DISCUSSION PAPER SERIES

IZA DP No. 6397

Decomposing the Composition Effect

Christoph Rothe

February 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Decomposing the Composition Effect

Christoph Rothe
Toulouse School of Economics
and IZA

Discussion Paper No. 6397
February 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.

IZA Discussion Paper No. 6397
February 2012

ABSTRACT
Decomposing the Composition Effect

This paper proposes a decomposition of the composition effect, i.e. the part of the observed between-group difference in the distribution of some economic outcome that can be explained by differences in the distribution of covariates. Our decomposition contains three types of components: (i) the "direct contributions" of each covariate due to between-group differences in the respective marginal distributions, (ii) several "two way" and "higher order" interaction effects due to the interplay between two or more covariates' marginal distributions, and (iii) a "dependence effect" accounting for between-group differences in dependence patterns among the covariates. Our methods can be used to decompose differences in arbitrary distributional features, like quantiles or inequality measures, and allows for general nonlinear relationships between the outcome and the covariates. It can easily be implemented in practice using standard econometric techniques. An application to wage data from the US illustrates the empirical relevance of the decomposition's components.

JEL Classification: C13, C18, C21, J31

Keywords: counterfactual distribution, decomposition methods

Corresponding author:

Christoph Rothe
Toulouse School of Economics (GREMAQ)
21 Allée de Brienne
31000 Toulouse
France
E-mail: rothe@cict.fr

1. Introduction

Understanding the factors accounting for the differences in the distributions of individuals' economic outcomes across two countries, time periods, or subgroups of the population is central in several fields of economic research, particularly in labor and development economics. For instance, the question why wage inequality has increased substantially in the US and other industrialized countries over the past decades has received enormous attention in the recent literature. Other examples include the study of the gender wage gap, wage differentials between natives and immigrants, and variations in health outcomes across several developing regions. The distributional aspect is often critical in these applications. Comparing real hourly wages among male US workers in 1985 and 2005, for example, one can observe that the median wage has remained approximately constant, but that the 90% and 10% quantile have increased by about 20% and 5%, respectively. There has thus been a substantial change in the overall shape of the wage distribution, which would not be revealed by a simple comparison of means.

With such applications in mind, a number of papers have proposed procedures to decompose between-group differences in economic outcomes into two components: a composition effect due to differences in observable covariates across groups, and a structure effect due to differences in the relationship that links the covariates to the outcome. The most popular example is certainly the Oaxaca-Blinder procedure (Oaxaca, 1973; Blinder, 1973), which can be used to decompose differences in mean outcomes when the data are generated by a simple linear model. More flexible methods that can be used to decompose general distributional features like quantiles or inequality measures, and allow for complex nonlinear relationships between the covariates and the outcome variable are proposed and studied by DiNardo, Fortin, and Lemieux (1996), Gosling, Machin, and Meghir (2000), Donald, Green, and Paarsch (2000), Barsky, Bound, Charles, and Lupton (2002), Machado and Mata (2005), Melly (2005), Chernozhukov, Fernandez-Val, and Melly (2009) and Rothe (2010b), among many others. See also Fortin, Lemieux, and Firpo (2011) for an extensive literature review. These papers differ mainly in the estimation techniques being used. The parameters that are being estimated are conceptually very similar.

When the data contain information about several individual characteristics, it is natural to ask whether one of them is "driving" the value of the composition effect. For example, one might wonder whether the composition part of the wage gap between natives and immigrants is mostly due to differences in distribution of, say, education. Such a *detailed decomposition* (Fortin et al., 2011) can easily be obtained for the mean of the outcome variable in a linear model by the Oaxaca-Blinder procedure. In more general settings, however, none of the above-mentioned more flexible methods is able to divide the composition effect into the contribution of each covariate in a similar fashion. That is, while they are able to quantify the extend to which e.g. differences in the composition of workers contribute to the native-immigrant wage gap for every quantile of the wage distribution, they are unable to determine the contribution of differences in the distribution of education alone.

Our paper addresses this issue by proposing a detailed decomposition of the composition effect that applies to any distributional feature, and does not rely on a particular structure of the data generating process. We first argue that when considering distributional features other than the mean, it is generally impossible to apportion the composition effect into components attributable to between-group differences in specific covariates such that these components add up to the full composition effect. Any attempt to define such a partition, even in the most basic empirical settings, has to rely on arbitrary choices. For example, when the data are generated by a simple linear model, between-group differences in the outcomes' variance can be due to the fact that the covariances among the explanatory variables differ across groups. Since the covariance of two random variables is the product of their respective standard deviations and the correlation coefficient, it depends on features of both covariates' marginal distributions, and on the dependence structure between them. Such terms can therefore not be attributed unambiguously to a specific covariate. Similar issues are present for essentially all distributional features that are commonly of interest in applications (with the exception of the mean). Since objects like quantiles or inequality measures are nonlinear transformations of the distribution of the outcome, they are generally not additively separable in the marginal distributions of the covariates, but contain "interaction terms" that stem from the interplay of two or more covariates. Moreover, they are also influenced by the dependence patterns among

the covariates across groups.

Our decomposition does not try to circumvent these issues by working with approximations or imposing strong restrictions on the data generating process, but takes them into account explicitly. In particular, we propose to split up the composition effect into three types of components: (i) the "direct contribution" of each covariate due to between-group differences in the respective marginal distributions, (ii) several "two way" and "higher order interaction effects" due to the interplay between two or more marginal distributions, and (iii) a "dependence effect" accounting for different dependence patterns among the covariates.

We show that such a decomposition is well-defined in a general setting by using results from copula theory (Nelsen, 2006). These results imply that the joint distribution function of the covariates can be separated into its marginals and a so-called copula function, that determines the dependence structure. Our decomposition exploits this representation by creating various counterfactual covariate distributions that share properties of both groups, and studying the resulting hypothetical outcome distributions. Similar techniques are used in Rothe (2012) to extend the notion of partial effects to unconditional distributions. Our decomposition is conceptually easy to implement, requiring only standard econometric methods to estimate copula functions and conditional CDFs. Valid standard errors can be computed via a classical bootstrap approach.

For the special case of the mean of the outcome in a linear regression model, our decomposition method reduces to the Oaxaca-Blinder procedure, and can thus be understood as a natural extension of this classical method to nonlinear models and general distributional features. Our decomposition also shares some similarities with approaches that use sequential conditioning arguments to define the impact of individual covariates on the outcome distribution, as e.g. in in DiNardo et al. (1996), Machado and Mata (2005), or Fortin et al. (2011). It also has the same aim as the so-called RIF decomposition described in Fortin et al. (2011). We review the aforementioned methods and their relationship to our approach in more detail in Section 5.

It should be stressed that our focus in this paper is exclusively on a decomposition of the composition effect. We do not address the related issue of deriving a decomposition of the structure effect, i.e. dividing between-group differences in the structural functions

that link the covariates and the outcome variable into components that can be attributed to individual covariates. Such a task already faces conceptional difficulties in simple linear models with discrete covariates, and does not seem possible for general nonlinear structural functions with interactions between the covariates.

The remainder of the paper is structured as follows. In the next section, we introduce a general setting for studying structure and composition effects. Section 3 illustrates the conceptional difficulties for decomposing the composition effect through a simple example. In Section 4, we describe our new detailed decomposition based on copulas. Section 5 compares our method to other techniques proposed in the literature. Section 6 shows how to implement our decomposition in practice, and Section 7 contains an empirical application to wage data from the US. Finally, Section 8 concludes.

2. Structure and Composition Effects

In this section, we introduce a general setting to study structure and composition effects, which covers the frameworks used in e.g. DiNardo et al. (1996), Gosling et al. (2000), Donald et al. (2000), Barsky et al. (2002), Machado and Mata (2005), Melly (2005), Chernozhukov et al. (2009), Rothe (2010b), and Fortin et al. (2011). Throughout this and the following three sections, we abstract from estimation issues and focus on defining population quantities of interest. We return to questions of empirical implementation in Section 6 and 7.

In our setting, we consider a population with two non-overlapping subgroups indexed by $i \in \{0, 1\}$. These two groups could e.g. be men and women, natives and immigrants or workers in two different countries or time periods. For any individual in group i, we observe an outcome variable Y_i and a d-dimensional vector of covariates X_i, with corresponding distribution functions denoted by F_Y^i and F_X^i, respectively. The outcome variable is generated through the nonseparable model

$$Y_i = m_i(X_i, \eta_i), \quad i \in \{0, 1\} \tag{2.1}$$

where $\eta_i \in \mathbb{R}^{d_\eta}$ is an unobserved error term, assumed to be independent of X_i, and m_i is the so-called structural function. Since we do not impose any restrictions on neither the distribution of the unobservables nor the specific functional form of m_i, writing the

relationship between Y_i and X_i as in (2.1) does not restrict the analysis in any sense. Since the function m_i and the distribution of η_i are clearly not separately identified, we assume without loss of generality that η_1 and η_0 have the same distribution. Our setting is fully nonparametric and thus very flexible, allowing e.g. for heteroskedasticity or skewness in the conditional distribution of Y_i given X_i, among many other patterns of dependence. In a typical application like the analysis of differences in wage distributions, Y_i would be an individuals' wage, X_i and η_i would correspond to individual observed and unobserved characteristics (from the perspective of the econometrician), and m_i would be the wage structure function.

