Additive Decompositions with Interaction Effects[1]

Martin Biewen (University of Tübingen, IZA Bonn, DIW Berlin)

This version: 15 November 2013

Abstract.

This note presents a general way to decompose differences over time or between objects into the ceteris paribus effects and the interaction effects of an arbitrary number of factors. The decomposition addresses the issue of interaction effects between factors which has been neglected in the decomposition literature. It has the additional advantage of being path-independent and aggregation consistent. A number of examples clarify the issues involved and demonstrate that interaction effects may be a feature of reality that is of particular interest.

JEL-Classification: J31, J71, C02, D31

Keywords: ceteris paribus effects, sequential decomposition, Shapley decomposition

Correspondence:

Martin Biewen, Department of Economics, University of Tübingen, Mohlstr. 36, 72074 Tübingen, Germany, Fax: +49-7071-295013, martin.biewen@uni-tuebingen.de

[1]This is a substantially revised version of Biewen (2012): Additive Decompositions with Interaction Effects, IZA Discussion Paper No. 6730. I am grateful to Andos Juhasz for 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

A typical question asked in distributional analysis is: what percentage of a change in inequality is due to factor 1, what percentage due to factor 2, and so on. However, if part of an overall effect is the genuine result of the joint presence of two or more factors, the idea of splitting up this total effect into disjunct parts contributed by each factor may appear questionable. Representative applications of this methodology include Juhn et al. (1993), DiNardo et al. (1996), and Daly/Valetta (2006), among many others. 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 a similar vein, Daly/Valetta (2006) decompose the increase in poverty and inequality in the US into parts contributed by the changes in the earnings distribution of men, changes in the labor force participation of women, and changes in family structures.

A 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 be path-dependent, i.e. it may depend on the order in which the different factors are added. The path-dependence of sequential decompositions may be easily remedied by averaging decomposition results over all possible decomposition sequences. This was originally proposed by Shorrocks (1999) and Chantreuil/Trannoy (1999), and is usually called 'Shapley decomposition' (due to its formal resemblance to the Shapley value from game theory). However, the Shapley decomposition does not address another disadvantage of sequential decompositions, namely that the decomposition forces an answer to the question of disentangling an overall effect into disjunct parts even if such an answer may be undesirable because part of the overall effect is the genuine result of the joint presence of more than one factor. As an 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]

It is not that the authors of the studies cited have not recognized the challenges posed to their decomposition schemes by the existence of interaction effects (e.g., Juhn et al., p. 429). However, to our best knowledge, no attempt has been made to explicitly address these challenges in a general and systematic way. 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 may be questionable, 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 may be used to detect lower and higher order interaction effects and thus display the part of an outcome that is the genuine result of the interplay between two or more factors.

The rest of this paper is organized as follows. Section 2 revisits decomposition schemes based on sequential orderings including the Shapley decomposition. Section 3 introduces the alternative decomposition scheme involving interaction effects and examines its relationship to sequential decompositions and the Shapley decomposition. Section 4 provides a number of examples in order to clarify some of the issues involved and in order to demonstrate that interaction effects may be a feature of reality that is of particular interest.

2 Sequential decompositions and Shapley decomposition

This section reviews the sequential decomposition schemes that are widely used in the literature as well as the Shapley decomposition which is defined as the average over all possible sequential decompositions. 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

[2]This 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.

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.

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

$$
\begin{aligned}
f_{11} - f_{00} &= (f_{10} - f_{00}) && (1)\\
&+ (f_{11} - f_{10}). && (2)
\end{aligned}
$$

Here, (1) measures the contribution of factor one, while (2) 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

$$
\begin{aligned}
f_{11} - f_{00} &= (f_{01} - f_{00}) && (3)\\
&+ (f_{11} - f_{01}). && (4)
\end{aligned}
$$

Here, the contribution of factor two (3) is measured first, while that of factor one (4) is measured second.

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

$$
\begin{aligned}
f_{11111...1} - f_{00000...0} &= (f_{10000...0} - f_{00000...0}) && (5)\\
&+ (f_{11000...0} - f_{10000...0}) && (6)\\
&+ (f_{11100...0} - f_{11000...0}) && (7)\\
&+ (f_{11110...0} - f_{11100...0}) && (8)\\
&+ \ \ldots && (9)\\
&+ (f_{11111...1} - f_{11111...0}) && (10)
\end{aligned}
$$

in obvious notation. Here, (5) measures the contribution of factor one, (6) measures that of factor two, and, finally, (10) that of factor m. Alternative sequential decompositions in the m-factor case are obtained by permutating the order in which the different factors are introduced. In total, there are $m \cdot (m-1) \cdot \ldots 2 \cdot 1 = m!$ possible sequential decompositions in the m-factor case.

