ESSAYS ON MICROECONOMETRICS AND FINANCE

A Dissertation
Submitted to
the Temple University Graduate Board

in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

by

Weige Huang
May, 2019

Examining Committee Members:

Brantly Callaway, Advisor, Department of Economics
Oleg Rytchkov, Department of Finance
Douglas A. Webber, Department of Economics
Irina Murtazashvili, External Member, Drexel University LeBow College of Business

ABSTRACT

ESSAYS ON MICROECONOMETRICS AND FINANCE

by

Weige Huang

This dissertation includes four chapters which are four separate papers on Mircoeconometrics and Finance. The first two chapters establish estimators which are useful to study distributional effects of a continuous treatment and local elasticities, respectively. The tools then are applied on to intergenerational income mobility. Thus, Chapter 1 and 2 are on Microeconomics. Chapter 3 and 4 are on Finance. In particular, I apply the Oaxaca-Blinder decomposition, distribution regression and recentered influence function regression to decompose the portfolio returns between North America and Europe.

Chapter 1, titled DISTRIBUTIONAL EFFECTS OF A CONTINUOUS TREATMENT WITH AN APPLICATION ON INTERGENERATIONAL MOBILITY (with Brantly Callaway), considers the effect of a continuous treatment on the entire distribution of outcomes after adjusting for differences in the distribution of covariates across different levels of the treatment. Our methodology encompasses dose response functions, counterfactual distributions, and "distributional policy effects" depending on the assumptions invoked by the researcher. We propose a three-step estimator that consists of (i) estimating the distribution of the outcome conditional on the treatment and other covariates using quantile regression; (ii) for each value of the treatment, averaging over a counterfactual distribution of the covariates holding the treatment fixed; (iii) manipulating the counterfactual distribution into a parameter of interest. We show that our estimators converge uniformly to Gaussian processes and that the empirical bootstrap can be used to conduct uniformly valid inference across a range of values of the treatment. We use our method to study intergenerational income

mobility where we consider distributional effects of parents' income on child's income such as (i) the fraction of children with income below the poverty line, (ii) the variance of child's income, and (iii) the inter-quantile range of child's income – all as a function of parents' income.

Chapter 2, tiled LOCAL INTERGENERATIONAL ELASTICITIES (with Brantly Callaway), proposes a "local" intergenerational mobility parameter (LIGE) that allows the effect of parents' income to vary across different values of parents' income. We also extend this result to an "adjusted" local intergenerational elasticity (ALIGE) which adjusts for differences in the distribution of observed characteristics at different values of parents' income. We develop the asymptotic properties of the LIGE and ALIGE, and apply them to study intergenerational mobility using data from the PSID. We find that the intergenerational elasticity is much larger for low values of parents' income (indicating *less* mobility) relative to high values of parents' income; adjusting for differences in characteristics reduces the local IGE at all values of parents' income as well as flattening it across different values of parents' income.

Chapter 3, titled DECOMPOSING DIFFERENCES IN PORTFOLIO RETURNS BETWEEN NORTH AMERICA AND EUROPE, decomposes differences in mean and a series of quantiles of portfolio returns between North America and Europe into Fama and French's five factors. We show that the differences in risk premia on factors, especially on market and size factors, account for most of the differences and the differences in factor risks seem to play an insignificant role in aggregate. The results from Blinder-Oaxaca decomposition show that the differences in market and size factor risk premia explain 71.9% and 22.8% of the overall mean difference, respectively. We also show that the roles that the risk premia on market and size factors play vary at different levels of portfolio returns, implying the market and size factor risk premia vary at different levels of portfolio returns. Also, we find that the risks on some factors seem to vary at different levels of portfolio returns.

Chapter 4, titled DECOMPOSING DIFFERENCES IN QUANTILE PORTFOLIO RETURNS BETWEEN NORTH AMERICA AND EUROPE USING RECENTERED INFLUENCE FUNCTION REGRESSION, decomposes quantile portfolio returns using recentered influence function regression. Chapter 1 decomposes the differences in quantile portfolio returns using distribution regressions. The main issue of using distribution regressions is that the decomposition results are path dependent. In this paper, we can obtain path independent decomposition results by combining the Oaxaca-Blinder decomposition and the recentered influence function regression method. We show that aggregate composition effects are all positive across quantiles and the market factor is the most significant factor which has detailed composition effect monotonically decreasing along quantiles. The main decomposition results are consistent with Chapter 3.

ACKNOWLEDGEMENTS

I am particularly grateful to my advisor, Brantly Callaway, for his advising on this research and his dedication to my work. I cannot find any words good enough to express my appreciation. Many thanks also go to my committee members, Oleg Rytchkov and Douglas A. Webber, whose dedications are very much appreciated. Thanks also go to my external committee member, Irina Murtazashvili and my colleagues Emmanuel Selorm Tsyawo and Afrouz Azadikhah Jahromi for their comments. I also thank Donald T. Wargo, Catherine Maclean, Charles E. Swanson and William Stull for their help in my job searching which saves me much time for completing this research. I also feel grateful to Michael Leeds, Mohsen Fardmanesh, Dai Zusai, Moritz Ritter, YoungJin Choi, Dimitrios Diamantaras, Yuan Yuan, Martin Lopez-Daneri, Pedro Silos, Janice Vincent, Linda Wyatt, Lulei Song, James M Kelly and all other colleagues in the Department of Economics, who either taught me courses or helped me through other works which are very helpful for completing this research.

To my wife, Weirong Wu and my parents, Ying Xu and Yi Huang

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER 1

DISTRIBUTIONAL EFFECTS OF A CONTINUOUS TREATMENT WITH AN APPLICATION ON INTERGENERATIONAL MOBILITY

1.1 Introduction

Researchers in economics often consider the effect of one variable (a treatment) on an outcome while perhaps adjusting for differences in other variables that are related to outcomes and that are distributed differently across different values of the treatment. While the case of a binary treatment has received much attention in the literature (see, for example, the review of Imbens and Wooldridge (2009)), this paper develops new methods for the case with a continuous treatment which has received considerably less attention.

Methods for dealing with a continuous treatment are likely to be of interest to empirical researchers across a variety of areas. To give some examples of applications with a continuous treatment, Imbens, Rubin, and Sacerdote (2001) study the effect of unearned income on labor market outcomes; Galvao and Wang (2015) consider an application on the effect of mother's weight gain during pregnancy on child's birth weight; Jasova, Mendicino, and Supera (2018) study the effect of the term structure of debt on banks' lending behavior; and local labor markets approaches common in labor economics often involve a continuous treatment (for example, Autor, Dorn, and Hanson 2013; Acemoglu and Restrepo 2017; Collins and Niemesh 2017).

This paper proposes simple, but flexible, semiparametric estimators of distributional effects[1] of a continuous treatment while adjusting for differences in the distribution of covariates across different levels of the treatment. Our procedure requires three steps. First, we estimate the distribution of the outcome conditional on the treatment and other observed characteristics by inverting quantile regression (Koenker and Bassett Jr 1978; Koenker 2005; Chernozhukov, Fernandez-Val, and Melly 2013) estimates of the conditional quantiles. The second step is to estimate the "counterfactual distribution" which involves integrating over the first step estimates while changing the distribution of observable characteristics. This step obtains the entire distribution of the outcome as a function of the continuous treatment after adjusting for differences in the distribution of covariates across different levels of the treatment. Finally, parameters of interest such as measures of the spread of the outcome and the probability of having a very low outcome (e.g. child's income being below the poverty line) are obtained as functions of the counterfactual distribution.

The literature on continuous treatment effects includes Hirano and Imbens (2004), Flores (2007), Florens, Heckman, Meghir, and Vytlacil (2008), Flores, Flores-Lagunes, Gonzalez, and Neumann (2012), Galvao and Wang (2015), and Kennedy, Ma, McHugh, and Small (2016). Within this literature, the only paper that we are aware of that looks

[1]We use the term distributional effects broadly here. It encompasses counterfactual distributions, dose response functions, treatment effects, and distributional policy effects with a continuous treatment depending on the assumptions invoked by the researcher. Dose response functions and treatment effects are defined in terms of potential outcomes, and we consider these under the assumption of unconfoundedness. Counterfactual distributions are similar, but do not rely on potential outcomes notation or the assumption of unconfoundedness. The counterfactual distributions that we consider fix the return to characteristics (for a particular value of the treatment) but change the distribution of the characteristics. Distributional policy effects compare (functionals of) the counterfactual distribution to (functionals of) the observed distribution of outcomes conditional on the treatment. They are closely related to composition effects in the literature on decompositions.

at distributional parameters is Galvao and Wang (2015) which proposes a weighting estimator of quantile dose response functions with a continuous treatment under the assumption of unconfoundedness. Our approach is different in that our estimators are based on first step quantile regression and do not require estimating conditional densities in the first step.[2] There are trade-offs to using quantile regression relative to weighting estimators based on conditional densities. Quantile regression imposes stronger parametric assumptions than nonparametrically estimating conditional densities, though it is much simpler to implement in practice; on the other hand, quantile regression is much more flexible (though perhaps somewhat harder to implement) than assuming a fully parametric model for a conditional density.[3]

We obtain the limiting processes for each of our parameters of interest and develop inference procedures using results from the empirical process literature (see, for example, Van Der Vaart and Wellner (1996) and Kosorok (2007)) and results on first step quantile regression estimators (Chernozhukov, Fernandez-Val, and Melly 2013). We show that the limiting processes can be approximated using the empirical bootstrap. In the context of intergenerational mobility, these results allow us to test functional hypotheses about income mobility such as (1) whether adjusting for covariates has any effect on any particular parameter (e.g. the percentage of children with income below the poverty line as a function of parents' income), or (2) whether parameters of interest are the same at all values of parents' income (e.g. the variance of child's income).

[2]Another primary difference between our approach and that of Galvao and Wang (2015) is that quantile dose response functions are not our primary object of interest. For studying intergenerational mobility, we found that several other parameters (discussed in detail in Section 1.2) that are functions of the counterfactual distribution are more useful. However, it seems that it would be possible to extend the results in Galvao and Wang (2015) to cover these parameters as well.

[3]Our paper is also related to the literature on decompositions with a continuous treatment. Ñopo (2008) and Ulrick (2012) provide decompositions for the mean with a continuous treatment. Ao, Calonico, and Lee (2017) consider decompositions with a multi-valued discrete treatment. Bowles and Gintis (2002), Groves (2005), Blanden, Gregg, and Macmillan (2007b), and Richey and Rosburg (2017) have decomposed intergenerational mobility parameters into parts that are explained by various background characteristics.

We use our method to study the effect of parents' income on child's income. It is well known that children from families with high income tend to have higher incomes than children from families with low income (see Solon (1992) and Solon (1999), among many others). However, much less is known about the distribution of child's income across parents' income levels. And learning about this distribution provides much more information to researchers and policy makers about the effect of parents' income on child's income.

To give an example, our baseline estimates suggest that a child whose parents' income is at the poverty line (we set this to be \$22,100 and discuss why below) has an income of \$33,800 on average. If this is all that a researcher knows about outcomes for children from families right at the poverty line, it could be the case that (i) the variance of these individuals' income is low implying that many of them have incomes very close to \$33,800, or (ii) the variance of these individuals' income is high implying that some of them have much higher incomes than \$33,800 and others have much lower incomes. In the first case, most children from low income families would be moving out of poverty and into the lower middle class; while in the second case, many children would remain in poverty while others might have substantially higher incomes. These two scenarios have quite different implications for our understanding of the effect of intergenerational income mobility; in particular, if a researcher is interested in the role that parents' income plays in transmitting poverty, only knowing average income as a function of parents' income is not enough.

From a methodological perspective, a key challenge is that parents' income is a continuous variable. There is a large literature on estimating counterfactual distributions with discrete groups which includes DiNardo, Fortin, and Lemieux (1996), Machado and Mata (2005), Firpo (2007), Firpo, Fortin, and Lemieux (2009), and Chernozhukov, Fernandez-Val, and Melly (2013) among others. One idea would be to divide parents' income into a small number of groups and use techniques from this literature. However, this approach would suffer from requiring us to choose cutoffs of parents' income in some ad hoc way (see Bhattacharya and Mazumder (2011) for similar arguments about the cutoffs required for transition matrices). Instead, we keep parents' income as a continuous variable and develop new tools to study counterfactual distributions with a continuous treatment.

The resulting counterfactual distribution is difficult to work with and not easy to directly understand because it is a function of both child's income and parents' income. Instead, we focus on various functionals of the counterfactual distribution that are functions only of parents' income. In particular, we consider (1) average child's income, (2) the fraction of children whose incomes are below the poverty line, (3) the variance of child's income, (4) the inter-quantile range of child's income, and several others – all as a function of parents' income. Each of these parameters can be plotted in two dimensions and the results are easy to interpret.

Like most of the intergenerational income mobility literature, we find a strong relationship between parents' income and child's income. Without adjusting for any differences in covariates, we find that (1) children from low income families have lower income on average than children from high income families; (2) children from low income families have higher income on average than their parents, while children from high income families have lower income on average than their parents; (3) children from low income families are much more likely to have income below the poverty line than children from high income families; (4) children from low income families are much less likely to be in the top 10% of income than children from high income families; (5) children from low income families may have somewhat higher variance in their earnings than children from high income families. The first two of these results are in common with almost all of the intergenerational mobility literature. The last three results are similar to existing results using transition matrices though our approach does not require specifying cutoffs of the continuous treatment and provides a straightforward way to incorporate adjusting for covariates.

A second motivation of our paper is to look at the role that background characteristics play in the transmission of income across generations. We find that background characteristics such as parents' education, race, and whether or not a child is from a single parent household, are strongly correlated with parents' income. We find that adjusting for covariates does not overturn any of the five main results above; however, overall, adjusting for differences in observed characteristics across parents' income levels does tend to reduce the effects of parents' income. Adjusting for differences in observed characteristics flattens somewhat the relationship between child's income and parents' income. It also reduces by about one quarter the estimated probability that a child's income will be below the poverty rate for children from families with income close to the poverty line. Taken together, our results suggest that differences in background characteristics explain some, but not all, of the differences in outcomes experienced by children whose parents had different incomes.

1.2 Parameters Of Interest

This section develops several distributional parameters of interest for the case of a continuous treatment. We are motivated by our application on intergenerational income mobility, but these parameters are likely to be of interest in other applications as well. This section also distinguishes between several classes of parameters: treatment effects, dose response functions, counterfactual distributions, and distributional policy effects. The differences are based primarily on (i) the particular application and (ii) whether or not the researcher wishes to invoke the assumption of unconfoundedness.

Our approach is different from existing work on intergenerational income mobility in three ways. First, we keep parents' income as a continuous variable and all of our results are "local"; that is, conditional on a particular value of parents' income. This setup is different from most work on intergenerational mobility that either estimates a single intergenerational mobility parameter or breaks the observations into several groups. Second, our method allows us to look at the entire distribution of child's income conditional on parents' income. This allows us to estimate parameters such as the fraction of children below the poverty line or the variance of child's income, both as a function of parents' income. These parameters provide much more information about outcomes of children given their parents' income than simply computing the average. Finally, we also are interested in comparing these parameters that can be obtained directly from the observed data to ones that result from "adjusting" the effect of parents' income for differences in observable characteristics. This section details the ideas behind each of these three contributions. Our starting point is that we have a sample of observations from the joint distribution (Y, T, X). In the application, Y is the log of child's income, T is the log of parents' income and X are additional covariates such as parents' education, child's birth year, gender, and race.

1.2.1 Identification

We use the following notation. Let Y denote an individual's outcome, T denote an individual's level of treatment, and X denote a $k \times 1$ vector of covariates. We let $\mathcal{Y}$, $\mathcal{T}$, and $\mathcal{X}$ denote the supports of Y, T, and X. Next, let $Y(t)$ denote an individual's potential outcome – the outcome that would occur for an individual if they experienced treatment level t. We consider the following assumptions.

Assumption 1 (Unconfoundedness)**.**

$$Y(t) \perp\!\!\!\perp T | X$$

Assumption 2 (Common Support)**.**

$f_{T|X}$ is uniformly bounded away from 0 and ∞ on $\mathcal{T}\mathcal{X}$.

Assumption 1 says that, conditional on covariates X, treatment is as good as randomly assigned. Unconfoundedness is also known as selection on observables or ignorability. Some version of Assumption 1 is invoked in much of the literature on continuous treatment effects (e.g. Hirano and Imbens (2004), Flores (2007), Flores, Flores-Lagunes, Gonzalez, and Neumann (2012), and Galvao and Wang (2015)). It is very closely related to the unconfoundedness assumption in the literature with a binary treatment (e.g. Rosenbaum and Rubin (1983), Heckman, Ichimura, and Todd (1997), Hirano, Imbens, and Ridder (2003), and Imbens and Wooldridge (2009)). Assumption 2 imposes a common support assumption. It says that, for all values of the covariates, there are individuals that experience each level of the treatment. This is a strong assumption. In the context of intergenerational mobility, for example, it requires that there be some very poor parents with very high education as well as some very rich parents with very low education. This type of assumption is common in the treatment effects literature though; and, in the case where a researcher is interested in effects at particular values of t, it could be weakened to hold only at those values of the treatment. Under Assumptions 1 and 2, it is straightforward to show that for any $t \in \mathcal{T}$,

$$P(Y(t) \leq y) = \int_{\mathcal{X}} F_{Y|T,X}(y|t,x)\, dF_X(x) \tag{1.2.1}$$

which says that the distribution of potential outcomes if all individuals were assigned the treatment level t can be obtained by integrating the distribution of Y conditional on X and T over the distribution of X for the entire population. In the continuous treatment effect literature, $P(Y(t) \leq y)$ is called the distribution dose response function. One can invert the distribution dose response function to obtain the quantile dose response function or one can integrate over the distribution to obtain the average dose response function. We discuss more parameters of interest in the next subsection.

Interestingly, even without Assumption 1, the term on the right hand side of Equation 1.2.1 has a useful interpretation. First, notice that the observed distribution of the outcome conditional on the treatment is given by:

$$F_{Y|T}(y|t) = \int_{\mathcal{X}} F_{Y|T,X}(y|t,x) \; dF_{X|T}(x|t) \tag{1.2.2}$$

that is, $F_{Y|T}(y|t)$ is the same as integrating the distribution of the outcome Y conditional on observed characteristics X and the treatment $T = t$ over the distribution of X conditional on $T = t$. One can also consider the *counterfactual distribution* of outcomes that individuals that experience treatment level t would experience if the returns to observed characteristics X were held constant but the distribution of covariates for individuals with treatment level t were manipulated to be the same as the distribution of covariates for the entire population. It is given by

$$F^C_{Y|T}(y|t) = \int_{\mathcal{X}} F_{Y|T,X}(y|t,x) \; \mathrm{d}F_X(x) \tag{1.2.3}$$

which is the same as Equation 1.2.1 and where we use the superscript C to indicate that it is a counterfactual distribution.[4] Notice that Equation 1.2.3 does not require Assumption 1 nor does it require potential outcomes notation. We use the terminology counterfactual distribution throughout the rest of the paper; however, depending on the application, a researcher may wish to invoke Assumption 1 with the payoff being that the resulting parameters may be interpreted as causal effects.

Although it is not equal to the observed distribution, all the terms on the right hand side are identified and one can estimate this counterfactual distribution by plugging in to the above equation. While it is possible to show that Equation 1.2.3 is equivalent to a weighting estimator (weighting estimators are developed in DiNardo,

[4]The counterfactual distribution mentioned above is not the only possible counterfactual distribution, though it is the most common. Other manipulations of the distribution of the covariates are possible (see the discussion in Rothe (2010) and Chernozhukov, Fernandez-Val, and Melly (2013)).

Fortin, and Lemieux (1996) and Firpo (2007) in the case where the treatment is binary and in Galvao and Wang (2015) in the continuous treatment case), we find it more natural to estimate the conditional distribution directly in Equation 1.2.3 which is more similar to the approaches taken in Machado and Mata (2005), Melly (2005), and Chernozhukov, Fernandez-Val, and Melly (2013), all in the case where the treatment is a discrete variable. The reason is that, with a continuous treatment variable, the weights are given by conditional density functions[5] which are more challenging to estimate than the conditional distribution function above. Relative to weighting estimators, our approach can be seen as a "regression-adjustment" approach (see Wooldridge (2010)[Section 21.3.2]) to estimating distributional effects with a continuous treatment.

In the context of intergenerational mobility, the counterfactual distribution of child's income is built by *fixing* the distribution of child's income conditional on parents' income and observed characteristics but *changing* the distribution of observed characteristics conditional on parents' income. In particular, we consider changing the distribution of observed covariates conditional on parents' income to be the distribution of covariates for the entire population. To give an example, suppose the only covariate is parents' education and that parents' education is positively related to parents' income and child's income. Further, suppose that we are interested in the distribution of child's income conditional on parents having low income. To obtain a counterfactual distribution, we fix the distribution of child's income conditional on both education and parents' income, but change the distribution of education to be that of the entire population – thus putting relatively more weight on the income of children with highly educated parents who had low income.

[5]With a binary (or even discrete) treatment, the weights depend on the propensity score which is much more straightforward to estimate – for example, one could use logit or probit.

1.2.2 Parameters Of Interest

The observed distribution $F_{Y|T}$ of the outcome conditional on the treatment and the counterfactual distribution $F^C_{Y|T}$ contain much useful information, but they suffer from being difficult to interpret or plot directly. In particular, they are both indexed by y and t which makes plots that vary both y and t three dimensional and difficult to easily interpret. Instead, we focus on estimating functionals of $F_{Y|T}$ and $F^C_{Y|T}$. This section covers these functionals.

Fraction of Individuals with "Low" Outcomes as a Function of the Treatment

The first parameter that we consider is the fraction of individuals whose outcome falls below a particular cutoff y_p as a function of the treatment variable. This is a particularly interesting parameter in the context of intergenerational income mobility. Set y_p equal to the poverty line. Then, this parameter is the fraction of children whose permanent income falls below the poverty line as a function of parents' income. This is given by

$$F_{Y|T}(y_p|t) \qquad \text{and} \qquad F^C_{Y|T}(y_p|t)$$

for the fraction below the poverty line coming from the observed data and from the counterfactual distribution, respectively. These are straightforward measures to plot, as a function of t, and if children with lower income parents are more likely to have permanent incomes below the poverty line, then one would expect that this line would be downward sloping.

For intergenerational income mobility, we are also interested in the fraction of children that have very high permanent income. Let y_R be some particular value of child's permanent income – later we set this to be the 90th percentile of income in the

U.S. in 2010. Then, the fraction of "rich" children conditional on parents' income is given by

$$1 - F_{Y|T}(y_R|t) \qquad \text{and} \qquad 1 - F^C_{Y|T}(y_R|t)$$

coming from the observed distribution and the counterfactual distribution, respectively.

Quantiles of the Outcome as a Function of the Treatment

One can obtain the quantiles of the outcome as a function of the treatment from the observed distribution and counterfactual distributions. For some $\tau \in (0,1)$, these are given by

$$Q_{Y|T}(\tau|t) = \inf\{y : F_{Y|T}(y|t) \geq \tau\} \qquad \text{and} \qquad Q^C_{Y|T}(\tau|t) = \inf\{y : F^C_{Y|T}(y|t) \geq \tau\}$$

for the quantiles of the observed distribution and the quantiles of the counterfactual distribution, both as a function of the treatment. Under Assumption 1, $Q^C_{Y|T}$ is called the quantile dose response function in Galvao and Wang (2015). The quantiles are also useful inputs into the remaining parameters of interest.

Average Outcome as a Function of the Treatment

The next parameter that we consider is the average outcome as a function of the treatment which is given by

$$E[Y|T=t] = \int_0^1 Q_{Y|T}(\tau|t)\ \mathrm{d}\tau \qquad \text{and} \qquad E^C[Y|T=t] = \int_0^1 Q^C_{Y|T}(\tau|t)\ \mathrm{d}\tau$$

where these depend on the observed distribution and counterfactual distribution, respectively.[6] Average child's income conditional on parents' income is closely related

[6] In practice, we trim out the uppermost and lowermost quantiles so that the integration is from ϵ to $1-\epsilon$ (for some small positive ϵ) though it is likely to be possible to integrate from 0 to 1 under some additional conditions as in Bhattacharya (2007), Barrett and Donald (2009), and Donald, Hsu,

to the Intergenerational Elasticity (IGE) that is very commonly estimated in the intergenerational mobility literature. IGE is the coefficient on the log of parents' income in the regression of log child's income on log parents' income. The slope of $E[Y|T = t]$ corresponds to the IGE though, in our case, the slope is not restricted to be constant.

Measures of Spread of the Outcome as a Function of the Treatment

Because our method obtains the entire observed distribution and counterfactual distribution of the outcome conditional on the treatment, we can study other features of these distributions than just their mean. In this section, we consider the variance of the outcome conditional on the treatment and the inter-quantile range of the outcome conditional on the treatment. For intergenerational mobility, these give measures of the spread of child's income conditional on parents' income.

Given the existing results in the intergenerational mobility literature, one would strongly suspect that child's income tends to increase with parents' income, at least on average. However, much less is known about the spread of child's income conditional on parents' income. It is possible that the distribution of child's income simply shifts to the right as parents' income increases. If the variance of child's income decreases with parents' income, that would suggest that having parents with high income increases income on average and increases the certainty of obtaining higher income. Decreasing variance would also suggest that the income of children from low income families is riskier. On the other hand, if the variance of child's income is increasing in parents' income, that would suggest that children from high income families are more likely to become very rich but also have some risk of having low incomes (and the reverse would be true for children of low income families).

The first measure of spread that we consider is the variance of the outcome as a

and Barrett (2012).

function of the treatment.[7] It is given by

$$Var(Y|T=t) = \int_0^1 \left(Q_{Y|T}(\tau|t) - E[Y|T=t]\right)^2 \, \mathrm{d}\tau$$

and

$$Var^C(Y|T=t) = \int_0^1 \left(Q^C_{Y|T}(\tau|t) - E^C[Y|T=t]\right)^2 \, \mathrm{d}\tau$$

The second measure of spread is an inter-quantile range which is given by

$$IQR(\tau_1, \tau_2; t) = Q_{Y|T}(\tau_1|t) - Q_{Y|T}(\tau_2|t) \qquad \text{and} \qquad IQR^C(\tau_1, \tau_2; t) = Q^C_{Y|T}(\tau_1|t) - Q^C_{Y|T}(\tau_2|t)$$

where $\tau_1 > \tau_2$. A typical example would be to look at the spread between the 90th percentile of child's income and 10th percentile of child's income conditional on parents' income being given by t.[8]

Distributional Policy Effects

Another interesting class of parameters to consider are those that determine how (functionals) of the distribution of outcomes change as a result of adjusting for covariates. Rothe (2010) terms these types of parameters "distributional policy effects," and they are given by the difference between (functionals) of $F_{Y|T}$ and $F^C_{Y|T}$.[9] We denote these by

$$\Delta_k(t) = \Gamma_k(F_{Y|T}) - \Gamma_k(F^C_{Y|T})$$

[7] In practice, for the variance, we trim out the uppermost and lowermost quantiles just like we did for the mean. See the discussion in Footnote 6.

[8] There are other parameters as well that we could consider given that the observed and counterfactual distributions are identified. For example, Barrett and Donald (2009) consider several varieties of Lorenz curves and Gini coefficients that would be of interest in the case where a researcher is interested in inequality as a function of the treatment.

[9] Distributional policy effects are also closely related to the composition effect in aggregate decompositions.

where k indexes some particular parameter and Γ_k denotes the functional that transforms a conditional distribution into the parameter of interest. For example, to examine the role that adjusting for differences in covariates plays in terms of the fraction of children with income below the poverty line, we can consider the parameter

$$\Delta^{POV}(t) = F_{Y|T}(y_p|t) - F^C_{Y|T}(y_p|t)$$

which is the difference in the poverty rates coming from the observed distribution and the counterfactual distribution for some particular value of parents' income t. Similarly, to assess the effect of covariates on average child's income conditional on parents' income, one can also consider the parameter

$$\Delta^E(t) = E[Y|T = t] - E^C[Y|T = t]$$

For some value $t \in \mathcal{T}$, $\Delta^E(t) > 0$ implies that adjusting for covariates lowers average income for children with parents with income t. If covariates, such as education, are positively related to parents' income and positively related to child's income, then one would expect that $\Delta^E(t)$ would be negative for small values of t and positive for large values of t.

Treatment Effects

If one imposes Assumption 1, then our results are closely related to treatment effects. Then, for example, the average treatment effect is given by

$$ATE(t, t') = E^C[Y|T = t] - E^C[Y|T = t']$$

and depends on two values of parents' income. Setting $t = t_{0.75}$ and $t' = t_{0.25}$ which represent the 75th percentile and 25th percentile of the treatment, respectively.

$ATE(t_{0.75}, t_{0.25})$ is how much a random individual's income would increase on average if they changed from having parents in the 25th percentile of the income distribution to the 75th percentile. Similarly,

$$DTE(t, t'; y_p) = F^C_{Y|T}(y_p|t) - F^C_{Y|T}(y_p|t')$$

is how much the fraction of individual's with income below the poverty line changes for t relative to t'.

1.2.3 Testing If Parameters Depend On The Level Of The Treatment

Each of the parameters mentioned above can be considered as a function of the treatment t. As a final step in our analysis, we are interested in testing whether the treatment has any effect on the parameters of interest. Let $\theta(t)$ denote a generic parameter of interest – this includes parameters obtained from the observed distribution or the counterfactual distribution. Then, we are interested in the null hypothesis that

$$\theta(t) = E[\theta(T)] \qquad \text{for all } t \in \mathcal{T}$$

Let $R_\theta(t) = \theta(t) - E[\theta(T)]$. We are interested in testing the following hypothesis

$$H_0 : R_\theta(t) = 0 \qquad \text{for all } t \in \mathcal{T} \tag{1.2.4}$$

To give an example, one could be interested in testing whether the variance of child's income changes with parents' income, both using the observed distribution and using the counterfactual distribution that adjusts for differences in the distribution of covariates across different levels of parents' income. This sort of test allows one to do exactly that.

1.3 Estimation

Estimation proceeds in three steps. In step 1, we estimate the distribution of the outcome Y conditional on the treatment T and possibly other observed characteristics X using quantile regression to obtain the conditional quantiles and then inverting to obtain the conditional distribution. For counterfactual distributions, step 2 involves integrating the conditional distribution over a counterfactual distribution of X conditional on T. In particular, we consider the counterfactual distribution $F^C_{X|T} = F_X$; that is, we set the distribution of X conditional on T to be equal to the distribution of X for the overall population for all values of T. With step 2 complete, we have a (counterfactual) distribution of Y conditional on T. The final step is to manipulate the (counterfactual) distribution into the particular parameters of interest given in Section 1.2.

1.3.1 Step 1: Estimating The Conditional Distribution

We estimate the conditional distribution function $F_{Y|T,X}$ using quantile regression (Koenker and Bassett Jr 1978; Koenker 2005; Chernozhukov, Fernandez-Val, and Melly 2013).[10] We make the following assumptions

Assumption 3. *For all $\tau \in \mathcal{T}$*

$$Q_{Y|T,X}(\tau|t,x) = P_1(t,x)'\alpha(\tau)$$

where $P_1(t,x)$ are functions of t and x (e.g. the leading special case is that $P_1(t,x)$ is given by the $(k+1) \times 1$ vector $(t,x')'$ though it can also include interactions, higher order terms, etc.) and α is a $dim(P_1(t,x)) \times 1$ vector of parameters indexed by τ.

[10]In the Supplementary Appendix, we consider an alternative approach based on first step distribution regression (Foresi and Peracchi 1995a; Chernozhukov, Fernandez-Val, and Melly 2013).

Assumption 4. *For all $\tau \in \mathcal{T}$*

$$Q_{Y|T}(\tau|t) = P_2(t)'\beta(\tau)$$

where $P_2(t)$ are functions of t and $\beta(\tau)$ is a $dim(P_2(t)) \times 1$ vector of parameters indexed by τ.

Assumptions 3 and 4 impose that the conditional quantiles are linear in parameters and can be estimated using standard quantile regression techniques. With the conditional quantiles in hand, the conditional distribution can be obtained by inverting the conditional quantiles. To implement the quantile regression estimator, we estimate the conditional quantiles over a fine, equally-spaced grid of S possible values for τ satisfying $0 < \tau_1 < \cdots < \tau_S < 1$. The estimated conditional quantiles are given by

$$\hat{Q}_{Y|T,X}(\tau|t,x) = P_1(t,x)'\hat{\alpha}(\tau) \qquad \text{and} \qquad \hat{Q}_{Y|T}(\tau|t) = P_2(t)'\hat{\beta}(\tau)$$

and which can each be inverted to obtain the conditional distributions by

$$\hat{F}_{Y|T,X}(y|t,x) = \frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\{\hat{Q}_{Y|T,X}(\tau_s|t,x) \leq y\} \qquad \text{and} \qquad \hat{F}_{Y|T}(y|t) = \frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\{\hat{Q}_{Y|T}(\tau_s|t) \leq y\}$$

$\hat{F}_{Y|T}$, given above, is the observed distribution which is one of our objects of interest; however, we still need to manipulate $\hat{F}_{Y|T,X}$ to be the counterfactual distribution of interest.

1.3.2 Step 2: Estimating Counterfactual Distributions

From the subsection above, we obtained an estimator of $F_{Y|T,X}$. For fixed y and t, estimating $F^C_{Y|T}(y|t)$ amounts to averaging over X while holding t fixed. That is,

$$\hat{F}^C_{Y|T}(y|t) = \frac{1}{n}\sum_{i=1}^{n} \hat{F}_{Y|T,X}(y|t, X_i)$$

which is the same as replacing the population distribution function in Equation 1.2.3 with the sample distribution function. We plug in the estimates $\hat{F}_{Y|T}$ and $\hat{F}^C_{Y|T}$ below to obtain estimates of particular parameters of interest.