Following the literature on decomposition methods, we assume that (2.1) describes a stable relationship that is not affected by changes in the distribution of the covariates. This is equivalent to assuming that the conditional CDF $F^i_{Y|X}$ of Y_i given X_i is invariant to counterfactual changes in the distribution of X_i. This condition makes it possible to construct the marginal distribution of the outcome variable in group i under a counterfactual experiment that exogenously changes the distribution of observable characteristics. Recall that by the law of total probability we have that

$$F^i_Y(y) = \int F^i_{Y|X}(y|x)dF^i_X(x) \text{ for } i \in \{0,1\}, \tag{2.2}$$

and thus for any d-dimensional CDF G the function $H(y) = \int F^i_{Y|X}(y,x)dG(x)$ can be interpreted as the CDF of the counterfactual random variable $m_i(Z,\eta_i)$, where $Z \sim G$ is a d-dimensional random vector that is independent of η_i. This idea, which is central to the decomposition literature, allows us to define the counterfactual distribution of the outcome under the structure $m_i(\cdot,\eta_i)$ of group i and the covariate distribution F^j_X of group j as

$$F^{i|j}_Y(y) = \int F^i_{Y|X}(y,x)dF^j_X(x). \tag{2.3}$$

For any CDF F we refer to objects of the form $\nu(F)$ in the following as a distributional feature, where $\nu : \mathcal{F} \to \mathbb{R}$ is a functional from the space of all one-dimensional distribution functions to the real line. Examples of distributional features include the mean, with $\nu : F \mapsto \int y dF(y)$, and the τ-quantile, with $\nu : F \mapsto F^{-1}(\tau)$, but also higher-order centered or uncentered moments, quantile-related statistics like interquantile ranges or

quantile ratios, and inequality measures such as the Gini coefficient. Using this notation, one can now decompose the observed difference

$$\Delta_O^\nu = \nu(F_Y^1) - \nu(F_Y^0)$$

between the distributional features $\nu(F_Y^1)$ and $\nu(F_Y^0)$ as follows:

$$\Delta_O^\nu = \Delta_S^\nu + \Delta_X^\nu \tag{2.4}$$

where

$$\Delta_S^\nu = \nu(F_Y^1) - \nu(F_Y^{0|1}) \text{ and } \Delta_X^\nu = \nu(F_Y^{0|1}) - \nu(F_Y^0).$$

Here Δ_X^ν is a composition effect, solely due to differences in the distribution of the covariates between the two groups, and Δ_S^ν is a structure effect, solely due to differences in the structural functions m_1 and m_0. Note that the order in which differences in structural functions and the covariate distributions are considered when defining Δ_X^ν and Δ_S^ν has to be taken into account when interpreting the parameters in practice.

3. Problems for Decomposing the Composition Effect

When the data contain information about several individual characteristics, it is natural to ask whether one of them is "driving" the value of the composition effect. For example, one might wonder whether the composition part of the wage gap between natives and immigrants is mostly due to differences in distribution of, say, education. More generally, it seems desirable to have a methodology that is able to apportion the composition effect into components attributable to each covariate. Such a *detailed decomposition* (Fortin et al., 2011) can easily be obtained for the mean of the outcome variable in a linear model by the Oaxaca-Blinder procedure, which certainly contributes to the popularity of the method. When considering more general distributional distributional features, however, like the variance or the quantiles of the outcomes, it is impossible to define a similar decomposition without making subjective choices. This is due to two reasons. First, it is generally not possible to express a distributional feature $\nu(F_Y^i) = \nu(\int F_{Y|X}^i(\cdot, x) dF_X^i(x))$ as a sum of terms that each depend on the marginal distribution of a single covariate only. Instead, explicit expressions for $\nu(F_Y^i)$ typically contain "interaction effects" resulting

from the interplay of two or more marginal distributions, which cannot be attributed unambiguously to a single covariate. Second, in general a distributional feature $\nu(F_Y^i)$ does not only depend on the marginal distributions of the covariates, but also on the dependence patterns among them. Dependence patterns describe the relationship among at least two covariates, and can therefore not be attributed unambiguously to a single covariate.

Both issues are a necessary consequence of considering features that are a nonlinear transformations of the distribution function of the outcome variable. They are present even for simple data generating processes like linear models. To see this, consider the following trivial example. Suppose that in group $i \in \{0, 1\}$ we have that $m_i(X_i, \eta_i) = X_{i1} + X_{i2} + \eta_i$, where $X_i \sim N(\mu_i, \Sigma_i)$ is bivariate normal with

$$\mu_i = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \end{pmatrix} \text{ and } \Sigma_i = \begin{pmatrix} \sigma_{i1}^2 & \rho_i \sigma_{i1} \sigma_{i2} \\ \rho_i \sigma_{i1} \sigma_{i2} & \sigma_{i2}^2 \end{pmatrix},$$

and $\eta_i \sim N(0, 1)$. Since the structural function is the same in both groups, we clearly have that $\Delta_S^\nu = 0$ for any functional ν, and thus any difference in the distribution of Y_1 and Y_0 reflects a composition effect. For instance, if $\nu(F_Y^i) = \text{Var}(Y_i)$ we have that $\Delta_X^\nu = \text{Var}(Y_1) - \text{Var}(Y_0)$ with $\text{Var}(Y_i) = 1 + \sigma_{i1}^2 + \sigma_{i2}^2 + 2\rho_i \sigma_{i1} \sigma_{i2}$. When $\nu(F_Y^i) = Q_Y^i(\tau)$ is the τ-quantile of the distribution of Y_i, we find that $\Delta_X^\nu = Q_Y^1(\tau) - Q_Y^0(\tau)$ with $Q_Y^i(\tau) = \mu_{i1} + \mu_{i2} + \Phi^{-1}(\tau)\sqrt{1 + \sigma_{i1}^2 + \sigma_{i2}^2 + 2\rho_i \sigma_{i1} \sigma_{i2}}$ and Φ the CDF of the standard normal distribution. In both cases, the object of interest is not additively separable in the parameters $(\mu_{i1}, \sigma_{i1}^2)$ and $(\mu_{i,2}, \sigma_{i,2}^2)$, which characterize the covariates' marginal distributions in this example. Moreover, in both cases the object of interest depends on the value of the correlation coefficient ρ_i, which determines the dependence structure among the covariates here.

The above discussion illustrates that a detailed decomposition with d elements, one for each covariate, that add up to the full composition effect will generally have to rely on arbitrary choices regarding which parts of the composition effect are attributed to specific covariates. To avoid making such choices, this paper considers a more general detailed decomposition containing additional terms that explicitly collect those parts of the composition effect that cannot be attributed to a particular covariate.

4. A New Detailed Decomposition Based on Copulas

We now propose a new decomposition of the composition effect which accounts for the two main problems listed in the previous section. In particular, our decomposition does not only contain terms accounting for the direct contribution of differences in the marginal distribution of individual covariates, but also several terms accounting for two-way and higher-order "interaction effects", and a term accounting for differences in the dependence structure of the two covariate distributions.

The copula function of the covariate distribution plays a central role in the development of our decomposition. Roughly speaking, copulas are functions that connect multivariate distribution functions to their univariate marginals (Trivedi and Zimmer, 2007). In particular, it follows from Sklar's Theorem (Sklar, 1959; Nelsen, 2006, Theorem 2.3.3) that the CDF of X_i can always be written as

$$F_X^i(x) = C^i(F_{X_1}^i(x_1), \ldots, F_{X_d}^i(x_d)) \text{ for } i \in \{0, 1\}, \tag{4.1}$$

where C^i is a copula function, i.e. a multivariate CDF with standard uniformly distributed marginals, and $F_{X_k}^i$ is the marginal distribution of the kth component of X_i. In this representation, the copula is uniquely determined on the range of the marginal distribution functions.[1] The copula function characterizes the joint distribution of ranks in the respective marginal distributions, and due to its uniqueness property it can be interpreted as the object that determines the patterns of dependence among the components of X_i. Equation (4.1) thus provides a way to disentangle the shape of the marginal CDFs and the dependence among the components of X_i.