The *Shapley decomposition* as proposed by Shorrocks (1999) and Chantreuil/Trannoy (1999) computes the contributions of the m factors by averaging over all possible $m!$ sequential orderings. It has a number of desirable properties. Apart from the fact that it provides an exact break-down of the overall effect into m contributions, it is path-independent and it suggests an interpretation of the different contributions as the marginal effect of each factor averaged over all possible situations (Shorrocks, 1999). A disadvantage of the decomposition is that it is uninformative about possible interaction effects between different factors. A further disadvantage is that it is generally not aggregation consistent implying that if one factor is broken up in a number of subfactors, this may change the contributions of the other factors (Shorrocks, 1999, Chantreuil/Trannoy, 1999). The Shapley decomposition is now widely used in many different areas, see e.g. Sastre/Trannoy (2002), Israeli (2007),[3] Bargain/Callan (2010), Devicienti (2010), and Okamoto (2011). As will become clear below, the Shapley decomposition provides a summary measure of the direct and the interaction effects of the different factors involved.

3 Additive decomposition with interaction effects

This section presents the alternative decomposition scheme involving interaction effects. In the two factor case, the decomposition of the difference $f_{11} - f_{00}$ is given by

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

$$+ (f_{01} - f_{00}) \tag{12}$$

$$+ [(f_{11} - f_{00}) - (f_{10} - f_{00}) - (f_{01} - f_{00})]\,. \tag{13}$$

Here, (11) represents the ceteris paribus effect of factor one, (12) the ceteris paribus effect of factor two, and (13) the interaction effect between the two factors. The ceteris paribus effects describe the effects that occur if each of the factors is changed in isolation. If the two separate changes do not add up to the overall change, this necessarily implies that the two factors interact in their effect on the outcome, i.e. the interaction effect is the part of the overall change that cannot be explained by changing both of the factors in isolation. The interaction effect (13) has other, more intuitive interpretations:

$$[(f_{11} - f_{00}) - (f_{10} - f_{00}) - (f_{01} - f_{00})] \tag{14}$$

$$= (f_{11} - f_{01}) - (f_{10} - f_{00}) \tag{15}$$

[3]Israeli (2007) also addresses the issue of interaction but in a slightly different context.

$$= (f_{11} - f_{10}) - (f_{01} - f_{00}) \tag{16}$$

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 (15), and to the effect of factor two varied by whether or not factor one is present (16) (this is similar to a cross derivative in an infinitesimal setting).

In the three factor case, the proposed decomposition is given by

$$\begin{aligned}
f_{111} - f_{000} = &(f_{100} - f_{000}) && (17)\\
&+ (f_{010} - f_{000}) && (18)\\
&+ (f_{001} - f_{000}) && (19)\\
&+ [(f_{110} - f_{000}) - (f_{100} - f_{000}) - (f_{010} - f_{000})] && (20)\\
&+ [(f_{101} - f_{000}) - (f_{100} - f_{000}) - (f_{001} - f_{000})] && (21)\\
&+ [(f_{011} - f_{000}) - (f_{010} - f_{000}) - (f_{001} - f_{000})] && (22)\\
&+ [(f_{111} - f_{000}) - (17) - (18) - (19) - (20) - (21) - (22)]\,. && (23)
\end{aligned}$$

Again, (17) - (19) represent the ceteris paribus effects of factors one to three. In general, the ceteris paribus effect comes closest to what one has in mind when talking about 'the effect' of a factor, i.e. it describes how much of the change in the outcome can be explained by changing just this factor and holding everything else constant. The ceteris paribus effects of the different factors thus deserve a special role in the decomposition. Contribution (20) is the two-way interaction effect between factor one and factor two. Contributions (21), (22) 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 (23). The three-way interaction effect int^3 is defined as

$$int^3 = total_3 - \sum_{i=1}^{3} cp_i - \sum_{k \in P_2} int_k^2, \tag{24}$$

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).