1.3.3 Step 3: Estimating Parameters Of Interest

Once the observed distribution and counterfactual distribution of the outcome conditional on the treatment have been estimated, one can estimate the parameters of interest considered in Section 1.2. Estimating the fraction of individual's with income below the poverty line is straightforward and given by

$$\hat{F}_{Y|T}(y_p|t) \qquad \text{and} \qquad \hat{F}^C_{Y|T}(y_p|t)$$

Estimating quantiles of the outcome conditional on the treatment can also be obtained simply by plugging in to the results in Section 1.2:

$$\hat{Q}_{Y|T}(\tau|t) = \inf\{y : \hat{F}_{Y|T}(y|t) \geq \tau\} \qquad \text{and} \qquad \hat{Q}^C_{Y|T}(\tau|t) = \inf\{y : \hat{F}^C_{Y|T}(y|t) \geq \tau\}$$

which simply inverts the counterfactual distribution of outcomes. Next, we can estimate $E[Y|T = t]$ by.

$$\hat{E}[Y|T = t] = \frac{1}{S}\sum_{s=1}^{S} \hat{Q}_{Y|T}(\tau_s|\ t) \qquad \text{and} \qquad \hat{E}^C[Y|T = t] = \frac{1}{S}\sum_{s=1}^{S} \hat{Q}^C_{Y|T}(\tau_s|\ t)$$

where $0 < \tau_1 < \tau_2 < \cdots < \tau_S < 1$ is the grid of values of τ given above. We can also estimate the conditional variance by plugging in

$$\hat{Var}(Y|T=t) = \frac{1}{S}\sum_{s=1}^{S}\left(\hat{Q}_{Y|T}\left(\tau_s|\ t\right) - \hat{E}[Y|T=t]\right)^2$$

and

$$\hat{Var}^C(Y|T=t) = \frac{1}{S}\sum_{s=}^{S}\left(\hat{Q}^C_{Y|T}\left(\tau_s|t\right) - \hat{E}^C[Y|T=t]\right)^2$$

Finally, estimates of the inter-quantile range are given by

$$I\hat{Q}R(\tau_1,\tau_2;t) = \hat{Q}_{Y|T}(\tau_1|t) - \hat{Q}_{Y|T}(\tau_2|t) \qquad \text{and} \qquad I\hat{Q}R^C(\tau_1,\tau_2;t) = \hat{Q}^C_{Y|T}(\tau_1|t) - \hat{Q}^C_{Y|T}(\tau_2|t)$$

1.3.4 Asymptotic Theory

This section develops asymptotic theory and inference procedures for the parameters discussed in Section 1.2. Our inference results are uniformly valid in the treatment T, and we derive the joint limiting distributions of parameters that depend on the observed distribution $F_{Y|T}$ and on the counterfactual distribution $F^C_{Y|T}$. We show that each of the parameters that we consider converges uniformly to a Gaussian process. These results allow us to test functional hypotheses such as (1) whether the results from adjusting for differences in other covariates X are different from the results obtained directly from the observed data at *any* value of the treatment, (2) whether any parameter of interest (such as the variance or inter-quantile range) of the outcome is constant across different values of the treatment, among others. We develop these asymptotic results using arguments from the empirical processes literature (see, for example Van Der Vaart and Wellner 1996; Kosorok 2007) and, in particular, they build off the theoretical results on first step quantile regression in Chernozhukov, Fernandez-Val, and Melly (2013).[11] For any discrete set of values of T, a Gaussian process is just a (multivariate) normal distribution, so our results also contain as special cases pointwise results. The second part of the results in this section shows that the empirical bootstrap is valid for conducting inference – both uniformly and pointwise. All proofs are contained in the Appendix. We make the following assumption,

Assumption 5. *(Random Sampling)*

$\{Y_i, T_i, X_i\}_{i=1}^n$ *are iid draws from the joint distribution* $F_{Y,T,X}$.

Several other standard assumptions for quantile regression and other technical conditions are collected in Assumption 1 in the Appendix. We use the following notation. Let $l^\infty(S)$ denote the space of all uniformly bounded functions on the set S equipped with the supremum norm denoted $\|\cdot\|_\infty$. Let $\mathcal{Y}$, $\mathcal{T}$, and $\mathcal{X}$ denote the supports of Y, T, and X, respectively. Let

$$\hat{G}^C_{Y|T}(y|t) = \sqrt{n}(\hat{F}^C_{Y|T}(y|t) - F^C_{Y|T}(y|t))$$

denote the empirical process of the counterfactual distribution of the outcome conditional on the treatment. Further, let

$$\hat{G}_{Y|T}(y|t) = \sqrt{n}(\hat{F}_{Y|T}(y|t) - F_{Y|T}(y|t))$$

denote the empirical process of the observed distribution of the outcome conditional on the treatment.

Theorem 1.1 establishes the joint limiting process for the observed distribution and counterfactual distribution.

[11]This is similar to other papers that also use quantile regression as a first step estimator and build off the same results in other contexts; e.g. Melly and Santangelo (2015) and Wuthrich (2015).

Theorem 1.1. *Let $\mathbb{S} = l^\infty(\mathcal{YT})^2$. Under Assumptions 2 to 4 and 6 and 1 (given in Appendix A.1)*

$$(\hat{G}_{Y|T}(y|t), \hat{G}^C_{Y|T}(y|t)) \rightsquigarrow (\mathbb{V}_{Y|T}(y|t), \mathbb{V}^C_{Y|T}(y|t))$$

in the space $\mathbb{S}$ where $(\mathbb{V}_{Y|T}, \mathbb{V}^C_{Y|T})$ is a tight Gaussian process indexed by (y,t) with mean 0 and where $\mathbb{V}_{Y|T}(y|t) = \mathbb{G}_{Y|T}(y|t)$ and $\mathbb{V}^C_{Y|T}(y|t) = \int_{\mathcal{X}} \mathbb{G}_{Y|T,X}(y|t,x)\ \mathrm{d}F_X(x) + \int_{\mathcal{X}} F_{Y|T,X}(y|t,x)\ \mathrm{d}\mathbb{G}_X(x)$ where $\mathbb{G}_{Y|T}$, $\mathbb{G}_{Y|T,X}$, and $\mathbb{G}_X$ are given in Lemma 1 in the Appendix.

Theorem 1.1 is an important building block for establishing the limiting processes of each of the parameters of interest in Section 1.2. It essentially follows from the results in Chernozhukov, Fernandez-Val, and Melly (2013) with relatively small differences related to establishing the joint limiting process. It should also be noted that our results hold uniformly in the treatment T though this does not require major changes in the theory. We will show next that each of the parameters of interest is a Hadamard differentiable function of either the counterfactual distribution or the observed distribution. Theorem 1.1 is also useful because it considers the joint limiting process of the observed distribution and the counterfactual distribution which allows one to consider uniform inference on the *difference* between particular parameters under the observed distribution and counterfactual distribution. It will also be important for testing whether or not a particular parameter changes across different values of the treatment.

The next corollary provides a general result for the limiting process of Hadamard differentiable functions of $F_{Y|T}$ and $F^C_{Y|T}$. This result covers all of the parameters of interest in Section 1.2.

Corollary 1. *Let $\mathbb{D} = l^\infty(\mathcal{YT})$ and consider the Hadamard differentiable map $\Gamma : \mathbb{D}_\Gamma \subset \mathbb{D} \mapsto l^\infty(\mathcal{T})$ with derivative Γ'_γ for $\gamma \in \mathbb{D}$. Let $\hat{G}_T(t) = \sqrt{n}(\Gamma(\hat{F}_{Y|T}(\cdot|t)) - \Gamma(F_{Y|T}(\cdot|t)))$*

and $\hat{G}^C_T(t) = \sqrt{n}(\Gamma(\hat{F}^C_{Y|T}(\cdot|t)) - \Gamma(F^C_{Y|T}(\cdot|t)))$. *Then,*

$$(\hat{G}_T(t), \hat{G}^C_T(t)) \rightsquigarrow (\Gamma'_{F_{Y|T}}, \Gamma'_{F^C_{Y|T}})$$

in the space $l^\infty(\mathcal{T})^2$.

In the Supplementary Appendix we show that each of the parameters that we consider in the paper is indeed Hadamard differentiable and give explicit expressions for each term in Corollary 1 (which depend on the particular parameter of interest). In addition, given the results in Theorem 1.1 and Corollary 1, the validity of the empirical bootstrap for conducting uniform inference follows using well known arguments (for example, Van Der Vaart and Wellner (1996) and Kosorok (2007)). Let $\theta(t)$ generically denote one of the parameters of interest in the preceding sections, for example, $F_{Y|T}(y_p|t)$ or $E^C[Y|T=t]$. Let $\hat{\theta}(t)$ denote an estimator of $\theta(t)$. In particular, we can construct uniformly valid confidence bands that cover the entire curve with $(1-\alpha)$ probability for any parameter of interest given by

$$\hat{C}_\theta(t) = \hat{\theta}(t) \pm \hat{c}_{1-\alpha}\hat{\Sigma}(t)^{1/2}/\sqrt{n}$$

where $\hat{c}_{1-\alpha}$ is a critical value satisfying

$$\lim_{n\to\infty} P(\theta(t) \in \hat{C}_\theta(t) \text{ for all } t \in \mathcal{T}) = 1-\alpha$$

Here, $\hat{\Sigma}(t)$ denotes a uniformly consistent estimator of $\Sigma(t)$, the asymptotic variance function of $\sqrt{n}(\hat{\theta}(t) - \theta(t))$, such as

$$\hat{\Sigma}(t) = \frac{q_{0.75}(t) - q_{0.25}(t)}{z_{0.75}(t) - z_{0.25}(t)} \tag{1.3.1}$$

which is the bootstrap interquartile range scaled by the interquartile range of the standard normal distribution (this is a uniformly consistent estimate of $\Sigma(t)$, see Chernozhukov and Fernández-Val (2005)).

Consider the following bootstrap procedure. For some large number B and for each $b = 1, \ldots, B$ compute

$$\hat{c}_b = \sup_{t \in \mathcal{T}} \hat{\Sigma}(t)^{-1/2} |\sqrt{n}(\hat{\theta}^b(t) - \hat{\theta}(t))|$$

where $\hat{\theta}^b(t)$ is the bootstrapped estimate of $\theta(t)$ using the b-th boostrapped sample. Then, setting $\hat{c}_{1-\alpha}$ to be the $(1-\alpha)$ quantile of $\{\hat{c}_b : 1 \leq b \leq B\}$ implies that $\hat{C}_\theta(t)$ asymptotically covers $\theta(t)$ for all values $t \in \mathcal{T}$ with probability $(1-\alpha)$.

We provide the theoretical justification for the above procedure in the Supplementary Appendix. Finally, and using similar arguments as above, one can establish the limiting process and prove the validity of the empirical bootstrap for testing whether or not parameters depend on the value of the treatment as in Section 1.2.3. We also provide the details for this procedure in the Supplementary Appendix.

1.4 Application On Intergenerational Income Mobility

1.4.1 Related Literature

The literature on intergenerational income mobility is vast and we briefly summarize some of the most relevant parts (a much more detailed review of the literature can be found in Black and Devereux (2011)). Our results are related to work that has used quantile regression to study intergenerational mobility (Eide and Showalter 1999; Grawe 2004). These papers show that the distribution of child's income conditional on parents' income narrows as parents' income increases. Our unconditional results can be compared directly with the results in those papers. However, our counterfactuals are fundamentally different than quantile regression specifications that include additional control variables. Richey and Rosburg (2016b) propose a similar counterfactual distribution to the one in the current paper in the context of a decomposition of intergenerational income mobility though they propose a first step distribution regression estimator and second step simulation estimator. The intergenerational elasticity (IGE), which is the slope coefficient from a regression of the log of child's income on the log of parents' income, has a long history in the intergenerational income mobility literature. But recent work has considered more complicated setups such as (1) transition matrices, (2) the probability that child's income is greater than parents' income, and (3) the correlation of the ranks of child's income and parents' income, among other ideas (Jantti et al. 2006; Bhattacharya and Mazumder 2011; Murtazashvili 2012; Chetty, Hendren, Kline, and Saez 2014; Chetty et al. 2014; Murtazashvili, Liu, and Prokhorov 2015; An, Le, and Xiao 2017; Chetty et al. 2017; Collins and Wanamaker 2017; Kitagawa, Nybom, and Stuhler 2017).

Of these, transition matrices are most closely related to our approach and have received considerable attention in the intergenerational income mobility literature (see Jantti et al. (2006), Bhattacharya and Mazumder (2011), Black and Devereux (2011), and Richey and Rosburg (2015), among others). In principle, one could use a transition matrix to calculate the probability that a child's income is below the poverty line for different values of parents' income. However, transition matrices typically pick cutoff points at particular quantiles of parents' income (e.g. at the 25th, 50th, and 75th percentiles) and look at quantiles of child's income as well. A key advantage of our approach is that it does not require choosing cutoffs like this. Another advantage of our approach is that it is straightforward to include covariates in the analysis. A final distinction is that because quantiles of income depend both on an individual's income and on the income of other individuals, transition matrices are relative mobility measures. On the other hand, calculating the probability that a child's income is below the poverty line as a function of parents' income is an absolute mobility measure as it does not depend on outcomes for other individuals.

1.4.2 Data

The data that we use comes from the Panel Study of Income Dynamics (PSID) which has been the primary database used in much of the literature on intergenerational mobility. Like the majority of the income mobility literature using the PSID, we use total family income (including both father's and mother's income) instead of individual income (Chadwick and Solon 2002; Mayer and Lopoo 2005; Bloome 2015).[12] The other main data issue in the intergenerational mobility literature is constructing measures of permanent income. Here we follow existing work and use averages of income over several years to construct the permanent income (Solon 1992; Zimmerman 1992; Mazumder 2005). We construct child's permanent income (our outcome variable) in their adulthood by averaging at least three family incomes conditional on being at least 25 years old and being the head or the spouse of a household. We measure the parents' family income (our treatment variable) by averaging at least three family incomes when the child is 16 years old or younger. Before we calculate these family incomes, we drop yearly family incomes less than $100. We also change all family incomes in all years into 2010 dollars using the CPI-U-RS series. Our sample consists of individuals whose ages are at least 1 in 1987 such that these individuals are at least 25 years old in 2011. Also, these individuals have to be less than 16 years old in 1970 to ensure that these individuals are sons or daughters at the very beginning of the survey. Finally, we drop the Survey of Economic Opportunity (SEO) part of the PSID sample; this is standard in the intergenerational mobility literature.

[12] The main alternative is to use only father's and son's income, but our approach offers several advantages. First, it seems likely that it is total family income that would affect a child's outcomes. Second, this approach allows us to keep daughters in the analysis; in particular, families with one spouse with high income and the other with low income (or out of the labor force) will be treated as high income families in our analysis rather than as low income families.

The covariates that we use in our analysis include child's gender and year of birth and the family head's gender, race, educational attainment, and veteran status. The main complication in obtaining the covariates of the family head is determining who is the family head, because the family head can change over time – for example, parents may divorce, remarry, or die over the course of their child's childhood. We set the family head characteristics as the mode of characteristics for the individual coded as the family head between the time that a child is born and reaches 16 years old. Our sample consists of 3,630 child-parent pairs.

Table A.1 provides summary statistics by quartile of parents' income. The 25th percentile of parents' income is $44,200, the median is $59,200, and the 75th percentile is $78,000. As expected, child's income is increasing in parents' income. On average, children from families in the 1st, 2nd, and 3rd quartiles have higher income than their parents; children from the fourth quartile have lower income on average than their parents.

There are some striking patterns in the data that are immediately noticeable, and most of these differences are most pronounced between the 1st and 2nd quartiles of parents' income. Parents in the first quartile are much more likely to be non-white than parents in the 2nd quartile (28% vs. 7%). Children from families in the 1st quartile are much less likely to have a male head (77% vs. 95%) which likely indicates that these children are from a single parent family. Finally, there are big differences across parents' income quartiles in education. 35% of family heads in the lowest quartile have less than a high school education. The corresponding quantities are 19%, 8%, and 5% for the 2nd, 3rd, and 4th quartiles, respectively. There are also big differences in the fraction of heads with at least a college degree – 7% in the first quartile, 17% in the 2nd, 34% in the 3rd, and 57% in the 4th. Taken together, the summary statistics suggest that child's income is positively correlated with parents' income. But child's income is also correlated with other background family characteristics – primarily education, race, and coming from a two-parent family – that are likely to also be important contributors to a child's income.

Table A.2 presents OLS regression results of the log of child's income on the log of parents' income as well as additional controls. These results are useful to compare with the existing literature as well as to serve as a prelude to our main results. Without additional controls, the estimated IGE is 0.609.[13] Adding demographic controls, as in specification (2) in the table, shrinks the estimated coefficient to 0.573. By far, the most important demographic control is a dummy variable for whether or not the race of a family is non-white. The third specification adds a control for year born which is likely to be important as older individuals have more work experience; it has the expected sign but the estimate of the IGE does not change much. The fourth column adds education controls. Once again, the estimated IGE shrinks considerably to 0.452; so, here, adding additional controls reduced the estimated IGE by about 26%. The coefficients on the family head having less than a high school education and on the family head having at least a college degree (having a high school degree but less than a college degree is the omitted group) are large in magnitude. These results suggest that controlling for covariates such as race and education mitigates the effect of parents' income on child's income, though parents' income is still an important determinant of child's income.

1.4.3 Main Results

Our main results are provided in Figures A.1 to A.5 below. Each one corresponds to one of our main parameters of interest, and they each follow the same pattern. The top left panel provides the parameter as a function of parents' income from the observed data. The top right panel provides the same parameter as a function of parents' income

[13]This estimate is towards the upper end of the range estimates of the IGE in the literature (Mazumder 2005; Black and Devereux 2011; Chetty, Hendren, Kline, and Saez 2014). However, recent work suggests that the IGE is larger using more recent periods, like in the current paper, than in earlier periods (Chetty et al. 2014; Davis and Mazumder 2017).

but using the counterfactual distribution which adjusts for differences in observed covariates across different levels of parents' income. The bottom left panel shows the difference between the parameter coming from the observed distribution and the one coming from the counterfactual distribution. And the bottom right panel tests whether the parameter coming from the counterfactual distribution is the same across all values of parents' income. Each panel provides uniform confidence bands for the parameter of interest. This allows us to reject any hypothesis of interest for the entire function if the band does not cover zero. We also impose that $Q_{Y|T,X}(\tau|t,x) = \alpha_1(\tau)t + x'\alpha_2(\tau)$ (and that X includes an intercept term) and that $Q_{Y|T}(\tau|t) = \beta_0(\tau) + \beta_1(\tau)t$, both for all values of τ.

Question 1: How much does adjusting for covariates matter?

The first part of our analysis considers very similar research questions as much existing work. We first focus on average child's income as a function of parents' income and its derivative which is a local version of the Intergenerational Elasticity (IGE) measure commonly reported in the intergenerational mobility literature.[14] Average child's income as a function of parents' income is reported in Figure A.1. As expected, child's income is increasing in parents' income. This result holds using the observed distribution (top left panel) or after adjusting for differences in covariates (top right panel). On average, children from families with lower income have higher income than their parents while children from higher income families tend to have lower incomes than their parents. They cross at $61,000 without adjusting for covariates and $59,900 after adjusting for covariates.

Most interestingly, however, is that we can reject that adjusting for covariates

[14]For the IGE, we take a numerical derivative of $E[Y|T]$ and $E^C[Y|T]$, respectively. For example, for the counterfactual IGE, we calculate $IGE^C(t) = \delta^{-1}(E^C[Y|T = t + \delta/2] - E^C[Y|T = t - \delta/2])$ for some small, fixed δ (we set this equal to 0.1 in practice). One difference between our results and most existing work is that our measure of the IGE is local, though Landersø and Heckman (2017) have considered a local IGE in previous work. However, the main departure here from existing work is that we also consider average child's income and local IGE *after adjusting for differences in covariates* across different levels of parents' income.

does not make a difference in the estimates. Adjusting for covariates tends to increase expected income of children from low income families and decrease expected income of children from high income families (see the bottom left panel of Figure A.1). This is in line with the results from the previous section where we saw that parents' income was strongly correlated with parents' education, parents' race, and having a male household head. It suggests that adjusting for differences in covariates decreases the strength of the relationship between parents' income and child's income.

As for our local IGE measures (the slope of average child's income as a function of parents' income), with or without adjusting for differences in covariates, they are roughly constant across all levels of parents' income. However, the level is quite different. Without adjusting for covariates, our local measure of IGE tends to be around 0.57 across all values of parents' income. Adjusting for covariates, the local IGE is around 0.42 across all value of parents' income. Taken together, these results suggest that, on average, the effect of increasing parents' income on child's income is roughly the same across all levels of parents' income and that taking into account differences in the distribution of covariates across different levels of parents' income tends to somewhat decrease the effect of parents' income – results that are in line with the existing literature on intergenerational income mobility. Next, we turn to looking at distributional effects of parents' income on child's income.

Question 2: What is the effect of parents' income on the distribution of child's income?

Average child's income conditional on parents' income only tells part of the story of their relationship. Our estimates in the previous section indicate that children from low income families have higher incomes than their parents on average. However, of course, not all children from families whose income takes a particular value have actual incomes equal to the average. Our methods allow us to look at these distributional parameters. First, we consider the effect of parents' income on the probability that a child's income is below the poverty line.

The results for the poverty rate are presented in Figure A.2. Without adjusting for covariates, 21.7% of children from families with incomes at the poverty line are estimated to have incomes below the poverty line themselves. After adjusting for covariates, only 16.3% are estimated to have incomes below the poverty line. At the median of parents' income, without adjusting for covariates 4.3% of children have income below the poverty line and slightly more, 4.7%, have income below the poverty line after adjusting for differences in observed characteristics (this difference is not statistically significant). For children of families in the 90th percentile of income, we estimate that only 1.0% have incomes below the poverty line without adjusting for covariates while 1.7% have incomes below the poverty line when we do adjust for covariates (this difference is not statistically significant). These results say that children from relatively poor families are much more likely to have incomes below the poverty line than children from middle or upper income families. This provides substantially more detail than simply looking at average child's income as a function of parents' income. In fact, children from relatively poor families do not just have lower incomes on average than children from other families, they are much more likely to have very low incomes themselves.

Similarly, children from low income families are much less likely to become "rich" than children from middle or high income families (we set the value to be considered "rich" at $132,923 which is the 90th percentile of income in the U.S. in 2010). Without adjusting for covariates, we estimate that 2.0% of children from families at the poverty line, 7.4% of children from families at the median, and 26.7% of children from families at the 90th percentile become rich. Adjusting for covariates does not make much difference except for children from families at the 90th percentile where the estimate is reduced to 20.1%.

Next, we consider how wide the distribution of child's income is as a function of parents' income. To do this, we examine the variance of child's income and the inter-quantile range of child's income. First, Figure A.4 plots the variance of child's income as function of parents' income. There are clear differences between the variance depending on whether or not the model adjusts for covariates. Without covariates, the variance of child's income is higher for children with low income parents relative to high income parents (see the top left panel). However, once one accounts for differences in covariates across parents' incomes, the variance flattens (see the top right panel and bottom right panel). Our results for the variance, however, are relatively imprecise and we cannot reject that adjusting for covariates has no effect nor can we reject that the results that adjust for covariates do not change across parents' incomes. On the other hand, we can reject that the variance is constant in the case where we do not adjust for covariates (results not shown in figure).

The inter-quantile range tells a similar story. These results are presented in Figure A.5 (in the figure, we set $\tau_1 = 0.9$ and $\tau_2 = 0.1$). Without covariates, it appears that the spread of child's income, as measured by the IQR, is decreasing in parents' income. But adjusting for covariates instead indicates that the IQR is flat across parents' incomes and that the differences are driven by differences for parents with very low income.

Treatment Effects

It seems unlikely that our estimates should be considered to be estimates of the causal effect of parents' income on child's income, but we briefly consider the relationship of our estimates to treatment effect estimates under Assumption 1.[15] We

[15]We suspect that, in the context of intergenerational income mobility, estimates of the average effect of moving from a low level of parents' income to a high level of parents' income using our approach are likely to overstate the causal effect of parents' income on child's income. This would be the case if children of high income parents have some latent characteristics (or their parents have some latent characteristics) that lead to higher income relative to children of low income parents even after conditioning on observables. One small piece of evidence related to this concerns parents' education. For education, we include three dummy variables – less than high school, high school

estimate that, on average, moving from the 25th percentile to the 75th percentile of parents' income increases child's income by 24.7 log points. Under the assumption of unconfoundedness, this should be interpreted as a causal effect. Similarly, we estimate that, under unconfoundedness, moving from the 25th percentile to the 75th percentile of parents' income decreases the probability of a child's income being below the poverty line by 4.2 percentage points (a 59% reduction).

Summary of Main Results

Our estimates of average child's income as a function of parents' income and of the local IGE are largely in line with the existing literature. Children from families with relatively low income have lower earnings than children from higher income families. This result holds, though is somewhat reduced, when differences in covariates such as race and education are accounted for.

More interestingly, we were able to estimate the entire distribution of child's income as a function of parents' income. We found that children from families with low incomes were much more likely to have incomes below the poverty line than children from higher income families; again, this was somewhat mitigated when adjustments were made for differences in background characteristics, but there were still substantial differences. We also found suggestive evidence that the variance of child's income was larger for children from low income families than from high income families, but adjusting for differences in covariates completely flattened the variance across parents' income levels.

---

graduate but not a college graduate, or a college graduate. Looking *within* these three groups, parents in the top quartile have more education than parents from the bottom quartile; for example, parents with a college degree from the top quartile are relatively more likely to have an advanced degree and parents with a high school degree are relatively more likely to have some college than parents in the bottom quartile. Likewise, we suspect that our estimates of the effect of parents' income on the probability of a child having income below the poverty line will overstate the causal effect of parents' income for similar reasons. It is less clear the direction of the bias for estimating the spread parameters, such as the variance of child's incomes, conditional on parents' income.

1.5 Conclusion

This paper has developed new tools to study distributional effects of a continuous treatment. We proposed a straightforward three step procedure to estimate these distributional effects that is based on first step quantile regression. Our procedure is easy to implement in practical applications and more flexible than making distributional assumptions about the treatment or outcome.

We applied these methods to study intergenerational income mobility. Our methods allow us to (1) study the entire distribution of child's income conditional on parents' income, (2) adjust for differences in observed characteristics among children who have parents with different income levels, and (3) treat parents' income as a continuous variable rather than splitting it into a small number of groups. These tools may be useful to researchers in other fields who are interested in distributional effects with a continuous treatment or are interested in the causal effect of a continuous treatment under the assumption of unconfoundedness.

CHAPTER 2

LOCAL INTERGENERATIONAL ELASTICITIES

2.1 Introduction

The intergenerational elasticity (IGE) is the most commonly reported measure of intergenerational income mobility (see, for example, Solon (1992)). It is the coefficient from the regression of the log of child's income on the log of parents' income. Large values of the IGE indicate a relative lack of mobility and small values indicate relatively high mobility. The IGE, however, is a global measure of intergenerational mobility and some researchers have explored how the IGE varies across different values of parents' income (e.g. Landersø and Heckman (2017)) as a local intergenerational mobility measure.[1] Researchers studying intergenerational mobility have also been interested in the role of other background characteristics (e.g. race and education) that are correlated with both parents' income and child's income in explaining intergenerational mobility (for example, Bowles and Gintis (2002), Blanden, Gregg, and Macmillan (2007a), and Richey and Rosburg (2017)). This paper develops new tools for estimating a local intergenerational elasticity after first adjusting for differences in the distribution

[1] Also relatedly, Bratsberg et al. (2007) and Björklund, Roine, and Waldenström (2012) group their data by percentiles of parents' income and calculate intergenerational elasticities within groups which is similar to our procedure; Murtazashvili (2012) uses a random coefficients model to allow the effect of parents' income to differ across individuals.

of characteristics across different values of parents' income.

The interpretation of local intergenerational income elasticities is somewhat subtle. First, they are local effects and should be interpreted as the effect on average child's income for *marginal* changes in parents' income. They do not indicate what would happen if parents' income changed dramatically. Also, local intergenerational elasticities are the effect on average child's income (possible after adjusting for differences in covariates) and do not answer questions like how parents' income affects the probability that child's income is below the poverty line.[2]

We propose a semiparametric estimator of the adjusted local intergenerational elasticity that allows for the effects of parents' income and covariates on child's income to change across different values of parents' income. We develop the asymptotic properties of a local linear estimator of the local intergenerational elasticity and our adjusted local intergenerational elasticity. Our estimators converge more slowly than parametric estimators though they do not suffer from the curse of dimensionality.

We apply our method to data from the Panel Study of Income Dynamics. Without adjusting for covariates, the local IGE is relatively large and tends to decrease with parents' income. Adjusting for covariates decreases the local IGE across all values of parents' income; however, there is still a strong relationship between child's income and parents' income. Adjusting for covariates also substantially flattens the local IGE across different values of parents' income.

2.2 Parameters Of Interest

Let Y denote the log of child's income, T denote the log of parents' income, and X denote a $k \times 1$ vector of covariates. Next, we define our two main objects of interest.

[2] Richey and Rosburg (2016a) and Callaway and Huang (2018a) consider how parents' income affects the entire distribution of outcomes though the approach in those papers is substantially different from that of the current paper.

Definition 1. *The* ***Local Intergenerational Elasticity (LIGE)*** *is given by*

$$LIGE(t) = \frac{\partial E[Y|T=t]}{\partial t}$$

$LIGE(t)$ measures the local effect of parents' income on average child's income at a particular value of parents' income t. This type of parameter has been considered in Landersø and Heckman (2017). We are also interested in the effect of parents' income on child's income after adjusting for differences in the distribution of observed characteristics that are related to child's income (e.g. parents with high income are likely to have relatively high education as well) across different values of parents' income. Note that here we are not attempting to establish the causal effect of parents' income; rather, we are trying to imagine what average child's income would be if the return to observed characteristics were held fixed but the distribution of characteristics was changed to be the same as the distribution of characteristics for all individuals in the population.

Next, note that the observed average child's income conditional on parents' income is given by

$$E[Y|T=t] = \int_{\mathcal{X}} E[Y|T=t, X=x] \; dF_{X|T}(x|t)$$

which holds by the law of iterated expectations and where $\mathcal{X}$ denotes the support of X. We consider the counterfactual average outcome conditional on parents' income where the return to characteristics and parents' income is held fixed but the distribution of observed characteristics (conditional on $T=t$) is changed to be the distribution of characteristics for the entire population; that is

$$E^C[Y|T=t] = \int_{\mathcal{X}} E[Y|T=t, X=x] \; dF_X(x)$$

Given this counterfactual, we define our main parameter of interest next.

Definition 2. *The **Adjusted Local Intergenerational Elasticity (ALIGE)** is given by*

$$ALIGE(t) = \frac{\partial E^C[Y|T=t]}{\partial t}$$

$ALIGE(t)$ corresponds to $LIGE(t)$ except that it occurs after adjusting for differences in the distribution of covariates across different values of parents' income. Our next aim is to develop a flexible model for $E[Y|X,T]$ in order to ultimately estimate the $ALIGE$. We make the following assumption

Assumption 6 (Smooth Coefficient Model)**.**

$$Y = X'\beta(T) + U$$

Assumption 6 is key for implementing our method. This type of semiparametric model is called a smooth coefficient model (see, for example, Li, Huang, Li, and Fu (2002) and Cai, Fan, and Li (2000) as well as Callaway and Huang (2018b) for a similar model in the context of decompositions with a continuous treatment). A leading alternative idea would be to estimate $E[Y|T,X]$ nonparametrically and plug in these estimates to obtain estimates of the ALIGE. With the moderate amount of data typically available in applications this approach is not likely to be feasible as it suffers from the curse of dimensionality.[3] Even in a case like ours where most of the covariates are discrete, splitting the sample for each possible combination of the discrete variables and then employing nonparametric estimation does not appear to be a feasible strategy either due to sample sizes being extremely small within some cells. It is straightforward to handle this case with our approach though. Another

[3]The curse of dimensionality would be somewhat mitigated from integrating out X; however, the fully nonparametric approach is still likely to be difficult to carry out in practice.

alternative would be to assume that the conditional expectation follows some particular parametric model, but it seems challenging in practice to specify the right functional form; in particular, for the derivative of $E[Y|T, X]$ to depend on X, the functional form must include interactions between T and X which may be difficult to choose appropriately. Our approach, on the other hand, is quite flexible. We allow the effect of covariates and parents' income to depend on the value of parents' income. For example, the effect of parents' education on child's income can vary across different values of parents' income.

The next result characterizes the ALIGE under Assumption 6.

Proposition 1. *Under Assumption 6,*

$$ALIGE(t) = E[X]' \frac{\partial \beta(t)}{\partial t}$$

Proof. First, notice that under Assumption 6,

$$\begin{aligned} E^C[Y|T = t] &= \int_{\mathcal{X}} x'\beta(t)\ dF_X(x) \\ &= E[X]'\beta(t) \end{aligned}$$

Taking the derivative with respect to t implies the result. □

Our approach exploits the uniqueness of T among the set of conditioning variables. Importantly, unlike the fully nonparametric approach, our approach will not suffer from the curse of dimensionality. Our estimator will converge at a slower rate than parametric estimators, but its rate will not slow down due to adding more covariates.

2.3 Estimation

Estimating the LIGE is relatively straightforward. We use local linear kernel regression and an estimate of the derivative is given by the (local) coefficient on the linear term (also notice that the results for the LIGE are a special case of the results for the ALIGE by taking X to only include a constant). For estimating the ALIGE, notice that a first order Taylor approximation of the model in Assumption 6 around t implies

$$Y \approx X'\beta(t) + (T-t)X'\frac{\partial \beta(t)}{\partial t} + U$$

Then, a local linear estimator of $(\beta(t), \partial\beta(t)/\partial t)$ is given by

$$\begin{pmatrix} \hat{\beta}(t) \\ \widehat{\frac{\partial \beta(t)}{\partial t}} \end{pmatrix} = \left(\mathbf{X}'\mathbf{K}(t)\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{K}(t)\mathbf{y}$$

where $\mathbf{X}$ is an $n \times 2k$ matrix (where k is the dimension of X) with the ith row given by $\mathbf{X}_i = (X_i', (T_i - t)X_i')$ and $\mathbf{K}(t)$ is an $n \times n$ diagonal matrix whose ith diagonal element is given by $K_h(T_i - t) = K((T_i - t)/h)$ where K is a kernel (satisfying some regularity conditions; in practice, we use a trimmed Gaussian kernel though other choices are possible) and h is a bandwidth.

Then, one can estimate the ALIGE as follows

$$\widehat{ALIGE}(t) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)' \frac{\widehat{\partial \beta(t)}}{\partial t}$$

In practice, we estimate $LIGE$ and $ALIGE$ over a grid of L possible values for t given by $t^* = (t_1, t_2, \ldots, t_L)$.

2.3.1 Asymptotic Theory

This section develops the limiting distribution of the LIGE and the ALIGE.

For estimating the LIGE, first let $Y = g(T) + \epsilon$ where $g(t) = E[Y|T = t]$. Under standard regularity conditions for local linear estimators (see, for example, Li and Racine (2007)[Theorem 2.7]),[4] one can show that

$$n^{1/2}h^{3/2}(\widehat{LIGE}(t) - LIGE(t)) = \frac{1}{\kappa_2 f_T(t)} n^{-1/2}h^{-3/2}\sum_{i=1}^{n}(T_i - t)K_h(T_i - t)\,\epsilon_i$$
$$\xrightarrow{d} N(0, V_L)$$

with $V_L = f_T^{-1}(t)\kappa_2^{-2}\kappa_{22}E(\epsilon^2|t)$ and where $f_T(t)$ is the marginal density of T, $\kappa_2 = \int v^2K(v)\,dv$, and $\kappa_{22} = \int v^2K^2(v)\,dv$.