To describe how the copula representation (4.1) can be used to decompose the compo-

[1]This means that if all components of X_i are continuously distributed, there exists a one-to-one relationship between the copula C^i and the multivariate CDF F_X^i. In the presence of at least one discrete component, we can uniquely determine $C^i(u_1, \ldots, u_d)$ for all $(u_1, \ldots, u_d) \in \text{Ran}(F_{X_1}^i) \times \ldots \times \text{Ran}(F_{X_d}^i)$. In this case, the definition of C^i can be made unique on the entire set $[0, 1]^d$ by assuming that the discrete components of X_i are generated through a threshold-crossing model with a continuously distributed latent variable. That is, suppose that $X_i = g(X_i^*) = (g_1(X_{i1}^*), \ldots, g_d(X_{id}^*))$ for some continuously distributed random vector X_i^* and a function $g(\cdot)$ that is weakly increasing in each of its arguments. For example, if the kth component of X_i is binary, we could have $X_{ik} = \mathbb{I}\{X_{ik}^* > 0\} = g_k(X_{ik}^*)$. With such a structure, C^i can be uniquely defined on $[0, 1]^d$ as the copula function of the joint CDF of the latent variables X_i^*.

sition effect, we introduce some further notation. In the following, every boldface letter denotes an element of the d-dimensional product set $\{0,1\}^d$. We also write $\mathbf{1_d} = (1,1,\ldots,1)$ and $\mathbf{0_d} = (0,0,\ldots,0)$, and for any $j \in \{1,\ldots,d\}$ we denote by $\mathbf{e_j}$ the jth unit vector, i.e. the d-dimensional vector whose jth component is equal to one, and whose remaining components are equal to zero. Furthermore, for any $\mathbf{k} \in \{0,1\}^d$ we define $|\mathbf{k}| = \sum_{l=1}^{d} \mathbf{k}_l$. Finally, we introduce a general notation for counterfactual outcome distributions, writing

$$F_Y^{i|j,\mathbf{k}}(y) = \int F_{Y|X}^{i}(y,x) dC^j(F_{X_1}^{\mathbf{k}_1}(x_1),\ldots,F_{X_d}^{\mathbf{k}_d}(x_d)) \tag{4.2}$$

for the distribution of the outcome in a counterfactual setting where the structure is as in group i, the covariate distribution has the copula function of group j, and the marginal distribution of the lth covariate is equal to the that in group $\mathbf{k}_l$. For example, we have that $F_Y^1 = F_Y^{1|1,\mathbf{1_d}}$ and $F_Y^0 = F_Y^{0|0,\mathbf{0_d}}$.

With the notation (4.2), we can now easily consider counterfactual settings that mix various features of the two groups' covariate distributions. As a first step, the composition effect Δ_X^ν can be decomposed into a *dependence effect* Δ_D^ν resulting from between-group differences in the copula functions, and a *total marginal distribution effect* Δ_M^ν resulting from differences in the marginal covariate distributions across the two groups:

$$\Delta_X^\nu = \Delta_D^\nu + \Delta_M^\nu \tag{4.3}$$

where

$$\Delta_D^\nu = \nu(F_Y^{0|1,\mathbf{1_d}}) - \nu(F_Y^{0|0,\mathbf{1_d}}) \text{ and } \Delta_M^\nu = \nu(F_Y^{0|0,\mathbf{1_d}}) - \nu(F_Y^{0|0,\mathbf{0_d}}).$$

In the simple example with bivariate normal covariates given in Section 3, the term Δ_D^ν would capture the contribution of differences in correlation coefficients across the two groups, whereas the term Δ_M^ν subsumes the contribution of different mean and variance parameters. Note that the order in which differences in the copula and the marginal CDFs are considered when defining Δ_D^ν and Δ_M^ν has to be taken into account when interpreting the parameters in practice.

In a second step, we then further decompose the total marginal distribution effect Δ_M^ν into several *partial marginal distribution effects* $\Delta_M^\nu(\mathbf{k})$, which account for between-group

differences in the marginal distributions of one or several covariates:

$$\Delta_M^\nu = \sum_{1 \leq |\mathbf{k}| \leq d} \Delta_M^\nu(\mathbf{k}), \tag{4.4}$$

where for any $\mathbf{k} \in \{0,1\}^d$ we define

$$\Delta_M^\nu(\mathbf{k}) = \sum_{1 \leq |\mathbf{m}| \leq |\mathbf{k}|} (-1)^{|\mathbf{k}|-|\mathbf{m}|} (\nu(F_Y^{0|0,\mathbf{m}}) - \nu(F_Y^{0|0,\mathbf{0_d}})).$$

For $|\mathbf{k}| = 1$, i.e. when $\mathbf{k} = \mathbf{e_j}$ is the jth unit vector, the term $\Delta_M^\nu(\mathbf{e_j}) = \nu(F_Y^{0|0,\mathbf{e_j}}) - \nu(F_Y^{0|0,\mathbf{0_d}})$ can be interpreted as the effect of conducting a counterfactual experiment in group zero that entails changing the marginal distribution of the jth covariate to its corresponding counterpart in group one, while holding everything else constant. It thus measures the direct contribution of between-group differences in the marginal distribution of the jth covariate to the composition effect. Note that these direct contributions are equal to Fixed Partial Distributional Policy Effects (FPPE) introduced in Rothe (2012), and thus have an interesting interpretation outside the present decomposition framework. For any $\mathbf{k}$ with $|\mathbf{k}| > 1$, the terms $\Delta_M^\nu(\mathbf{k})$ capture the contributions to the composition effect of "$|\mathbf{k}|$-way interaction effects" between the marginal distributions for which respective component of $\mathbf{k}$ is equal to one. Consider for example the special case $d = 2$, in which the only interaction term $\Delta_M^\nu(\mathbf{1_2})$ is given by

$$\Delta_M^\nu(\mathbf{1_2}) = (\nu(F_Y^{0|0,\mathbf{1_2}}) - \nu(F_Y^{0|0,\mathbf{0_2}})) - \Delta_M^\nu(\mathbf{e_1}) - \Delta_M^\nu(\mathbf{e_2}).$$

Here the first term on the right-hand side measures the *joint* contribution of differences in the marginal distribution of covariates between the two groups, whereas the second and third summand provide an adjustment for the *direct* effect of differences in one specific marginal distribution alone, thus leading to the interpretation of $\Delta_M^\nu(\mathbf{1_2})$ as a "pure" interaction effect. A similar interpretation applies in higher-dimensional settings.

As a further illustration, one can explicitly calculate the components of our detailed decomposition for the simple setting described in Section 3, where in group $i \in \{0,1\}$ the structural function is given by $m_i(X_i, \eta_i) = X_{i1} + X_{i2} + \eta_i$, the covariates $X_i \sim N(\mu_i, \Sigma_i)$ are bivariate normal with

$$\mu_i = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \end{pmatrix} \text{ and } \Sigma_i = \begin{pmatrix} \sigma_{i1}^2 & \rho_i \sigma_{i1} \sigma_{i2} \\ \rho_i \sigma_{i1} \sigma_{i2} & \sigma_{i2}^2 \end{pmatrix},$$

and $\eta_i \sim N(0,1)$. Here the the copula C^i is a bivariate Gaussian copula with one-dimensional parameter ρ_i. For the case that $\nu(F_Y^i) = \mathrm{Var}(Y_i)$, we find that the dependence structure and total marginal distribution effects are given by

$$\begin{aligned}
\Delta_D^{\mathrm{Var}} &= 2(\rho_1 - \rho_0)\sigma_{11}\sigma_{12} \text{ and} \\
\Delta_M^{\mathrm{Var}} &= (\sigma_{11}^2 - \sigma_{01}^2) + (\sigma_{12}^2 - \sigma_{02}^2) + 2\rho_0(\sigma_{11}\sigma_{12} - \sigma_{01}\sigma_{02}),
\end{aligned}$$

and the partial marginal distribution effects are

$$\begin{aligned}
\Delta_M^{\mathrm{Var}}(\mathbf{e_1}) &= (\sigma_{11}^2 - \sigma_{01}^2) + 2\rho_0\sigma_{02}(\sigma_{11} - \sigma_{01}), \\
\Delta_M^{\mathrm{Var}}(\mathbf{e_2}) &= (\sigma_{12}^2 - \sigma_{02}^2) + 2\rho_0\sigma_{01}(\sigma_{12} - \sigma_{02}), \text{ and} \\
\Delta_M^{\mathrm{Var}}(\mathbf{1_2}) &= -2\rho_0(\sigma_{01}\sigma_{12} + \sigma_{02}\sigma_{11}).
\end{aligned}$$