This decomposition idea 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

$$int^4 = total_4 - \sum_{i=1}^{4} cp_i - \sum_{k \in P_2} int_k^2 - \sum_{k \in P_3} int_k^3, \tag{25}$$

where the last term stands for all possible three-way interaction effects between three factors chosen out of the four. In the general case with m factors,

$$int^m = total_m - \sum_{i=1}^{m} cp_i - \sum_{k \in P_2} int_k^2 - \sum_{k \in P_3} int_k^3 - \ldots - \sum_{k \in P_{m-1}} int_k^{m-1}, \tag{26}$$

yielding the general decomposition formula

$$total_m = \sum_{i=1}^{m} cp_i + \sum_{k \in P_2} int_k^2 + \sum_{k \in P_3} int_k^3 + \ldots + \sum_{k \in P_{m-1}} int_k^{m-1} + int^m. \tag{27}$$

Apart from incorporating interaction effects, decomposition (27) has number of desirable properties.[4] First, it is *path-independent* because all factors are treated symmetrically. 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. Third, the decomposition is *aggregation consistent* in the sense that the decompositions of the joint influence of a subset of factors appear as a part of the larger decomposition involving all factors. More precisely, if two or more factors are combined, the contributions in the aggregated decomposition result by summing elements of the disaggregated decomposition. This is evident from the generic case of three factors (17) to (23), in which one may combine two of the factors in order to form an aggregated factor. For example, if one combines factors one and two, then the ceteris paribus effect of the combined factor one/two is given by $cp_1 + cp_2 + int_{\{1,2\}}^2$, the ceteris paribus effect of the other factor by cp_3, and the interaction effect between the combined factor one/two and factor three by

[4]It seems that a recent paper by Rothe (2012) introduced (independently from this paper) a similar definition of interaction effects in the context of decomposing the composition effect in between-group decompositions. Note however, that the factors are defined somewhat differently in Rothe's paper. An important further difference is that Rothe specifically considers the composition effect in a between-group decomposition, while the formula given here is formulated in terms of general and unspecified counterfactual outcomes, independently of how they are generated (e.g., they may be generated by simulation, see empirical example below).

$int^2_{\{13\}} + int^2_{\{23\}} + int^3_{\{123\}}$. A final advantage of the decomposition is that it is *comprehensive* in the sense that it involves all counterfactual states of the world.[5] This property is shared by the Shapley decomposition but not by individual sequential decompositions which will be insensitive to changes in counterfactuals that do not appear in the decomposition. The property of comprehensiveness appears to be desirable as the decomposition should, in some way, reflect all the different counterfactuals of the problem under investigation.

It has to be stressed that the objective of this paper is not to claim that the Shapley decomposition is wrong or unreasonable. The aim here is rather to show that explicitly considering interaction effects may lead to a more informative analysis than averaging over possibly heterogenous decomposition sequences. Note that the decomposition with interaction effects is also more informative in a formal sense than any given sequential decomposition and than the Shapley decomposition. This holds in the sense that the contributions of any sequential decomposition and (therefore of the Shapley decomposition) may be computed from the elements of the decomposition with interaction effects because the latter allows one to compute the effects of jointly changing any subset of factors together on which all sequential decompositions (and therefore the Shapley decomposition) are based (see (5) to (10)).

In order to give further justification to the specific definition of interaction effects provided above, it is helpful to note that 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, and that the contributions of a sequential decomposition scheme are path-independent if and only if all interaction effects *as defined above* are zero.[6] This validates the definition of the interaction effects given above, as it reproduces the intuition that path-independence is equivalent to the absence of interaction effects. Notice that the above statement is rather strong as it requires that *all* interaction effects (i.e. also the ones of a higher order) have to be zero for path independence.

A likely consequence is that the larger and more prevalent interaction effects are, the more

[5]This property can be seen as follows. A decomposition with two factors involves all possible counterfactuals for these two factors (see equations (11) to (13)). 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 (23)). 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.

[6]See Appendix.

problems with path-dependence in a sequential decomposition may occur. The two statements 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. This justifies the practice of trying out different factor orderings when computing sequential decompositions, see DiNardo et al. (1996), Hyslop/Mare (2005) or Daly/Valetta (2006). Examples where the contribution of individual factors may be quite dependent on the order in which the different factors are introduced can be found in Biewen (2001) and Biewen/Juhasz (2012). This is then indicative of interaction effects. Biewen (2001) provides an extreme example of path dependence (or interaction effects, respectively). The article examines distributional change in East Germany following the reunification of the country in 1990 which led both to a drastic rise in unemployment and to a change in the conditional income structure (i.e. income conditional on household characteristics such as employment, unemployment, family composition etc.) in the sense that income differentials between employed and unemployed individuals grew in a drastic way. This leads to the situation that introducing higher unemployment in the (near-socialist) base situation of 1990 has almost no effect (because there were no big income differentials between employed and unemployed individuals), while it has an extreme effect under the new (capitalist) conditional income structure which followed a couple of years later and in which the income differences between employed and unemployed individuals were huge.