Next, for the ALIGE, consider

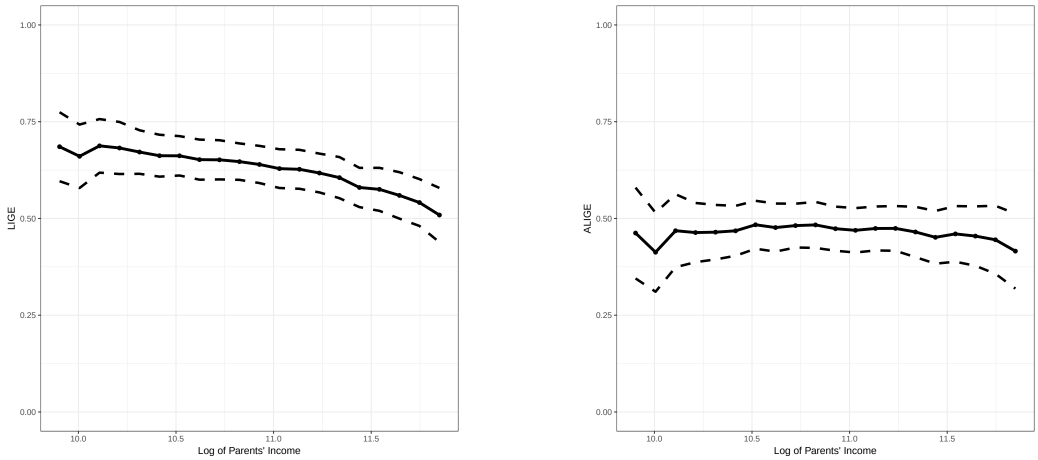
$$\widetilde{ALIGE}(t) = E[X]'\frac{\widehat{\partial\beta(t)}}{\partial t}$$

$\widehat{ALIGE}$ and $\widetilde{ALIGE}$ are asymptotically equivalent because $n^{-1}\sum_{i=1}^{n} X_i$ converges to $E[X]$ faster than the terms that we estimate nonparametrically. Thus, the asymptotic behavior of $\widehat{ALIGE}$ is driven by the behavior of the local linear term. Under standard regularity conditions for smooth coefficient models (see Cai, Fan, and Yao (2000) and Li, Huang, Li, and Fu (2002)) one can therefore show that

$$\begin{aligned} &n^{1/2}h^{3/2}(\widehat{ALIGE}(t) - ALIGE(t)) \\ &\quad = n^{-1/2}h^{-3/2}\sum_{i=1}^{n} E[X]'\left[\kappa_2 f_T(t)E[XX'|t]\right]^{-1} X_i(T_i - t)K_h(T_i - t)U_i \\ &\quad \xrightarrow{d} N(0, V_A) \end{aligned}$$

[4]Practically, the most important regularity condition is on the bandwidth. We use cross-validation to choose the bandwidth. In practice, this will "undersmooth" for the derivative term in the local linear estimator leading to the bias term going to zero asymptotically.

Figure 2.1: The LIGE and ALIGE as a function of parents' income



LIGE ALIGE

Notes: The left panel plots the LIGE and the right panel plots the ALIGE. In each panel, the dashed lines are pointwise 95% confidence intervals computed using the wild bootstrap with 500 iterations. *Sources:* Panel Study of Income Dynamics, as described in text

where $V_A = f_T^{-1}(t)\kappa_2^{-2}\kappa_{22} E[X]' E[XX'|t]^{-1} E[XX'U^2|t] E[XX'|t]^{-1} E[X]$. In practice, we carry out pointwise inference using the wild bootstrap.

2.4 Application

We use data from Callaway and Huang (2018a) which comes from the Panel Study of Income Dynamics (PSID). The data consists of 3,630 child-parent measures of permanent income along other characteristics including child's gender and birth year, and the family head's gender, race, educational attainment, and veteran status. See Callaway and Huang (2018a) for a detailed discussion of this data as well as summary statistics for the dataset. Regressing the log of child's income on the log of parents' income results in an estimated IGE of 0.603.

Our main results are presented in Figure 2.1. We estimate the LIGE and ALIGE over a grid of twenty equally spaced values of t ranging from $\log(20,000)$ (roughly equal to the poverty line) to $\log(140,000)$. Without adjusting for differences in covariates, the LIGE is equal to 0.69 for children whose parents' income was $20,000. It declines substantially in parents' income. For children whose parents income was $140,000, the LIGE is 0.51. These results suggest that the effect of parents' income on child's income varies across different values of parents' income; standard measures of intergenerational mobility such as the IGE cannot show this type of heterogeneity.

Adjusting for covariates somewhat diminishes estimates of local intergenerational elasticities. However, even after adjusted for differences in observed covariates across different values of parents' income, the link between child's and parents' income is still strong. For children whose parents' income was $20,000, the ALIGE is estimated to be 0.46. Interestingly, adjusting for covariates also substantially flattens estimated local intergenerational elasticities. The point estimates of the ALIGE are very similar across all values of parents' income (perhaps somewhat declining) and we cannot reject that the ALIGE is constant for all values of parents' income.

2.5 Conclusion

This paper has developed new "local" measures of intergenerational elasticities that allow for the researcher to adjust for differences in the distribution of characteristics across different values of parents' income. We developed a flexible semiparametric estimator of the adjusted local IGE and studied its properties. We found that our adjusting for covariates decreased the local IGE at all values of parents' income and tended to flatten the local IGE as well.

There are many interesting possible extensions of the current framework. First, it would be interesting to develop formal tests that the LIGE or ALIGE are constant across all values of parents' income which could be accomplished using results from the specification testing literature (for example, Hardle and Mammen (1993) and Zheng (1996), among many others); or, relatedly, to develop uniform confidence bands for the LIGE or ALIGE. Second, one could develop tests for whether the LIGE and ALIGE are equal at particular values of parents' income or across all values of parents' income. Finally, it would be relatively straightforward using our approach to examine the role of each covariate in *explaining* the difference between the LIGE and ALIGE. This would allow one to decompose the gap between the LIGE and ALIGE into parts due to, for example, differences between the distribution of race and education at particular values of parents' income and the overall distribution of race and education. We leave these extensions to future work.

CHAPTER 3

DECOMPOSING DIFFERENCES IN PORTFOLIO RETURNS BETWEEN NORTH AMERICA AND EUROPE

3.1 Introduction

We observe that there exist economically large differences in mean returns on portfolios formed from sorts on size and value, profitability, or investment between North America and Europe (Fama and French (2017)), although the differences are mostly insignificant in statistical terms due to large standard deviations (Table 4.2). The mean differences in factors between the two regions are also large in practical terms and the correlations between factors in North America are quite different from the ones in Europe (Table 4.1). Applying Fama and French (2015) (FF(2015)) five-factor asset pricing model on the portfolios examined in the paper, we also find that the factor exposures are different between North America and Europe (not shown).

One interesting question raised is that what are the most important explanations accounting for these differences? To explore this question, we decompose these differences into FF(2015) five factors using Blinder-Oaxaca decomposition method (Oaxaca (1973) and Blinder (1973)).[1] More specifically, we decompose these differences

[1] Blinder-Oaxaca decomposition method is widely used in labor economics, see e.g., . Blinder-

into aggregate composition and structure components. Using the language of labor economics, the differences are attributed to two parts, one of which is the aggregate composition component and the other of which is the structure component. The aggregate composition component is due to the differences in the distributions of factors, i.e., factor risk premia, between North America and Europe. The structure component is due to the differences in the structures of factors (factor exposures), i.e., factor risks. Further, the detailed decomposition helps answer the following questions. Which factors are behind most of differences? To what extent have the differences in portfolio returns because of the differences in market factor risk premium or in risk on investment factor?

We also observe that the differences in portfolio returns between North America and Europe are not normally distributed. Thus, we decompose the differences in a series of quantile portfolio returns between North America and Europe. The decomposition for a sequence of quantile differences helps explore some interesting questions. For instance, do the roles that the factors play in explaining the differences vary at different levels of portfolio returns? More specifically, does the aggregate composition component vary at different levels of portfolio returns? Does the composition component associated to market factor vary at different levels of portfolio returns? Is the structure component linked to investment factor changing at different levels of portfolio returns? These questions are of great interesting because they are equivalent to asking the following questions. Does the role that the factor risk premia play in explaining differences in quantile portfolio returns vary at different levels of portfolio returns? Does the role that the market factor risk premium plays vary at different levels of portfolio returns?

Oaxaca decomposition can be found in the financial economics literature, e.g., Wang and Hanna (2007), Alesina, Lotti, and Mistrulli (2013), Robb, Fairlie, and Robinson (2014), Shin and Hanna (2015), Aristei and Gallo (2016), Brown and Previtero (2014), Kabir and Shakur (2014), Montecino, Epstein, et al. (2015) among others. Füss, Gietzen, and Rindler (2011) decompose the bond spreads over the course of the crisis to study the impact of changes in risk perception. However, we are not aware of any other papers which decompose the equity returns. Thus, we add decomposition analysis to the literature on empirical asset pricing.

Does the role that the investment factor risk plays vary at different levels of portfolio returns? The answers to these questions could have implications for the question that whether the factor risks or risk premia vary at different levels of portfolio returns. Thus, the examinations on these questions could shed further light on empirical asset pricing by adding to the literature on variations of risks or risk premia.[2]

To decompose the quantiles, one usually needs to deal with nonlinear models. However, in the conventional Blinder-Oaxaca decomposition the dependent variable is linear with the coefficients and thus the detailed decomposition is path independent. Gomulka and Stern (1990), Fairlie (2005) and Bauer and Sinning (2008) extends the Blinder-Oaxaca decomposition technique to nonlinear models like probit and logit. For the nonlinear case, the decomposition could be sensitive to the order of decomposition, namely the results will be path dependent. Decomposing differences in quantile portfolio returns is also related to the literature on decomposing general distributional statistics using flexible methods, see e.g., 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), Rothe (2010), and Chernozhukov, Fernandez-Val, and Melly (2013), Richey and Rosburg (2016a) among others. Fortin, Lemieux, and Firpo (2011) offers a comprehensive overview on the decomposition methods in economics.

The components in the decomposition for quantile differences can be computed by a series of counterfactual distributions based on distribution regressions. Distribution regression is a continuum of binary regressions. Chernozhukov, Fernandez-Val, and Melly (2013) show that distribution regression provides a flexible model for the entire conditional distribution and also establish the central limit theorems. In the paper, we use distribution regressions to decompose differences in a series of quantile portfolio

[2]See, e.g., Gilbert, Hrdlicka, and Kamara (2018), Graham and Harvey (2018) and among others. Gilbert, Hrdlicka, and Kamara (2018) show exposures to SMB and HML vary with firms' earnings announcement month and Graham and Harvey (2018) show risk premia are higher during recessions and higher during periods of uncertainty.

returns.

Also, in the process of decomposing quantile differences, we estimate a series of counterfactual distributions which are of interesting in their own right. To better predict the returns and describe precisely the risks, one needs to estimate the higher-order multidimensional structure of the portfolio returns (Rothschild and Stiglitz (1971)).[3] With the entire distribution of portfolio returns, one can estimate any higher-order structure. Also, with a series of counterfactual distributions one is able to test functional hypotheses such as no-effect, positive effect, or stochastic dominance had the marginal distribution of the market factor in North America been replaced by the one in Europe. These tests are of interest in their own right. For size limitation, however, we focus on the decomposition analysis in the paper.

Notice that FF(2015) five-factor model is linear, which can not capture the nonlinear relationship between portfolio returns and factors. In the paper, we use distribution regressions to estimate the entire distribution of portfolio returns and then quantiles. Distribution regressions help capture the non-linear relationship between portfolio returns and factors. Distribution regressions estimate well the quantiles of returns on portfolios studied in the paper (not shown). Also, we show that the observed quantile differences in portfolio returns between North America and Europe fit the true quantile differences well, especially in the low quantiles. Thus, we believe distribution regressions do a good job in describing the entire distribution of portfolio returns.

In the paper, we use Blinder-Oaxaca decomposition method to decompose differences in mean portfolio returns. The decomposition components for differences in a series of quantile portfolio returns are computed via the observed distributions and a series of counterfactual distributions based on distribution regressions. We also compute the components in the decomposition for mean differences using distribution

[3]The mean-variance analysis based on conditional first and second moments of asset returns are sufficient to inform investors' choice only under special assumptions, such as multivariate normality of asset returns or quadratic utility function of investors. However, the returns on portfolios examined in this paper are not normally distributed (not shown).

regressions. We compare the decomposition results for mean differences from two methods, in which the former are path independent and the later are path dependent.

We find that the differences in mean portfolio returns are mostly due to the differences in risk premia on factors, especially market and size factors, and the differences in factor risks seem to play an insignificant role in aggregate.[4] The decompositions for differences in a series of quantile portfolio returns also show that the quantile differences are mostly due to the differences in risk premia on factors, especially market and size factors, and the differences in factor risks in aggregate explain little of the quantile differences. We also find that the roles that the market and size factor risk premia play in explaining differences vary at different levels of portfolio returns. The results are robust to the changes in the structures used as reference and the decomposition orders. The roles that the risks on some factors play also seem to vary at different levels of portfolio returns. These findings imply that the factor risks and risk premia on factors vary at different levels of portfolio returns, shedding further light on empirical asset pricing.

We organize the rest of the paper as follows. Section 4.2 describes FF(2015) five-factor asset pricing model. Section 3.3 shows the decomposition for mean differences using Blinder-Oaxaca decomposition method. Section 3.4 describes the decomposition for differences in general distributional statistics. In Section 3.5, we present the decomposition for differences in quantiles and inference processes. Section 4.4 describes data and show the main results. Section 3.7 presents several robustness checks. We conclude in Section 4.6.

[4]We note an important limitation of decomposition from Fortin, Lemieux, and Firpo (2011): "A second important limitation is that while decompositions are useful for quantifying the contribution of various factors to a difference or change in outcomes in an accounting sense, they may not necessarily deepen our understanding of the mechanisms underlying the relationship between factors and outcomes."

3.2 Fama And French's Five-factor Model

There is much evidence that average stock returns are related to overall market performance (Sharpe (1964), Lintner (1965), and Breeden (1979)), firm's size (Banz (1981)), value (Basu (1983) and Rosenberg, Reid, and Lanstein (1985)), profitability (Haugen and Baker (1996) and Novy-Marx (2013)) and investment (Titman, Wei, and Xie (2004), Cooper, Gulen, and Schill (2008), and Aharoni, Grundy, and Zeng (2013)). [5] Motivated by the dividend discount valuation model, FF(2015) test a five-factor model that adds profitability and investment factors into the market, size and value-growth factors of Fama and French (1993) three-factor model. Using international data, Fama and French (2017) study international markets including North America and Europe, and show the five-factor model largely absorbs the patterns in average returns.

The FF(2015) five-factor time-series regression is

$$R_{it} = a_i + b_i Mkt_t + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{it} \qquad (3.2.1)$$

where e_{it} is a zero-mean residual and

- R_{it} is the return on portfolio i for month t in excess of risk-free rate (the one-month US Treasury bill rate).
- Mkt_t is the value-weight market portfolio return minus the risk-free rate.
- SMB_t (size factor), HML_t (value factor), RMW_t (profitability factor) and CMA_t (investment factor) are differences between the returns on diversified portfolios of small and big stocks, high and low B/M stocks, stocks with robust and

[5] There are more research including Breeden, Gibbons, and Litzenberger (1989), Reinganum (1981), Cohen, Gompers, and Vuolteenaho (2002), Fairfield, Whisenant, and Yohn (2003), Fama and French (2006), Fama and French (2008), Hou, Xue, and Zhang (2015), and Fama and French (2016) and others.

weak profitability, and stocks with low and high (conservative and aggressive) investment, respectively.

- a_i is α for portfolio i and b_i, s_i, h_i, r_i and c_i are factor exposures.

Remark 3.1. In empirical asset pricing, expected factor returns are factor risk premia and factor exposures measures factor risks.

Remark 3.2. The factors and the portfolios in the left hand side of the regression can be constructed using various sorts. In the paper, we use the factors from 2×3 sorts on size and B/M, profitability or investment. The 75 portfolios used to decompose are portfolios constructed from 5×5 sorts for each region. The detail will be present in Section 4.4.

Remark 3.3. One concern about the factor model is that the role of size and book-to-market factors in the three-factor model is spurious, arising only because size and book-to-market portfolios are used for both dependent and explanatory returns. Fama and French (1993) use split-sample tests to address this concern.

Remark 3.4. Factors are meant to mimic the underlying risk factors in returns related to the variable sorted. For example, the value factor mimics the risk factor in returns related to book-to-market equity ratio. The value, profitability, and investment factors are different mixes of size, value, profitability, and investment effects in returns because of the correlations between the size, value, profitability, and investment variables used to construct factors.[6]

For the sake of clarity in decomposition, we simplify the notation by dropping the subscript i and t and adding the subscript g, $g = 0, 1$, in which 0 denotes North America and 1 represents Europe. Then, we have

$$Y_g = X_g' \beta_g + e_g \tag{3.2.2}$$

[6]High B/M value stocks tend to have low profitability and investment, and low B/M growth stocks, especially large low B/M stocks, tend to be profitable and invest aggressively. (Fama and French (1995))

where Y_g denotes the excess portfolio returns on a portfolio in region g, X_g is a vector of a constant and five factors in region g and e_g is error term. Hereafter, we mean the excess portfolio returns when the portfolio returns are mentioned. The distinction will be made whenever emphasis is necessary.

3.3 Blinder-Oaxaca Decomposition

This section shows how to decompose the difference in mean portfolio returns between North America and Europe using conventional Blinder-Oaxaca decomposition method. We run regression 3.2.2 for each portfolio in North America and Europe separately.[7] With estimates $\hat{\beta}_g$ for each portfolio and using the structures in North America, $\hat{\beta}_0$, as reference, the observed overall mean difference in portfolio returns between North America and Europe can be decomposed as follows

$$\begin{aligned}
\hat{\Delta}_O &= \hat{E}(Y_0) - \hat{E}(Y_1) \\
&= \bar{X}_0'\hat{\beta}_0 - \bar{X}_1\hat{\beta}_1 \\
&= \underbrace{(\bar{X}_0' - \bar{X}_1')\hat{\beta}_0}_{\hat{\Delta}_C} + \underbrace{\bar{X}_1'(\hat{\beta}_0 - \hat{\beta}_1)}_{\hat{\Delta}_S} \\
&= \sum_{j=1}^{6} \underbrace{(\bar{X}_{0j} - \bar{X}_{1j})\hat{\beta}_{0j}}_{\hat{\Delta}_{C_j}} + \sum_{j=1}^{6} \underbrace{\bar{X}_{1j}(\hat{\beta}_{0j} - \hat{\beta}_{1j})}_{\hat{\Delta}_{S_j}}
\end{aligned}$$

where

- $\hat{\Delta}_O$ is the observed overall mean difference in the portfolio returns between North America and Europe.
- $\bar{X}_g$ is the average factor returns in region g.

[7] We show the decomposition results by averaging out the results for all portfolios.

- $\hat{\Delta}_C = (\bar{X}_0' - \bar{X}_1')\hat{\beta}_0$ and $\hat{\Delta}_S = \bar{X}_1'(\hat{\beta}_0 - \hat{\beta}_1)$ are the aggregate composition and structure components, respectively.

- $\hat{\Delta}_{C_j} = (\bar{X}_{0j} - \bar{X}_{1j})\hat{\beta}_{0j}$ and $\hat{\Delta}_{S_j} = \bar{X}_{1j}(\hat{\beta}_{0j} - \hat{\beta}_{1j})$ are the detailed composition and structure components linked to the jth factor, respectively.

We can see that the aggregate composition component is due to the differences in average factor returns and the aggregate structure component is due to the differences in factor exposures. The detailed composition component associated to jth factor is due to the difference in averages of jth factor returns and the detailed structure component is due to the difference in jth factor exposures. That is, the composition components are due to differences in factor risk premia and the structure components are due to differences in factor risks. We note that $\hat{\Delta}_{C_1}$ is equal to zero because the first element in each factor vector X_g is a constant. Using Blinder-Oaxaca decomposition, the decomposition results for the mean difference are path independent, that is, the detailed decomposition results are not affected by the order of decomposition. We also decompose the mean difference using distribution regression below, in which the results are path dependent. We compare two results.

3.4 Decomposition For General Distributional Statistics

In this section, we first introduce several kinds of counterfactul distributions which are useful in decomposition. We then present the processes of decomposing the difference in general distributional statistics including quantile as a special case.

3.4.1 Counterfactual Distributions

We take as the primary building blocks of counterfactuals the conditional distribution of portfolio returns in region g, $F_{Y_g|X_g}(y|x)$, and the distribution of X_g. Notice that the observed distribution of returns on a portfolio in region g is given by

$$F_{Y\langle g|g\rangle}(y) = \int F_{Y_g|X_g}(y|x)dF_{X_g}(x) \tag{3.4.1}$$

that is, $F_{Y\langle g|g\rangle}(y)$ is the same as integrating the conditional distribution of portfolio returns in region g over the distribution of X_g.

We obtain a counterfactual of $F_{Y\langle g|g\rangle}(y)$ as we change the distribution of X_g or the conditional distribution, $F_{Y_g|X_g}(y|x)$. Let $F_{Y\langle 0|1\rangle}(y)$ represent the counterfactual distribution of returns on a portfolio which has the distribution of X_1 and the conditional distribution, $F_{Y_0|X_0}(y|x)$. We obtain $F_{Y\langle 0|1\rangle}(y)$ by integrating the conditional distribution of portfolio returns in North America with respect to the distribution of factors in Europe as follows

$$F_{Y\langle 0|1\rangle}(y) \equiv \int F_{Y_0|X_0}(y|x)dF_{X_1}(x). \tag{3.4.2}$$

We can see the difference between this counterfactual distribution, $F_{Y\langle 0|1\rangle}(y)$, and the observed distribution, $F_{Y\langle 1|1\rangle}(y)$, is only on the conditional distributions used to compute the distributions. Or, the difference between this counterfactual distribution, $F_{Y\langle 0|1\rangle}(y)$, and the observed distribution, $F_{Y\langle 0|0\rangle}(y)$, is only on the distributions of X used to compute the distributions.

In this paper, we decompose differences in mean and quantiles of portfolio returns. Especially in the quantile cases, we need to sequentially estimate the detailed components and thus suffer from path dependent problems. Therefore, we first select the order in which the decomposition is performed. For the main results in this paper, we calculate the detailed components in the following order: the constant, market, size, value, profitability and investment factors. Accordingly, we have five composition components and six structure components.[8] To compute detailed decomposition components of differences in quantile portfolio returns, we define the following counterfactual distributions.

Let $X_{(0,1,j)}, j = 1 \cdots 6$ denote a vector of factors in which the first j factors of X_0 have been replaced with the first j factors of X_1, holding the last $6 - j$ factors of X_0 constant. For instance, $X_{(0,1,3)}$ denotes a vector of factors in which the constant term, market factor, and size factor are from Europe and the value, profitability and investment factors from North America. Note that $X_{(0,1,6)} = X_1$ and $X_{(1,0,6)} = X_0$. Let $F_{X_{(0,1,j)}}(x)$ represent the joint distribution of factors $X_{(0,1,j)}$.

Let $F_{Y\langle 0|(0,1,j)\rangle}(y)$ represent the counterfactual distribution of returns on a portfolio which has the conditional distribution, $F_{Y_0|X_0}(y|x)$, and the distribution of $X_{(0,1,j)}$. $F_{Y\langle 0|(0,1,j)\rangle}(y)$ can be estimated by integrating $F_{Y_0|X_0}(y|x)$ over $X_{(0,1,j)}$ as follows

$$F_{Y\langle 0|(0,1,j)\rangle}(y) \equiv \int F_{Y_0|X_0}(y|x) dF_{X_{(0,1,j)}}(x). \tag{3.4.3}$$

Let $F_{Y_{(0,1,j)}|X}(y|x)$ denote the counterfactual conditional distribution of returns on a portfolio which has the structures associated to the first j factors in $F_{Y_0|X_0}(y|x)$ replaced by the structures associated to the first j factors in $F_{Y_1|X_1}(y|x)$, holding constant the structures associated to the last $6-j$ factors in $F_{Y_0|X_0}(y|x)$. For instance, $F_{Y_{(0,1,3)}|X}(y|x)$ denotes the conterfactual conditional distribution of returns on a portfolio that has the structures linked to the constant term, market and size factors in North America replaced by the structures linked to the ones in Europe, holding constant the structures associated to the value, profitability and investment factors in North America. Note that $F_{Y_{(0,1,6)}|X}(y|x) = F_{Y_1|X_1}(y|x)$ and $F_{Y_{(1,0,6)}|X}(y|x) = F_{Y_0|X_0}(y|x)$.

Let $F_{Y\langle (0,1,j)|1\rangle}(y)$ represent the counterfactual distribution of returns on a portfolio

[8]Note that the first composition component linked to the constant term is zero.

which has the conditional distribution, $F_{Y_{(0,1,j)}|X}(y|x)$, and the distribution of X_1. $F_{Y\langle(0,1,j)|1\rangle}(y)$ can be obtained by integrating $F_{Y_{(0,1,j)}|X}(y|x)$ over X_1 as follows

$$F_{Y\langle(0,1,j)|1\rangle}(y) \equiv \int F_{Y_{(0,1,j)}|X}(y|x)dF_{X_1}(x). \tag{3.4.4}$$

3.4.2 Decomposing Differences In General Distributional Statistics

Using the structures in North America as reference, i.e., using $F_{Y\langle 0|1\rangle}(y)$ to compute aggregate components, the overall difference in a distributional statistic of portfolio returns between North America and Europe can be decomposed as follows

$$\begin{aligned}
\Delta_O^\nu &= \nu\left(F_{Y\langle 0|0\rangle}(y)\right) - \nu\left(F_{Y\langle 1|1\rangle}(y)\right) \\
&= \underbrace{\left(\nu\left(F_{Y\langle 0|0\rangle}(y)\right) - \nu\left(F_{Y\langle 0|1\rangle}(y)\right)\right)}_{\Delta_C^\nu} + \underbrace{\left(\nu\left(F_{Y\langle 0|1\rangle}(y)\right) - \nu\left(F_{Y\langle 1|1\rangle}(y)\right)\right)}_{\Delta_S^\nu} \\
&= \sum_{j=1}^{6} \underbrace{\left(\nu\left(F_{Y\langle 0|(0,1,j-1)\rangle}(y)\right) - \nu\left(F_{Y\langle 0|(0,1,j)\rangle}(y)\right)\right)}_{\Delta_{C_j}^\nu} \\
&+ \sum_{j=1}^{6} \underbrace{\left(\nu\left(F_{Y\langle(0,1,j-1)|1\rangle}(y)\right) - \nu\left(F_{Y\langle(0,1,j)|1\rangle}(y)\right)\right)}_{\Delta_{S_j}^\nu}
\end{aligned} \tag{3.4.5}$$

where

- $\nu\left(F_{Y\langle\cdot|\cdot\rangle}(y)\right)$ is a distributional statistic of the distribution, $F_{Y\langle\cdot|\cdot\rangle}(y)$.
- Δ_C^ν and Δ_S^ν are aggregate composition and structure components, respectively.
- $\Delta_{C_j}^\nu$ and $\Delta_{S_j}^\nu$ are detailed composition and structure components linked to the jth factor.

Notice that $\Delta_{C_1}^\nu$ is always equal to zero because $F_{Y\langle 0|(0,1,0)\rangle}(y) = F_{Y\langle 0|(0,1,1)\rangle}(y)$ for the first element in the factor vector is a constant term. Also, $\nu\left(F_{Y\langle 0|(0,1,0)\rangle}(y)\right) = \nu\left(F_{Y\langle 0|0\rangle}(y)\right)$ and $\nu\left(F_{Y\langle 0|1\rangle}(y)\right) = \nu\left(F_{Y\langle 0|(0,1,6)\rangle}(y)\right) = \nu\left(F_{Y\langle (0,1,0)|1\rangle}(y)\right)$.

Remark 3.5. To compute aggregate decomposition components, we select the structures in North America as reference, that is, we use $F_{Y\langle 0|1\rangle}(y)$ to compute aggregate components. Alternatively, one can select the structures in Europe as reference, i.e., using $F_{Y\langle 1|0\rangle}(y)$ to compute aggregate components. The decomposition results are possibly different. In this paper, we show the main results using the structures in North America as reference and the results using the structures in Europe as reference are shown as a robustness check.

Remark 3.6. We compute detailed components sequentially based on a series of counterfactual distributions. This approach suffers from the inevitable path dependent problem, that is, the decomposition results depend on the order in which the decomposition is performed. In the main results, we calculate detailed components following the order: the constant term, market, size, value, profitability and investment factors. The order chosen is consistent with the order of the factors being added into FF(2015) five-factor model. In the empirical asset pricing literature, researchers first study market factor in CAPM. Fama and French (1993) introduce the size and value factors and FF(2015) add the profitability and investment factors. We also show the results of switching the decomposition order as robustness checks.

Remark 3.7. We focus on decomposing the differences in quantile portfolio returns, that is, ν is a quantile function. One can readily extend to decompose the differences in general distributional statistics of the distribution of portfolio returns, such as variance, between two regions. In the paper, we also decompose the differences in mean portfolio returns by distribution regressions. Note that the results from distribution regressions are path dependent.

3.5 Decomposition Of Differences In Quantiles And Inference

In this section, we firstly introduce distribution regression, which is useful for estimating the entire distribution of an outcome variable. Secondly, we use distribution regressions to estimate a series of observed (counterfactual) distributions shown in Equation 3.4.5, which are used to compute distributional statistics including quantiles. Thirdly, we present decomposition for the difference in quantile and mean portfolio returns. Lastly, we describe the inference processes.

3.5.1 Distribution Regression

In the literature, there are two main approaches to estimate entire distribution, which are distribution regression and quantile regression. We use distribution regression in this paper.[9] Distribution regression consists of a continuum of binary regressions to the data. Foresi and Peracchi (1995b) propose a fixed number of binary regressions to partially describe the conditional distribution of equity excess returns. Chernozhukov, Fernandez-Val, and Melly (2013) show that a continuum of binary regressions provide a coherent and flexible model for the entire conditional distribution and also establish the central limit theorems.

The main idea of distribution regression is to estimate a continuum of binary response models using $\mathbb{1}\{Y \leq y\}$ as the dependent variable while varying y. To

[9] To estimate entire distribution, quantile regression (Koenker and Bassett Jr (1978) and Koenker (2005)) could be a reasonable alternative approach. The results in this paper would be similar had we used quantile regression. In the first step, obtain estimates of conditional quantiles. Step 2 uses plug-in to compute a series of conditional (counterfactual) quantiles. Step 3 inverts conditional quantiles to obtain the conditional distributions. From there, decomposition for differences in other distributional statistics would be exactly the same. Our approach models the conditional distributions using distribution regression and involves inverting the distributions. For more about quantile regression and its applications in estimating distribution, see Buchinsky (1994), Gosling, Machin, and Meghir (2000), Machado and Mata (2005), Melly (2005) and among others. See also Chapter 20 by Linton and Xiao in Koenker, Chernozhukov, He, and Peng (2017) for quantile regression applications in finance.

implement the distribution regression estimator, we estimate a series of logit models over a fine grid of possible values for y, that is

$$\begin{aligned} F_{Y_g|X_g}(y|x) =& E[\mathbb{1}\{Y_g \leq y\}|X_g = x] \\ =& \Lambda(x'\beta_g(y)) \end{aligned}$$

where Λ is a known link function – we use the logistic link function, $\Lambda(u) = \frac{e^u}{1+e^u}$, though one could make some other choice here. $\beta_g(y)$ are unknown parameters corresponding to each y, i.e., the parameters $\beta_g(y)$ change as y changes. $\mathbb{1}\{Y_g \leq y\}$ is an indicator function that equals one if $Y_g \leq y$ is true and zero otherwise. The estimated conditional distribution is given by

$$\hat{F}_{Y_g|X_g}(y|x) = \Lambda(x'\hat{\beta}_g(y)) \tag{3.5.1}$$

For fixed y, estimating $F_{Y\langle g|g\rangle}(y)$ amounts to average $\hat{F}_{Y_g|X_g}(y|x)$ over X_g. That is,

$$\hat{F}_{Y\langle g|g\rangle}(y) = \frac{1}{n}\sum_{i=1}^{n} \hat{F}_{Y_g|X_g}(y|X_{gi}) \tag{3.5.2}$$

which is the same as replacing the population distribution function in Equation 3.4.1 with the sample distribution function.

Since the estimated distribution obtained may be nonmonotonic in y, we apply the monotonization method of Chernozhukov, Fernández-Val, and Galichon (2010) based on rearrangement.[10] The rearranged distribution is given by

$$\hat{F}^r_{Y\langle g|g\rangle}(y) = \inf\left\{u : \int_0^1 \mathbb{1}\{\hat{F}_{Y\langle g|g\rangle}(y) \leq u\}dv \geq y\right\}.$$

[10]For practical and computational purposes, it is helpful to think of the rearrangement as sorting (Chernozhukov, Fernández-Val, and Galichon (2010), p. 1098). In practice, we use rearranged estimators of the distributions for all the results, but we omit this discussion throughout the rest of this paper for the sake of clarity.

3.5.2 Estimating Counterfactual Distributions

With the estimates $\hat{\beta}(y)$ from Equation 3.5.1, estimating $F_{Y\langle 0|(0,1,j)\rangle}(y)$ amounts to average $F_{Y_0|X_0}(y|X)$ over $X_{(0,1,j)}$. That is,

$$\hat{F}_{Y\langle 0|(0,1,j)\rangle}(y) = \frac{1}{n}\sum_{i=1}^{n} \hat{F}_{Y_0|X_0}(y|X_{(0,1,j)i}) \tag{3.5.3}$$

which is the same as replacing the population distribution function in Equation 3.4.3 with the sample distribution function.

Similarly, with the estimates $\hat{\beta}(y)$ from Equation 3.5.1 and thus the constructed $F_{Y_{(0,1,j)}|X_0}(y|x)$, estimating $F_{Y\langle(0,1,j)|1\rangle}(y)$ amounts to average $F_{Y_{(0,1,j)}|X}(y|x)$ over the distribution of X_1. That is,

$$\hat{F}_{Y\langle 0|(0,1,j)\rangle}(y) = \frac{1}{n}\sum_{i=1}^{n} \hat{F}_{Y_{(0,1,j)}|X}(y|X_{1i}) \tag{3.5.4}$$

which is the same as replacing the population distribution function in Equation 3.4.4 with the sample distribution function.

3.5.3 Decomposing Quantile Differences

With the observed and counterfactual distributions from Equations 3.5.2, 3.5.3 and 3.5.4, one can compute any distributional statistics. Then, the decomposition components can be computed by plug-in method, that is, plugging the estimates of distributional statistics into Equation 3.4.5. In the paper, we decompose the difference in τ-quantile of portfolio returns between North America and Europe. We begin by inverting each distribution, $\hat{F}_{Y\langle\cdot|\cdot\rangle}$, to obtain the corresponding estimate of τ-quantile, $Q_{\langle\cdot|\cdot\rangle}(\tau)$, by the following equation

$$\hat{Q}_{\langle\cdot|\cdot\rangle}(\tau) = \inf\{Q : \hat{F}^{r}_{Y\langle\cdot|\cdot\rangle}(Q) \geq \tau\}, \quad \tau \in (0,1). \tag{3.5.5}$$

Then, the components of decomposition for differences in τ-quantile of portfolio returns can be computed by plugging $\hat{Q}_{\langle\cdot|\cdot\rangle}(\tau)$ into Equation 3.4.5, where ν is a quantile function: $\nu\left(\hat{F}_{Y\langle\cdot|\cdot\rangle}(y)\right) = \hat{Q}_{\langle\cdot|\cdot\rangle}(\tau)$.

3.5.4 Decomposing Mean Difference

To decompose the difference in mean portfolio returns between North America and Europe using distribution regressions, we need to estimate $E(Y\langle\cdot|\cdot\rangle)$.

Consider a grid of equally spaced values of τ given by $0 < \tau_1 < \tau_2 < \cdots < \tau_M < 1$. Then, we estimate $E(Y\langle\cdot|\cdot\rangle)$ by

$$\hat{E}(Y\langle\cdot|\cdot\rangle) = \frac{1}{M}\sum_{m=1}^{M}\hat{Q}_{\langle\cdot|\cdot\rangle}(\tau_m)$$

The decomposition follows straightforward as Equation 3.4.5 by replacing $\nu\left(\hat{F}_{Y\langle\cdot|\cdot\rangle}(y)\right)$ with $\hat{E}(Y\langle\cdot|\cdot\rangle)$. We compare the results to the ones from Blinder-Oaxaca decomposition.