For $\nu(F_Y^i) = Q_Y^i(\tau)$ the distributional feature of interest, we find that

$$\begin{aligned}
\Delta_D^{Q(\tau)} &= \Phi^{-1}(\tau)\left(\sqrt{1 + \sigma_{11}^2 + \sigma_{12}^2 + 2\rho_1\sigma_{11}\sigma_{12}} - \sqrt{1 + \sigma_{11}^2 + \sigma_{12}^2 + 2\rho_0\sigma_{11}\sigma_{12}}\right) \text{ and} \\
\Delta_M^{Q(\tau)} &= \Phi^{-1}(\tau)\left(\sqrt{1 + \sigma_{11}^2 + \sigma_{12}^2 + 2\rho_0\sigma_{11}\sigma_{12}} - \sqrt{1 + \sigma_{01}^2 + \sigma_{02}^2 + 2\rho_0\sigma_{01}\sigma_{02}}\right) \\
&\quad + (\mu_{11} + \mu_{12}) - (\mu_{01} + \mu_{02}),
\end{aligned}$$

and that the partial marginal distribution effects are

$$\begin{aligned}
\Delta_M^{Q(\tau)}(\mathbf{e_1}) &= \Phi^{-1}(\tau)\left(\sqrt{1 + \sigma_{11}^2 + \sigma_{02}^2 + 2\rho_0\sigma_{11}\sigma_{02}} - \sqrt{1 + \sigma_{01}^2 + \sigma_{02}^2 + 2\rho_0\sigma_{01}\sigma_{02}}\right) \\
&\quad + (\mu_{11} - \mu_{01}), \\
\Delta_M^{Q(\tau)}(\mathbf{e_2}) &= \Phi^{-1}(\tau)\left(\sqrt{1 + \sigma_{01}^2 + \sigma_{12}^2 + 2\rho_0\sigma_{01}\sigma_{12}} - \sqrt{1 + \sigma_{01}^2 + \sigma_{02}^2 + 2\rho_0\sigma_{01}\sigma_{02}}\right) \\
&\quad + (\mu_{12} - \mu_{02}), \text{ and} \\
\Delta_M^{Q(\tau)}(\mathbf{1_2}) &= \Phi^{-1}(\tau)\left(\sqrt{1 + \sigma_{11}^2 + \sigma_{12}^2 + 2\rho_0\sigma_{11}\sigma_{12}} + \sqrt{1 + \sigma_{01}^2 + \sigma_{02}^2 + 2\rho_0\sigma_{01}\sigma_{02}}\right) \\
&\quad - \Phi^{-1}(\tau)\left(\sqrt{1 + \sigma_{11}^2 + \sigma_{02}^2 + 2\rho_0\sigma_{11}\sigma_{02}} + \sqrt{1 + \sigma_{01}^2 + \sigma_{12}^2 + 2\rho_0\sigma_{01}\sigma_{12}}\right)
\end{aligned}$$

Of course, other distributional features could be considered as well.

In a setting with several explanatory variables, the large number of subcomponents of Δ_M^ν often makes it infeasible to report all of them in empirical applications. In most cases, we expect the "direct" marginal distribution effects $\Delta_M^\nu(\mathbf{e_j})$ to be the most important

(and most interesting) elements of the decomposition. If that turns out to be the case, one possibility to keep the presentation of findings clear and concise is to report these quantities together with the "aggregate interaction effect" $\sum_{2\leq|\mathbf{k}|\leq d}\Delta_M^{\nu}(\mathbf{k})$, instead of computing (and reporting) every single two-way or higher-order interaction term.

5. Relationship to other Approaches

In this section, we compare our copula decomposition to several related approaches that have been proposed in the literature. We consider the classical Oaxaca-Blinder procedure, methods based on so-called sequential conditioning arguments, and a recently proposed method based on the RIF regression approach in Firpo, Fortin, and Lemieux (2009).

5.1. Approaches based on Linear Models. In the context of the linear regression model, the Oaxaca-Blinder procedure, due to Oaxaca (1973) and Blinder (1973), is arguably the most widely used approach of decomposing inter-group differences in mean levels of an outcome variable into a structure effect and a composition effect. It can also used to obtain a detailed decomposition of the composition effect. Using our notation, the Oaxaca-Blinder procedure assumes that $m_i(x,\eta_i)=\beta_{i0}+\sum_{j=1}^{d}x_j\beta_{i,j}+\eta_i$, and considers the case that $\nu(F_Y^i)=\int ydF_Y^i(y)=\mathbb{E}(Y_i)$. Under these conditions, the composition effect can be written as

$$\Delta_X^{\mathbb{E}}=\sum_{j=1}^{d}(\mathbb{E}(X_{1,j})-\mathbb{E}(X_{0,j}))\beta_{0j},$$

and the summands on the right hand side are the elements of the detailed decomposition, with the jth summand being the contribution of the jth covariate.

The reason that none of the problems for decomposing the composition effect listed in Section 3 are apparent for the Oaxaca-Blinder procedure is due to the special structure of the framework. The additive separability of covariates in the data generating process is preserved by the linearity of the functional $F\mapsto\int ydF(y)$ that maps a CDF into the corresponding expectation. Under these particular circumstances, the composition effect can be apportioned unambiguously into contributions of each covariate. However, this is not at odds with our approach. In fact, our method encompasses the Oaxaca-Blinder procedure as a special case, as we have $\Delta_M^{\mathbb{E}}(\mathbf{e_j})=(\mathbb{E}(X_{1j})-\mathbb{E}(X_{0j}))\beta_{0j}$, $\Delta_M^{\mathbb{E}}(\mathbf{k})=0$ for

$\mathbf{k} \neq \mathbf{e_j}$ and $\Delta_D^{\mathbb{E}} = 0$ in this setting. Our approach can thus be understood as a natural extension of the Oaxaca-Blinder decomposition to nonlinear DGPs and general features of the outcome distribution.

5.2. Approaches based on Sequential Conditioning. There exists an extensive literature that has proposed measures to quantify the impact of individual covariates on between-group differences in general distributional features. Examples include DiNardo et al. (1996), Machado and Mata (2005), Altonji, Bharadwaj, and Lange (2008) or Fortin et al. (2011), among many others. To define the respective object of interest, these papers make use of sequential conditioning arguments. Generally speaking, the idea that they share is to write the joint distribution of the covariates as the product of the marginal distribution of its mth component and the conditional distribution of the mth component given the remaining ones, and then manipulate one of the two factors in a counterfactual experiment.

More formally, let X_{-m} denote the $(d-1)$-dimensional subvector of X that does not contain the mth component, and define a counterfactual outcome CDF by

$$G_Y^{i,j,k|m}(y) = \int F_{Y|X}^i(y,x) dF_{X_m|X_{-m}}^j(x_m|x_{-m}) dF_{X_{-m}}^k(x_{-m}).$$

With this notation, one way to define the impact of the mth covariate is through a change in the conditional distribution of X_m given X_{-m}:

$$\widetilde{\Delta}_M^{\nu}(m) = \nu(G_Y^{0,1,0|m}) - \nu(F_Y^0).$$

Another possibility is to define the effect through a counterfactual change in the marginal distribution of X_m while holding the conditional distribution constant:

$$\widetilde{\Delta}_M^{\nu*}(m) = \nu(G_Y^{0,0,1|m}) - \nu(F_Y^0).$$

As argued in Rothe (2012), neither type of parameter can be interpreted as the impact of between-group differences in the marginal distribution of X_m alone: the counterfactual experiment used to define $\widetilde{\Delta}_M^{\nu}(m)$ changes the conditional distribution of X_m given X_{-m}, and thus affects the dependence structure among the covariates, and the experiment used to define $\widetilde{\Delta}_M^{\nu*}(m)$ changes the marginal distribution of X_{-m}. Both types of parameters

differ in this respect from our "direct" marginal distribution effects $\Delta_M^\nu(\mathbf{e_m})$. As they are defined through an experiment that holds the copula function constant, the latter can be interpreted as reflecting differences in the marginal distribution of X_m alone. A further difference to our decomposition is that neither $\sum_{m=1}^d \widetilde{\Delta}_M^\nu(m)$ nor $\sum_{m=1}^d \widetilde{\Delta}_M^{\nu*}(m)$ are equal to the composition effect Δ_M^ν in general. When defining a detailed decomposition of the composition effect based on one of these approaches, one would thus end up with an unexplained residual quantity that is difficult to interpret.