4 Examples

The purpose of this section is to apply the above framework to a number of examples in order to highlight some of the issues involved and to demonstrate that explicitly considering interaction effects may lead to a more informative analysis than considering sequential decompositions or averaging over potentially heterogenous paths (as done in the Shapley decomposition).

4.1 The Oaxaca-Blinder decomposition

In the following, it will be shown that the variant of the famous Oaxaca-Blinder decomposition that involves an interaction term is a special case of the general decomposition scheme described above. Starting with Oaxaca (1973) and Blinder (1973), economists have been asking the question of how to decompose differences between groups or over time into a 'characteristics' and a 'returns' effects.[7] 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 extensions.

The Oaxaca-Blinder decomposition is a two-factor scenario ('characteristics' and 'returns'). To fit the decomposition into the framework defined above, define

$$f_{11} = \overline{x}_M \beta_M, \quad f_{10} = \overline{x}_M \beta_F, \quad f_{01} = \overline{x}_F \beta_M, \quad f_{00} = \overline{x}_F \beta_F, \tag{28}$$

where $\overline{x}_M$, $\overline{x}_F$ 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 (1) to (4):

$$\begin{aligned} f_{11} - f_{00} &= (f_{10} - f_{00}) + (f_{11} - f_{10}) \\ &= \overline{x}_M \beta_M - \overline{x}_F \beta_F = (\overline{x}_M - \overline{x}_F)\beta_F + \overline{x}_M(\beta_M - \beta_F) \end{aligned} \tag{29}$$

$$\begin{aligned} f_{11} - f_{00} &= (f_{01} - f_{00}) + (f_{11} - f_{01}) \\ &= \overline{x}_M \beta_M - \overline{x}_F \beta_F = \overline{x}_F(\beta_M - \beta_F) + (\overline{x}_M - \overline{x}_F)\beta_M \end{aligned} \tag{30}$$

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 (11) to (13)) is given by

$$\begin{aligned} f_{11} - f_{00} &= (f_{10} - f_{00}) + (f_{01} - f_{00}) + [(f_{11} - f_{00}) - (f_{10} - f_{00}) - (f_{01} - f_{00})] \\ &= (\overline{x}_M - \overline{x}_F)\beta_F + \overline{x}_F(\beta_M - \beta_F) + (\overline{x}_M - \overline{x}_F)(\beta_M - \beta_F) \qquad (31) \\ &= \Delta\overline{x}\beta_F + \overline{x}_F \Delta\beta + \Delta\overline{x}\Delta\beta. \qquad (32) \end{aligned}$$

This is the variant of the Oaxaca-Blinder decomposition involving an interaction term (see Winsborough/Dickenson, 1971, and Blinder, 1973, footnote 3). Surprisingly, this variant of the Oaxaca-Blinder decomposition is rarely used in economics.[8] Its validity is obvious from figure 1.

[7] See Fortin et al. (2011) for a comprehensive overview.

[8] An exception is Daymont/Andrisani (1984).

— Figure 1 about here —

Decomposition (31) asks why area $ACGI = \overline{x}_M\beta_M$ is larger than area $DEGH = \overline{x}_F\beta_F$. It is easily seen that the difference between the two is composed of $ABDE = \Delta\overline{x}\beta_F$, $EFHI = \overline{x}_F\Delta\beta$, and $BCEF = \Delta\overline{x}\Delta\beta$. Why do women have lower average wages than men? One part of the difference is due to their less favorable characteristics ($= \Delta\overline{x}\beta_F$), another one due to their lower returns ($= \overline{x}_F\Delta\beta$), and a third part can only explained by both of these factors together ($= \Delta\overline{x}\Delta\beta$). This latter part would be zero if either $\Delta\overline{x} = 0$ or $\Delta\beta = 0$. It may be questionable to assign this interaction term to either the 'characteristics' or the 'returns' effect as it is done in sequential decompositions (29) and (30).[9] The question of whether or not to treat the interaction effect as a separate contribution touches subtle issues when the framework is used in order to measure 'discrimination' (see Jones/Kelley, 1984). Generally, it seems 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 $\Delta\overline{x}$ and $\Delta\beta$ are. This is the reason why in infinitesimal settings (where $\Delta\overline{x}, \Delta\beta \rightarrow 0$), 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 (32) 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 appears questionable 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 may be relevant to the concept of constructing price indices.