3.5.5 Inference

The significance test is previously ignored in most decomposition analyses in economics. Chernozhukov, Fernandez-Val, and Melly (2013) provide estimation and inference procedures for the entire marginal counterfactual distribution and its functionals based on regression methods including distribution regression and quantile regression. They also drive a functional central limit theorem and prove the validity of bootstrap for the entire empirical coefficient process of distribution regression and related functionals.

In this paper, we calculate a series of counterfactual distributions and quantiles which are functionals of distributions. In the spirit of Chernozhukov, Fernandez-Val, and Melly (2013), we complete the inference processes by bootstrapping.[11]

3.6 Data And Results

We use the Fama and French (2017) dataset which can be downloaded at Kenneth R. French's personal website. We decompose the differences in monthly excess portfolio returns between two regions, North America (United States and Canada), and Europe (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom). The dataset ranges from July 1990 to November 2017 (329 months). Hereafter, we simply call portfolio returns instead of monthly excess portfolio returns. The distinction will be made whenever emphasis is necessary.

The five factors used in the paper are portfolios constructed from 2×3 sorts on market capitalization (*Size*) and book-to-market equity ratio (B/M), operating profitability (OP) or investment (*Inv*). [12] At the end of each June, stocks are allocated to small and big *Size* groups. Stocks are allocated independently to three B/M groups using B/M breakpoints.[13] The intersections of the two sorts produce 6 value-weight *Size*-B/M portfolios. The value factor, HML, is the average of the value-weight returns on the two high value stock portfolios of the 2×3 sorts minus the average of the value-weight returns on the two low value stock portfolios. The profitability and investment factors, RMW and CMA, are constructed in the same way as HML except the second sort is on either OP or *Inv* and it is robust minus weak or conservative

[11] For inference in decomposing differences in mean portfolio returns using Blinder-Oaxaca decomposition, we also use bootstraps to obtain the standard deviations of decomposition components.

[12] We choose factors from 2×3 sorts because the five-factor model's performance is not sensitive to the way its factors are defined. We do not include the other common factors, e.g., the momentum factor of Carhart (1997) and liquidity factor of Pástor and Stambaugh (2003) because these factors have regression slopes close to zero so produce trivial changes in model performance when the portfolios used here are sorted on *Size* and B/M, OP or *Inv*. See FF(2015).

[13] In 2×3 sorts, big stocks are those in the top 90% of market capitalization for the region, and small stocks are in the bottom 10%. The B/M, OP and *Inv* breakpoints are the 30th and 70th percentiles of B/M, OP, and *Inv* for the big stocks of the region, respectively.

minus aggressive. The size factor, SMB, is the average of the value-weight returns on the nine small stock portfolios of the three 2×3 sorts minus the average of the value-weight returns on the nine big stock portfolios.[14]

The 75 portfolios used to decompose are 25 *Size-B/M*, 25 *Size-OP*, and 25 *Size-Inv* portfolios constructed from 5×5 sorts for each region. At the end of each June, stocks are allocated to five *Size* groups using *Size* breakpoints. Stocks are allocated independently to five *B/M* groups using *B/M* breaks. The intersections of the two sorts produce 25 value-weight *Size-B/M* portfolios. The 25 *Size-Inv* and *Size-OP* portfolios are constructed in the same way as the *Size-B/M* portfolios except the second sort is on either *OP* or *Inv*.[15]

3.6.1 Summary Statistics For Factor Returns

Table 4.1a reports the summary statistics for five factors. For the market, size and value factors, there exist big differences in average factor returns between North America and Europe, which are 0.16%, 0.1% and -0.14% per month, respectively. These differences are large in economic terms though not statistically significantly different from zero. The average returns on RMW and CMA for North America and Europe are very close.

Table 4.1b shows the correlations in factors within and between two regions. In both North America and Europe, the correlations between Mkt and RMW or CMA are negative. Mkt is negatively correlated with HML in North America (-0.23) but positively in Europe (0.18). However, Mkt is positively correlated with SMB in North

[14] B is book equity at the end of the fiscal year ending in year $t-1$ and M is market capitalization at the end of December of year $t-1$, adjusted for changes in shares outstanding between the measurement of B and the end of December. *OP* is measured with accounting data for the fiscal year ending in year $t-1$ and is revenues minus cost of goods sold, minus selling, general, and administrative expenses, minus interest expense all divided by *B*. *Inv* is the change in total assets from the fiscal year ending in year $t-2$ to the fiscal year ending in $t-1$, divided by $t-2$ total assets.

[15] The *Size* breakpoints for 5×5 sorts are the 3rd, 7th, 13th, and 25th percentiles of the region's aggregate market capitalization. The *B/M, OP* and *Inv* breaks are the quintile of *B/M, OP*, and *Inv* for the big stocks of the region, respectively.

America but negatively with SMB in Europe. In North America, SMB is negatively correlated with HML, RMW and CMA. However, the correlation between SMB and HML, RMW or CMA is very small in Europe. HML is highly and positively correlated with RMW and CMA in North America, but highly and negatively correlated with RMW although still highly and positively correlated with CMA in Europe. The correlation between RMW and CMA is 0.35 in North America versus -0.18 in Europe.

The rightmost block of Table 4.1b reports the correlations in factors between North America and Europe. Mkt, HML, and CMA are highly and positively correlated with correlations 0.8, 0.6 and 0.57, respectively. The correlation between Mkt and SMB, RMW or CMA is negative. SMB is negatively correlated with CMA (-0.12). HML is positively correlated with CMA (0.52) and negatively with RMW (-0.15). RMW is positively correlated with RMW (0.38).

Table 3.1: Summary statistics for factors

(a) Average and Standard Deviation

	North America					Europe					Difference				
	Mkt	SMB	HML	RMW	CMA	Mkt	SMB	HML	RMW	CMA	Mkt	SMB	HML	RMW	CMA
Average	0.67	0.17	0.20	0.34	0.26	0.51	0.07	0.34	0.40	0.21	0.16	0.10	-0.14	-0.06	0.05
SD	4.20	2.76	3.23	2.41	2.65	4.90	2.17	2.38	1.51	1.82	2.96	2.94	2.61	2.55	2.20

(b) Correlations

	North America					Europe					Between				
	Mkt	SMB	HML	RMW	CMA	Mkt	SMB	HML	RMW	CMA	Mkt	SMB	HML	RMW	CMA
Mkt	1.00	0.20	-0.23	-0.37	-0.44	1.00	-0.17	0.18	-0.26	-0.30	0.80	-0.26	0.04	-0.22	-0.35
SMB	0.20	1.00	-0.10	-0.42	-0.14	-0.17	1.00	0.01	-0.05	0.02	-0.26	0.31	0.03	-0.03	-0.12
HML	-0.23	-0.10	1.00	0.38	0.78	0.18	0.01	1.00	-0.54	0.54	0.04	0.03	0.60	-0.15	0.52
RMW	-0.37	-0.42	0.38	1.00	0.35	-0.26	-0.05	-0.54	1.00	-0.18	-0.22	-0.03	-0.15	0.22	0.38
CMA	-0.44	-0.14	0.78	0.35	1.00	-0.30	0.02	0.54	-0.18	1.00	-0.35	-0.12	0.52	0.38	0.57

Note: The table shows summary statistics and correlations between factors. Panel a shows the summary statistics and Panel b shows the correlations in factors within and between North America and Europe.

3.6.2 Summary Statistics For Portfolio Returns

Table 4.2 shows means and standard deviations of returns on 75 portfolios studied in the paper. None of the portfolio returns is statistically different from zero although most of them are economically large. Table 3.3 shows summary statistics for the differences in portfolio returns between North America and Europe. We see most of differences are economically significant although none is statistically different from zero.

One might doubt the normality of the portfolio returns then using t-statistic to test significances is not a good idea. We use Kolmogorov–Smirnov tests to check the normality of portfolio returns.[16] For every portfolio, the null hypothesis that the distribution of differences is normal is rejected at the confidence level of 95%. Thus, we use Wilcoxon tests to test significances of the differences and find that the differences are not significantly different from zero.[17] Suppose the differences are not economically significant, it is still interesting to decompose the differences and examine the decomposition components. The non-normality of the distribution of the differences makes it necessary to decompose the differences in quantile returns.

[16] See Massey Jr (1951), Lilliefors (1967), Lilliefors (1969), Friedman and Rafsky (1979), and Stephens (1974) for Kolmogorov–Smirnov test.

[17] The results of Kolmogorov–Smirnov tests and Wilcoxon tests are not shown and are offered upon request.

Table 3.2: Summary statistics for portfolio returns

(a) Mean

	North America					Europe				
	Small	2	3	4	Big	Small	2	3	4	Big
Low B/M	0.43	0.40	0.79	0.82	0.64	-0.08	0.29	0.34	0.49	0.34
2	0.60	0.62	0.68	0.70	0.66	0.37	0.49	0.56	0.53	0.53
3	0.91	0.79	0.78	0.81	0.63	0.45	0.55	0.54	0.59	0.57
4	0.88	0.80	0.78	0.77	0.65	0.59	0.71	0.54	0.55	0.64
High B/M	1.16	0.86	0.89	0.86	0.57	0.75	0.76	0.74	0.65	0.54
Low Inv	1.21	0.88	0.91	0.90	0.74	0.55	0.59	0.64	0.64	0.57
2	1.12	0.90	0.90	0.94	0.65	0.71	0.77	0.67	0.63	0.58
3	0.99	0.88	0.90	0.85	0.65	0.71	0.76	0.70	0.64	0.48
4	0.96	0.87	0.79	0.88	0.62	0.65	0.64	0.48	0.63	0.45
High Inv	0.54	0.36	0.54	0.51	0.55	0.16	0.37	0.30	0.39	0.45
Low OP	0.84	0.44	0.59	0.57	0.25	0.15	0.25	0.26	0.20	0.16
2	1.06	0.78	0.78	0.79	0.57	0.67	0.57	0.56	0.55	0.55
3	1.05	0.95	0.81	0.93	0.63	0.75	0.70	0.73	0.71	0.56
4	1.07	1.01	0.90	0.78	0.71	0.90	0.75	0.63	0.69	0.47
High OP	1.09	1.14	1.06	0.93	0.72	0.76	0.96	0.81	0.70	0.60

(b) Standard deviation

	North America					Europe				
	Small	2	3	4	Big	Small	2	3	4	Big
Low B/M	7.96	7.31	6.88	6.44	4.53	5.49	5.63	5.71	5.39	4.84
2	6.82	6.49	5.70	5.06	4.10	5.27	5.25	5.23	4.99	4.77
3	6.13	5.47	5.02	4.57	4.21	4.95	5.02	5.14	5.01	5.21
4	5.39	4.92	4.72	4.58	4.12	4.84	5.06	5.18	5.35	5.42
High B/M	5.28	5.16	4.88	4.76	5.23	4.83	5.30	5.59	5.80	6.34
Low Inv	6.42	5.68	5.09	4.76	4.05	4.98	5.22	5.46	5.20	4.85
2	4.97	4.76	4.36	4.21	3.69	4.46	4.82	5.04	5.00	4.81
3	4.92	4.78	4.65	4.38	4.05	4.56	4.80	4.89	4.85	5.17
4	5.29	5.42	5.23	4.87	4.77	4.76	5.14	5.17	5.23	5.45
High Inv	6.80	6.92	7.26	6.64	5.82	5.69	5.74	5.94	6.07	5.35
Low OP	6.68	6.65	6.77	6.26	5.60	5.18	5.43	5.56	5.52	6.09
2	4.91	4.94	4.87	4.69	4.89	4.67	5.00	5.09	5.14	5.53
3	4.98	4.99	4.60	4.40	4.31	4.81	5.02	5.07	5.25	5.08
4	5.34	5.27	4.93	4.45	4.07	4.79	5.09	5.17	5.08	4.99
High OP	5.66	5.44	5.40	4.83	3.96	4.94	5.32	5.33	5.28	4.76

Note: The table shows summary statistics for 75 portfolios in North America and Europe, respectively. The top panel shows means of monthly portfolio returns in excess of the one-month Treasury bill rate. The bottom panel shows standard deviations.

We calculate the average portfolio returns by averaging out the portfolio returns for all portfolios in North America and Europe, respectively. Figure 3.1 plots time series of the average portfolio returns in North America against Europe. It clearly shows that the portfolio returns between North America and Europe are positively correlated. The correlation is 0.747. That simply means that the portfolio returns in Europe are high when the portfolio returns in North America are high. Positive perfect correlation, that is all portfolio returns lining on the 45-degree line, means that the rank of the portfolio return for month t in the distribution of portfolio returns in North America is the same as the rank of the portfolio return for month t in the distribution of portfolio returns in Europe.

This is related to rank invariance in the quantile treatment effect literature. Rank invariance requires that the rank of a counterfactual portfolio return in the counterfactual distribution remains the same as the rank of the portfolio return in the observed distribution.[18] Rank invariance is important because it is required for the decomposition for differences in quantile portfolio returns. It is also important for analyses of the decomposition results in the quantile differences from the perspective of asset pricing at the end of the paper. Thus, we implicitly assume that the rank invariance assumption holds without statistically testing it.

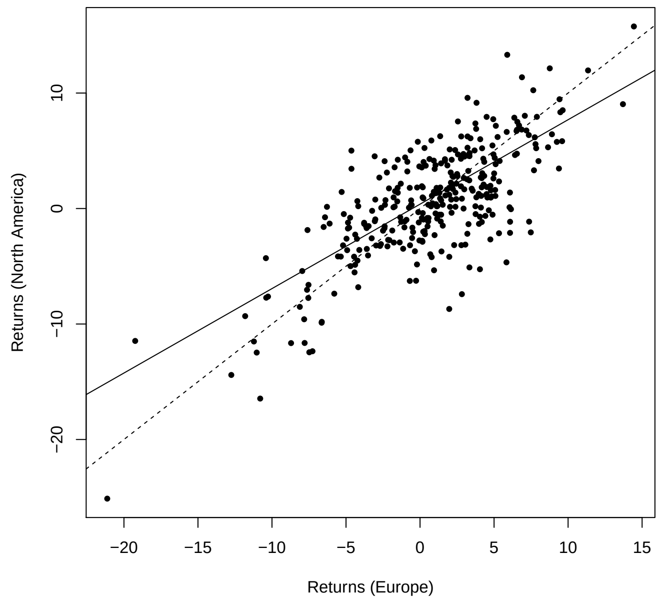
Table 3.3: Summary statistics for differences in portfolio returns

	Mean					Standard deviation				
	Small	2	3	4	Big	Small	2	3	4	Big
Low B/M	0.52	0.11	0.45	0.33	0.30	5.83	5.60	4.91	4.41	3.27
2	0.23	0.14	0.12	0.17	0.13	5.28	5.22	4.44	3.67	3.30
3	0.45	0.24	0.24	0.22	0.06	4.68	4.41	4.06	3.45	3.18
4	0.29	0.09	0.23	0.21	0.00	4.48	4.13	3.85	3.70	3.62
High B/M	0.41	0.10	0.15	0.21	0.04	4.35	4.29	4.22	4.10	4.37
Low Inv	0.66	0.29	0.26	0.26	0.17	5.11	4.35	4.16	3.54	3.25
2	0.41	0.13	0.22	0.31	0.06	4.26	4.13	3.83	3.48	3.14
3	0.28	0.12	0.20	0.21	0.17	4.26	4.28	3.78	3.58	3.26
4	0.31	0.23	0.32	0.24	0.17	4.35	4.38	4.08	3.75	3.76
High Inv	0.38	0.00	0.24	0.12	0.10	5.01	5.12	5.08	4.40	4.55
Low OP	0.69	0.19	0.33	0.36	0.09	5.10	5.02	5.14	4.70	4.36
2	0.39	0.21	0.23	0.23	0.02	4.40	4.31	3.99	3.78	3.74
3	0.30	0.25	0.08	0.22	0.06	4.36	4.20	3.79	3.47	3.17
4	0.16	0.26	0.27	0.09	0.25	4.43	4.47	4.16	3.56	3.36
High OP	0.32	0.18	0.25	0.23	0.11	4.49	4.61	4.27	3.72	3.30

Note: The table shows summary statistics for means of differences in monthly excess portfolio returns between North America and Europe.

[18] In the quantile treatment effect literature, rank invariance is as known as rank preservation. See, e.g., Heckman, Smith, and Clements (1997), Chernozhukov and Hansen (2005), Chernozhukov and Hansen (2006), and Chernozhukov and Hansen (2008), Horowitz and Lee (2007), Chernozhukov, Imbens, and Newey (2007), Abadie, Angrist, and Imbens (2002), Frölich and Melly (2013), Firpo (2007), Firpo and Pinto (2016), Imbens and Newey (2009), and Dong and Shen (2018).

Figure 3.1: Average portfolio returns in North America against Europe



Note: The figure plots time series of the average portfolio returns in North America against Europe. The true line is the regression fitted line and the dashed line is the 45 degree line.

3.6.3 Decomposition Results

Before showing the main results, we first note that we choose the structures in North America as reference and the order of decomposition is the constant, market, size, value, profitability and investment factors. In Section 3.7, we show robustness checks by changing the reference or the decomposition order. We also note that the decomposition results are averaged out over the decomposition results for all portfolios. That is, we do decomposition for each portfolio and average out the results. The decomposition results for each portfolio are offered upon request.

We first show the true and observed overall differences in portfolio returns between North America and Europe. We note that we decompose the observed overall differences in the paper. We then compare the decomposition results in the observed overall mean difference using Blinder-Oaxaca decomposition and distribution regressions.

Lastly, we present the decomposition results in the observed quantile differences across a sequence of quantiles.

3.6.3.1 True And Observed Overall Differences

Figure 3.2 shows the observed and true overall quantile and mean differences in portfolio returns between North America and Europe.[19] The observed overall mean (OLS) difference and the true overall mean difference coincide with each other, both of which are 0.23%. The standard deviation of the observed mean (OLS) difference is 0.045%. The observed overall mean (DR) difference is much lower, which is 0.11% (see Table 3.4). The observed and true overall quantile differences in portfolio returns are quite close in the lower quantiles. In the higher quantiles, however, the observed ones is lower than the true ones.[20]

The overall quantile differences jump down in the quantiles below 10%. The differences then start to climb up and go down across quantiles. We observe that the differences in the lower quantiles are mostly negative and the differences in the middle quantiles are positive. We recall that the overall mean difference in portfolio returns between North America and Europe is positively 0.23% per month. This positiveness in mean difference is mostly contributed by the positive quantile differences in the middle quantiles. The negative quantile differences in the low quantiles mean that the portfolios in North America have higher down-side risk. This higher down-side risk may partially explain the higher mean returns on portfolios in North America. Higher

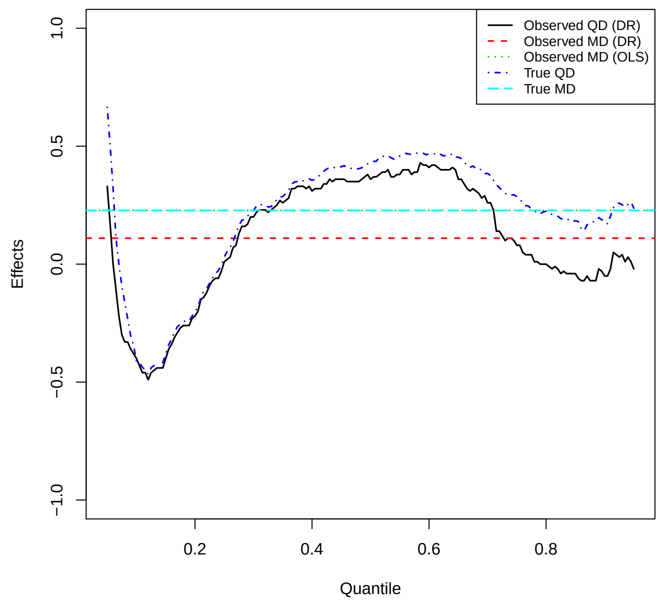
[19]The observed overall quantile and mean (DR) differences are obtained by following steps: Firstly, estimate the quantile and mean returns by distribution regressions for each portfolio. Secondly, compute the differences between two regions for each portfolio. Thirdly, average out the differences for all portfolios. The observed overall mean difference (OLS) is obtained by the same way as the observed mean difference (DR) except it is computed by ordinary least squares. The true overall quantile and mean differences are computed by following steps: Firstly, sort the portfolio returns for each portfolio and directly compute the quantiles and mean. Secondly, compute the differences in mean and quantile returns between two regions for each portfolio. Lastly, average out the differences in mean and quantile returns for all portfolios.

[20]For the sake of clarity, we do not show the confidence interval for the observed overall quantile differences. The confidence interval contains zero across all quantiles. The confidence intervals for aggregate decomposition components shown below are also not plotted for the same reason.

risks are compensated by higher returns.

In the paper, we decompose the observed overall differences.

Figure 3.2: Observed and True overall quantile and mean differences



Note: The figure plots the overall differences in quantiles and mean of portfolio returns between North America and Europe. QD and MD denote quantile and mean differences, repectively. DR and OLS represent methods used to compueted Observed QD and MD, which are distribution regressions and ordinary least squares, respectively.

3.6.3.2 Decomposing Mean Difference

Table 3.4 shows the decomposition results for the mean difference using Blinder-Oaxaca decomposition and distribution regressions. We observe that using Blinder-Oaxaca decomposition 89.5% of the overall difference, 0.228%, comes from the aggregate composition component, 0.204%. Using distribution regressions, the composition component, 0.16%, over explains the overall difference, 0.11%. The aggregate composition component from Blinder-Oaxaca decomposition is both statistically and economically significant. The aggregate structure component is weakly statistically different from zero but economically trivial. The composition component from dis-

Table 3.4: Decomposition results for mean difference (in percentage)
North America: MSHRC

	Mean (BO)	SD	Mean (DR)	SD
Overall	0.228	0.045	0.11	0.37
Agg_c	0.204	0.043	0.16	0.36
Agg_s	0.024	0.012	-0.05	0.06
Mkt_c	0.163	0.039	0.11	0.34
SMB_c	0.052	0.015	0.06	0.11
HML_c	-0.010	0.007	-0.03	0.03
RMW_c	0.001	0.006	0.01	0.01
CMA_c	-0.002	0.006	0.00	0.01
Constant_s	0.010	0.012	0.10	0.15
Mkt_s	0.005	0.002	-0.11	0.11
SMB_s	0.000	0.002	-0.02	0.04
HML_s	0.009	0.004	-0.01	0.02
RMW_s	0.007	0.004	0.00	0.03
CMA_s	-0.006	0.003	0.00	0.06

Note: The table shows the deocomposition results for the mean difference in portfolio returns between North America and Europe. "BO" and "DR" denote the approches used, which are Blinder-Oaxaca decomposition and distribution regressions, respectively. "Overall" represents the observed overall mean differences. "Agg" denotes aggregate component. "c" and "s" represent composition and structure components, respectively. For instance, "Agg_c" denotes aggregate composition component and "Mkt_c" represents the detailed composition component associated to market factor. "Constant_s" denotes the structure component linked to constant term. "North America: MSHRC" represents that the structure in North America are used as reference and the order of decomposition is the constant, market, size, value, profitability, investment factors.

tribution regressions are economically significant but statistically insignificant. The structure component, however, is not only statistically insignificant but also economically trivial and negative.

For the detailed composition components from Blinder-Oaxaca decomposition, the composition component linked to market factor is 0.163% per month, which explains 71.9% of the overall difference and 79.9% of the aggregate composition component. It is economically and statically significant. The second largest is the composition component linked to size factor which is 0.052% per month and explaining 22.8% of the overall difference and 25.5% of the aggregate composition component. Although it is statistically significant, it is economically smaller relative to market factor. The other components are all economically insignificant although the structure components linked to market, value and investment factors are statistically significant.

For the detailed decomposition results from distribution regressions, we observe that the composition component linked to market factor is 0.11% per month, explaining the entire overall difference and 68.8% of the aggregate composition component. The composition component linked to size factor is 0.06%. The structure component linked to market factor is -0.11%. The other components are economically trivial. Notice that none of the components is statistically significant.

In sum, the results in the mean difference from two approaches seem to be different in terms of the sizes of the decomposition components.[21] However, the results from both approaches show that the mean difference in portfolio returns is mainly contributed by the differences in factor risk premia. More specifically, the differences in the market and size factor risk premia play a significant role in explaining the mean difference in portfolio returns and the differences in the factor risks play an insignificant role in aggregate.

3.6.3.3 Decomposing Quantile Differences

Aggregate components

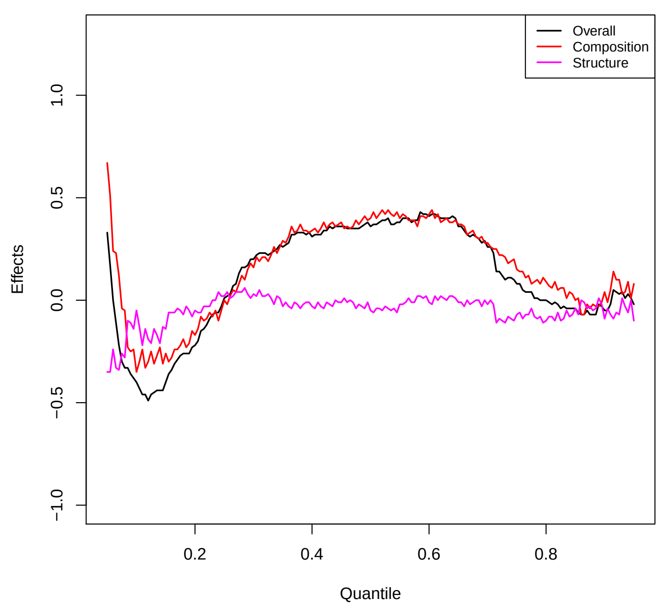
Figure 3.3 shows the aggregate decomposition results for a series of quantile differences using distribution regressions. It is obvious to observe that the aggregate composition components play a significant role in explaining the observed overall quantile differences in portfolio returns. The composition components vary across quantiles and follow tightly the pattern of the overall quantile differences, first jumping

[21]The sizes are different mainly because we are decomposing the observed overall mean difference instead of the true overall mean difference. Also, the results from distribution regressions are path dependent and the ones from Blinder-Oaxaca decomposition are path independent.

down in the low quantiles and climbing up then going down across quantiles. However, the structure components seem to move around zero and are trivial relative to the composition components.

In sum, we find that the differences in quantiles of portfolio returns between North America and Europe are mostly contributed by the differences in the distributions of factors, i.e., factor risk premia. The differences in the structures, i.e., factor risks, seem to play a trivial role in aggregate. Next we will look at the detailed decomposition to further examine the role of each factor.

Figure 3.3: Aggregate decomposition across quantiles
North America: MSHRC



Note: The figure shows the aggregate decomposition decomponents across a series of quantiles. “Overall”, “Composition” and “Structure” denotes the observed overall quantile differences, aggregate composition and structre components, respectively. “North America: MSHRC” represents that the structures in North America are used as reference and the order of decomposition is the constant, market, size, value, profitability and investment factors.

Detailed composition components

The top panel of Figure 3.4 shows the detailed composition components for quantile differences across quantiles and Figure 3.5 shows each component with its 95% confidence interval separately. We observe that the composition components across quantiles are mostly contributed by the components associated to market and size factors. The composition components linked to value, profitability and investment factors are economically trivial and statistically insignificant.

In the extreme low quantiles, the positive quantile differences are mainly from the composition component linked to market factor. It is interesting to further examine the "anomaly", i.e., the big jump, in the extreme low quantiles. Except in the extreme low quantiles, the composition component linked to market factor seems to go down smoothly across quantiles and the composition component associated to size factor goes up monotonically across quantiles. However, the speeds are different, which leads to the aggregate composition component first going up in the low quantiles and going down in the high quantiles.

The positive composition component linked to market factor implies higher returns on market factor in North America than the ones in Europe in the lower quantiles and the negative component in the higher quantiles means lower returns on market factor. We recall the mean difference in market factor between North America and Europe is 0.16% per month (see Table 4.1a). This positive mean difference in market factor is mainly contributed by the positive differences in the lower quantiles.

The composition component linked to size factor is negative in the low quantiles but positive in the high quantiles. The component seems to be statistically significant. This implies that the size factor returns in North America are lower than in Europe in the low quantiles but higher in the high quantiles. More specifically, when the portfolio has low returns (in the low quantile returns) the difference in size factor returns between North America and Europe is negative and when the portfolio has high returns (in the high quantile returns) the difference in size factor is positive.

In sum, the aggregate composition component, i.e., the contribution of the difference in factor risk premia, is mainly contributed by the differences in market and size factor premia. When the portfolio returns are low, the market factor risk premium positively contributes to the quantile differences and the size factor risk premium negatively contributes to the difference. When the portfolio returns are high, the market factor risk premium negatively contributes to the quantile differences and the size factor risk premium positively contributes to the difference. The variations in the differences in factor risk premia across quantiles imply variations in factor risk premia themselves. That is, factor risk premia vary at different levels of portfolio returns, adding to the literature on variations in risk premia.

Detailed structure components

Figure 3.4 (Panel b) and 3.6 show the detailed structure component across quantiles. We recall that the aggregate structure component is trivial relative to the aggregate composition component. The components associated to size, value and profitability factors are all insignificant. The component associated to market factor is flat across quantiles except in the extreme quantiles. The large and positive market component in the extreme low quantiles implies that the market factor risk in North America is much higher than in Europe when the portfolio returns are low. And the large and negative market component in the extreme high quantiles implies that the market factor risk in North America is much lower than in Europe when the portfolio returns are high. It is interesting to further explore about the large difference in market factor risk between two regions in the extreme quantile portfolio returns.

The component linked to investment factor seems to play a statistically and economically significant role, which starts negative and monotonically goes up across quantiles. This means that the investment factor risk in North America is lower than in Europe when the portfolio returns are low and the investment factor risk in North America is higher than in Europe when the portfolio returns are high. This variation in the difference in investment factor risk across quantiles implies the variation in investment factor risk itself in North America and Europe. That is, the investment factor risk varies at different levels of portfolio returns, adding to the literature on variations in risks.

In sum, the market and investment factor risks seem to play a significant role in explaining the overall quantile difference although in aggregate the difference in factor risks do not seem to play an important role. Also, the roles of factor risks vary at different levels of portfolio returns, implying variations in the factor risks themselves.

3.7 Robustness

In this section, we show some results for robustness checks. We need robustness checks for the following reasons. Firstly, the aggregate decomposition results from two approaches depend on the structures used as reference. Secondly, although the detailed decomposition using Blinder-Oaxaca decomposition is path independent, the detailed decomposition results from distribution regressions are path dependent. That is, the order of computing detailed components could affect the decomposition results.

We refer to the main results shown above as the reference results. We summarize the main differences between the robustness checks and the reference results. We offer the results in detail upon request.

3.7.1 North America: Size, Market, Value, Profitability And Investment

This section shows the results, in which the structures in North America are used as reference and the order of decomposition is the constant, size, market, value, profitability and investment factors. In this case, we do not change the reference structures. Thus, the results for mean difference using Blinder-Oaxaca decomposition will be the same as the ones in the reference results because the results from Blinder-Oaxaca decomposition are path independent. The results for the mean difference using distribution regressions in this case are quite similar to the reference results.

For the quantile differences, by construction if the results are path dependent, the detailed components linked to market and size factors will be significantly different from the ones in the reference results. However, in this case we do not see significant differences in the detailed components associated to market and size factors. That is, the reference results are robust to this change in the decomposition order.

3.7.2 North America: Profitability, Investment, Market, Size And Value

The results, in which the structures in North America are used as reference and the order of decomposition is the constant, profitability, investment, market, size and value factors, are also quite similar to the reference results. The exception is that the structure component associated to value factor seems to be significant and the structure component linked to investment factor is not significant any more (Figure 3.7).

3.7.3 Europe: Market, Size, Value, Profitability And Investment

This section shows the results in which the structures in Europe are used as reference and the order of decomposition is the constant, market, size, value, profitability and investment factors. Table 3.5 shows the decomposition results for the mean difference using both Blinder-Oaxaca decomposition and distribution regressions. Comparing to the reference results shown in Table 3.4, we observe that the results from Blinder-Oaxaca decomposition are quite close to the reference results. Comparing the results from distribution regressions to the reference results, we can come to the same mian conclusions as the ones drawn from the reference results although the sizes of the components are different.

Table 3.5: Decomposition results for mean difference (in percentage)
Europe: MSHRC

	Mean (BO)	SD	Mean (DR)	SD
Overall	0.228	0.045	0.11	0.43
Agg_c	0.209	0.043	0.26	0.42
Agg_s	0.019	0.013	-0.15	0.06
Mkt_c	0.161	0.039	0.14	0.39
SMB_c	0.053	0.014	0.11	0.12
HML_c	-0.006	0.007	0.00	0.02
RMW_c	0.002	0.004	0.03	0.02
CMA_c	0.000	0.005	-0.01	0.02
Constant_s	0.010	0.013	-0.04	0.17
Mkt_s	0.007	0.003	-0.07	0.44
SMB_s	-0.001	0.002	0.01	0.09
HML_s	0.005	0.004	-0.09	0.05
RMW_s	0.006	0.004	-0.03	0.03
CMA_s	-0.008	0.004	-0.02	0.03

Note: The table shows the deocomposition results for the mean difference in portfolio returns between North America and Europe. “BO” and “DR” denote the approches used, which are Blinder-Oaxaca decomposition and distribution regressions, respectively. “Overall” represents the observed overall mean differences. ”Agg” denotes aggregate component. “c” and “s” represent composition and structure components, respectively. For instance, “Agg_c” denotes aggregate composition component and “Mkt_c” represents the detailed composition component associated to market factor. “Constant_s” denotes the structure component linked to constant term. “Europe: MSHRC” represents that the structures in Europe are used as reference and the order of decomposition is the constant, market, size and value, profitability and investment factors.

For the quantile differences, we observe the detailed composition components are quite similar, both of which show that the market and size factors play a significant role (see Panel a of Figure 3.9 and 3.10). Comparing the aggregate components for the quantile differences shown in Figure 3.3 and 3.8, however, we observe that the aggregate composition component in this case does not follow the pattern of the overall quantile differences as tightly as in reference results, especially in the high quantiles.

Also, the detailed structure components in this case are quite different from the ones in the reference results (Figure 3.9 and 3.11). In this case, the structure component associated to the constant is negative in the low quantiles but positive in the reference results. The components linked to market and size factors are also different from the reference results. The component linked to investment factor is not significant any more.

In sum, the main conclusions drawn from the reference results are mostly robust to changing the reference structures and the composition order in terms of that the differences in portfolio returns between North America and Europe are mostly explained by the differences in factor risk premia and the differences in factor risks play a relatively trivial role. Further, the differences are mostly explained by the differences in market and size factor risk premia between two regions. Also, the factor risk premia vary at different levels of portfolio returns.