In order to avoid the last problem, one could use a modified sequential conditioning argument and define another counterfactual outcome distribution

$$H_Y^{i|\mathbf{k}}(y) = \int F_{Y|X}^i(y,x) dF_{X_1|X_2,\ldots,X_d}^{\mathbf{k}_1}(x_1|x_2,\ldots,x_d)$$
$$\times\, dF_{X_2|X_3,\ldots,X_d}^{\mathbf{k}_2}(x_2|x_3,\ldots,x_d) \times \ldots \times dF_{X_d}^{\mathbf{k}_d}(x_d),$$

for any $\mathbf{k} \in \{0,1\}^d$. If we now let $\mathbf{h_m} = (1,1,\ldots,1,0,0,\ldots,0)$ be a d-vector whose first m components are equal one, and whose remaining $d-m$ components are equal to zero, the composition effect can be written as

$$\Delta_X^\nu = \sum_{m=1}^d (\nu(H_Y^{0|\mathbf{h_m}}) - \nu(H_Y^{0|\mathbf{h_{m-1}}})).$$

While now the components of the decomposition "add up" to the composition effect, such an approach has the disadvantage of being "path dependent", in the sense that the results depend on the ordering of the random variables in the vectors X_i, which is of course arbitrary. However, proceeding like this could be useful is there is a natural causal ordering among the covariates, e.g. if they are chosen sequentially by the individuals.

5.3. Approach based on RIF Regressions. In this subsection, we contrast the decomposition methodology proposed in this paper with a related procedure based on the RIF regression techniques in Firpo et al. (2009), which is outlined in e.g. Fortin et al. (2011, Section 5.2) and Firpo, Fortin, and Lemieux (2007, 2011). To describe the approach, we need to introduce some notation. For any CDF F, denote the influence function (Van der Vaart, 2000, p. 291) of the distributional feature $\nu(F)$ by $\mathrm{IF}(\cdot;\nu,F)$, and let $\mathrm{RIF}(\cdot;\nu,F) = \nu(F) + \mathrm{IF}(\cdot;\nu,F)$ be the corresponding recentered influence function. Recall that any influence function satisfies that $\int \mathrm{IF}(y;\nu,F)dF(y) = 0$, so that

the expectation of $\mathrm{RIF}(Y;\nu,F)$ with $Y \sim F$ is the distributional feature of interest. Next, consider the coefficients γ_i^ν of a least-squares approximation of the RIF by a linear function of the covariates, i.e.

$$\gamma_i^\nu = \operatorname*{argmin}_\gamma \mathbb{E}((\mathrm{RIF}(Y_i;\nu,F_Y^i) - \sum_{k=0}^d X_{i,k}\gamma_{i,k})^2)$$

where $X_{i,0} \equiv 1$ denotes a constant, and the coefficients $\gamma_{i|j}^\nu$ of a weighted linear least-squares approximation, i.e.

$$\gamma_{i|j}^\nu = \operatorname*{argmin}_\gamma \mathbb{E}(\Psi^{j|i}(X_i)(\mathrm{RIF}(Y_i|\nu,F_Y^{i|j}) - \sum_{k=0}^d X_{i,k}\gamma_k)^2),$$

where $F_Y^{i|j}$ is as defined in (2.3) above, and $\Psi^{j|i}(x) = dF_X^j(x)/dF_X^i(x)$ is the between-group ratio of the covariates' density functions. Here the weights $\Psi^{j|i}(X_i)$ are chosen such that the weighted distribution of X_i is equal to that of X_j. Noting that

$$\nu(F_Y^i) = \sum_{k=0}^d \mathbb{E}(X_{i,k})\gamma_{i,k}^\nu \quad \text{and} \quad \nu(F_Y^{i|j}) = \sum_{k=0}^d \mathbb{E}(X_{j,k})\gamma_{i|j,k}^\nu, \tag{5.1}$$

one can rewrite the usual structure and composition effect in the decomposition (2.4) as

$$\Delta_S^\nu = \sum_{k=0}^d \mathbb{E}(X_{1,k})(\gamma_{1,k}^\nu - \gamma_{0|1,k}^\nu) \quad \text{and} \quad \Delta_X^\nu = \widetilde{\Delta}_{X,p}^\nu + \widetilde{\Delta}_{X,e}^\nu, \tag{5.2}$$

respectively, where

$$\begin{aligned} \widetilde{\Delta}_{X,p}^\nu &= \sum_{k=1}^d (\mathbb{E}(X_{1,k}) - \mathbb{E}(X_{0,k}))\gamma_{0,k}^\nu, \\ \widetilde{\Delta}_{X,e}^\nu &= \sum_{k=1}^d \mathbb{E}(X_{1,k})(\gamma_{0|1,k}^\nu - \gamma_{0,k}^\nu). \end{aligned} \tag{5.3}$$

The summands $(\mathbb{E}(X_{1,k}) - \mathbb{E}(X_{0,k}))\gamma_{0,k}^\nu$ on the right-hand side of (5.3) are interpreted by Firpo et al. (2011) as the direct contribution of differences in the kth covariate distribution to the composition effect, and $\widetilde{\Delta}_{X,e}^\nu$ is interpreted as a specification error resulting from the fact that the conditional expectation $h_i(x;\nu,F_Y^i) = \mathbb{E}(\mathrm{RIF}(Y_i;\nu,F_Y^i)|X_i = x)$ is not a linear function of x for most features ν. Sample counterparts of these quantities can easily be estimated in practice, requiring only a (weighted) linear regression of a suitably

transformed version of the dependent variable on the covariates, and the calculation of sample means.[2]

For the special case that the data generating process is a simple linear model and $\nu(F_Y^i) = \mathbb{E}(Y_i)$, the RIF decomposition coincides with the Oaxaca-Blinder procedure, as $\text{RIF}(y; \nu, F) = y$ in this case. For other distributional features like variances or quantiles, however, the elements $(\mathbb{E}(X_{1,k}) - \mathbb{E}(X_{0,k}))\gamma_{0,k}^{\nu}$ of the detailed RIF decomposition are generally difficult to interpret in practice. For example, it is immediate that the RIF decomposition suggests by construction that the kth covariate does not contribute "directly" to the composition effect as long as $E(X_{1,k}) = E(X_{0,k})$, irrespective of the properties of the data generating process. More generally, the interpretation of the RIF decomposition is problematic whenever the covariates affect the respective feature of the outcome distribution through channels other than their means, which is true for essentially all common distributional features other than the mean. Another issue is that the "specification error" $\widetilde{\Delta}_{X,e}^{\nu}$ incurred through the linear prediction step can be of substantial magnitude even in simple settings. These points are further illustrated in the following example.

Example 2. *Suppose that for each $i \in \{0, 1\}$ the data are generated by a simple linear model as $Y_i = \beta_{i,0} + \sum_{k=1}^{d} X_{i,k}\beta_{i,k} + \eta_i$ with $\eta_1 \overset{d}{=} \eta_0$, that within each group the covariates are fully independent of each other and normally distributed, and that the variance is the distributional feature of interest, i.e. $\nu(F_Y^i) = Var(Y_i)$. In this case, the composition effect is easily seen to be*

$$\Delta_X^{Var} = \sum_{k=1}^{d} (\, Var(X_{1,k}) - \, Var(X_{0,k}))\beta_{0,k}^2.$$

Due to the independence of the covariates, the problems discussed in Section 3 do not appear in this case, and the composition effect can be written as the sum of d terms that each depend on the marginal distribution of a single covariate only. One would thus expect any reasonable decomposition procedure to report the "direct" contribution of the kth covariate to the composition effect to be equal to $(\, Var(X_{1,k}) - \, Var(X_{0,k}))\beta_{0,k}^2$ in this

[2]Note that since one has to estimate $\nu(F_Y^0)$, $\nu(F_Y^0)$ and $\nu(F_Y^{0|1})$ in order to estimate the corresponding RIF, the linear least-squares approximation step is unnecessary if one is only interested in Δ_S^{ν} and Δ_X^{ν}.

case. However, this is not true for the RIF decomposition. Since here $RIF(Y_0;\nu,F_Y^0) = (Y_0-\mathbb{E}(Y_0))^2 = (\eta_0+\sum_{k=1}^d(X_{0,k}-\mathbb{E}(X_{0,k}))\beta_{0,k})^2$*, it follows from the symmetry properties of the covariate distribution that* $\gamma_{0,0}^{Var} = Var(Y_0)$ *and* $\gamma_{0,k}^{Var} = 0$ *for* $k \geq 1$*, and thus we find that*

$$\widetilde{\Delta}_{X,p}^{Var} = 0 \quad \textit{and} \quad \widetilde{\Delta}_{X,e}^{Var} = \sum_{k=1}^{d}(\,Var(X_{1,k}) - Var(X_{0,k}))\beta_{0,k}^2 = \Delta_X^{Var}.$$

Applying the RIF decomposition would thus suggests that there is no *"direct" contribution from* any *of the covariates to the composition effect. The entire composition effect would be attributed to the "specification error"* $\widetilde{\Delta}_{X,e}^{Var}$*, irrespective of the actual between-group differences in the covariates' distributions.* □