[9] The comparison between (29) and (31) implies that in (29), the interaction term is assigned to the 'returns' effect. Similarly, the comparison between (30) and (31) implies that in (30), the interaction effect is assigned to the 'characteristics' effect.

4.2 An example with three-way interactions

Following up on the arguments given in the previous section, the next example further illustrates the differences between sequential decompositions and the idea of explicitly considering interaction effects. The example is shown in 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

$$
\begin{aligned}
x_1y_1z_1 - x_0y_0z_0 = & [(x_1y_0z_0 - x_0y_0z_0)] && (33)\\
& + [(x_0y_1z_0 - x_0y_0z_0)] && (34)\\
& + [(x_0y_0z_1 - x_0y_0z_0)] && (35)\\
& + [(x_1y_1z_0 - x_0y_0z_0) - (x_1y_0z_0 - x_0y_0z_0) - (x_0y_1z_0 - x_0y_0z_0)] && (36)\\
& + [(x_1y_0z_1 - x_0y_0z_0) - (x_1y_0z_0 - x_0y_0z_0) - (x_0y_0z_1 - x_0y_0z_0)] && (37)\\
& + [(x_0y_1z_1 - x_0y_0z_0) - (x_0y_1z_0 - x_0y_0z_0) - (x_0y_0z_1 - x_0y_0z_0)] && (38)\\
& + [(x_1y_1z_1 - x_0y_0z_0) - (33) - (34) - (35) - (36) - (37) - (38)]\,, && (39)
\end{aligned}
$$

which is just decomposition (17) to (23). 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 will 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 $(x_0y_0z_1 - x_0y_0z_0)$. The Shapley decomposition will then average over all possible assignments of interaction effects to one of the ceteris paribus effects.

— Figure 2 about here —

It should be noted that decomposition formula (27) is more general than the example given in figure 2. Formula (27) 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.

4.3 Wage differentials between the services sector and other sectors of the economy

The following example illustrates the empirical relevance of an interaction effect in a Oaxaca-Blinder type decomposition. In the example, wage differences between the services sector and other sectors of the economy are decomposed into a characteristics, a coefficient, and an interaction effect as in decomposition (31).[10] The 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 does not seem adequate 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.

4.4 Decomposition of distributional change

The following example illustrates how the decomposition formula described above may be used to trace out possible interaction effects between different factors. The example considers three factors explaining changes in the distribution of equivalized incomes in Germany between 1999/2000 (= period 0) and 2005/2006 (= period 1). The three factors are: changes due to shifts in the labor market returns to household characteristics (= factor 1), changes in the tax system (= factor 2), and all 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

[10] The data of this example and the example in the next section come from the German Socio-Economic Panel.

security contributions, divided by an equivalence scale in order to arrive at a measure of personal income for each household member.[11]

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. 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

$$\widehat{\Delta} y_{lab}^{01} = z_0{}'\hat{\beta}_1 - z_0{}'\hat{\beta}_0 \tag{40}$$

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

$$\widehat{\Delta} y_{lab}^{10} = z_1{}'\hat{\beta}_0 - z_1{}'\hat{\beta}_1. \tag{41}$$

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

$$y_{110} = y_{gross,0} + \widehat{\Delta} y_{lab}^{01} + y_{transf,0} - y_{sscontr,0} - \ tax_1(y_{gross,0} + \widehat{\Delta} y_{lab}^{01}), \tag{42}$$

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

$$y_{000} = y_{gross,0} + y_{transf,0} - y_{sscontr,0} - \ tax_0(y_{gross,0}) \tag{43}$$

$$y_{001} = y_{gross,1} + \widehat{\Delta} y_{lab}^{10} + y_{transf,1} - y_{sscontr,1} - \ tax_0(y_{gross,1} + \widehat{\Delta} y_{lab}^{10}) \tag{44}$$

$$y_{010} = y_{gross,0} + y_{transf,0} - y_{sscontr,0} - \ tax_1(y_{gross,0}) \tag{45}$$

[11] The setup for this example is the same as in Biewen/Juhasz (2012). See the more detailed descriptions there.