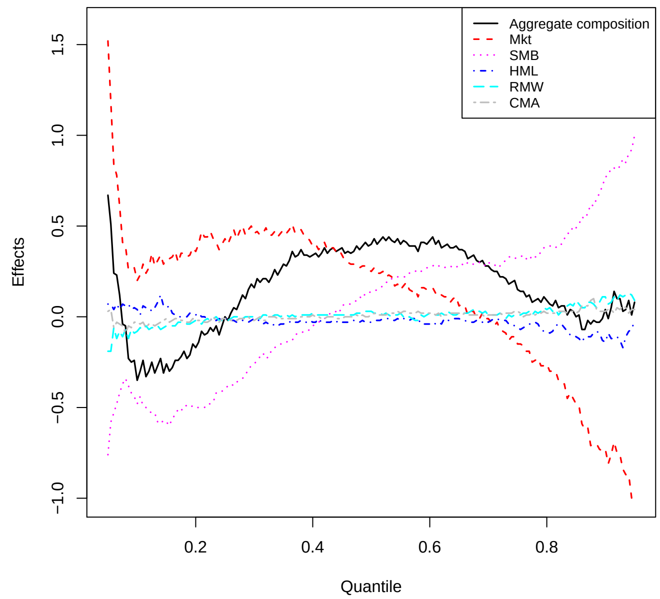
We note that the significance of the detailed structure components associated to some factors are not that robust to the changes in the reference structures and the decomposition order. However, we are inclined to conclude that the factor risks are likely to vary at different levels of portfolio returns.

3.8 Conclusion

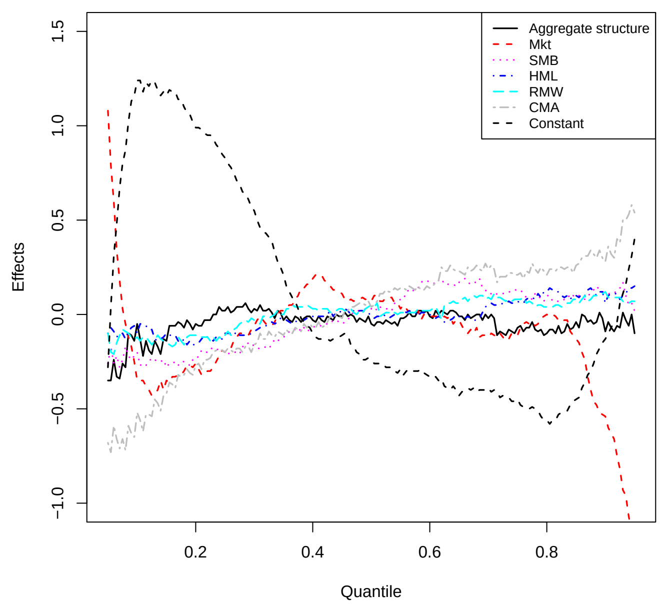
In the paper, we decompose the large differences in mean portfolio returns between North America and Europe into Fama and French's five factors. We show that the mean differences are mainly contributed by the differences in risk premia on factors, especially on market and size factors. The differences in factor risks do not seem to play an important role in explaining the mean differences. We also decompose a series of differences in quantiles of portfolio returns. Again, we find that the quantile differences are also mainly contributed by the differences in risk premia on factors, especially on market and size factors and the differences in factor risks seem to play an insignificant role in aggregate. The detailed decomposition for the aggregate composition component further shows that the roles that the market and size factors play in explaining the quantile differences vary at different levels of portfolio returns. These findings imply the market and size factor risk premia vary at different levels of portfolio returns. The detailed decomposition for the aggregate structure component helps us find that the differences in risks on some factors possibly play a role in explaining the quantile differences too although they are not that robust to changing the reference structures and the decomposition order. Also, the roles that the factor risks play seems to vary at different levels of portfolio returns too. These findings imply that the factor risks could vary at different levels of portfolio returns.

These findings shed further light on empirical asset pricing in terms of that one needs to think about differences in factor risk premia and risks in different markets and the variation of factor risk premia and risks at different levels of portfolio returns. Also, the paper adds to the literature on portfolio evaluation which compares two similar portfolios in different regions. In future studies, we will decompose the differences in portfolio returns between normal time periods and financial crises or recessions to further explore the roles of factor risks and risk premia.

Figure 3.4: Detailed decomposition
North America: MSHRC



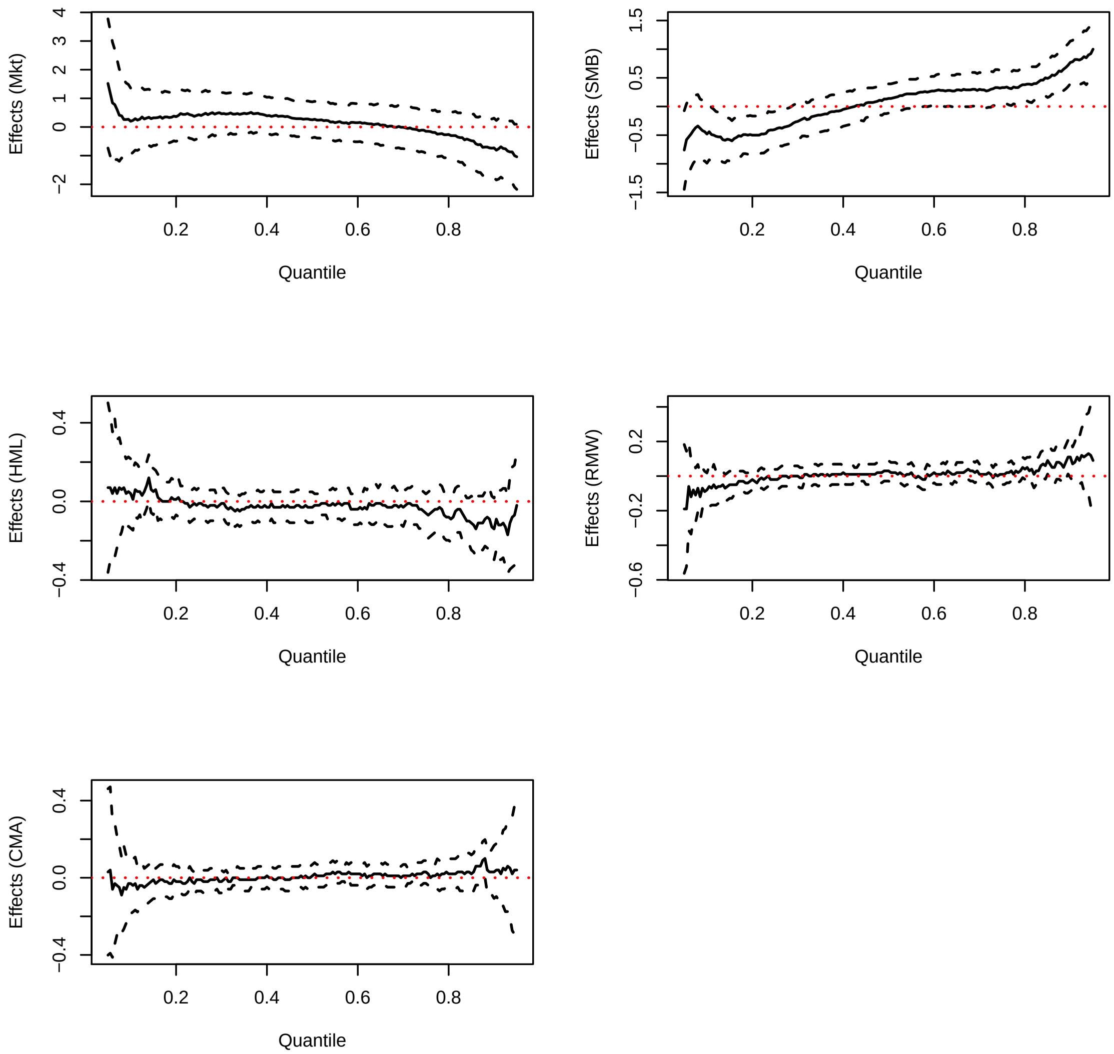
(a) Detailed composition components across quantiles



(b) Detailed structure components across quantiles

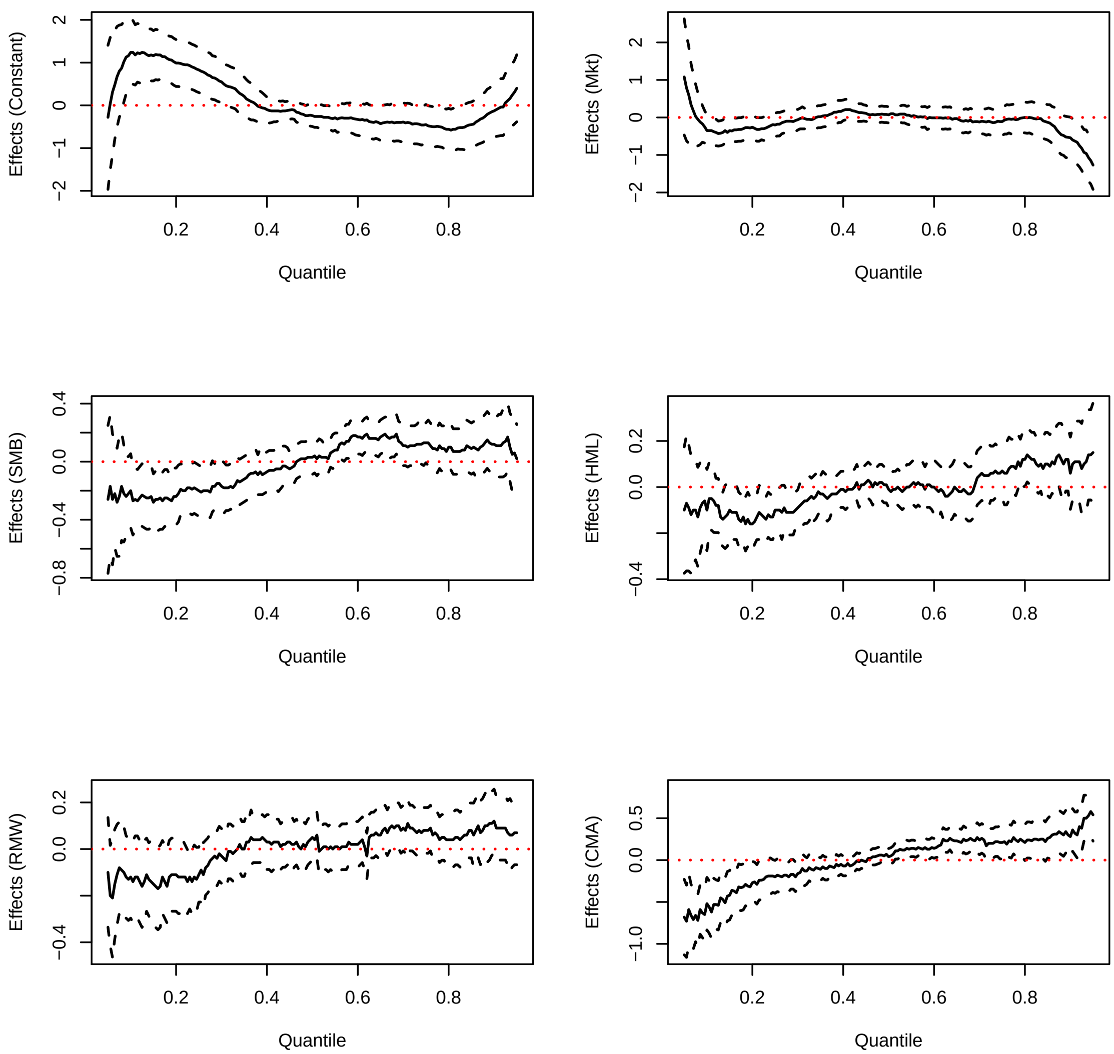
Note: The figure shows the detailed decomposition results. The top panel shows the detailed composition components across a series of quantiles. The bottom panel shows the detailed structure components across a series of quantiles. "North America: MSHRC" represents that the structures in North America are used as reference and the order of decomposition is the constant, market, size, value, profitability and investment factors.

Figure 3.5: Detailed composition components across quantiles
North America: MSHRC



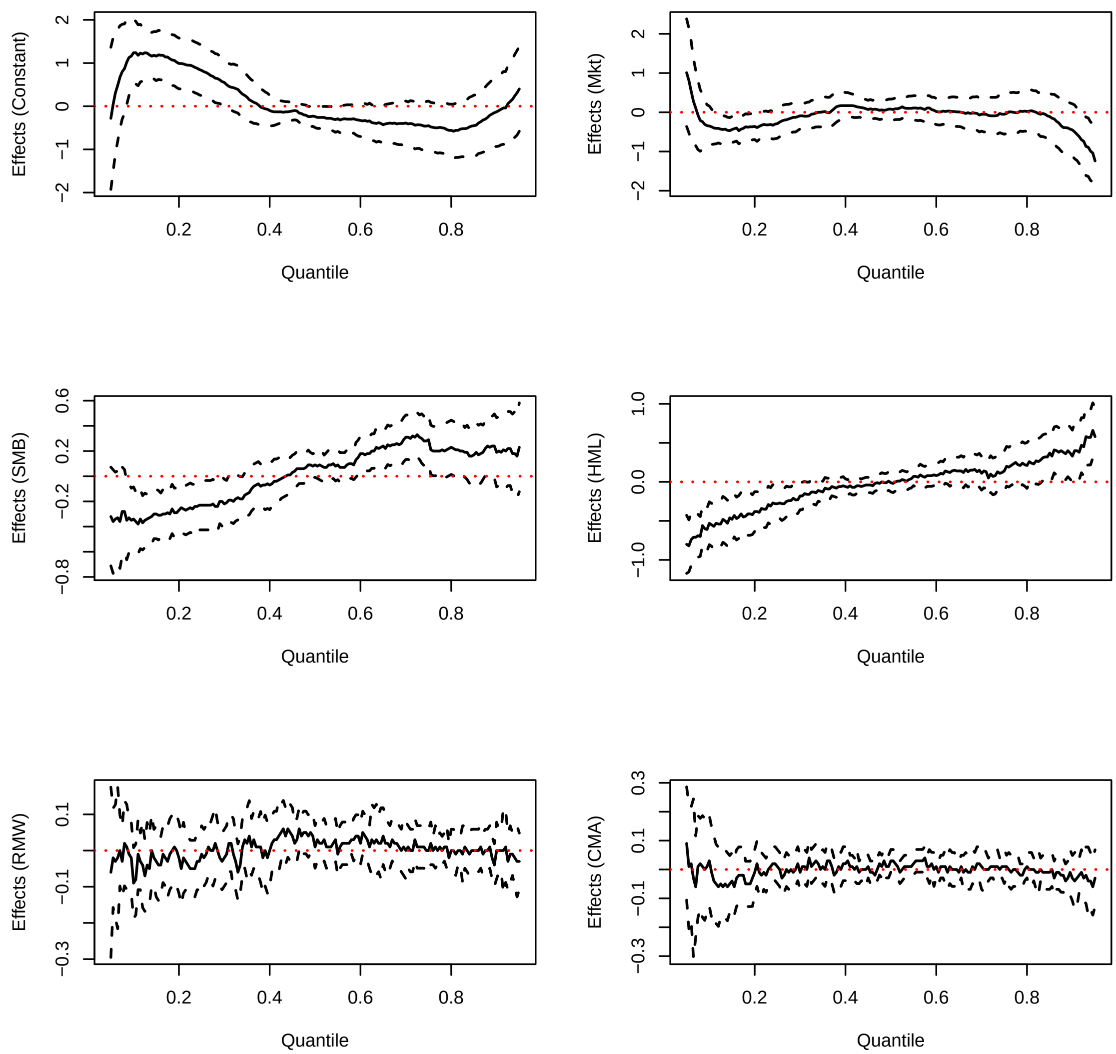
Note: The figure shows separately across quantiles the detailed composition component associated to each factor with 95% confidence interval. "North America: MSHRC" represents that the structures in North America are used as reference and the order of decomposition is the constant, market, size, value, profitability and investment factors.

Figure 3.6: Detailed structure components across quantiles
North America: MSHRC



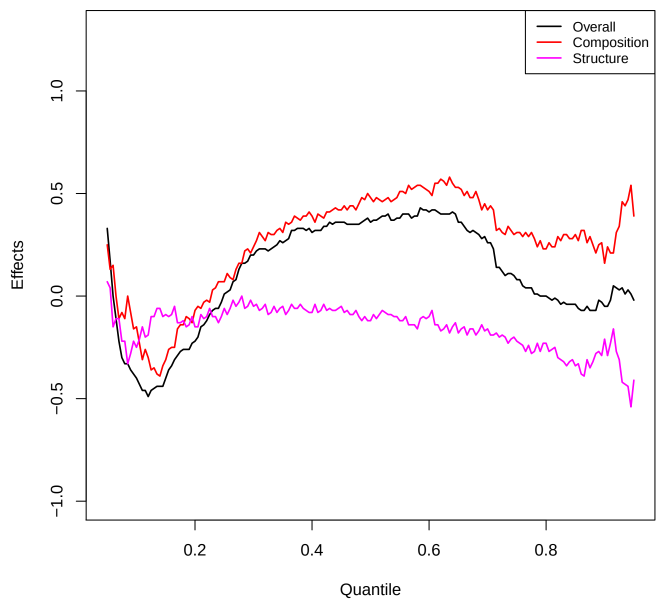
Note: The figure shows separately across quantiles the detailed structure component associated to each factor with 95% confidence interval. "North America: MSHRC" represents that the structures in North America are used as reference and the order of decomposition is the constant, market, size, value, profitability and investment factors.

Figure 3.7: Detailed structure components across quantiles
North America: RCMSH



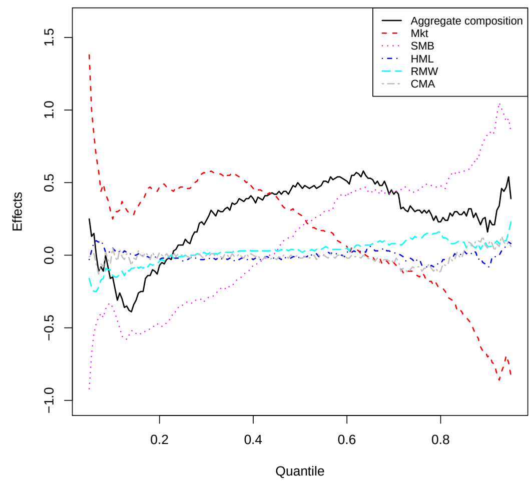
Note: The figure shows separately across quantiles the detailed structure component associated to each factor with 95% confidence interval. "North America: RCMSH" represents that the structures in North America are used as reference and the order of decomposition is the constant, profitability, investment, market, size and value factors.

Figure 3.8: Aggregate decomposition across quantiles
Europe: MSHRC

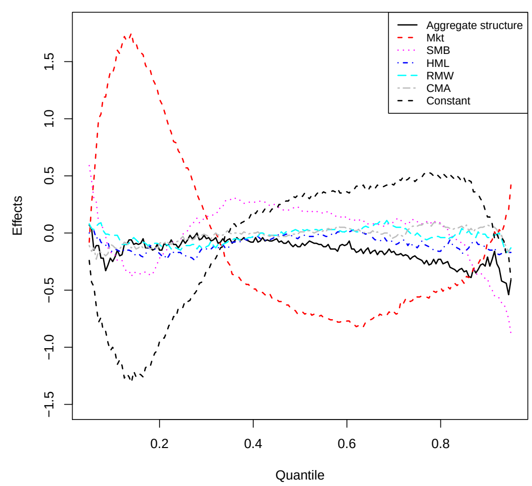


Note: The figure shows the aggregate decomposition decomponents across a series of quantiles. "Overall", "Composition" and "Structure" denotes the observed overall quantile differences, aggregate composition and structre components, respectively. "Europe: MSHRC" represents that the structures in Europe are used as reference and the order of decomposition is the constant, market, size and value, profitability and investment factors.

Figure 3.9: Detailed decomposition
Europe: MSHRC



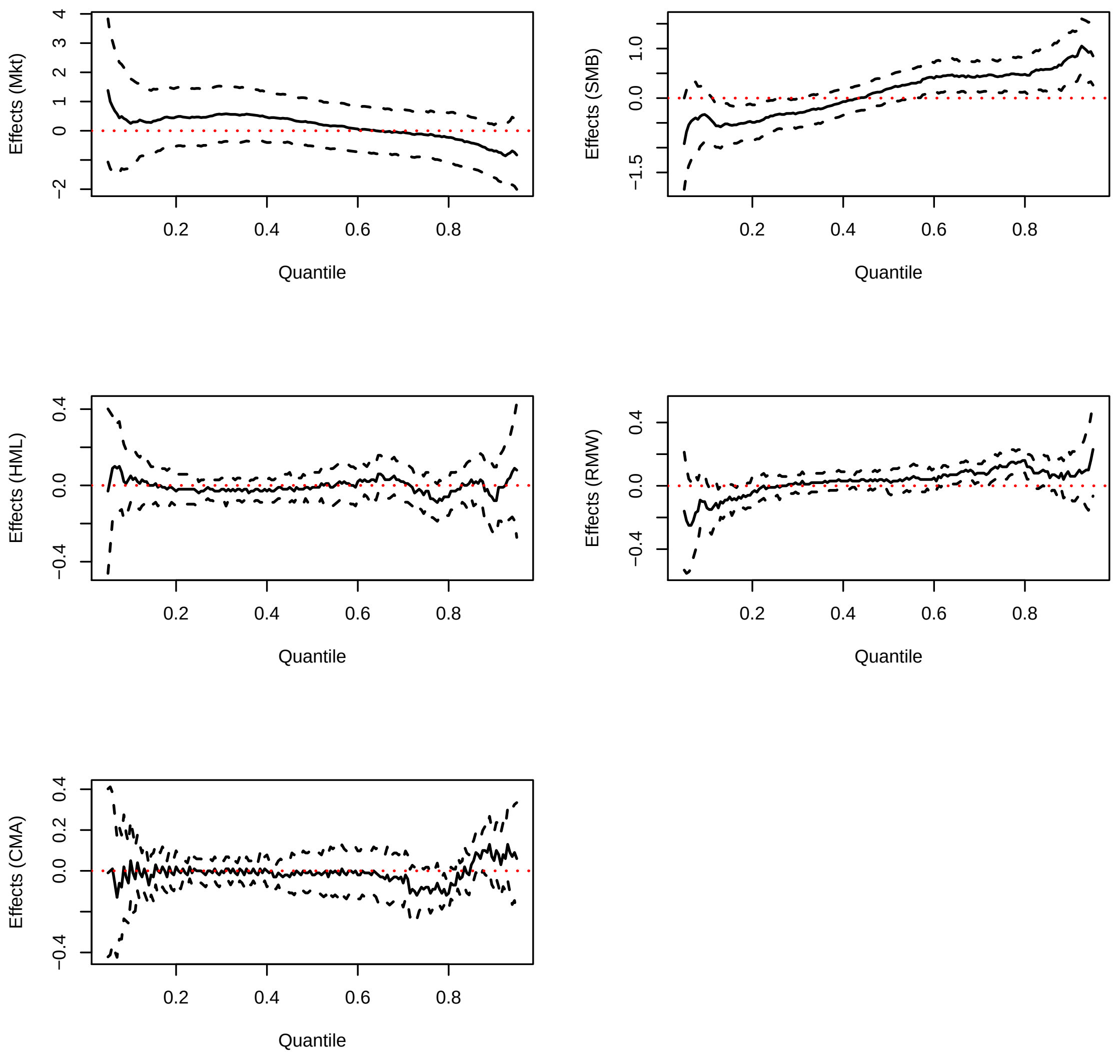
(a) Detailed composition components across quantiles



(b) Detailed structure components across quantiles

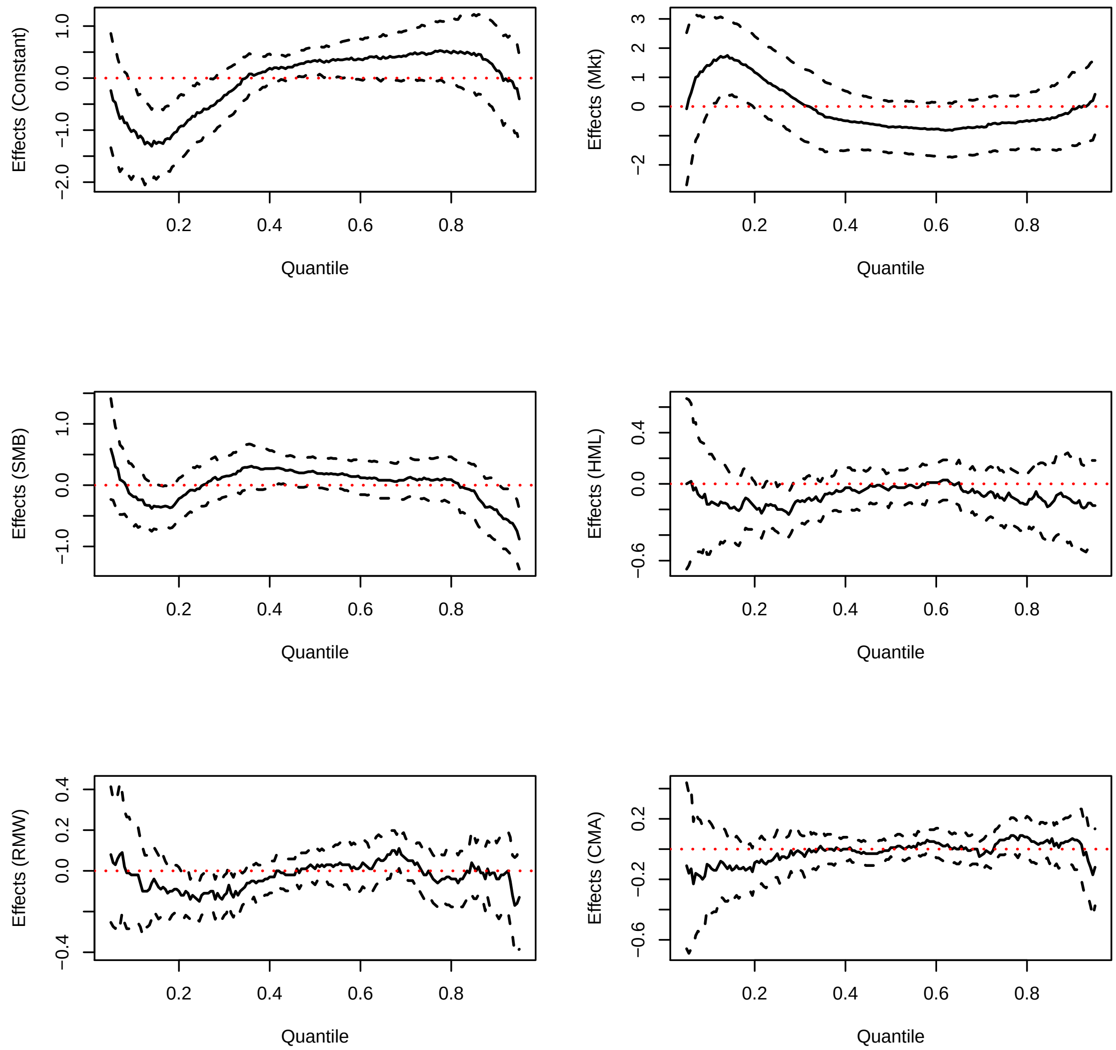
Note: The figure shows the detailed decomposition results. The top panel shows the detailed composition components across a series of quantiles. The bottom panel shows the detailed structure components across a series of quantiles. "Europe: MSHRC" represents that the structures in Europe are used as reference and the order of decomposition is the constant, market, size, value, profitability and investment factors.

Figure 3.10: Detailed composition components across quantiles
Europe: MSHRC



Note: The figure shows separately across quantiles the detailed composition component associated to each factor with 95% confidence interval. "Europe: MSHRC" represents that the structures in Europe are used as reference and the order of decomposition is the constant, market, size, value, profitability and investment factors.

Figure 3.11: Detailed structure components across quantiles
Europe: MSHRC



Note: The figure shows separately across quantiles the detailed structure component associated to each factor with 95% confidence interval. "Europe: MSHRC" represents that the structures in Europe are used as reference and the order of decomposition is the constant, market, size, value, profitability and investment factors.

CHAPTER 4

DECOMPOSING DIFFERENCES IN QUANTILE PORTFOLIO RETURNS BETWEEN NORTH AMERICA AND EUROPE USING RECENTERED INFLUENCE FUNCTION REGRESSION

4.1 Introduction

Huang (2018) is the first one to decompose the differences between two portfolio returns as far as we are aware. Huang (2018) shows that the market factor contributes the most to the quantile differences in the returns of portfolios of North America and Europe by combining the Oaxaca-Blinder decomposition and distribution regression. As Huang (2018) points out their results about quantile differences suffer the path dependent problem. The main disadvantage of using distribution regression to decompose the quantile differences is that the results are path dependent, that is, the order of decomposition affects the estimates of the detailed components. Thus, a more attractive approach should be able to decompose the quantile differences and also be path independent. In this paper, we apply the recentered influence function

(RIF) regressions method, which is able to obtain path independent estimates of the detailed components for the quantile differences.[1] RIF regression is developed by Firpo, Fortin, and Lemieux (2009) (see two interesting applications, Borah and Basu (2013) and Maclean, Webber, and Marti (2014)).[2] Combining RIF regression with traditional Oaxaca-Blinder decomposition (Oaxaca (1973) and Blinder (1973)), we do the detailed decomposition for the differences in quantile portfolio returns and obtain the path independent results.

For the purpose of comparing the results, we use the same data as Huang (2018). We will briefly describe the data in Section 4.4. We show that aggregate composition effects across all quantiles are positive and the market factor is the most significant factor. This results are consistent with Huang (2018). The detailed composition effect linked to the market factor is monotonically decreasing along quantiles. We also show that the aggregate structure effect for 5th quantile is positive and the aggregate composition effect is much more greater than structure effect for 95th quantile, which are not consistent with Huang (2018).

We organize the rest of the paper as follows. Section 4.2 provides a brief literature review, especially focusing on the Fama and French's five-factor model. Decomposition using the RIF-regression method is presented in Section 4.3. Section 4.4 describes the data briefly. Section 4.5 presents the decomposition results. We conclude in Section 4.6.

[1] As stated in Fortin, Lemieux, and Firpo (2011), RIF regressions and the distribution regression approach of Chernozhukov, Fernandez-Val, and Melly (2013) are closely connected, both of which are estimated for explaining the determinants of the proportion of outcome variable less than a certain value. The main difference is that distribution regressions invert globally the estimates of models for proportions into quantiles but RIF regressions locally invert the proportions estimates into quantiles. Thus, although decomposition results from RIF regressions are path independent, the limitation is that we do not know how good the approximation is in the locally invertions of proportion into quantiles.

[2] RIF is also named as unconditional quantile regression when the recentered function is for quantile. Thus, in this paper we interchangeably use RIF and unconditional quantile regression.

4.2 Five-factor Model

There is much evidence that average stock returns are related to the market (Sharpe (1964), Lintner (1965), Breeden (1979)), size, value[3], (Banz (1981), Basu (1983), Rosenberg, Reid, and Lanstein (1985)), profitability (Novy-Marx (2013)) and investment (Aharoni, Grundy, and Zeng (2013)).[4] Motivated by the evidence of Novy-Marx (2013) and Titman, Wei, and Xie (2004) and the dividend discount model, Fama and French (2015) establish the five-factor model by adding profitability and investment factors into the Fama and French (1993) three-factor model. Using international data, Fama and French (2017) show the empirical robustness of regional five-factor model in explaining the monthly excess portfolio returns, especially in North America and Europe capital markets. Fama and French's five-factor model is as follows,

$$R_{it} = a_i + b_i R_{Mt} + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{it} \tag{4.2.1}$$

In this equation, R_{it} is the return in excess of riskfree rate on portfolio i for period t.[5] R_{Mt} is the excess return on the value-weighted market portfolio, SMB_t (Size factor) is the return on a diversified portfolio of small stocks minus the return on a diversified portfolio of big stocks, HML_t (value factor) is the difference between the returns of high and low B/M stocks, RMW_t (profitability factor) is the difference between the returns on diversified portfolios of stocks with robust and weak profitability, and CMA_t (investment factor) is the difference between the returns on diversified portfolios of the stocks of low and high (conservative and aggressive) investment firms.[6]

[3]Size is measured by market capitalization, price times shares outstanding. Value is book-to-market equity ratio, B/M.

[4]See also Breeden, Gibbons, and Litzenberger (1989), Reinganum (1981), Haugen and Baker (1996), Cohen, Gompers, and Vuolteenaho (2002), Fairfield, Whisenant, and Yohn (2003), Titman, Wei, and Xie (2004),Fama and French (2006), Fama and French (2008), Hou, Xue, and Zhang (2015), Fama and French (2016)

[5]For monthly data, the riskfree rate is one-month Treasury bill rate.

[6]The details of the constructions of the portfolios and factors could be found in Fama and French

4.3 RIF-regression Method

A RIF-regression is basically a standard regression with which the dependent variable is replaced by the recentered influence function of the statistic of interest. RIF-regression is quite related to distribution regression (Foresi and Peracchi (1995b) and Chernozhukov, Fernandez-Val, and Melly (2013)) as we show below.[7] In this paper, we are interested in the recentered influence function for quantiles. The recentered influence function for quantiles can be obtained by recentering the influence function. The influence function corresponding to the τth quantile of a variable, Y, is given by $(\tau - \mathbb{1}\{Y \leq Q_\tau\})/f_Y(Q_\tau)$, where $\mathbb{1}\{Y \leq Q_\tau\}$ is an indicator function which is equal to 1 if Y is less or equal to τth quantile of the unconditional distribution of Y (Q_τ), otherwise zero. $f_Y(Q_\tau)$ is the density of the marginal distribution of Y. Then, the recentered influence function for τth quantile can be written as

$$RIF(y, Q_\tau) = Q_\tau + \frac{\tau - \mathbb{1}\{y \leq Q_\tau\}}{f_Y(Q_\tau)} = c_{1,\tau} \cdot \mathbb{1}\{y \leq Q_\tau\} + c_{2,\tau}, \tag{4.3.1}$$

where $c_{1,\tau} = -1/f_Y(Q_\tau)$ and $c_{2,\tau} = Q_\tau + \tau/f_Y(Q_\tau)$. Except the constants $c_{1,\tau}$ and $c_{2,\tau}$, we can see that RIF is just an indicator variable $\mathbb{1}\{y \leq Q_\tau\}$, which makes the RIF-regression very similar to the distribution regression. To estimate RIF, we first compute the sample quantile $\hat{Q}_\tau$ and estimate the density at $\hat{Q}_\tau$ using kernel methods. Then, by plugging in the estimates of the sample quantile $\hat{Q}_\tau$ and the density at $\hat{Q}_\tau$ into the equation 4.3.1, an estimate of the RIF of each observation is obtained.

In the paper, Y_{it} is the excess portfolio return on portfolio i in period t, $Y_{it} = R_{it}$. We can estimate $R\hat{I}F_{it}(Y_{it}, Q_\tau)$ by using $R\hat{I}F_{it}(Y_{it}, Q_\tau) = \hat{Q}_\tau + \frac{\tau - \mathbb{1}\{Y_{it} \leq \hat{Q}_\tau\}}{\hat{f}_Y(\hat{Q}_\tau)}$. We then

(1993), Fama and French (2015) and Fama and French (2017).