While the last example is certainly a polar case, it should nevertheless be a reason for concern. Since the elements of the RIF decomposition can deviate by potentially large amounts from intuitive population quantities in simple linear models, one would certainly not expect them to have systematically better properties in more complex settings. To understand the underlying cause and magnitude of the problem, note that the equality (5.1) does not imply a linear relationship between the feature of interest and the means of the covariates that is invariant to changes in the distribution of X_1 and X_0. The identity is not a consequence of a special property of the RIF, but follows mechanically from the fact that the residuals from any least-squares approximation have mean zero by construction. What follows however from the results in Firpo et al. (2009) is that the relationship

$$\nu(F_Y^i) = \sum_{k=1}^{d}\mathbb{E}(X_{0,k})\theta_{0,k}^{\nu},$$

with $\theta_{0,k}^{\nu} = \mathbb{E}(\partial h_0(X_0;\nu,F_Y^0)/\partial X_{0,k})$ the average derivative of the conditional expectation of the RIF, is indeed *locally* valid for infinitesimal location shifts in the distribution of the covariates. One can thus show that the order of magnitude of the approximation error incurred by the RIF decomposition tends to be smaller if for every $k \in \{1,\ldots,d\}$ (i) the coefficient $\gamma_{0,k}^{\nu}$ is a good approximation of the average derivatives $\theta_{0,k}^{\nu}$, (ii) the relationship between the distributions of $X_{1,k}$ and $X_{0,k}$ can be well approximated by a location shift, i.e. $F_{X_k}^1(x) \approx F_{X_k}^0(x-\bar{e}_k)$ with $\bar{e}_k = \mathbb{E}(X_{1,k})-\mathbb{E}(X_{0,k})$, and (iii) the location shift parameter $\bar{e}_k$ is "sufficiently small" in absolute value.

Under these fairly restrictive conditions, one can also show that the elements of the detailed RIF decomposition can be interpreted as approximations of our "direct" partial composition effects $\Delta_M(\mathbf{e_k})$. To see this, recall that these terms were defined as

$$\Delta_M^{\nu}(\mathbf{e_k}) = \nu\left(\int F_{Y|X}^0(y,x)dC(F_{X_1}^0(x_1),\ldots,F_{X_k}^1(x_k),\ldots,F_{X_d}^0(x_d))\right) - \nu(F_Y^0)$$

in Section 4. It then follows under appropriate smoothness conditions on the functional ν, the conditional CDF $F_{Y|X}^0$ and the copula function C^0 that

$$\begin{aligned}
\Delta_M^{\nu}(\mathbf{e_k}) &\approx \nu\left(\int F_{Y|X}^0(y,(x_1,\ldots,x_k+\bar{e}_k,\ldots,x_d))dF_X^0(x)\right) - \nu(F_Y^0) \\
&\approx \bar{e}_k \frac{\partial}{\partial \bar{e}_k}\nu\left(\int F_{Y|X}^0(y,(x_1,\ldots,x_k+\bar{e}_k,\ldots,x_d))dF_X^0(x)\Big|_{\bar{e}=0}\right) \\
&= (\mathbb{E}(X_{1,k}) - \mathbb{E}(X_{0,k}))\theta_{i,k}^{\nu} \\
&\approx (\mathbb{E}(X_{1,k}) - \mathbb{E}(X_{0,k}))\gamma_{0,k}^{\nu},
\end{aligned}$$

where the errors incurred in the first, second, and forth step of the derivation are of the order $O(\sup_{x\in\mathbb{R}}|F_{X_k}^1(x) - F_{X_k}^0(x-\bar{e}_k)|)$, $O(\|\bar{e}_k\|^2)$ and $O(\|\theta_{0,k}^{\nu} - \gamma_{0,k}^{\nu}\|)$, respectively. Here the first and second step exploit the smoothness conditions, the third step follows from an application of the chain rule as in Rothe (2010a), and the fourth step is immediate. As illustrated above, the error terms can be of substantial absolute magnitude even in simple settings. Since the RIF decomposition at best provides a first-order approximation to our copula decomposition, and the latter can be computed without error, the method seems to be only useful in very particular settings with severe computational constraints.

6. Model Specification, Estimation, and Inference

The focus of this paper has so far been on defining a detailed decomposition of the composition effect in terms of population quantities. Proceeding like this is instructive, as it clarifies which objects we want to learn about in principle. Of course, in any empirical application these quantities are unknown, and have to be estimated from the data. In this section, we illustrate how this can be done using only standard econometric techniques. We do not formally derive theoretical properties of the resulting estimators, such as consistency and asymptotic normality, as they follow directly from standard arguments.

To see how our detailed decomposition (4.3)–(4.4) can be estimated in practice, recall that for any distributional feature ν the observed difference Δ_O^ν, the structure effect Δ_S^ν, the dependence effect Δ_D^ν, and the various partial marginal distribution effects $\Delta_M^\nu(\mathbf{k})$ can all be expressed in terms of objects of the form $\nu(F_Y^{i|j,\mathbf{k}})$, where

$$F_Y^{i|j,\mathbf{k}}(y) = \int F_{Y|X}^i(y,x) dC^j(F_{X_1}^{\mathbf{k}_1}(x_1), \ldots, F_{X_d}^{\mathbf{k}_d}(x_d))$$

as defined in equation (4.2). To obtain estimates of the elements of our decomposition, we can therefore simply use a plug-in approach and replace the terms $\nu(F_Y^{i|j,\mathbf{k}})$ at every occurrence with their sample counterparts $\nu(\widehat{F}_Y^{i|j,\mathbf{k}})$, where

$$\widehat{F}_Y^{i|j,\mathbf{k}}(y) = \int \widehat{F}_{Y|X}^i(y,x) d\widehat{C}^j(\widehat{F}_{X_1}^{\mathbf{k}_1}(x_1), \ldots, \widehat{F}_{X_d}^{\mathbf{k}_d}(x_d)),$$

with $\hat{F}_{Y|X}^i$, $\hat{C}^i$ and $\widehat{F}_{X_m}^i$ being suitable estimates of the conditional CDFs $F_{Y|X}^i$, the copula functions C^i, and the marginal CDFs $F_{X_m}^i$ of the m-th component of X_i, for $i = 0, 1$ and $m = 1, \ldots, d$, respectively. That is, our estimates are given by

$$\begin{aligned}
&\widehat{\Delta}_O^\nu = \nu(\widehat{F}_Y^{1|1,\mathbf{1_d}}) - \nu(\widehat{F}_Y^{0|0,\mathbf{0_d}}), \quad \widehat{\Delta}_S^\nu = \nu(\widehat{F}_Y^{1|1,\mathbf{1_d}}) - \nu(\widehat{F}_Y^{0|1,\mathbf{1_d}}),\\
&\widehat{\Delta}_X^\nu = \nu(\widehat{F}_Y^{0|1,\mathbf{1_d}}) - \nu(\widehat{F}_Y^{0|0,\mathbf{0_d}}), \quad \widehat{\Delta}_D^\nu = \nu(\widehat{F}_Y^{0|1,\mathbf{1_d}}) - \nu(\widehat{F}_Y^{0|0,\mathbf{1_d}}), \text{ and}\\
&\widehat{\Delta}_M^\nu(\mathbf{k}) = \nu(\widehat{F}_Y^{0|0,\mathbf{k}}) - \nu(\widehat{F}_Y^{0|0,\mathbf{0_d}}),
\end{aligned}$$

for any distributional feature ν.

We generally assume that the data available to the econometrician are two i.i.d. samples of size N_i from the distribution of (Y_i, X_i) for $i \in \{0, 1\}$. In such a classical cross-sectional setting, estimates of the just-mentioned unknown functions can in principle be obtained by a variety of different methods. For the univariate distribution functions $F_{X_m}^i$, the most straightforward estimator is arguably the usual empirical CDF, which is given by

$$\widehat{F}_{X_m}^i(x_m) = \frac{1}{N_i} \sum_{l=1}^{N_i} \mathbb{I}\{X_{i,m,l} \leq x_m\}.$$

For the higher-dimensional objects, i.e. the conditional CDFs and the copula functions, several nonparametric, semiparametric and fully parametric procedures have been proposed in the literature. Since most studies that make use of decomposition methods use data sets containing a large number of individual socio-demographic characteristics, the

substantial sample size requirements of nonparametric methods in such high-dimensional settings make their application hardly attractive. It therefore seems advisable to consider flexible parametric specifications, which have been applied successfully in the empirical literature.

