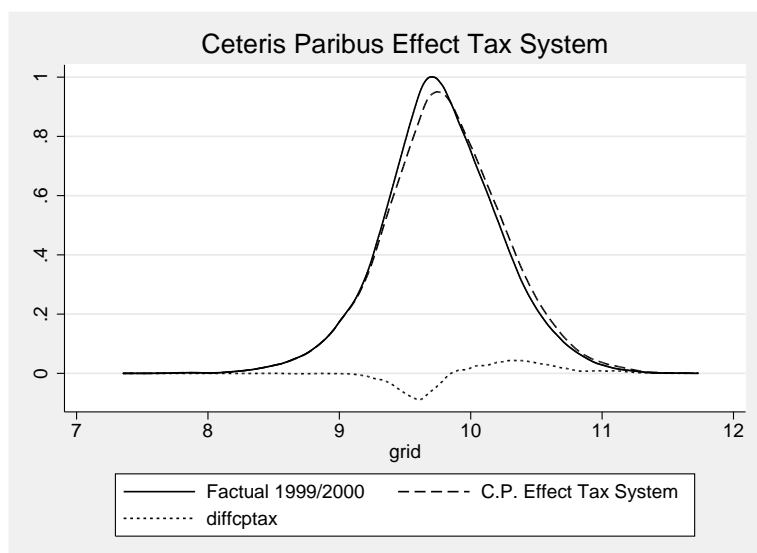
Notation: $x_0y_0z_0 =$ reference situation, $x_1y_1z_1 =$ target situation

Figure 3 – Counterfactual income distribution if only labor market returns are changed (dashed line) vs. factual distribution (bold line).



Source: GSOEP, own calculations. The graph shows the density of log equivalized incomes.

Figure 4 – Counterfactual income distribution if only the tax system is changed (dashed line) vs. factual distribution (bold line).



Source: GSOEP, own calculations. The graph shows the density of log equivalized incomes.

9 Tables

Table 1 – Oaxaca-Blinder decomposition of wage differentials between the services sector and other sectors of the economy

Average wage other sectors	2.7258	(0.0072)
Average wage services sector	2.5803	(0.0115)
Difference	0.1454	(0.0136)
Characteristics effect	0.0834	(0.0086)
Coefficients effect	0.0830	(0.0115)
Interaction effect	-0.0210	(0.0040)

Source: German Socio-Economic Panel, 2005. Standard errors shown in parentheses.

Standard errors and point estimates were computed as described in Jann (2008).

Table 2 – Coefficients and endowments in wage example

	Coefficient		Average endowment	
<i>Other sectors</i>				
Education	0.0721	(0.0022)	12.8174	(0.0417)
Experience	0.0321	(0.0029)	16.6536	(0.1280)
Experience squared	-0.0007	(0.0000)	349.715	(4.5973)
Tenure	0.0131	(0.0008)	11.114	(0.1319)
Female	-0.1689	(0.0126)	0.4489	(0.0074)
Constant	1.4666	(0.0390)	-	-
Observations	4356	-	-	-
<i>Services sector</i>				
Education	0.0783	(0.0036)	12.6519	(0.0535)
Experience	0.0314	(0.0045)	15.3221	(0.1652)
Experience squared	-0.0007	(0.0001)	303.473	(5.7538)
Tenure	0.0185	(0.0014)	8.8625	(0.1584)
Female	-0.2643	(0.0198)	0.5252	(0.0099)
Constant	1.3103	(0.0613)	-	-
Observations	2519	-	-	-

Source: German Socio-Economic Panel, 2005. Standard errors shown in parentheses.

Table 3 – Decomposition with interaction effects

	Absolute	Percentage of overall change	
<i>total</i>	.0298	100	-
<i>cp</i> ₁ (Returns)	.0105	35.09	(6.50)
<i>cp</i> ₂ (Tax system)	.0075	25.10	(3.93)
<i>cp</i> ₃ (Other factors)	.0147	49.41	(12.38)
<i>int</i> ₁₂	-.0036	-12.08	(3.31)
<i>int</i> ₁₃	.0005	1.75	(5.53)
<i>int</i> ₂₃	.0009	3.22	(10.67)
<i>int</i> ₁₂₃	-.0007	-2.51	(5.07)

Source: German Socio-Economic Panel, 2005. The decomposition decomposes the change in income inequality between 1999/2000 and 2005/2006 as measured by the Theil coefficient, $Theil_{2005/2006} - Theil_{1999/2000} = .1303 - .1005 = .0298$, into different components. The bootstrap standard errors shown in parentheses correctly take into account the longitudinal sample design, the clustering of observations within households, and stratification.

Table 4 – All possible sequential decompositions

	1,2,3	1,3,2	2,1,3
Returns	35.09 (6.50)	35.09 (6.50)	23.00 (5.85)
Tax system	13.02 (4.22)	13.73 (10.75)	25.10 (3.93)
Other factors	51.88 (7.48)	51.17 (11.27)	51.88 (7.48)
	2,3,1	3,1,2	3,1,2
Returns	22.24 (7.35)	36.84 (6.18)	22.24 (7.35)
Tax system	25.10 (3.93)	13.73 (10.75)	28.33 (11.58)
Other factors	52.64 (8.66)	49.41 (12.38)	49.41 (12.38)

Source: German Socio-Economic Panel, 2005. The entries of the table represent percentages of the overall change $Theil_{2005/2006} - Theil_{1999/2000} = .1303 - .1005 = .0298$. Each ordering represents one possible sequential decomposition. For example, the sequence '3,1,2' means that factor 3 (= all other factors) are changed first, then factor 1 (= labor market returns), and then factor 2 (= tax system). The bootstrap standard errors shown in parentheses correctly take into account the longitudinal sample design, the clustering of observations within households, and stratification.