[7] A distribution regression is simply a linear regression of $\mathbb{1}\{y \leq Q_\tau\}$ on X when the linear probability model is used as the link function.

regress $R\hat{I}F_{it}(Y_{it}, Q_\tau)$ on five factors as follows

$$R\hat{I}F_{it}(Y_{it}, Q_\tau) = a_i + b_i R_{Mt} + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{it}. \quad (4.3.2)$$

We simplify the model above by dropping the time and portfolio subscription and using g indicating North America or Europe. It follows

$$R\hat{I}F_g = \alpha_g + X'_g \beta_g + \epsilon_g \quad (4.3.3)$$

Then the RIF-regression version of the Oaxaca-Blinder decomposition for any unconditional τ-quantile follows

$$\begin{aligned}
\hat{\Delta}^\tau_O &= R\bar{I}F_a - R\bar{I}F_u \\
&= (R\bar{I}F_a - R\bar{I}F^c_a) + (R\bar{I}F^c_a - R\bar{I}F_u) \\
&= [(\hat{\alpha}_a + \bar{X}'_a\hat{\beta}_a) - (\hat{\alpha}_a + \bar{X}'_u\hat{\beta}_a)] + [(\hat{\alpha}_a + \bar{X}'_u\hat{\beta}_a) - (\hat{\alpha}_u + \bar{X}'_u\hat{\beta}_u)] \\
&= \hat{\Delta}^\tau_C + \hat{\Delta}^\tau_S \qquad (4.3.4) \\
&= [(\bar{X}_{a1} - \bar{X}_{u1})\hat{\beta}_{a1} + \cdots + (\bar{X}_{a5} - \bar{X}_{u5})\hat{\beta}_{a5}] \\
&+ [(\hat{\alpha}_a - \hat{\alpha}_u) + \bar{X}_{u1}(\hat{\beta}_{a1} - \hat{\beta}_{u1}) + \cdots + \bar{X}_{u5}(\hat{\beta}_{a5} - \hat{\beta}_{u5})] \\
&= [\hat{\Delta}^1_C + \cdots + \hat{\Delta}^5_C] + [\hat{\Delta}^\alpha_S + \hat{\Delta}^1_S + \cdots + \hat{\Delta}^5_S]
\end{aligned}$$

Where $R\bar{I}F_a$ and $R\bar{I}F_u$ are the estimated mean $R\hat{I}F$ of North America and Europe. $R\bar{I}F^c_a$ is equal to $(\hat{\alpha}_a + \bar{X}'_u\hat{\beta}_a)$, which is $R\bar{I}F_a$ except with $\bar{X}'_a$ replaced by $\bar{X}'_u$. $\hat{\Delta}^\tau_C$ and $\hat{\Delta}^\tau_S$ are the estimates of aggregate composition and structure effects, equal to $[(\hat{\alpha}_a + \bar{X}'_a\hat{\beta}_a) - (\hat{\alpha}_a + \bar{X}'_u\hat{\beta}_a)]$ and $[(\hat{\alpha}_a + \bar{X}'_u\hat{\beta}_a) - (\hat{\alpha}_u + \bar{X}'_u\hat{\beta}_u)]$ respectively. The unexplained component, which is linked to the constant term, is estimated by $\hat{\alpha}_a - \hat{\alpha}_u$. The detailed composition and structure effects can be computed directly. For instance, the composition and structure components linked to Market factor, $\hat{\Delta}^1_C$ and $\hat{\Delta}^1_S$, can be obtained by $(\bar{X}_{a1} - \bar{X}_{u1})\hat{\beta}_{a1}$ and $\bar{X}_{u1}(\hat{\beta}_{a1} - \hat{\beta}_{u1})$ respectively. The same applies to the components linked to other factors. Although the results from RIF-regression are path independent, they could be different while using another group as reference. For this reason, we substitute $R\bar{I}F_a^c$ with $R\bar{I}F_u^c = (\hat{\alpha}_u + \bar{X}_a'\hat{\beta}_u)$ as robustness checks. That is, we show the main results using NA as reference group and the robustness results using Europe as reference group.

4.4 Data

In the paper, North America (NA) includes United States and Canada, and Europe contains Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. To compare our results with Huang (2018), we choose the five factors based on 2×3 sorts and 75 Size-B/M, Size-OP, and Size-Inv portfolios of North America and Europe respectively. The dataset is from Fama and French (2017), which can be downloaded from Kenneth R. French's personal website. The dataset includes the five factor returns and the monthly excess returns on the 5×5 Size-B/M, Size-OP, and Size-Inv portfolios of North America and Europe, ranging from July 1990 to November 2017 (329 months). We briefly describe as follows how the 75 portfolios of NA and Europe are constructed. At the end of June each year, stocks are allocated to five Size groups (Small to Big) using as breakpoints the 3rd, 7th, 13th, and 25th percentiles of the region's aggregate market capitalization. Stocks are allocated independently to five B/M groups (Low B/M to High B/M) by the quintile of B/M for the big stocks of the region. The intersections of the two sorts produce 25 Size-B/M portfolios. The 25 Size-Inv or Size-OP portfolios are constructed in the same way as in the Size-B/M portfolios except the second sort is on either profitability (robust minus weak) or investment (conservative minus aggressive). More details can be found in Fama and French (2017).

Table 4.1: Summary statistics for factor returns

(a) Mean and standard error

	North America					Europe					Difference				
	Mkt	SMB	HML	RMW	CMA	Mkt	SMB	HML	RMW	CMA	Mkt	SMB	HML	RMW	CMA
Mean	0.67	0.17	0.20	0.34	0.26	0.51	0.07	0.34	0.40	0.21	0.16	0.10	-0.14	-0.06	0.05
SE	0.23	0.15	0.18	0.13	0.15	0.27	0.12	0.13	0.08	0.10	0.16	0.16	0.14	0.14	0.12

(b) Correlations of factor returns in and across NA and Europe

	North America					Europe					Between				
	Mkt	SMB	HML	RMW	CMA	Mkt	SMB	HML	RMW	CMA	Mkt	SMB	HML	RMW	CMA
Mkt.RF	1.00	0.20	-0.23	-0.37	-0.44	1.00	-0.17	0.18	-0.26	-0.30	0.80	-0.26	0.04	-0.22	-0.35
SMB	0.20	1.00	-0.10	-0.42	-0.14	-0.17	1.00	0.01	-0.05	0.02	-0.26	0.31	0.03	-0.03	-0.12
HML	-0.23	-0.10	1.00	0.38	0.78	0.18	0.01	1.00	-0.54	0.54	0.04	0.03	0.60	-0.15	0.52
RMW	-0.37	-0.42	0.38	1.00	0.35	-0.26	-0.05	-0.54	1.00	-0.18	-0.22	-0.03	-0.15	0.22	0.38
CMA	-0.44	-0.14	0.78	0.35	1.00	-0.30	0.02	0.54	-0.18	1.00	-0.35	-0.12	0.52	0.38	0.57

We report the summary statistics for the five factors and 75 portfolios studied in Table 4.1 and 4.2, respectively. We do not describe the statistics in detail. We summarize that the differences in the factor distributions and the correlations between factors make the decomposition interesting. The summary statistics in detail can be found in Huang (2018).

4.5 Decomposition Results

We first run unconditional quantile regression for each portfolio in North America and Europe then decompose the estimated quantile differences for each corresponding portfolio. Lastly, we average the results over all the portfolios. Due to the big standard errors, the decomposition results are not statistically significant. Therefore, we choose to not report the confidence intervals in the main text. However, the results with 95% confidence intervals are shown in the Appendix B.1 and B.2.

Figure 4.1 shows the aggregate decomposition results. We obverse that the average overall differences across quantiles plummet in the lower quantiles and then climb from 20th through 70th quantiles and then decrease slowly after. For all quantiles, the composition effects are positive. This means that the differences in the distribution

Table 4.2: Summary statistics for portfolio returns

(a) Mean

	North America					Europe				
	Small	2	3	4	Big	Small	2	3	4	Big
Low B/M	0.43	0.40	0.79	0.82	0.64	-0.08	0.29	0.34	0.49	0.34
2	0.60	0.62	0.68	0.70	0.66	0.37	0.49	0.56	0.53	0.53
3	0.91	0.79	0.78	0.81	0.63	0.45	0.55	0.54	0.59	0.57
4	0.88	0.80	0.78	0.77	0.65	0.59	0.71	0.54	0.55	0.64
High B/M	1.16	0.86	0.89	0.86	0.57	0.75	0.76	0.74	0.65	0.54
Low Inv	1.21	0.88	0.91	0.90	0.74	0.55	0.59	0.64	0.64	0.57
2	1.12	0.90	0.90	0.94	0.65	0.71	0.77	0.67	0.63	0.58
3	0.99	0.88	0.90	0.85	0.65	0.71	0.76	0.70	0.64	0.48
4	0.96	0.87	0.79	0.88	0.62	0.65	0.64	0.48	0.63	0.45
High Inv	0.54	0.36	0.54	0.51	0.55	0.16	0.37	0.30	0.39	0.45
Low OP	0.84	0.44	0.59	0.57	0.25	0.15	0.25	0.26	0.20	0.16
2	1.06	0.78	0.78	0.79	0.57	0.67	0.57	0.56	0.55	0.55
3	1.05	0.95	0.81	0.93	0.63	0.75	0.70	0.73	0.71	0.56
4	1.07	1.01	0.90	0.78	0.71	0.90	0.75	0.63	0.69	0.47
High OP	1.09	1.14	1.06	0.93	0.72	0.76	0.96	0.81	0.70	0.60

(b) Standard deviation

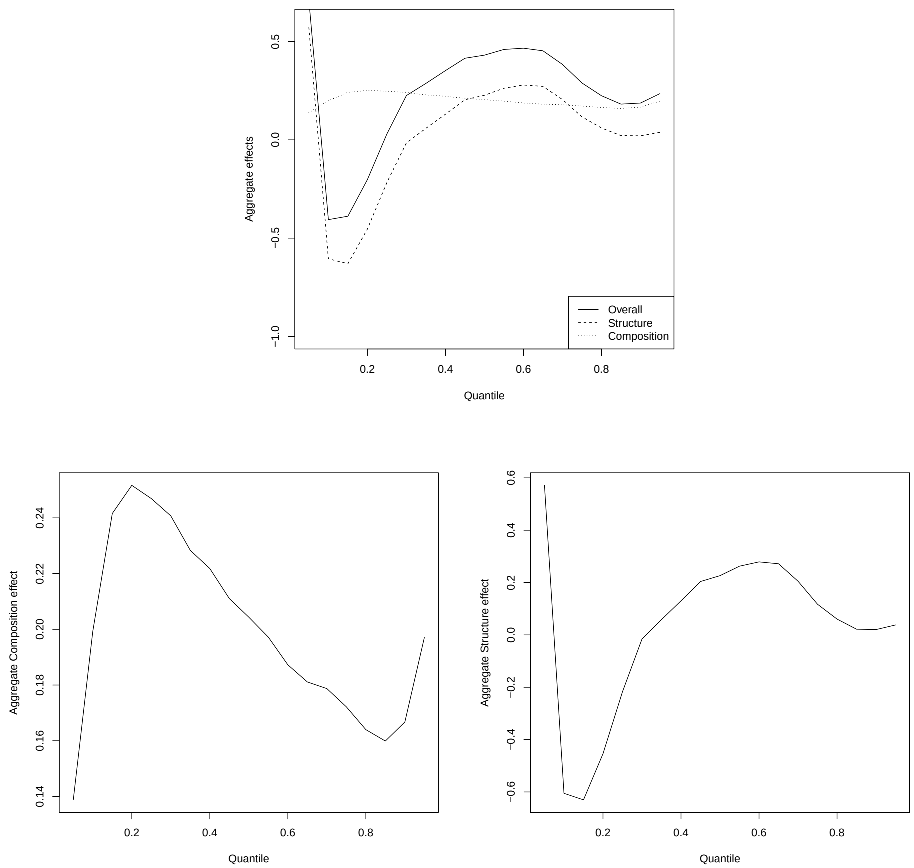
	North America					Europe				
	Small	2	3	4	Big	Small	2	3	4	Big
Low B/M	7.96	7.31	6.88	6.44	4.53	5.49	5.63	5.71	5.39	4.84
2	6.82	6.49	5.70	5.06	4.10	5.27	5.25	5.23	4.99	4.77
3	6.13	5.47	5.02	4.57	4.21	4.95	5.02	5.14	5.01	5.21
4	5.39	4.92	4.72	4.58	4.12	4.84	5.06	5.18	5.35	5.42
High B/M	5.28	5.16	4.88	4.76	5.23	4.83	5.30	5.59	5.80	6.34
Low Inv	6.42	5.68	5.09	4.76	4.05	4.98	5.22	5.46	5.20	4.85
2	4.97	4.76	4.36	4.21	3.69	4.46	4.82	5.04	5.00	4.81
3	4.92	4.78	4.65	4.38	4.05	4.56	4.80	4.89	4.85	5.17
4	5.29	5.42	5.23	4.87	4.77	4.76	5.14	5.17	5.23	5.45
High Inv	6.80	6.92	7.26	6.64	5.82	5.69	5.74	5.94	6.07	5.35
Low OP	6.68	6.65	6.77	6.26	5.60	5.18	5.43	5.56	5.52	6.09
2	4.91	4.94	4.87	4.69	4.89	4.67	5.00	5.09	5.14	5.53
3	4.98	4.99	4.60	4.40	4.31	4.81	5.02	5.07	5.25	5.08
4	5.34	5.27	4.93	4.45	4.07	4.79	5.09	5.17	5.08	4.99
High OP	5.66	5.44	5.40	4.83	3.96	4.94	5.32	5.33	5.28	4.76

Note: The table shows the summary statistics for the 75 portfolios in North America and Europe, respectively. The top panel shows means of monthly portfolio returns in excess of the one-month Treasury bill rate. The bottom panel shows the standard deviations.

of factors between North America and Europe contribute positively to the quantile portfolio returns differences. More specifically, the aggregate composition effects increase in the lower quantiles while the overall difference decreases. The composition effects start to decrease slowly after reaching the maximum at about 25th quantile.

Also, the aggregate structure effects follow the pattern of the overall differences and play a significant role in explaining the differences in portfolio returns. These seem to be inconsistent with Huang (2018) which says that the composition effects explain the most part of the differences and the structure effects play an insignificant role in aggregate. However, the detailed decomposition results below show that the structure effects linked to five factors together do not seem to play an significant role. The seemingly significant aggregate structure effects are mostly contributed by the constant term.

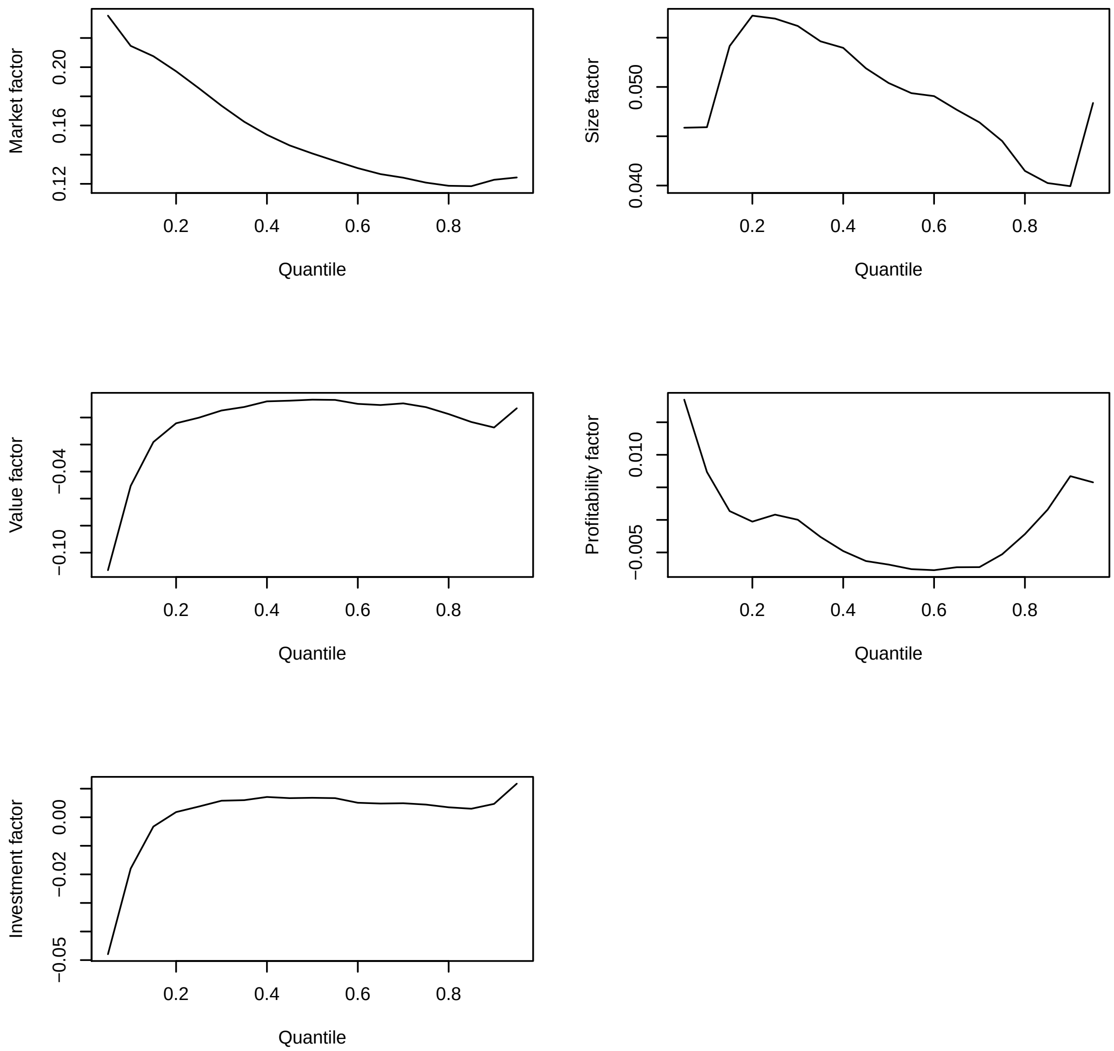
Figure 4.1: Aggregate decomposition effects



Notes: The figure plots the aggregate decomposition results across quantiles. The top panel plots the overall difference, structure and composition effects across quantiles, respectively. The left bottom panel plots the aggregate composition effects across quantiles. The right bottom panel plots the aggregate structure effects across quantiles.

Figure 4.2 shows the detailed composition effects across quantiles. We observe obviously that the most of aggregate composition effects are contributed by the detailed composition effect related to the market factor. The composition effect linked to market factor monotonically decreases across quantiles. This result could have important implications to investments and risk management and are worth further research. The left middle panel shows that the value factor contributes significantly to the composition effects in the lower quantiles. The distribution difference in other factors between North America and Europe do not seem to have economically significant effects on the overall quantile return differences.

Figure 4.2: Detailed composition effects

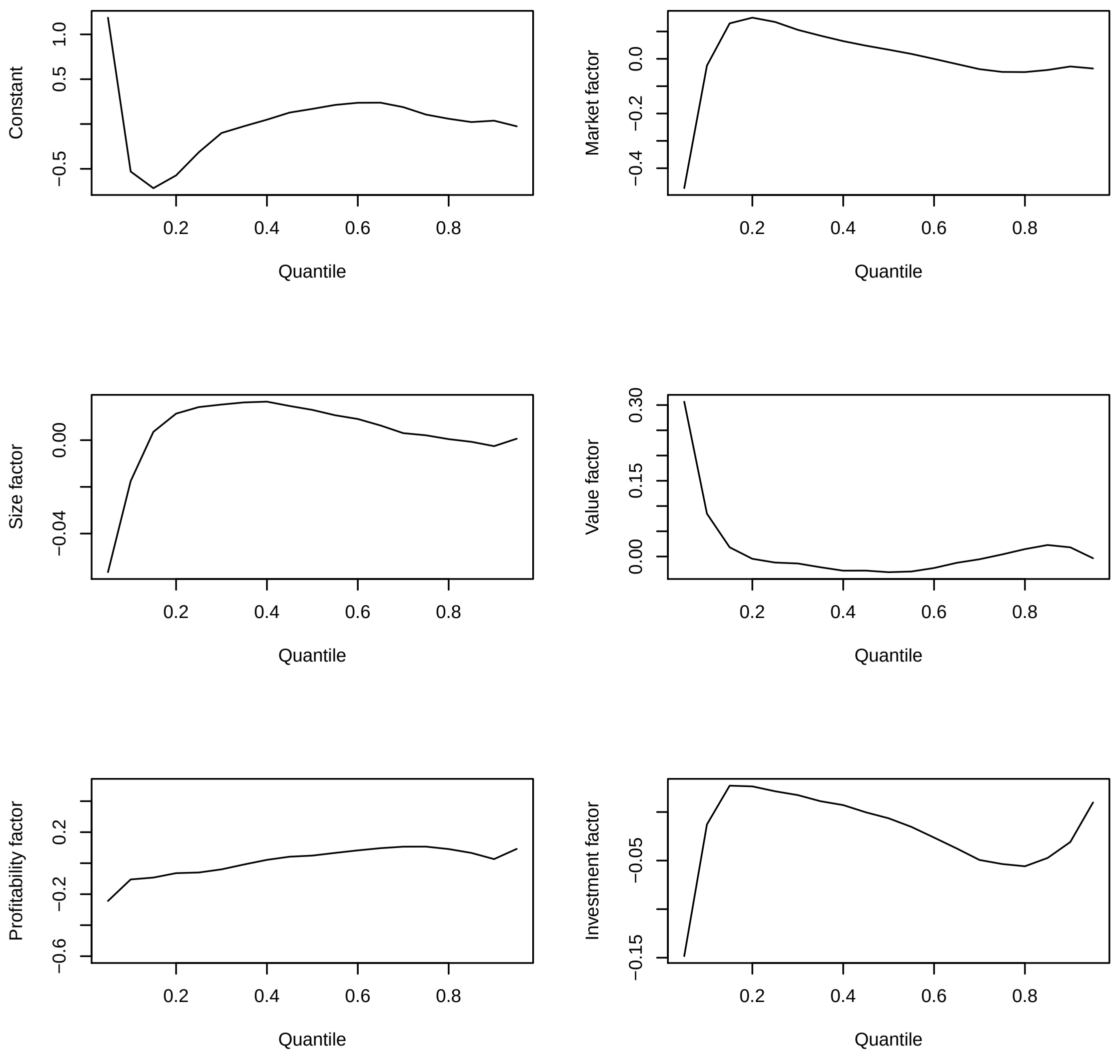


Notes: The figure plots the detailed composition effects across quantiles.

Figure 4.3 shows the detailed structure effects across quantiles. We observe that the very high and positive aggregate structure effects in the low quantiles are mostly explained by the differences in the constant terms. Also, except in the low quantiles the detailed structure effects do not seem to be economically significant.

To sum, the decomposition results using RIF-regression are quite different from the ones using distribution regressions.

Figure 4.3: Detailed structure effects



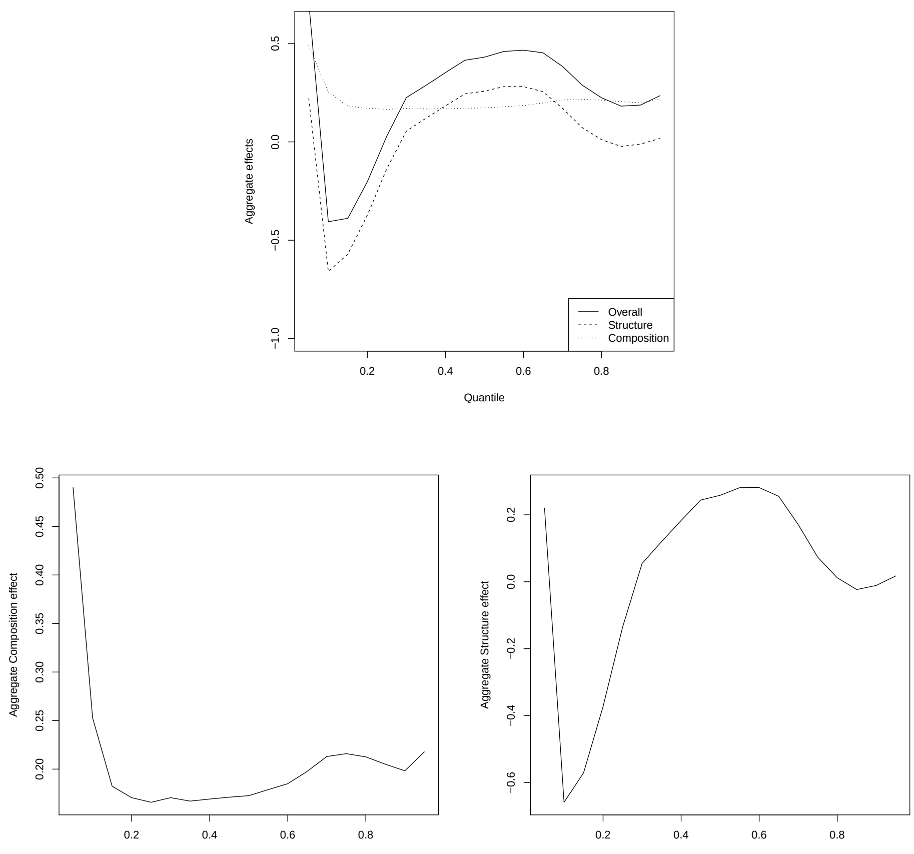
Notes: The figure plots detailed composition effects across quantiles.

4.5.1 Robustness

Although the decomposition results using RIF-regression are path independent, these results can be different by switching the reference group. In the main results shown above, we use NA as reference group. Here we show the results by using Europe as reference group. The results are shown in Figure 4.4 - 4.6.

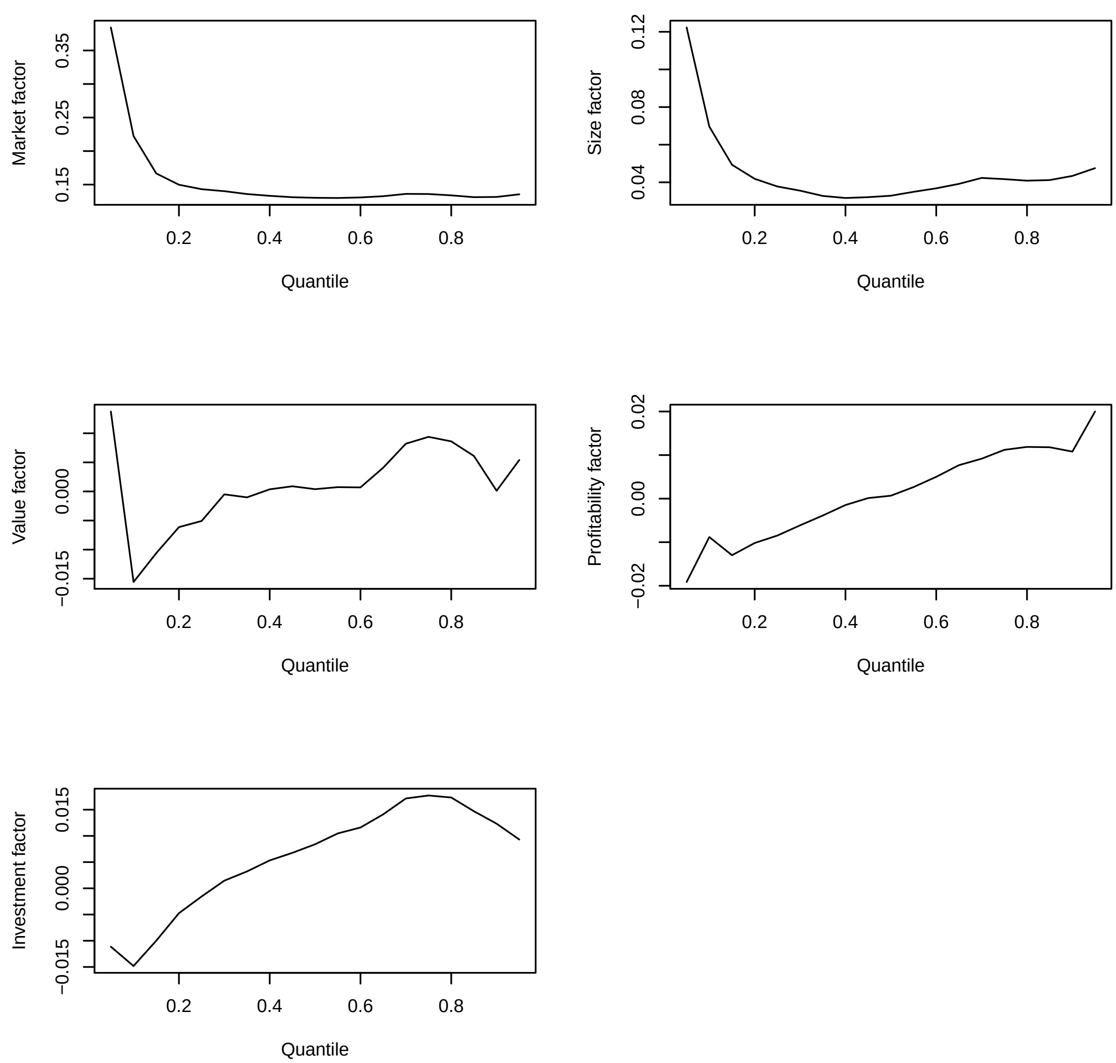
To be better compare with the main results shown above, Figure 4.7 - 4.9 show the differences between the aggregate and detailed decomposition effects using NA and Europe as reference groups across quantiles. We can see that except in the low quantiles, the differences between two results are close to zero. That is, the decomposition results using RIF-regression are robust to switching the reference group.

Figure 4.4: Aggregate decomposition effects



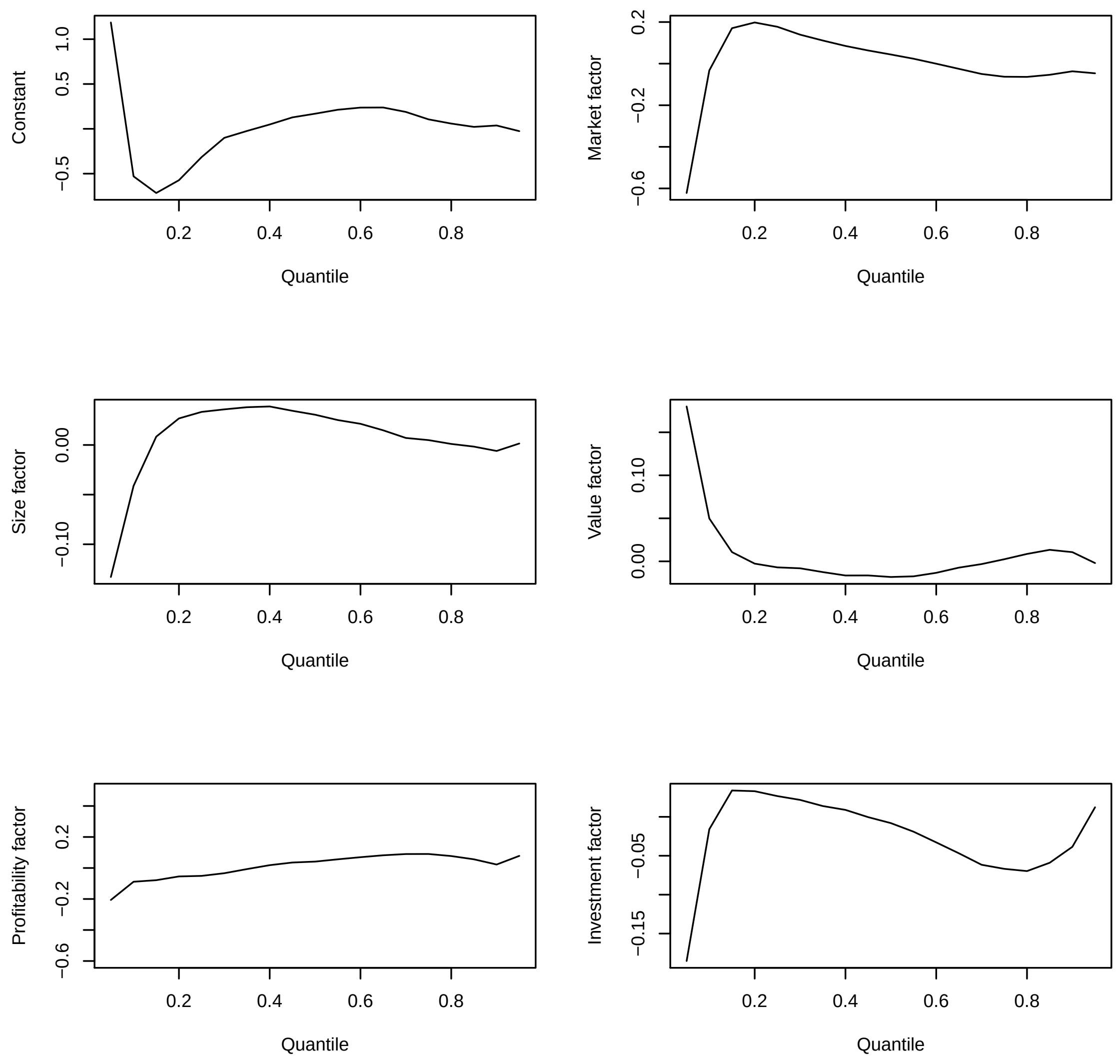
Notes: The figure plots the aggregate decomposition results across quantiles. The top panel plots the overall difference, structure and composition effects across quantiles, respectively. The left bottom panel plots the aggregate composition effects across quantiles. The right bottom panel plots the aggregate structure effects across quantiles.

Figure 4.5: Detailed composition effects



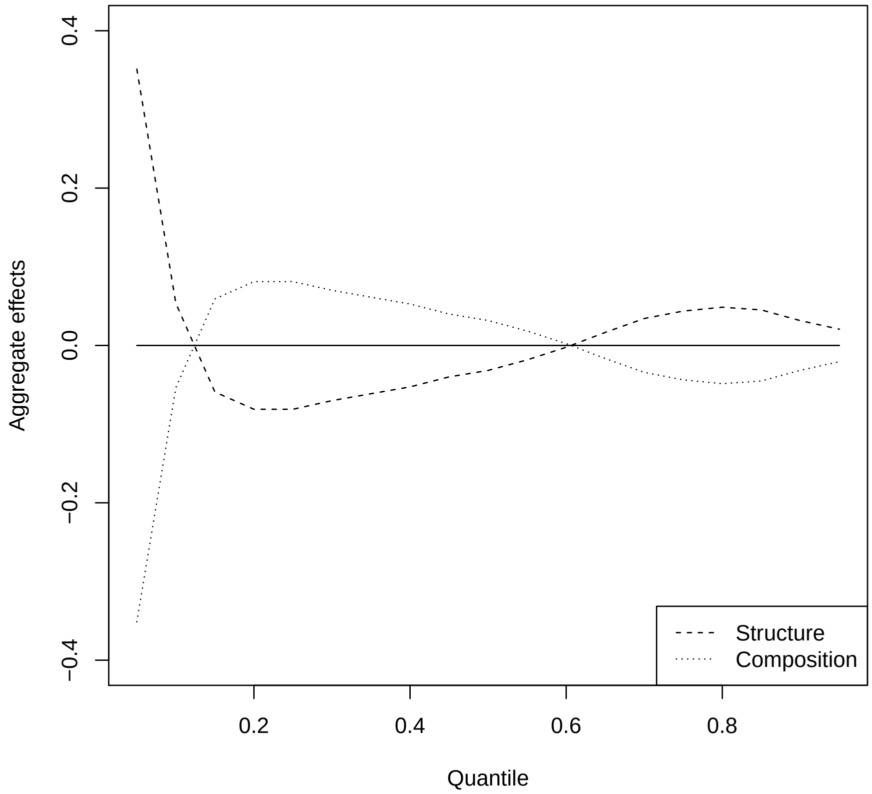
Notes: The figure plots the detailed composition effects across quantiles.