One popular approach to obtain estimates of a conditional CDF, used e.g. in Machado and Mata (2005), is to model the conditional quantile function by a linear quantile regression model (Koenker and Bassett, 1978; Koenker, 2005), and then invert the corresponding estimated quantile function. In this case, we have

$$\widehat{F}^i_{Y|X}(y,x) = \int_0^1 \mathbb{I}\{x\widehat{\theta}_i(\tau) \leq y\} d\tau,$$

where $\widehat{\theta}_i(\tau)$ are the estimated coefficients from a τ-quantile regression of Y_i on X_i. Another method that has recently attracted considerable interest is distributional regression (Foresi and Peracchi, 1995). When using this technique, the conditional CDF is modeled by a series of binary response models with varying "cutoffs". The resulting estimate of the conditional CDF is

$$\widehat{F}^i_{Y|X}(y,x) = \Phi(x\widehat{\theta}_i(y)),$$

where $\Phi(\cdot)$ is the standard normal CDF (or some other strictly increasing link function), and $\widehat{\theta}_i(y)$ is the maximum likelihood estimate of a Probit regression of the indicator variable $\mathbb{I}\{Y_i \leq y\}$ on the covariates X_i. Chernozhukov et al. (2009) derive asymptotic properties for these and several other parametric conditional CDF estimators, establishing classical properties like $\sqrt{N_i}$-consistency and asymptotic normality under standard regularity conditions. Rothe and Wied (2012) consider specification testing in these types of models.

To obtain estimates of the copula functions C^i for $i \in \{0,1\}$, it is also advisable to use one of the numerous flexible parametric specifications that have been proposed in the literature. Different copula models are able to generate different types of dependence patters, and thus the analyst should chose a specification that is considered flexible enough to encompass the relationship between the covariates in the respective group. The extensive reviews of the properties of various copula models in e.g. Nelsen (2006) or Trivedi and Zimmer (2007) are a useful guidance for this choice. In empirical applications with many covariates that are not expected to have the same pairwise dependence patters,

a suitable solution seems to be to use the Gaussian copula

$$C_\Sigma(u) = \Phi_\Sigma(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)),$$

with Φ_Σ the CDF of a d-variate standard normal distribution with correlation matrix Σ and Φ the standard normal CDF. When the sample size is small relative to the number of the covariates, so that estimating the $d(d-1)/2$ correlation parameters is not feasible, more parsimonious specifications with stronger shape restrictions could be considered as well. Again, we refer to Nelsen (2006) or Trivedi and Zimmer (2007) for an extensive review. For any common copula model, the respective dependence parameters can be estimated by Maximum Likelihood or Minimum Distance methods implemented in the usual econometric software packages, and estimates can be shown to be $\sqrt{N_i}$-consistent and asymptotically normal under standard regularity conditions (e.g. Genest, Ghoudi, and Rivest, 1995).

The estimated elements of our detailed decomposition (4.3)–(4.4) can be shown to be consistent and asymptotically normal under standard regularity conditions, using essentially the same arguments employed by Chernozhukov et al. (2009) for estimates of the structure and composition effect. Roughly speaking, the reasoning is as follows. First, one shows that the estimates of the marginal CDFs, the conditional CDFs, and the copula functions are not only pointwise asymptotically normal, but converge to a Gaussian process when taken as a function. Second, it then follows from the continuous mapping theorem that the estimated counterfactual CDFs $F_Y^{i|j,\mathbf{k}}$ also converge to a Gaussian process. Finally, asymptotic normality of $\nu(F_Y^{i|j,\mathbf{k}})$ follows the the functional delta method under a common smoothness condition on the functional ν. Since the asymptotic variance of the estimates typically has a complicated form, standard errors can be obtained via the usual nonparametric bootstrap. The validity of such an approach follows again from arguments in Chernozhukov et al. (2009).

7. An Empirical Application

In this section, we provide a small-scale empirical application that illustrates how our copula decomposition technique can be used in practice. Using data from the Current Population Survey (CPS), we decompose differences in various features of the 1985 and

Table 1: Descriptive Statistics

	1983–1985					2003–2005				
	Mean	SD	Q10	Q50	Q90	Mean	SD	Q10	Q50	Q90
Log Hourly Wage	1.78	0.52	1.05	1.82	2.46	1.85	0.58	1.11	1.83	2.65
Part-time status	0.09	–	–	–	–	0.09	–	–	–	–
Nonwhite	0.11	–	–	–	–	0.13	–	–	–	–
Union coverage	0.27	–	–	–	–	0.15	–	–	–	–
Married	0.67	–	–	–	–	0.62	–	–	–	–
Education	12.87	2.91	10.00	12.00	17.00	13.42	2.79	11.00	13.00	18.00
Experience	17.20	12.31	3.00	14.00	36.00	19.47	11.49	4.00	19.00	35.00

2005 distribution of wages among male workers in the United States. There is now extensive evidence that during this period wage inequality in the United States has been rising substantially in the top end of the wage distribution, but has slightly decreased in the bottom end, leading to what is often called a polarization of the US labor market (Autor, Katz, and Kearney, 2006; Lemieux, 2008).

We use a data set from Fortin et al. (2011), which was extracted from the 1983–1985 and 2003–2005 Outgoing Rotation Group (ORG) supplements of the CPS. We also refer to Lemieux (2006) for details on its construction. Our data contain information on 232,784 and 170,693 males, respectively, that were employed in the relevant periods. Workers in the 1983–1985 and 2003–2005 sample play the role of our group 0 and 1. The outcome variable of interest is the log hourly wage, measured in constant 1985 dollars. The covariates are years of education, years of potential labor market experience, and dummies for union coverage, race, marital status, and part-time status. All observations are weighted by the product of the number of hours worked and their respective CPS sample weights.

Some descriptive statistics are given in Table 1. Results on wages confirm the general picture that was found before in the literature. Over the sample period, the 90% quantile of the wage distribution has been rising substantially, while the 10% quantile and the median exhibit only a moderate increase or have remained approximately constant, respectively. There is thus a large increase in wage inequality as measured by the difference between the 90% and the 10% quantile, but this is the effect of an increase in inequality

in the right tail of the distribution. Regarding the covariates, the most striking feature is certainly the decline in union coverage from 27% in 1983–1985 to only 15% in 2003–2005. Average (potential) labor market experience also increased substantially by more than two years, and the average years of education also raised by a considerable amount of about half a year. Changes in other explanatory variables are less pronounced.

To estimate the various elements of our detailed decomposition, we proceed as described in Section 6. We model the copula functions C^i by a Gaussian copula, and the conditional CDFs $F^i_{Y|X}$ by a distributional regression model with a Gaussian link function. Compared to an approach based on quantile regression, distributional regression has the advantage that it is not affected by heaping in the distribution of wages, and seems to be better suited to capture the somewhat irregular behavior of the conditional wage distribution around the level of the minimum wage (Chernozhukov et al., 2009; Rothe and Wied, 2012). In addition to the covariates mentioned above, we use quadratic terms in education and experience and a full set of interaction terms for estimating the conditional CDFs. Standard errors are calculated via the nonparametric bootstrap, using $B = 200$ replications. Due to the large sample sizes, sampling variation in our estimates is mostly negligible.

Table 2 and Table 3 present the results of our decomposition for various measures of location and spread, respectively. Row by row, we report estimates of the observed change Δ^ν_O, the usual structure and composition effect Δ^ν_S and Δ^ν_X, our dependence and marginal distribution effects Δ^ν_D and Δ^ν_M, the direct contributions $\Delta^\nu_M(\mathbf{e_j})$ for each of the six covariates, and the sum $\sum_{2\leq|\mathbf{k}|\leq 6} \Delta^\nu_M(\mathbf{k})$ of all "higher-order" interaction terms $\Delta^\nu_M(\mathbf{k})$ with $|\mathbf{k}| > 1$. All estimates and corresponding standard errors have been multiplied by 100 to improve readability. Estimates of the observed change Δ^ν_O are very close to the respective descriptive statistics that can be calculated directly from the data, which indicates that our parametric model provides a reasonable fit. Comparing the observed change to the structure and composition effect, we can see that changes in labor force composition alone can explain a substantial part of changes in the mean wage, but they do not offer an explanation for the differential change in the quantiles. For example, when considering the median as the distributional feature of interest, composition effects alone would predict a large upward shift of about the same magnitude as the observed

Table 2: Estimated Decomposition of Differences in Distribution of Log Hourly Wages of Workers in 2003–2005 and 1983–1985 using CPS Data.