$$y_{011} = y_{gross,1} + \widehat{\Delta} y_{lab}^{10} + y_{transf,1} - y_{sscontr,1} - \ tax_1(y_{gross,1} + \widehat{\Delta} y_{lab}^{10}) \quad (46)$$

$$y_{100} = y_{gross,0} + \widehat{\Delta} y_{lab}^{01} + y_{transf,0} - y_{sscontr,0} - \ tax_0(y_{gross,0} + \widehat{\Delta} y_{lab}^{01}) \quad (47)$$

$$y_{101} = y_{gross,1} + y_{transf,1} - y_{sscontr,1} - \ tax_0(y_{gross,1}) \quad (48)$$

$$y_{111} = y_{gross,1} + y_{transf,1} - y_{sscontr,1} - \ tax_1(y_{gross,1}). \quad (49)$$

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 are shown in table 3 for the case of the Theil coefficient.[12] 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. The decomposition also shows that all other interaction effects, i.e. those with the other unmeasured factors as well as the three-way interaction effect are economically and statistically insignificant. For comparison, table 4 shows all possible sequential decompositions along with the Shapley decomposition. As a result of the interaction effect, the sequential decomposition results are quite path-dependent for the given case. The effect of the changes in labor market returns on increasing inequality varies between 22.24 and 35.09 percent, while that of changes in the tax system varies between 13.02 and 28.33 percent. The Shapley decomposition provides a summary measure of the different paths amounting to 29.08 percent the changes in labor market returns, and to 19.83 percent for those in the tax system.

— 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

[12] Results for other inequality indices are very similar.

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 (the reason is that the 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). 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 sum of 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.

— Figures 3 and 4 about here —

As argued above, an account of the rise in inequality between 1999/2000 and 2005/2006 that does not mention the interaction between changes in the tax system and changes in labor market returns may be seen as being incomplete. For example, for policy makers it will be of considerable interest to know that and to what extent the distributional effects of their tax reforms depended on how much remunerations in the labor market changed. Similarly, as a description of what 'caused' the inequality increase between the two periods, the summary information provided by the Shapley decomposition may be misleading because the 'average impact' of a given factor assumes that other factors are also changed to a certain effect (because interactions are partly included). This may not be what one has in mind when asking for 'the effect' of a given factor. By contrast, the decomposition with interaction effects leaves it to the reader to decide what precise effect she is interested in.

— Table 4 about here —

5 Conclusion

This note 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 some specific manner to individual factors or averaging over possibly heterogenous sequences, they can be reported separately and receive their own contribution in an additive decomposition. An example analyzing distributional change has shown that interaction effects may be an interesting feature of reality, e.g., because the effect of a policy may depend on how other factors have changed, or because they allow one to better understand the precise mechanics of distributional change.

6 References

Bargain, O., T. Callan (2010): Analysing the effects of tax-benefit reforms on income distribution: a decomposition approach, *Journal of Economic Inequality 8*, pp. 1-21.

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., Juhasz, A., 2012. Understanding Rising Income Inequality in Germany, 1999/2000 to 2005/2006, *Review of Income and Wealth 58*, pp. 622-647.

Blinder, A. (1973): Wage Discrimination: Reduced Form and Structural Estimates, *Journal of Human Resources 8*, pp. 436 - 455.

Chantreuil, F., Trannoy, A. (1999): Inequality Decomposition Values: The Trade-Off between Marginality and Consistency, *THEMA Discussion Paper*, Université Cergy-Pontoise, later published in: *Journal of Economic Inequality 11*, 2013, 83 - 98.

Daly, M.C., R. Valetta (2006): Inequality and Poverty in United States: The Effects of Rising Dispersion of Men's Earnings and Changing Family Behaviour, *Economica 73*, pp. 75-98.

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.

Devicienti, F. (2010): Shapley-value decompositions of changes in wage distributions: a note, *Journal of Economic Inequality 8*, pp. 35-45.

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.

Okamoto, M. (2011): Source decomposition of changes in income inequality: the integral-based approach and its approximation by the chained Shapley-value approach, *Journal of Economic Inequality 9*, pp. 145-181

Israeli, O. (2007): A Shapley-based decomposition of the R-square of a linear regression, *Journal of Economic Inequality*, Vol. 5, pp. 199-212.

Rothe, C. (2012): Decomposing the Composition Effect, *IZA Discussion Paper No. 6397*, Institute for the Study of Labor, Bonn.

Sastre, M., A. Trannoy (2002): Shapley inequality decomposition by factor components: Some methodological issues, *Journal of Economics 9*, p. 51-89.

Shorrocks, A., (1999). Decomposition Procedures for Distributional Analysis: A Unified Framework based on the Shapley Value, mimeo, University of Essex, later published in: *Journal of Economic Inequality 11*, 2013, 99 - 126.

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.