Figure 4.6: Detailed structure effects



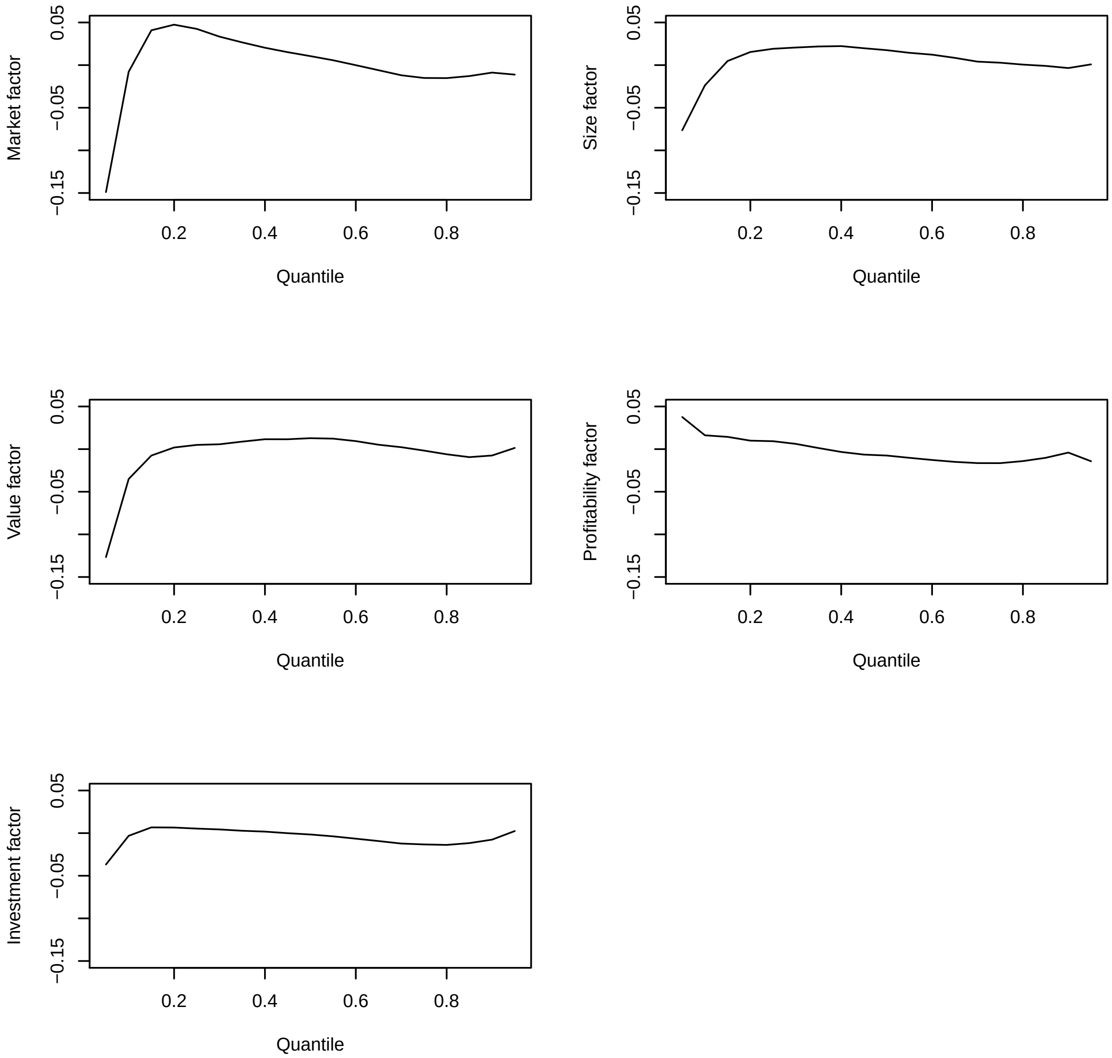
Notes: The figure plots the detailed composition effects across quantiles.

Figure 4.7: Differences in aggregate effects



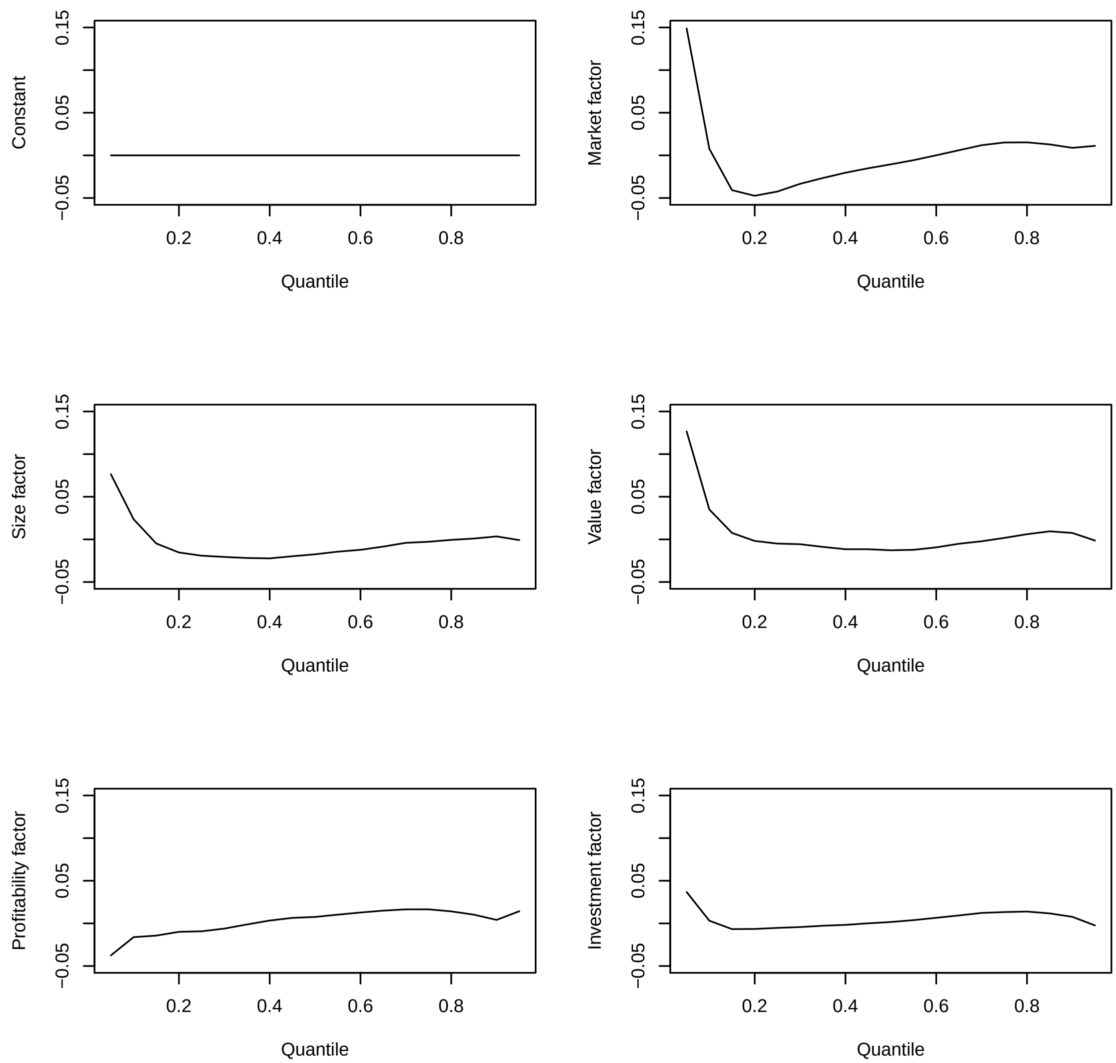
Notes: The figure plots the differences between the aggregate effects using NA and Europe as reference groups across quantiles.

Figure 4.8: Differences in detailed composition effects



Notes: The figure plots the differences between the detailed composition effects using NA and Europe as reference groups across quantiles.

Figure 4.9: Differences in detailed structure effects



Notes: The figure plots the differences between the detailed structure effects using NA and Europe as reference groups across quantiles.

4.6 Conclusion

The paper applies unconditional quantile regression to decompose the quantile difference in the portfolio returns between North America and Europe. The decomposition results are path dependent, which is complementary to the path dependent results from distribution regressions in Huang (2018). We show that the composition effects are all positive across all quantiles. This is inconsistent with Huang (2018). Although the aggregate structure effects are economically significant, they are mostly contributed by the differences in the constant terms. That is, the differences in the factor risks together do not seem to play a significant role in explaining the quantile differences in portfolio returns. The composition effect linked to the market factor is economically large and contributes most to the overall different. These are consistent with Huang (2018). Also except in the low quantiles, the effects linked to the other factors do not seem to play a significant role in explaining the differences.

BIBLIOGRAPHY

[1] Abadie, Alberto, Joshua Angrist, and Guido Imbens. "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings". *Econometrica* 70.1 (2002), pp. 91–117.

[2] Acemoglu, Daron and Pascual Restrepo. "Robots and jobs: Evidence from US labor markets". Working Paper. 2017.

[3] Aharoni, Gil, Bruce Grundy, and Qi Zeng. "Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis". *Journal of Financial Economics* 110.2 (2013), pp. 347–357.

[4] Alesina, Alberto F, Francesca Lotti, and Paolo Emilio Mistrulli. "Do women pay more for credit? Evidence from Italy". *Journal of the European Economic Association* 11.suppl_1 (2013), pp. 45–66.

[5] An, Yonghong, Wang Le, and Ruli Xiao. "Your American Dream is not mine! A new approach to estimating intergenerational mobility elasticities". Working Paper. 2017.

[6] Ao, Wallice, Sebastian Calonico, and Ying-Ying Lee. "Multivalued treatments and decomposition analysis: An application to the WIA program". Working Paper. 2017.

[7] Aristei, David and Manuela Gallo. "Does gender matter for firms' access to credit? Evidence from international data". *Finance Research Letters* 18 (2016), pp. 67–75.

[8] Autor, David, David Dorn, and Gordon H Hanson. "The China syndrome: Local labor market effects of import competition in the United States". *American Economic Review* 103.6 (2013), pp. 2121–68.

[9] Banz, Rolf W. "The relationship between return and market value of common stocks". *Journal of financial economics* 9.1 (1981), pp. 3–18.

[10] Barrett, Garry F and Stephen G Donald. "Statistical inference with generalized Gini indices of inequality, poverty, and welfare". *Journal of Business & Economic Statistics* 27.1 (2009), pp. 1–17.

[11] Barsky, Robert, John Bound, Kerwin Ko' Charles, and Joseph P Lupton. "Accounting for the black–white wealth gap: a nonparametric approach". *Journal of the American statistical Association* 97.459 (2002), pp. 663–673.

[12] Basu, Sanjoy. "The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence". *Journal of financial economics* 12.1 (1983), pp. 129–156.

[13] Bauer, Thomas K and Mathias Sinning. "An extension of the Blinder–Oaxaca decomposition to nonlinear models". *AStA Advances in Statistical Analysis* 92.2 (2008), pp. 197–206.

[14] Bhattacharya, Debopam. "Inference on inequality from household survey data". *Journal of Econometrics* 137.2 (2007), pp. 674–707.

[15] Bhattacharya, Debopam and Bhashkar Mazumder. "A nonparametric analysis of black–white differences in intergenerational income mobility in the United States". *Quantitative Economics* 2.3 (2011), pp. 335–379.

[16] Björklund, Anders, Jesper Roine, and Daniel Waldenström. "Intergenerational top income mobility in Sweden: Capitalist dynasties in the land of equal opportunity?" *Journal of Public Economics* 96.5 (2012), pp. 474–484.

[17] Black, Sandra E and Paul Devereux. "Recent developments in intergenerational mobility". *Handbook of Labor Economics*. Ed. by Ashenfelter, Orley and David Card. Vol. 4.B. Amsterdam: North-Holland, 2011. Chap. 16, pp. 1487–1541.

[18] Blanden, Jo, Paul Gregg, and Lindsey Macmillan. "Accounting for intergenerational income persistence: Noncognitive skills, ability and education". *The Economic Journal* 117.519 (2007), pp. C43–C60.

[19] Blanden, Jo, Paul Gregg, and Lindsey Macmillan. "Accounting for intergenerational persistence". *Economic Journal* 117 (2007), pp. C43–C60.

[20] Blinder, Alan S. "Wage discrimination: reduced form and structural estimates". *Journal of Human resources* (1973), pp. 436–455.

[21] Bloome, Deirdre. "Income inequality and intergenerational income mobility in the United States". *Social Forces* 93.3 (2015), pp. 1047–1080.

[22] Borah, Bijan J and Anirban Basu. "Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence". *Health economics* 22.9 (2013), pp. 1052–1070.

[23] Bowles, Samuel and Herbert Gintis. "The inheritance of inequality". *Journal of economic Perspectives* 16.3 (2002), pp. 3–30.

[24] Bratsberg, Bernt et al. "Nonlinearities in intergenerational earnings mobility: Consequences for cross-country comparisons". *The Economic Journal* 117.519 (2007), pp. C72–C92.

[25] Breeden, Douglas T. "An intertemporal asset pricing model with stochastic consumption and investment opportunities". *Journal of financial Economics* 7.3 (1979), pp. 265–296.

[26] Breeden, Douglas T, Michael R Gibbons, and Robert H Litzenberger. "Empirical tests of the consumption-oriented CAPM". *The Journal of Finance* 44.2 (1989), pp. 231–262.

[27] Brown, Jeffrey R and Alessandro Previtero. “Procrastination, present-biased preferences, and financial behaviors”. *Unpublished Manuscript, University of Illinois at Urbana-Champaign and University of Western Ontario* (2014).

[28] Buchinsky, Moshe. “Changes in the US wage structure 1963-1987: Application of quantile regression”. *Econometrica: Journal of the Econometric Society* (1994), pp. 405–458.

[29] Cai, Zongwu, Jianqing Fan, and Runze Li. “Efficient estimation and inferences for varying-coefficient models”. *Journal of the American Statistical Association* 95.451 (2000), pp. 888–902.

[30] Cai, Zongwu, Jianqing Fan, and Qiwei Yao. “Functional-coefficient regression models for nonlinear time series”. *Journal of the American Statistical Association* 95.451 (2000), pp. 941–956.

[31] Callaway, Brantly and Weige Huang. “Distributional effects of a continuous treatment with an application on intergenerational mobility”. Working Paper. 2018.

[32] Callaway, Brantly and Weige Huang. “Semiparametric estimation of Oaxaca-Blinder decompositions with continuous groups”. Working Paper. 2018.

[33] Carhart, Mark M. “On persistence in mutual fund performance”. *The Journal of finance* 52.1 (1997), pp. 57–82.

[34] Chadwick, Laura and Gary Solon. “Intergenerational income mobility among daughters”. *The American Economic Review* 92.1 (2002), pp. 335–344.

[35] Chernozhukov, Victor and Iván Fernández-Val. “Subsampling inference on quantile regression processes”. *Sankhyā: The Indian Journal of Statistics* (2005), pp. 253–276.

[36] Chernozhukov, Victor, Iván Fernández-Val, and Alfred Galichon. “Quantile and probability curves without crossing”. *Econometrica* 78.3 (2010), pp. 1093–1125.

[37] Chernozhukov, Victor, Ivan Fernandez-Val, and Blaise Melly. “Inference on counterfactual distributions”. *Econometrica* 81.6 (2013), pp. 2205–2268.

[38] Chernozhukov, Victor and Christian Hansen. “An IV model of quantile treatment effects”. *Econometrica* 73.1 (2005), pp. 245–261.

[39] Chernozhukov, Victor and Christian Hansen. “Instrumental quantile regression inference for structural and treatment effect models”. *Journal of Econometrics* 132.2 (2006), pp. 491–525.

[40] Chernozhukov, Victor and Christian Hansen. “Instrumental variable quantile regression: A robust inference approach”. *Journal of Econometrics* 142.1 (2008), pp. 379–398.

[41] Chernozhukov, Victor, Guido W Imbens, and Whitney K Newey. “Instrumental variable estimation of nonseparable models”. *Journal of Econometrics* 139.1 (2007), pp. 4–14.

[42] Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. "The fading American dream: Trends in absolute income mobility since 1940". *Science* 356.6336 (2017), pp. 398–406.

[43] Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. "Where is the land of opportunity? The geography of intergenerational mobility in the United States". *The Quarterly Journal of Economics* 129.4 (2014), pp. 1553–1623.

[44] Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. "Is the United States still a land of opportunity? Recent trends in intergenerational mobility". *The American Economic Review* 104.5 (2014), pp. 141–147.

[45] Cohen, Randolph B, Paul A Gompers, and Tuomo Vuolteenaho. "Who underreacts to cash-flow news? Evidence from trading between individuals and institutions". *Journal of financial Economics* 66.2 (2002), pp. 409–462.

[46] Collins, William J and Gregory T Niemesh. "Unions and the great compression of wage inequality in the United States at mid-century". Working Paper. 2017.

[47] Collins, William J and Marianne H Wanamaker. "Up from slavery? African American intergenerational economic mobility since 1880". Working Paper. 2017.

[48] Cooper, Michael J, Huseyin Gulen, and Michael J Schill. "Asset growth and the cross-section of stock returns". *The Journal of Finance* 63.4 (2008), pp. 1609–1651.

[49] Davis, Jonathan and Bhashkar Mazumder. "The decline in intergenerational mobility after 1980". Working Paper. 2017.

[50] DiNardo, John, Nicole M Fortin, and Thomas Lemieux. "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach". *Econometrica* 64.5 (1996), pp. 1001–1044.

[51] Donald, Stephen G, David A Green, and Harry J Paarsch. "Differences in wage distributions between Canada and the United States: An application of a flexible estimator of distribution functions in the presence of covariates". *The Review of economic studies* 67.4 (2000), pp. 609–633.

[52] Donald, Stephen G, Yu-Chin Hsu, and Garry F Barrett. "Incorporating covariates in the measurement of welfare and inequality: methods and applications". *The Econometrics Journal* 15.1 (2012), pp. C1–C30.

[53] Dong, Yingying and Shu Shen. "Testing for Rank Invariance or Similarity in Program Evaluation". *The Review of Economics and Statistics* 100.1 (2018), pp. 78–85. eprint: `https://doi.org/10.1162/REST_a_00686`. URL: `https://doi.org/10.1162/REST_a_00686`.

[54] Eide, Eric R and Mark H Showalter. "Factors affecting the transmission of earnings across generations: A quantile regression approach". *Journal of Human Resources* (1999), pp. 253–267.

[55] Fairfield, Patricia M, J Scott Whisenant, and Teri Lombardi Yohn. "Accrued earnings and growth: Implications for future profitability and market mispricing". *The accounting review* 78.1 (2003), pp. 353–371.

[56] Fairlie, Robert W. "An extension of the Blinder-Oaxaca decomposition technique to logit and probit models". *Journal of economic and social measurement* 30.4 (2005), pp. 305–316.

[57] Fama, Eugene F and Kenneth R French. "Common risk factors in the returns on stocks and bonds". *Journal of financial economics* 33.1 (1993), pp. 3–56.

[58] Fama, Eugene F and Kenneth R French. "Size and book-to-market factors in earnings and returns". *The Journal of Finance* 50.1 (1995), pp. 131–155.

[59] Fama, Eugene F and Kenneth R French. "Profitability, investment and average returns". *Journal of Financial Economics* 82.3 (2006), pp. 491–518.

[60] Fama, Eugene F and Kenneth R French. "Dissecting anomalies". *The Journal of Finance* 63.4 (2008), pp. 1653–1678.

[61] Fama, Eugene F and Kenneth R French. "A five-factor asset pricing model". *Journal of Financial Economics* 116 (2015), pp. 1–22.

[62] Fama, Eugene F and Kenneth R French. "Dissecting anomalies with a five-factor model". *The Review of Financial Studies* 29.1 (2016), pp. 69–103.

[63] Fama, Eugene F and Kenneth R French. "International tests of a five-factor asset pricing model". *Journal of Financial Economics* 123.3 (2017), pp. 441–463.

[64] Firpo, Sergio. "Efficient semiparametric estimation of quantile treatment effects". *Econometrica* 75.1 (2007), pp. 259–276.

[65] Firpo, Sergio, Nicole M Fortin, and Thomas Lemieux. "Unconditional quantile regressions". *Econometrica* 77.3 (2009), pp. 953–973.

[66] Firpo, Sergio and Cristine Pinto. "Identification and estimation of distributional impacts of interventions using changes in inequality measures". *Journal of Applied Econometrics* 31.3 (2016), pp. 457–486.

[67] Florens, Jean-Pierre, James J Heckman, Costas Meghir, and Edward Vytlacil. "Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects". *Econometrica* 76.5 (2008), pp. 1191–1206.

[68] Flores, Carlos A. "Estimation of dose-response functions and optimal doses with a continuous treatment". Working Paper. 2007.

[69] Flores, Carlos A, Alfonso Flores-Lagunes, Arturo Gonzalez, and Todd C Neumann. "Estimating the effects of length of exposure to instruction in a training program: the case of job corps". *Review of Economics and Statistics* 94.1 (2012), pp. 153–171.

[70] Foresi, Silverio and Franco Peracchi. "The conditional distribution of excess returns: An empirical analysis". *Journal of the American Statistical Association* 90.430 (1995), pp. 451–466.

[71] Foresi, Silverio and Franco Peracchi. "The conditional distribution of excess returns: An empirical analysis". *Journal of the American Statistical Association* 90.430 (1995), pp. 451–466.

[72] Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. "Decomposition methods in economics". *Handbook of labor economics* 4 (2011), pp. 1–102.

[73] Friedman, Jerome H and Lawrence C Rafsky. "Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests". *The Annals of Statistics* (1979), pp. 697–717.

[74] Frölich, Markus and Blaise Melly. "Unconditional quantile treatment effects under endogeneity". *Journal of Business & Economic Statistics* 31.3 (2013), pp. 346–357.

[75] Füss, Roland, Thomas Gietzen, and Philipp B Rindler. "Did the 2007-08 Financial Crisis Change Risk Perception?" (2011).

[76] Galvao, Antonio F and Liang Wang. "Uniformly semiparametric efficient estimation of treatment effects with a continuous treatment". *Journal of the American Statistical Association* 110.512 (2015), pp. 1528–1542.

[77] Gilbert, Thomas, Christopher Hrdlicka, and Avraham Kamara. "The structure of information release and the factor structure of returns". *Journal of Financial Economics* 127.3 (2018), pp. 546–566.

[78] Gomulka, Joanna and Nicholas Stern. "The employment of married women in the United Kingdom 1970-83". *Economica* (1990), pp. 171–199.

[79] Gosling, Amanda, Stephen Machin, and Costas Meghir. "The changing distribution of male wages in the UK". *The Review of Economic Studies* 67.4 (2000), pp. 635–666.

[80] Graham, John R and Campbell R Harvey. "The Equity Risk Premium in 2018" (2018).

[81] Grawe, Nathan D. "Reconsidering the use of nonlinearities in intergenerational earnings mobility as a test for credit constraints". *Journal of Human Resources* 39.3 (2004), pp. 813–827.

[82] Groves, Melissa Osborne. "Personality and the intergenerational transmission of economic status". *Unequal chances: Family background and economic success* (2005), pp. 208–231.

[83] Hardle, Wolfgang and Enno Mammen. "Comparing nonparametric versus parametric regression fits". *The Annals of Statistics* 21.4 (1993), pp. 1926–1947.

[84] Haugen, Robert A and Nardin L Baker. "Commonality in the determinants of expected stock returns". *Journal of Financial Economics* 41.3 (1996), pp. 401–439.

[85] Heckman, James J, Jeffrey Smith, and Nancy Clements. "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts". *The Review of Economic Studies* 64.4 (1997), pp. 487–535.

[86] Heckman, James, Hidehiko Ichimura, and Petra Todd. "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme". *The Review of Economic Studies* 64.4 (1997), pp. 605–654.

[87] Hirano, Keisuke and Guido W Imbens. "The propensity score with continuous treatments". *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* 226164 (2004), pp. 73–84.

[88] Hirano, Keisuke, Guido Imbens, and Geert Ridder. "Efficient estimation of average treatment effects using the estimated propensity score". *Econometrica* 71.4 (2003), pp. 1161–1189.

[89] Horowitz, Joel L and Sokbae Lee. "Nonparametric instrumental variables estimation of a quantile regression model". *Econometrica* 75.4 (2007), pp. 1191–1208.

[90] Hou, Kewei, Chen Xue, and Lu Zhang. "Digesting anomalies: An investment approach". *The Review of Financial Studies* 28.3 (2015), pp. 650–705.

[91] Huang, Weige. "Decomposing Differences in Quantile Portfolio Returns Between North America and Europe Using Recentered Influence Function Regression" (2018).

[92] Imbens, Guido W and Whitney K Newey. "Identification and estimation of triangular simultaneous equations models without additivity". *Econometrica* 77.5 (2009), pp. 1481–1512.

[93] Imbens, Guido W, Donald B Rubin, and Bruce I Sacerdote. "Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players". *American Economic Review* 91.4 (2001), pp. 778–794.

[94] Imbens, Guido and Jeffrey Wooldridge. "Recent developments in the econometrics of program evaluation". *Journal of Economic Literature* 47.1 (2009), pp. 5–86.

[95] Jantti, Markus et al. "American exceptionalism in a new light: A comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States". Working Paper. 2006.

[96] Jasova, Martina, Caterina Mendicino, and Dominik Supera. "Rollover risk and bank lending behavior: Evidence from unconventional central bank liquidity". Working Paper. 2018.

[97] Kabir, M Humayun and Shamim Shakur. "Nonlinear decomposition analysis of risk aversion and stock-holding behaviour of US households". *Applied Financial Economics* 24.7 (2014), pp. 495–503.

[98] Kennedy, Edward H, Zongming Ma, Matthew D McHugh, and Dylan S Small. “Non-parametric methods for doubly robust estimation of continuous treatment effects”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2016).

[99] Kitagawa, Toru, Martin Nybom, and Jan Stuhler. “Measurement error and rank correlations”. Working Paper. 2017.

[100] Koenker, Roger. *Quantile Regression*. Cambridge University Press, 2005.

[101] Koenker, Roger and Gilbert Bassett Jr. “Regression quantiles”. *Econometrica* (1978), pp. 33–50.

[102] Koenker, Roger, Victor Chernozhukov, Xuming He, and Limin Peng. *Handbook of Quantile Regression*. CRC Press, 2017.

[103] Kosorok, Michael R. *Introduction to Empirical Processes and Semiparametric Inference*. Springer Science & Business Media, 2007.

[104] Landersø, Rasmus and James J Heckman. “The scandinavian fantasy: The sources of intergenerational mobility in denmark and the U.S.” *The Scandinavian Journal of Economics* 119.1 (2017), pp. 178–230.

[105] Li, Qi, Cliff J Huang, Dong Li, and Tsu-Tan Fu. “Semiparametric smooth coefficient models”. *Journal of Business & Economic Statistics* 20.3 (2002), pp. 412–422.

[106] Li, Qi and Jeffrey Scott Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 2007.

[107] Lilliefors, Hubert W. “On the Kolmogorov-Smirnov test for normality with mean and variance unknown”. *Journal of the American statistical Association* 62.318 (1967), pp. 399–402.

[108] Lilliefors, Hubert W. “On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown”. *Journal of the American Statistical Association* 64.325 (1969), pp. 387–389.

[109] Lintner, John. “The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets”. *The review of economics and statistics* (1965), pp. 13–37.

[110] Machado, José AF and José Mata. “Counterfactual decomposition of changes in wage distributions using quantile regression”. *Journal of applied Econometrics* 20.4 (2005), pp. 445–465.

[111] Maclean, Johanna Catherine, Douglas A Webber, and Joachim Marti. “An application of unconditional quantile regression to cigarette taxes”. *Journal of Policy Analysis and management* 33.1 (2014), pp. 188–210.

[112] Massey Jr, Frank J. “The Kolmogorov-Smirnov test for goodness of fit”. *Journal of the American statistical Association* 46.253 (1951), pp. 68–78.

[113] Mayer, Susan E and Leonard M Lopoo. "Has the intergenerational transmission of economic status changed?" *Journal of Human Resources* 40.1 (2005), pp. 169–185.

[114] Mazumder, Bhashkar. "Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data". *Review of Economics and Statistics* 87.2 (2005), pp. 235–255.

[115] Melly, Blaise. "Decomposition of differences in distribution using quantile regression". *Labour economics* 12.4 (2005), pp. 577–590.

[116] Melly, Blaise and Giulia Santangelo. "The changes-in-changes model with covariates". Working Paper. 2015.

[117] Montecino, Juan Antonio, Gerald Epstein, et al. "Did Quantitative Easing Increase Income Inequality?" *Institute for New Economic Thinking* (2015).

[118] Murtazashvili, Irina. "An alternative measure of intergenerational income mobility based on a random coefficient model". *Journal of Applied Econometrics* 27.7 (2012), pp. 1161–1173.

[119] Murtazashvili, Irina, Di Liu, and Artem Prokhorov. "Two-sample nonparametric estimation of intergenerational income mobility in the United States and Sweden". *Canadian Journal of Economics/Revue canadienne d'économique* 48.5 (2015), pp. 1733–1761.

[120] Ñopo, Hugo. "An extension of the Blinder–Oaxaca decomposition to a continuum of comparison groups". *Economics Letters* 100.2 (2008), pp. 292–296.

[121] Novy-Marx, Robert. "The other side of value: The gross profitability premium". *Journal of Financial Economics* 108.1 (2013), pp. 1–28.

[122] Oaxaca, Ronald. "Male-female wage differentials in urban labor markets". *International economic review* (1973), pp. 693–709.

[123] Pástor, L'uboš and Robert F Stambaugh. "Liquidity risk and expected stock returns". *Journal of Political economy* 111.3 (2003), pp. 642–685.

[124] Reinganum, Marc R. "A new empirical perspective on the CAPM". *Journal of financial and quantitative analysis* 16.4 (1981), pp. 439–462.

[125] Richey, Jeremiah and Alicia Rosburg. "Decomposing economic mobility transition matrices". Working Paper. 2015.

[126] Richey, Jeremiah and Alicia Rosburg. "Decomposing joint distributions via reweighting functions: An application to intergenerational economic mobility" (2016).

[127] Richey, Jeremiah and Alicia Rosburg. "Understanding intergenerational economic mobility by decomposing joint distributions". Working Paper. 2016.

[128] Richey, Jeremiah and Alicia Rosburg. "Changing roles of ability and education in US intergenerational mobility". *Economic Inquiry* 55.1 (2017), pp. 187–201.

[129] Robb, Alicia, Robert W Fairlie, and David Robinson. “Black & white: Racial differences in new firm financing”. *Academy of Management Proceedings*. Vol. 2014. 1. Academy of Management Briarcliff Manor, NY 10510. 2014, p. 14271.

[130] Rosenbaum, Paul and Donald Rubin. “The central role of the propensity score in observational studies for causal effects”. *Biometrika* 70.1 (1983), pp. 41–55.

[131] Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein. “Persuasive evidence of market inefficiency”. *The Journal of Portfolio Management* 11.3 (1985), pp. 9–16.

[132] Rothe, Christoph. “Nonparametric estimation of distributional policy effects”. *Journal of Econometrics* 155.1 (2010), pp. 56–70.

[133] Rothschild, Michael and Joseph E Stiglitz. “Increasing risk II: Its economic consequences”. *Journal of Economic theory* 3.1 (1971), pp. 66–84.

[134] Sharpe, William F. “Capital asset prices: A theory of market equilibrium under conditions of risk”. *The journal of finance* 19.3 (1964), pp. 425–442.

[135] Shin, Su Hyun and Sherman D Hanna. “Decomposition analyses of racial/ethnic differences in high return investment ownership after the Great Recession”. *Journal of Financial Counseling and Planning* 26.1 (2015), pp. 43–62.

[136] Solon, Gary. “Intergenerational income mobility in the United States”. *The American Economic Review* (1992), pp. 393–408.

[137] Solon, Gary. “Intergenerational mobility in the labor market”. *Handbook of labor economics* 3 (1999), pp. 1761–1800.

[138] Stephens, Michael A. “EDF statistics for goodness of fit and some comparisons”. *Journal of the American statistical Association* 69.347 (1974), pp. 730–737.

[139] Titman, Sheridan, KC John Wei, and Feixue Xie. “Capital investments and stock returns”. *Journal of financial and Quantitative Analysis* 39.4 (2004), pp. 677–700.

[140] Ulrick, Shawn W. “The Oaxaca decomposition generalized to a continuous group variable”. *Economics Letters* 115.1 (2012), pp. 35–37.

[141] Van Der Vaart, Aad W and Jon A Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, 1996.

[142] Wang, Cong and Sherman D Hanna. “Racial/ethnic disparities in stock ownership: A decomposition analysis”. *Consumer Interests Annual* 53 (2007).

[143] Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. MIT press, 2010.

[144] Wuthrich, Kaspar. “Semiparametric estimation of quantile treatment effects with endogeneity”. Working Paper. 2015.

[145] Zheng, John Xu. “A consistent test of functional form via nonparametric estimation techniques”. *Journal of Econometrics* 75.2 (1996), pp. 263–289.

[146] Zimmerman, David J. "Regression toward mediocrity in economic stature". *The American Economic Review* (1992), pp. 409–429.

Appendix A

DISTRIBUTIONAL EFFECTS OF A CONTINUOUS TREATMENT WITH AN APPLICATION ON INTERGENERATIONAL MOBILITY

A.1 Additional Assumptions

Assumption 1.

(i) $\mathcal{YTX}$, which denotes the Cartesian product of the supports of Y, T, and X, is a compact subset of $\mathbb{R}^{2+k}$ where k is the dimension of X.

(ii) Y is continuously distributed with conditional density $f_{Y|T,X}(y|t,x)$ uniformly bounded away from 0 and ∞ and continuous in $(y,t,x) \in \mathcal{YTX}$.

(iii) The support $\mathcal{T}$ of T is the compact interval $[t_{min}, t_{max}]$ with density $f_T(t)$ bounded away from 0 and ∞ on $\mathcal{T}$.

(iv) For $\mathcal{U} = [\epsilon, 1-\epsilon] \subset (0,1)$, $F_{Y|T}$ and $F^C_{Y|T}$ admit positive continuous densities $f_{Y|T}$ and $f^C_{Y|T}$ on an interval $[a,b]$ containing an ε-enlargement of the sets $\{Q_{Y|T}(\tau|t) : \tau \in \mathcal{U}\}$ and $\{Q^C_{Y|T}(\tau|t) : \tau \in \mathcal{U}\}$, respectively.

(v) $E\|P_1(T,X)\|^{2+\varepsilon} < \infty$ and $E\|P_2(T)\|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$.

(vi) Let $J_1(\tau) = E\left[f_{Y|T,X}(P_1(T,X)'\alpha(\tau)|T,X)P_1(T,X)P_1(T,X)'\right]$. Also, let

$J_2(\tau) = E\left[f_{Y|T}(P_2(T)'\beta(\tau)|T)P_2(T)P_2(T)'\right]$. The minimum eigenvalues of $J_1(\tau)$ and $J_2(\tau)$ are uniformly bounded away from zero.