	Mean		Q90		Q50		Q10	
Observed Difference Δ_O^ν	6.227	(0.034)	18.667	(0.040)	-0.047	(0.046)	4.747	(0.029)
Structure Effect Δ_S^ν	0.880	(0.036)	8.260	(0.042)	-6.345	(0.052)	2.927	(0.041)
Composition Effect Δ_X^ν	5.348	(0.031)	10.407	(0.032)	6.297	(0.046)	1.820	(0.040)
Dependence Effect Δ_D^ν	-0.335	(0.032)	1.051	(0.034)	-0.639	(0.046)	-1.411	(0.043)
Marginal Distr. Effect Δ_M^ν	5.682	(0.037)	9.356	(0.035)	6.936	(0.049)	3.231	(0.043)
"Direct" Marginal Distribution Effect ($\Delta_M^\nu(\mathbf{e_j})$) attributable to between-group differences in ...								
Part-time status	-0.041	(0.036)	0.029	(0.038)	-0.048	(0.048)	-0.120	(0.042)
Nonwhite	-0.230	(0.030)	-0.181	(0.038)	-0.271	(0.038)	-0.238	(0.038)
Union coverage	-2.542	(0.031)	0.796	(0.039)	-4.006	(0.039)	-2.964	(0.048)
Married	-0.531	(0.027)	-0.272	(0.032)	-0.644	(0.038)	-0.577	(0.036)
Education	4.373	(0.030)	3.956	(0.030)	5.749	(0.038)	4.386	(0.053)
Experience	4.297	(0.033)	4.325	(0.035)	5.768	(0.042)	3.755	(0.050)
Interaction ($\sum_{\mathbf{k}\geq 2} \Delta_M^\nu(\mathbf{k})$)	0.357	(0.129)	0.703	(0.140)	0.388	(0.167)	-1.012	(0.162)

Standard errors in parenthesis. All numbers have been multiplied by 100 to improve readability.

change in means. However, this is being offset by a negative structure effect of similar magnitude, and thus the observed change in medians is approximately zero.

When decomposing the composition effect, one can see that changes in the dependence pattern among the covariates play a substantial role for the measures of spread we consider in Table 3. Depending on the distributional feature of interest, they can explain about 20%–40% of the composition effect alone. For the measures of location in Table 2, the effect is much less pronounced. With the exception of the 90% quantile, the dependence effect is negative for all measures of location, while the composition effect is positive in all cases. When considering estimates of the "direct" marginal distribution effects, we find that changes in the distribution of education and experience have strong positive impact on all quantiles, with slightly larger magnitudes for the median relative to the two extreme quantiles. As a consequence, these variables contribute only moderately to the increase in inequality. Changes in union coverage rates are estimated to have a strong negative impact on the median wage, a somewhat less negative one on the 10% quantile, and a small positive effect on the 90% quantile. This covariate is thus the

Table 3: Estimated Decomposition of Differences in Distribution of Log Hourly Wages of Workers in 2003–2005 and 1983–1985 using CPS Data.

	Variance		Q90-10		Q90-50		Q50-10	
Observed Difference Δ_O^ν	7.767	(0.017)	13.920	(0.043)	18.714	(0.034)	-4.795	(0.035)
Structure Effect Δ_S^ν	4.891	(0.018)	5.333	(0.049)	14.605	(0.040)	-9.272	(0.033)
Composition Effect Δ_X^ν	2.877	(0.014)	8.586	(0.046)	4.109	(0.035)	4.477	(0.033)
Dependence Effect Δ_D^ν	0.822	(0.014)	2.462	(0.046)	1.689	(0.032)	0.772	(0.035)
Marginal Distr. Effect Δ_M^ν	2.054	(0.014)	6.125	(0.043)	2.420	(0.031)	3.705	(0.033)
"Direct" Marginal Distribution Effect ($\Delta_M^\nu(\mathbf{e_j})$) attributable to between-group differences in...								
Part-time status	0.057	(0.015)	0.149	(0.048)	0.078	(0.030)	0.072	(0.040)
Nonwhite	0.024	(0.017)	0.057	(0.049)	0.090	(0.034)	-0.033	(0.029)
Union coverage	1.264	(0.017)	3.759	(0.057)	4.802	(0.035)	-1.043	(0.039)
Married	0.173	(0.017)	0.304	(0.049)	0.372	(0.037)	-0.068	(0.032)
Education	-0.058	(0.015)	-0.431	(0.055)	-1.793	(0.029)	1.362	(0.035)
Experience	0.189	(0.018)	0.570	(0.057)	-1.443	(0.035)	2.014	(0.036)
Interaction ($\sum_{\mathbf{k}\geq 2} \Delta_M^\nu(\mathbf{k})$)	0.404	(0.068)	1.715	(0.198)	0.315	(0.126)	1.400	(0.134)

Standard errors in parenthesis. All numbers have been multiplied by 100 to improve readability.

single most important one for explaining changes in overall and top-end inequality, as it accounts for about 25% of the observed change in each the 90%-10% and 90%-50% quantile differences alone. Between-group differences in other explanatory variables are found to be of relatively minor importance, as are the interaction effects between the covariates marginal distributions.

8. Concluding Remarks

In this paper, we have proposed a detailed decomposition of the composition effect. The method does not rely on a particular structure of the data generating process, and can be used to decompose differences in essentially all features of the distribution of outcomes. Our decomposition contains three types of components: i) the "direct contribution" of each covariate due to between-group differences in the respective marginal distributions, ii) several "two way" and "higher order" interaction effects due to the interplay between two or more covariates' marginal distributions, and iii) a "dependence effect" accounting for different dependence patterns among the covariates. We argue that all three types

are necessary for understanding differences in general distributional features even for simple data generating processes, and illustrate their empirical relevance through an application to US wage data. Our decomposition can be implemented using only standard econometric techniques for estimating conditional distribution and copula functions, and valid inference can be carried out via the bootstrap.

References

Altonji, J., P. Bharadwaj, and F. Lange (2008): "Changes in the characteristics of american youth: Implications for adult outcomes," *NBER Working paper*.

Autor, D., L. Katz, and M. Kearney (2006): "The polarization of the US labor market," *American Economic Review*, 96, 189–194.

Barsky, R., J. Bound, K. Charles, and J. Lupton (2002): "Accounting for the black-white wealth gap," *Journal of the American Statistical Association*, 97, 663–673.

Blinder, A. (1973): "Wage discrimination: reduced form and structural estimates," *Journal of Human Resources*, 8, 436–455.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2009): "Inference on Counterfactual Distributions," *Working Paper*.

DiNardo, J., N. Fortin, and T. Lemieux (1996): "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach," *Econometrica*, 64, 1001–1044.

Donald, S., D. Green, and H. Paarsch (2000): "Differences in wage distributions between Canada and the United States: An application of a flexible estimator of distribution functions in the presence of covariates," *Review of Economic Studies*, 67, 609–633.

Firpo, S., N. Fortin, and T. Lemieux (2007): "Decomposing wage distributions using recentered influence function regressions," *NBER Working Paper*.

——— (2009): "Unconditional quantile regressions," *Econometrica*, 77, 953–973.

——— (2011): "Occupational tasks and changes in the wage structure," *Working Paper*.

Foresi, S. and F. Peracchi (1995): "The Conditional Distribution of Excess Returns: An Empirical Analysis." *Journal of the American Statistical Association*, 90, 451–466.

Fortin, N., T. Lemieux, and S. Firpo (2011): "Decomposition Methods in Economics," *Handbook of Labor Economics*, 4, 1–102.

Genest, C., K. Ghoudi, and L. Rivest (1995): "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions," *Biometrika*, 82, 543–552.

Gosling, A., S. Machin, and C. Meghir (2000): "The Changing Distribution of Male Wages in the UK," *Review of Economic Studies*, 67, 635–666.

Koenker, R. (2005): *Quantile regression*, Cambridge University Press.

Koenker, R. and G. Bassett (1978): "Regression quantiles," *Econometrica*, 46, 33–50.

Lemieux, T. (2006): "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *American Economic Review*, 96, 461–498.

——— (2008): "The changing nature of wage inequality," *Journal of Population Economics*, 21, 21–48.

Machado, J. and J. Mata (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20, 445–465.

Melly, B. (2005): "Decomposition of differences in distribution using quantile regression," *Labour Economics*, 12, 577–590.

Nelsen, R. (2006): *An introduction to copulas*, Springer Verlag.

Oaxaca, R. (1973): "Male-female wage differentials in urban labor markets," *International Economic Review*, 14, 693–709.

Rothe, C. (2010a): "Identification of unconditional partial effects in nonseparable models," *Economics Letters*, 109, 171 – 174.

——— (2010b): "Nonparametric Estimation of Distributional Policy Effects," *Journal of Econometrics*, 155, 56–70.

——— (2012): "Partial Distributional Policy Effects," *Econometrica*, to appear.

Rothe, C. and D. Wied (2012): "Misspecification testing in a class of conditional distributional models," *Working Paper*.

Sklar, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 8, 229–231.

Trivedi, P. and D. Zimmer (2007): *Copula modeling: an introduction for practitioners*, Now Publishers Inc.

Van der Vaart, A. (2000): *Asymptotic Statistics*, Cambridge University Press.