7 Appendix

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. □

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

Proof. If the contributions of a sequential decomposition scheme are path-independent, then they are equal to the ceteris paribus effects (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.

$$f_{1,1,(0,\dots,0)} - f_{0,1,(0,\dots 0)} = f_{1,0,(0,\dots,0)} - f_{0,0,(0,\dots 0)} \tag{50}$$

(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 (50) means that the two-way interactions between the given factor and any other factor are equal to zero (see (15)). This holds for all factors and all two-way interactions. Then, for any three-way interaction

$$int^3 = total_3 - \sum_{i=1}^{3} cp_i - \sum_{k \in P_2} int_k^2 = total_3 - \sum_{i=1}^{3} cp_i = 0 \tag{51}$$

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,

$$int^j = total_j - \sum_{i=1}^{j} cp_i - \sum_{k \in P_2} int_k^2 - \sum_{k \in P_3} int_k^3 - \ldots - \sum_{k \in P_{j-1}} int_k^{j-1} = 0 \tag{52}$$

for all remaining $j = 4, \ldots, m$ (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.

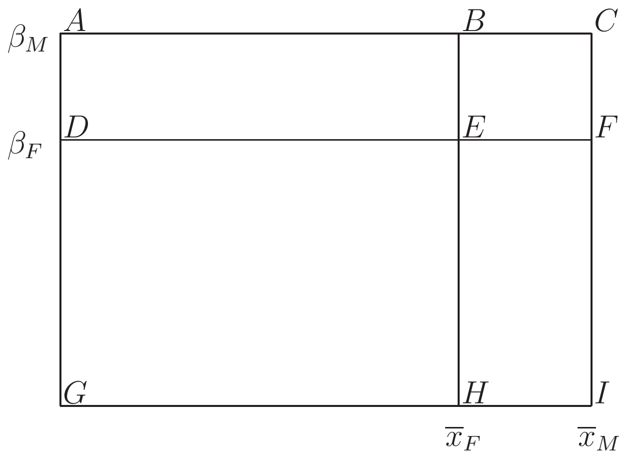
On the other hand, if all the interaction effects are zero, then for any number of factors $j \leq m$

$$total_j = \sum_{i=1}^{j} cp_i + interactions = \sum_{i=1}^{j} cp_i. \tag{53}$$

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 (16)). 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 (53)). 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 $4, 5, \ldots, m$ 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. □

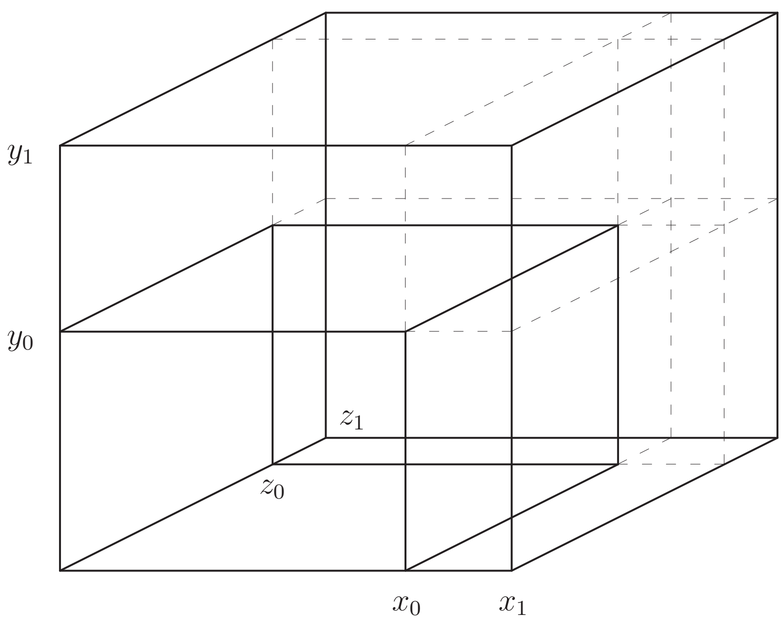
8 Figures

Figure 1 – Illustration of Oaxaca-Blinder decomposition with interaction term



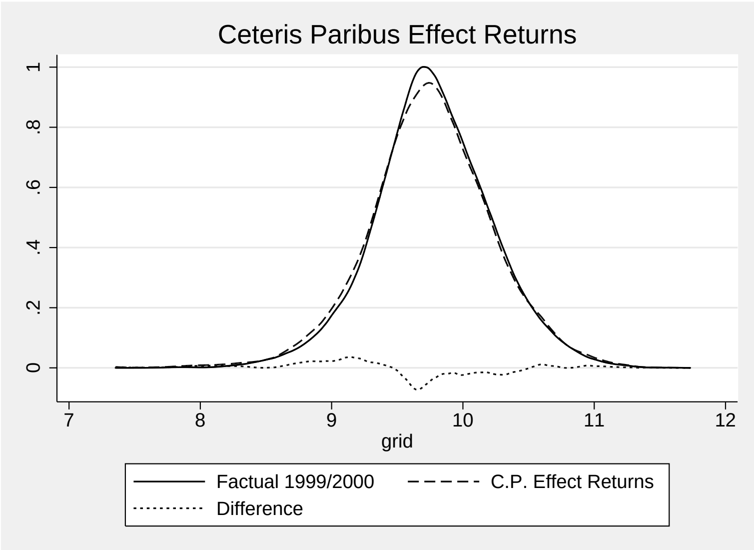
Notation: $\overline{x}_M, \overline{x}_F$ represent mens' and womens' average characteristics,
β_M, β_F represent their regression coefficients

Figure 2 – Illustration of the decomposition with three factors involving interaction effects



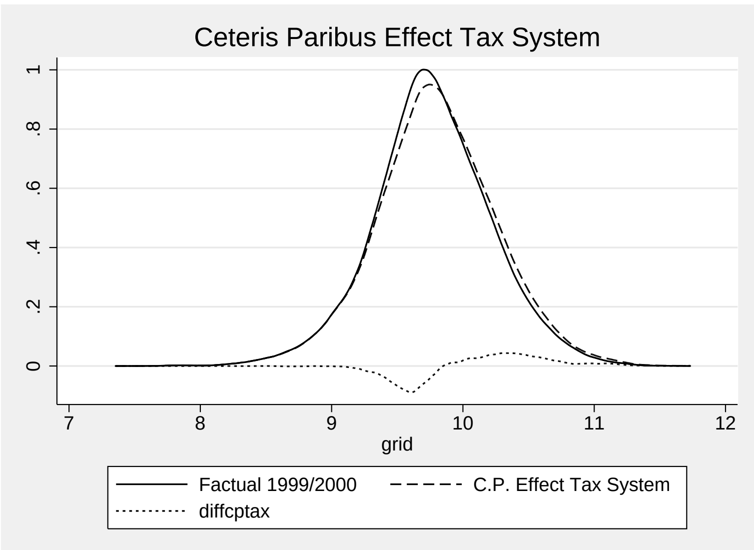
Notation: $x_0 y_0 z_0$ = reference situation, $x_1 y_1 z_1$ = target situation

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	
Other sectors				
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	-	-	-
Services sector				
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.

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

Table 3 – Decomposition with interaction effects

	Absolute	Percentage of overall change	
$total$	.0298	100	-
cp_1 (Returns)	.0105	35.09	(6.50)
cp_2 (Tax system)	.0075	25.10	(3.93)
cp_3 (Other factors)	.0147	49.41	(12.38)
int_{12}	-.0036	-12.08	(3.31)
int_{13}	.0005	1.75	(5.53)
int_{23}	.0009	3.22	(10.67)
int_{123}	-.0007	-2.51	(5.07)

Source: German Socio-Economic Panel. 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 take into account the longitudinal sample design, the clustering of observations within households, and stratification.

Table 4 – All possible sequential decompositions and Shapley decomposition

	1,2,3		1,3,2	
Returns	35.09	(6.50)	35.09	(6.50)
Tax system	13.02	(4.22)	13.73	(10.75)
Other factors	51.88	(7.48)	51.17	(11.27)
	2,3,1		3,1,2	
Returns	22.24	(7.35)	36.84	(6.18)
Tax system	25.10	(3.93)	13.73	(10.75)
Other factors	52.64	(8.66)	49.41	(12.38)
	2,1,3		3,1,2	
Returns	23.00	(5.85)	22.24	(7.35)
Tax system	25.10	(3.93)	28.33	(11.58)
Other factors	51.88	(7.48)	49.41	(12.38)
	Shapley (=average)			
Returns	29.08	(5.68)		
Tax system	19.83	(6.56)		
Other factors	51.06	(8.57)		

Source: German Socio-Economic Panel. For example, the sequence '3,1,2' means that factor 3 (= all other factors) is changed first, then factor 1 (= labor market returns), and then factor 2 (= tax system). The bootstrap standard errors shown in parentheses take into account the longitudinal sample design, the clustering of observations within households, and stratification.