A.2 Proofs

A.2.1 Proof Of Theorem 1.1

Let

$$\hat{G}_{Y|T,X}(y|t,x) = \sqrt{n}(\hat{F}_{Y|T,X}(y|t,x) - F_{Y|T,X}(y|t,x)) \quad \text{and} \quad \hat{G}_{Y|T}(y|t) = \sqrt{n}(\hat{F}_{Y|T}(y|t) - F_{Y|T}(y|t))$$

and let

$$\hat{G}_X(x) = \sqrt{n}(\hat{F}_X(x) - F_X(x))$$

which are the empirical processes of the conditional distribution of the outcome and the distribution of other observable characteristics.

Also, let $\xi = (y, t, x, \bar{y}, \bar{t}, \bar{x})$ and $W = (Y, T, X)$ and let

$$\psi_1(W, \xi) = f_{Y|T,X}(y|t,x)P_1(t,x)'J_1(F_{Y|T,X}(y|t,x))^{-1}H_1(Y,T,X,F_{Y|T,X}(y|t,x))$$

where

$$H_1(Y,T,X,\tau) = (\mathbb{1}\{Y \leq P_1(T,X)'\alpha(\tau)\} - \tau))\, P_1(T,X)$$

Next, let

$$\psi_2(W, \xi) = f_{Y|T}(\bar{y}|\bar{t})P_2(\bar{t})'J_2(F_{Y|T}(\bar{y}|\bar{t}))^{-1}H_2(Y,T,F_{Y|T}(y|t))$$

where

$$H_2(Y,T,\tau) = \left(\mathbb{1}\{Y \leq P_2(T)'\beta(\tau)\} - \tau)\right) P_2(T)$$

and let

$$\psi_3(W,\xi) = \mathbb{1}\{X \leq \bar{x}\} - F_X(\bar{x})$$

The first result establishes the joint limiting distribution of $\hat{G}_{Y|T}$, $\hat{G}_{Y|T,X}$, and $\hat{G}_X$.

Lemma 1. *Let $\mathbb{S} = l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{YT}) \times l^\infty(\mathcal{X})$. Under Assumptions 2 to 4 and 6 and Assumption 1,*

$$(\hat{G}_{Y|T,X}(y|t,x), \hat{G}_{Y|T}(\bar{y}|\bar{t}), \hat{G}_X(x)) \rightsquigarrow (\mathbb{G}_{Y|T,X}, \mathbb{G}_{Y|T}, \mathbb{G}_X)$$

in the space $\mathbb{S}$ and where $(\mathbb{G}_{Y|T,X}, \mathbb{G}_X)$ is a tight Gaussian process with mean 0 with covariance function $V(\xi_1,\xi_2)$ defined on $\mathbb{S}$ and given by

$$V(\xi_1,\xi_2) = E\left[\psi(W,\xi_1)\psi(W,\xi_2)'\right]$$

where $\psi(W,\xi) = (\psi_1(W,\xi), \psi_2(W,\xi), \psi_3(W,\xi))'$

Proof. The result follows immediately under Assumptions 6 and 1 and from the results in Chernozhukov, Fernandez-Val, and Melly (2013). □

Before proving the main result, we consider the following result first.

Lemma 2. *Consider the map $\psi : \mathbb{D}_\psi \subset \mathbb{D} = l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{X}) \mapsto l^\infty(\mathcal{YT})$ given by*

$$\psi(\Lambda) = \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot,x)\ \mathrm{d}\Lambda_2(x)$$

for $\Lambda = (\Lambda_1, \Lambda_2) \in \mathbb{D}$*. Then, under Assumptions 2 to 4 and 6 and Assumption 1, the map* ψ *is Hadamard differentiable at* Λ_0 *tangentially to* $\mathbb{D}$ *with derivative at* Λ_0 *in* $\lambda = (\lambda_1, \lambda_2) \in \mathbb{D}$ *given by*

$$\psi'_{\Lambda_0}(\lambda) = \int_{\mathcal{X}} \lambda_1(\cdot|\cdot, x)\ \mathrm{d}\Lambda_{20}(x) + \int_{\mathcal{X}} \Lambda_{10}(\cdot|\cdot, x)\ \mathrm{d}\lambda_2(x)$$

Proof. Consider any sequence $t_k > 0$ and $\Lambda_k \in \mathbb{D}$ for $k = 1, 2, 3, \ldots$ with $t_k \downarrow 0$ and

$$\lambda_{1k} = \frac{\Lambda_{1k} - \Lambda_1}{t_k}$$
$$\lambda_{2k} = \frac{\Lambda_{2k} - \Lambda_2}{t_k}$$

with $(\lambda_{1k}, \lambda_{2k}) \to (\lambda_1, \lambda_2) \in \mathbb{D}$ as $k \to \infty$.

Then,

$$\begin{aligned}
\frac{\psi(\Lambda_k) - \psi(\Lambda)}{t_k} - \psi'_{\Lambda}(\lambda) &= \int_{\mathcal{X}} \Lambda_{1k}(\cdot|\cdot, x)\ \mathrm{d}\Lambda_{2k}(x)/t_k - \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot, x)\ \mathrm{d}\Lambda_2(x)/t_k \\
&\quad - \int_{\mathcal{X}} \lambda_1(\cdot|\cdot, x)\ \mathrm{d}\Lambda_2(x) - \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot, x)\ \mathrm{d}\lambda_2(x) \\
&= \int_{\mathcal{X}} \frac{\Lambda_{1k}(\cdot|\cdot, x) - \Lambda_1(\cdot|\cdot, x)}{t_k}\ \mathrm{d}(\Lambda_{2k}(x) - \Lambda_2(x)) \\
&\quad + \int_{\mathcal{X}} \frac{\Lambda_{1k}(\cdot|\cdot, x) - \Lambda_1(\cdot|\cdot, x)}{t_k}\ \mathrm{d}\Lambda_2(x) \\
&\quad + \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot, x)\ \mathrm{d}(\Lambda_{2k}(x) - \Lambda_2(x))/t_k \\
&\quad - \int_{\mathcal{X}} \lambda_1(\cdot|\cdot, x)\ \mathrm{d}\Lambda_2(x) - \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot, x)\ \mathrm{d}\lambda_2(x) \\
&= t_k \int_{\mathcal{X}} \lambda_{1k}(\cdot|\cdot, x)\ \mathrm{d}\lambda_{2k}(x) \\
&\quad + \int_{\mathcal{X}} (\lambda_{1k}(\cdot|\cdot, x) - \lambda_1(\cdot|\cdot, x)\ \mathrm{d}\Lambda_2(x) \\
&\quad + \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot, x)\ \mathrm{d}(\lambda_{2k} - \lambda_2)(x) \\
&\to 0 \text{ as } k \to \infty
\end{aligned}$$

where, in the last equation, the first line is $O(t_k)$ which converges to 0 as $k \to \infty$, and the second and third terms converge to 0 because $(\lambda_{1k}, \lambda_{2k}) \to (\lambda_1, \lambda_2)$. $\square$

Lemma 3. *Consider the map* $\phi : \mathbb{D}_\phi \subset l^\infty(\mathcal{YT}) \times l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{X}) \mapsto l^\infty(\mathcal{YT})^2$ *given by*

$$\phi(\Gamma) = (\Gamma_1, \psi(\Gamma_2, \Gamma_3))$$

in $\Gamma = (\Gamma_1, \Gamma_2, \Gamma_3) \in l^\infty(\mathcal{YT}) \times l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{X})$ *and the map* $\psi : \mathbb{D}_\psi \subset l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{X}) \mapsto l^\infty(\mathcal{YT})$ *is given in Lemma 2. Then, under Assumptions 2 to 4 and 6 and Assumption 1, the map* ϕ *is Hadamard differentiable at* Γ_0 *tangentially to* $l^\infty(\mathcal{YT}) \times l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{X})$ *with derivative at* Γ_0 *in* $\gamma = (\gamma_1, \gamma_2, \gamma_3) \in \mathbb{D}$ *given by*

$$\begin{aligned}\phi'_{\Gamma_0}(\gamma) &= (\gamma_1, \psi'_{(\Gamma_{20}, \Gamma_{30})}(\gamma_2, \gamma_3)) \\ &= \left(\gamma_1, \int_{\mathcal{X}} \gamma_2(\cdot|\cdot, x)\ \mathrm{d}\Gamma_{30}(x) + \int_{\mathcal{X}} \Gamma_{20}(\cdot|\cdot, x)\ \mathrm{d}\gamma_3(x)\right)\end{aligned}$$

Proof. The result follows immediately from Lemma 2. $\square$

Proof of Theorem 1.1

Lemma 3 implies

$$(\hat{G}_{Y|T}(y|t), \hat{G}^C_{Y|T}(\bar{y}|\bar{t})) \rightsquigarrow (\mathbb{G}_{Y|T}, \mathbb{G}^C_{Y|T})$$

indexed by $(y, t, \bar{y}, \bar{t})$ in $\mathbb{S} = l^\infty(\mathcal{YT})^2$ and where $\mathbb{G}_{Y|T}$ is given in Lemma 1 and

$$\mathbb{G}^C_{Y|T} = \int_{\mathcal{X}} \mathbb{G}_{Y|T,X}(\cdot|\cdot, x)\ \mathrm{d}F_X(x) + \int_{\mathcal{X}} F_{Y|T,X}(\cdot|\cdot, x)\ \mathrm{d}\mathbb{G}_X(x)$$

($\mathbb{G}_{Y|T,X}$ and $\mathbb{G}_X$ are given in Lemma 1). Then, the process given in Theorem 1.1 is given by setting $\bar{y} = y$ and $\bar{t} = t$.

A.3 Tables And Figures

Notes: Summary statistics for the main dataset used in the paper. Each column provides average values of available variables by parents' income quartile. Standard errors are given in parentheses beneath the average. The row "Cutoff" is the maximum value of parents' income in that quartile (i.e. the dividing line between parents' income across two columns).

Sources: Panel Study of Income Dynamics, as described in text

Table A.1: Summary Statistics

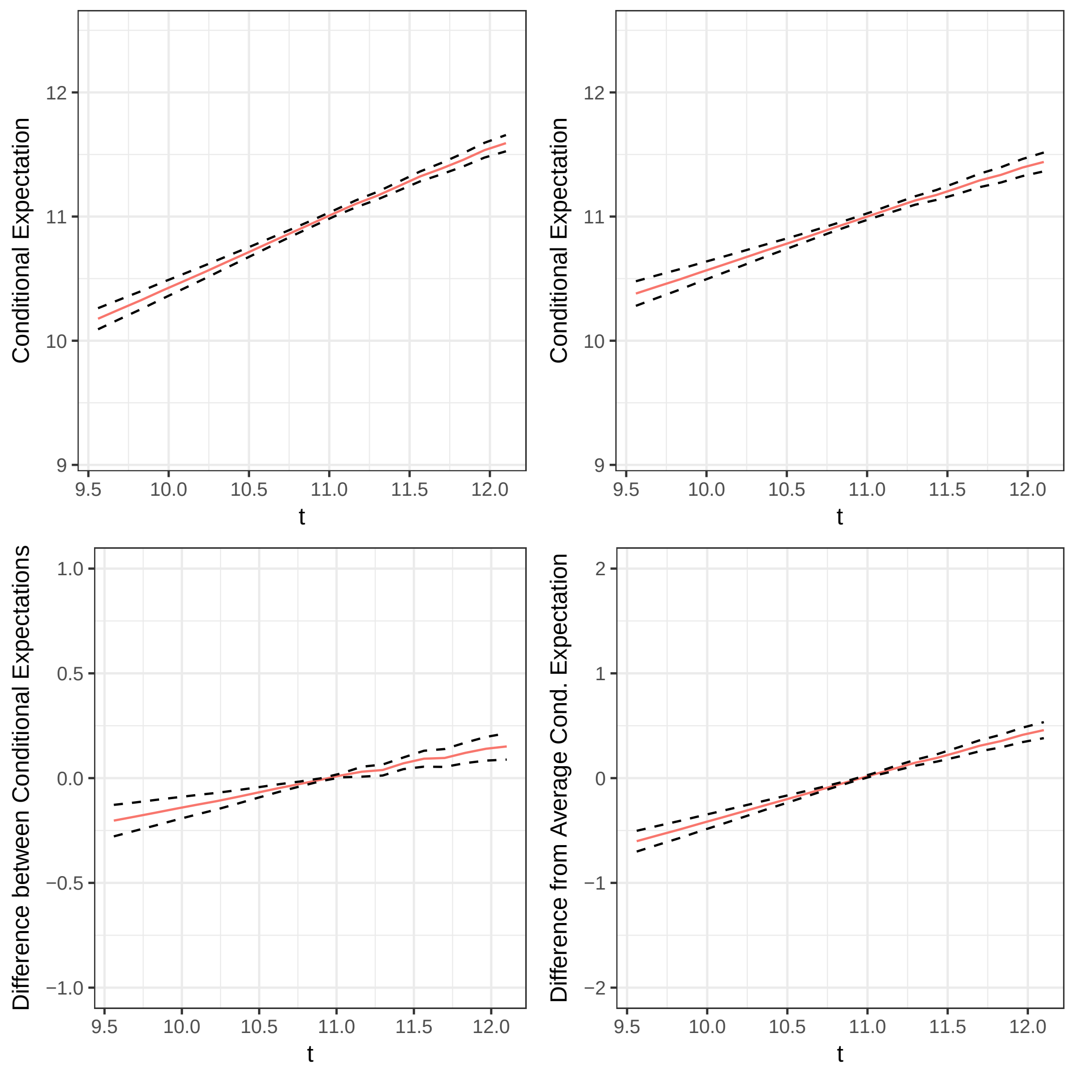
	Q1	Q2	Q3	Q4	All
Parents' Income (1000s)	32.53	51.47	67.97	107.49	64.87
	(0.291)	(0.144)	(0.183)	(1.289)	(0.568)
Child's Income (1000s)	45.96	64.02	74.8	96.8	70.4
	(0.935)	(1.518)	(1.491)	(2.393)	(0.888)
Head White	0.72	0.93	0.95	0.96	0.89
	(0.015)	(0.008)	(0.007)	(0.006)	(0.005)
Head Non-White	0.28	0.07	0.05	0.04	0.11
	(0.015)	(0.008)	(0.007)	(0.006)	(0.005)
Child Male	0.49	0.47	0.48	0.5	0.48
	(0.017)	(0.017)	(0.017)	(0.017)	(0.008)
Head Male	0.77	0.95	0.97	0.99	0.92
	(0.014)	(0.007)	(0.006)	(0.003)	(0.004)
Year Born	1970.69	1970.29	1970.64	1969.32	1970.23
	(0.324)	(0.309)	(0.32)	(0.345)	(0.163)
Head Veteran	0.26	0.37	0.47	0.47	0.39
	(0.014)	(0.016)	(0.017)	(0.017)	(0.008)
Head Less than HS	0.35	0.19	0.08	0.05	0.17
	(0.016)	(0.013)	(0.009)	(0.007)	(0.006)
Head HS	0.58	0.64	0.58	0.39	0.55
	(0.016)	(0.016)	(0.016)	(0.016)	(0.008)
Head College	0.07	0.17	0.34	0.57	0.29
	(0.008)	(0.013)	(0.016)	(0.016)	(0.007)
Cutoff	44.24	59.17	78.01	434.44	
N	908	907	907	908	3630

Table A.2: Intergenerational Elasticity (IGE) Estimates

	Dependent variable:			
	Log Child's Income			
	(1)	(2)	(3)	(4)
Log Parents' Income	0.609***	0.573***	0.559***	0.452***
	(0.023)	(0.024)	(0.024)	(0.027)
Head Non-White		−0.245***	−0.253***	−0.236***
		(0.040)	(0.040)	(0.039)
Male		0.025	0.022	0.023
		(0.019)	(0.019)	(0.019)
Head Male		−0.076*	−0.056	−0.019
		(0.043)	(0.043)	(0.043)
Year Born			−0.009***	−0.013***
			(0.001)	(0.001)
Head Veteran				−0.006
				(0.021)
Head Less Than HS Educ.				−0.225***
				(0.030)
Head College Educ.				0.104***
				(0.024)
Constant	4.282***	4.761***	22.282***	31.110***
	(0.255)	(0.261)	(1.984)	(2.211)

Notes: Results come from regressions of the log of child's income on the log of parents' income and additional controls using the full sample of 3,630 observations. Standard errors are heteroskedasticity robust.
Sources: Panel Study of Income Dynamics, as described in text

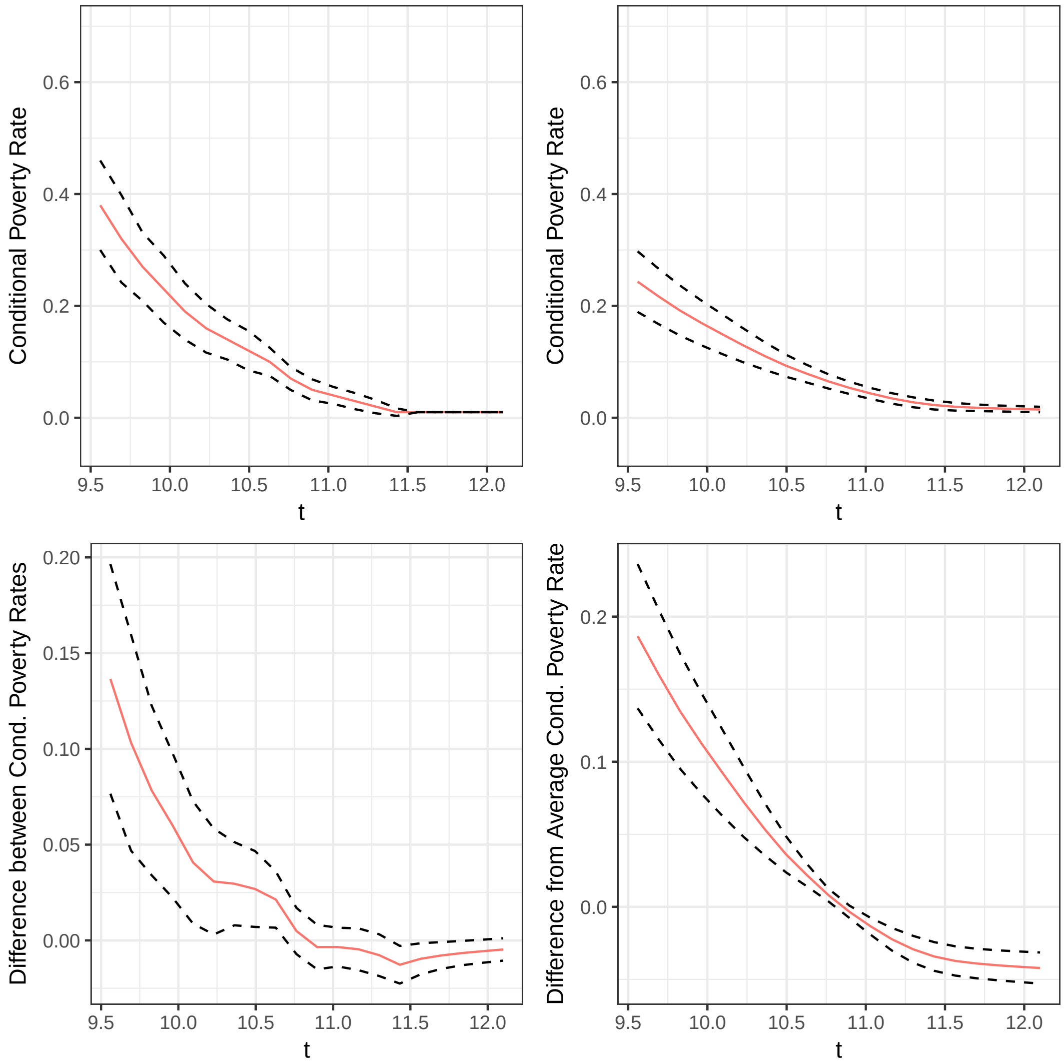
Figure A.1: Expected child's income conditional on parents' income



Notes: The top left panel plots average child's income as a function of parents' income with no adjustments for other covariates. The top right panel adjusts for differences in the covariates family head's race, family head's gender, gender of child, child's birth year, family head's veteran status, and family head's education (dummy variables for less than high school degree, high school degree but less than college degree, and college degree or more). The bottom left panel plots the difference between the estimates that do not adjust for covariates and that do adjust for covariates (i.e. the difference between the top left and top right panels as a function of parents' income). The bottom right panel plots the difference between the results that adjust for covariates and the average over t of the same results, as discussed in the text. In each panel, the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations as described in the text.

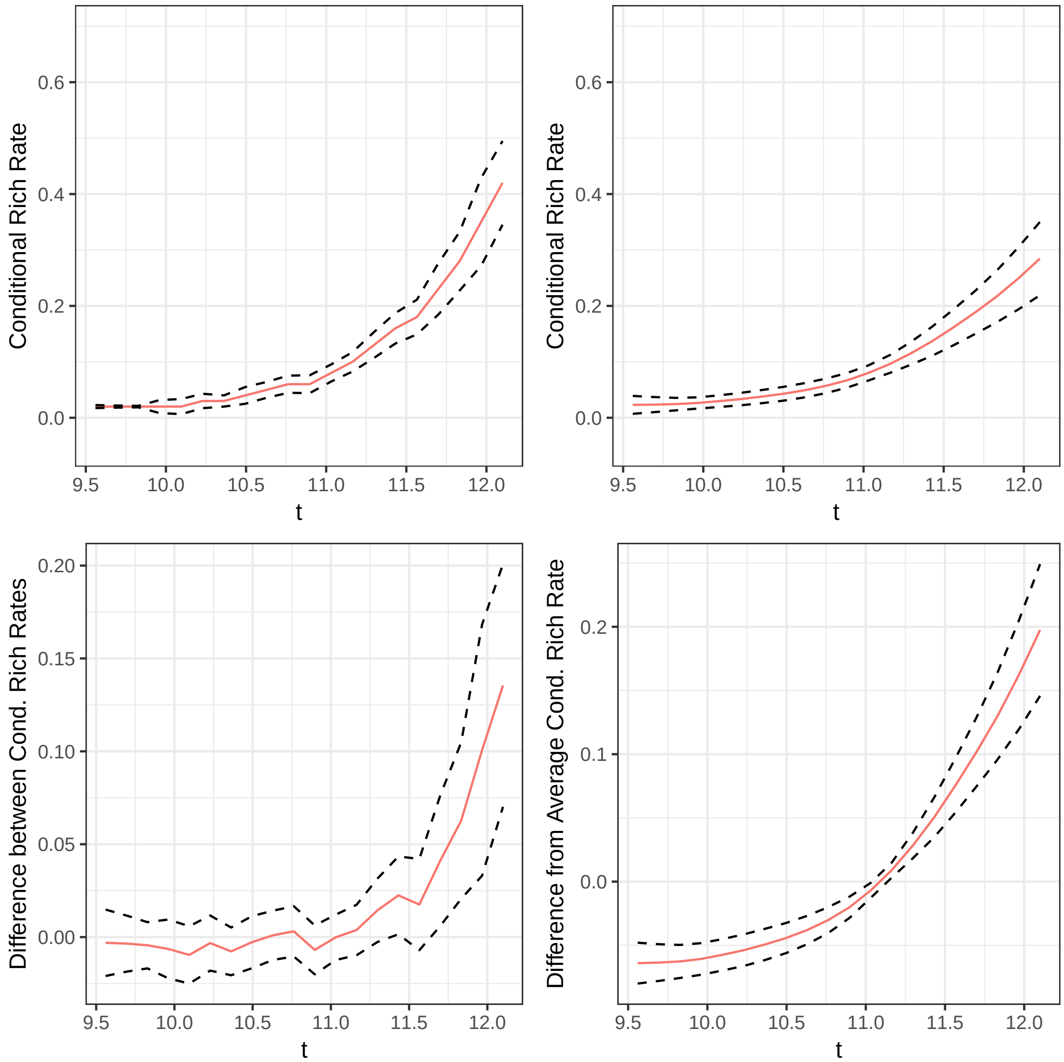
Sources: Panel Study of Income Dynamics, as described in text

Figure A.2: Fraction of children below the poverty line



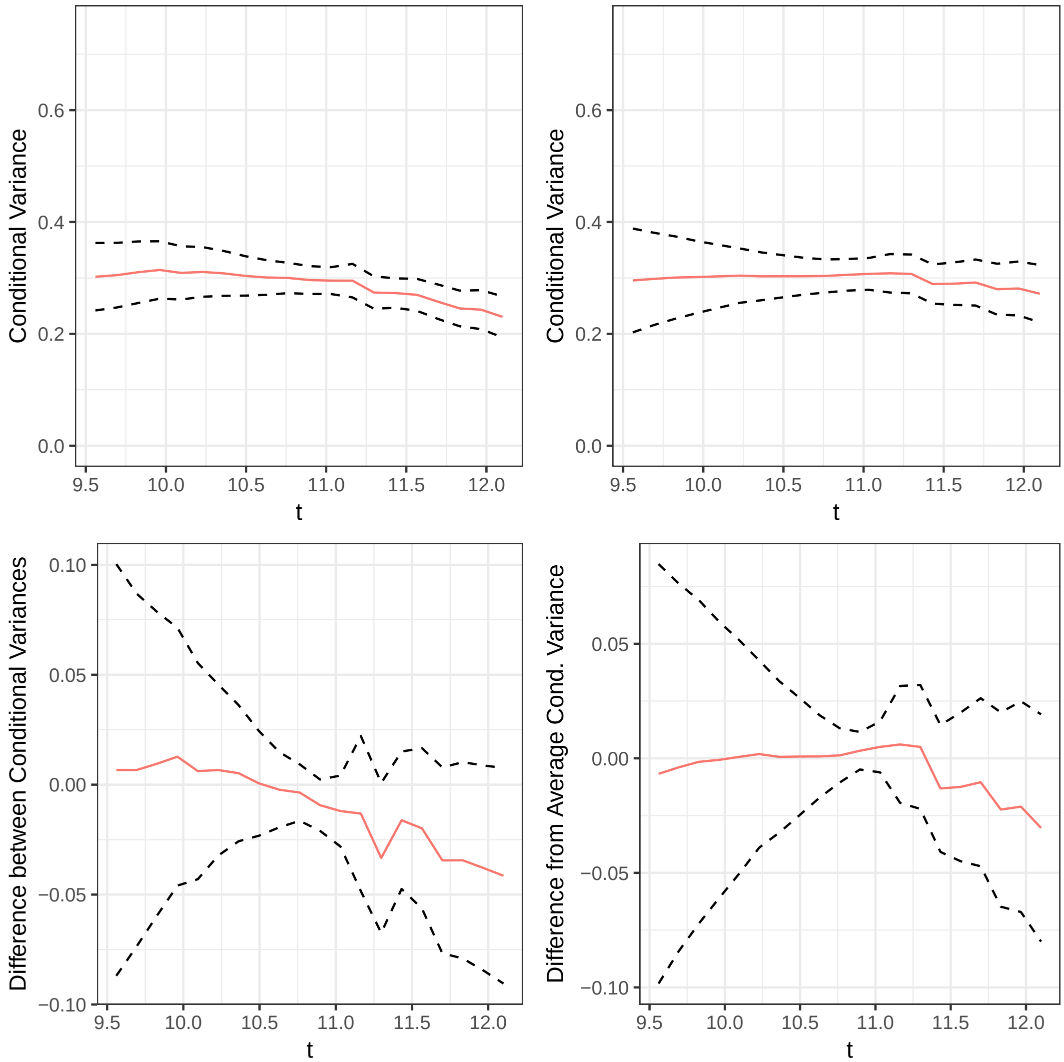
Notes: The top left panel plots the fraction of children below the poverty line as a function of parents' income with no adjustments for other covariates. The poverty line is set to be $22,113 which is the poverty line for a family with two adults and two children in 2010. The top right panel adjusts for differences in the covariates family head's race, family head's gender, gender of child, child's birth year, family head's veteran status, and family head's education (dummy variables for less than high school degree, high school degree but less than college degree, and college degree or more). The bottom left panel plots the difference between the estimates that do not adjust for covariates and that do adjust for covariates (i.e. the difference between the top left and top right panels as a function of parents' income). The bottom right panel plots the difference between the results that adjust for covariates and the average over t of the same results, as discussed in the text. In each panel, the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations as described in the text.
Sources: Panel Study of Income Dynamics, as described in text

Figure A.3: Fraction of "rich" children



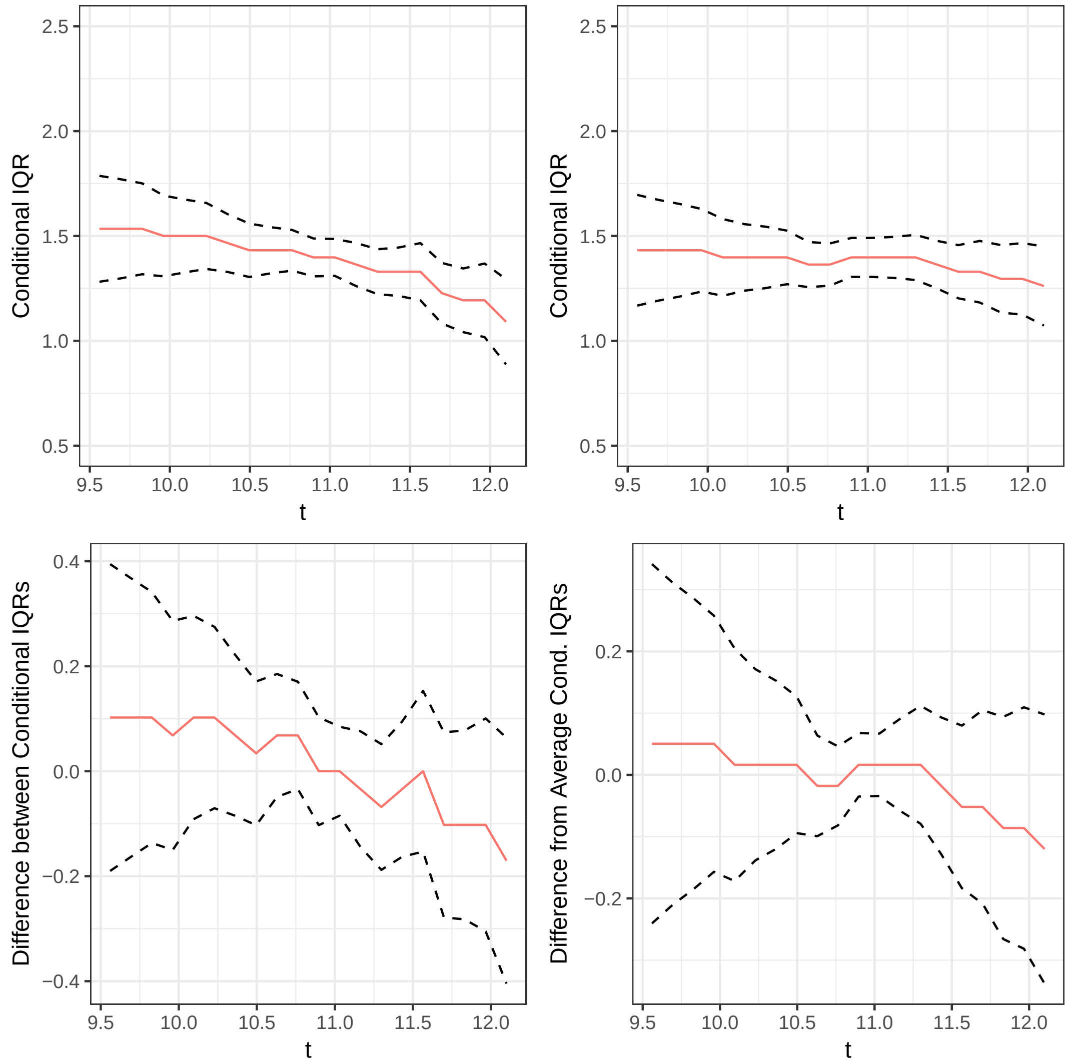
Notes: The top left panel plots the fraction of "rich" children as a function of parents' income with no adjustments for other covariates where "rich" is defined as having income above the 90th percentile of income in the U.S. in 2010 which is $132,923. The top right panel adjusts for differences in the covariates family head's race, family head's gender, gender of child, child's birth year, family head's veteran status, and family head's education (dummy variables for less than high school degree, high school degree but less than college degree, and college degree or more). The bottom left panel plots the difference between the estimates that do not adjust for covariates and that do adjust for covariates (i.e. the difference between the top left and top right panels as a function of parents' income). The bottom right panel plots the difference between the results that adjust for covariates and the average over t of the same results, as discussed in the text. In each panel, the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations as described in the text.
Sources: Panel Study of Income Dynamics, as described in text

Figure A.4: Variance of child's income conditional on parents' income



Notes: The top left panel plots the variance of child's income as a function of parents' income with no adjustments for other covariates. The top right panel adjusts for differences in the covariates family head's race, family head's gender, gender of child, child's birth year, family head's veteran status, and family head's education (dummy variables for less than high school degree, high school degree but less than college degree, and college degree or more). The bottom left panel plots the difference between the estimates that do not adjust for covariates and that do adjust for covariates (i.e. the difference between the top left and top right panels as a function of parents' income). The bottom right panel plots the difference between the results that adjust for covariates and the average over t of the same results, as discussed in the text. In each panel, the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations as described in the text.
Sources: Panel Study of Income Dynamics, as described in text

Figure A.5: Inter-quantile ranges



Notes: The top left panel plots the inter-quantile Range as a function of parents' income with no adjustments for other covariates for $\tau_1 = 0.9$ and $\tau_2 = 0.1$ (these are the values of τ_1 and τ_2 used in each panel). The top right panel adjusts for differences in the covariates family head's race, family head's gender, gender of child, child's birth year, family head's veteran status, and family head's education (dummy variables for less than high school degree, high school degree but less than college degree, and college degree or more). The bottom left panel plots the difference between the estimates that do not adjust for covariates and that do adjust for covariates (i.e. the difference between the top left and top right panels as a function of parents' income). The bottom right panel plots the difference between the results that adjust for covariates and the average over t of the same results, as discussed in the text. In each panel, the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations as described in the text.
Sources: Panel Study of Income Dynamics, as described in text

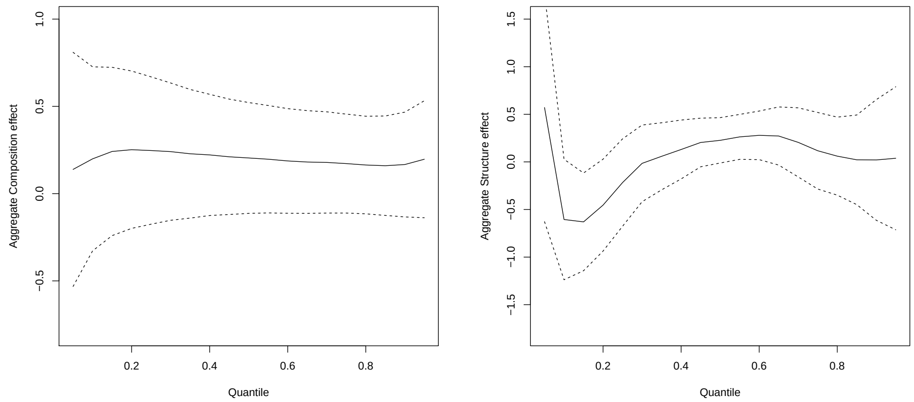
Appendix B

DECOMPOSING DIFFERENCES IN QUANTILE PORTFOLIO RETURNS BETWEEN NORTH AMERICA AND EUROPE USING RECENTERED INFLUENCE FUNCTION REGRESSION

B.1 Results With Confidence Intervals

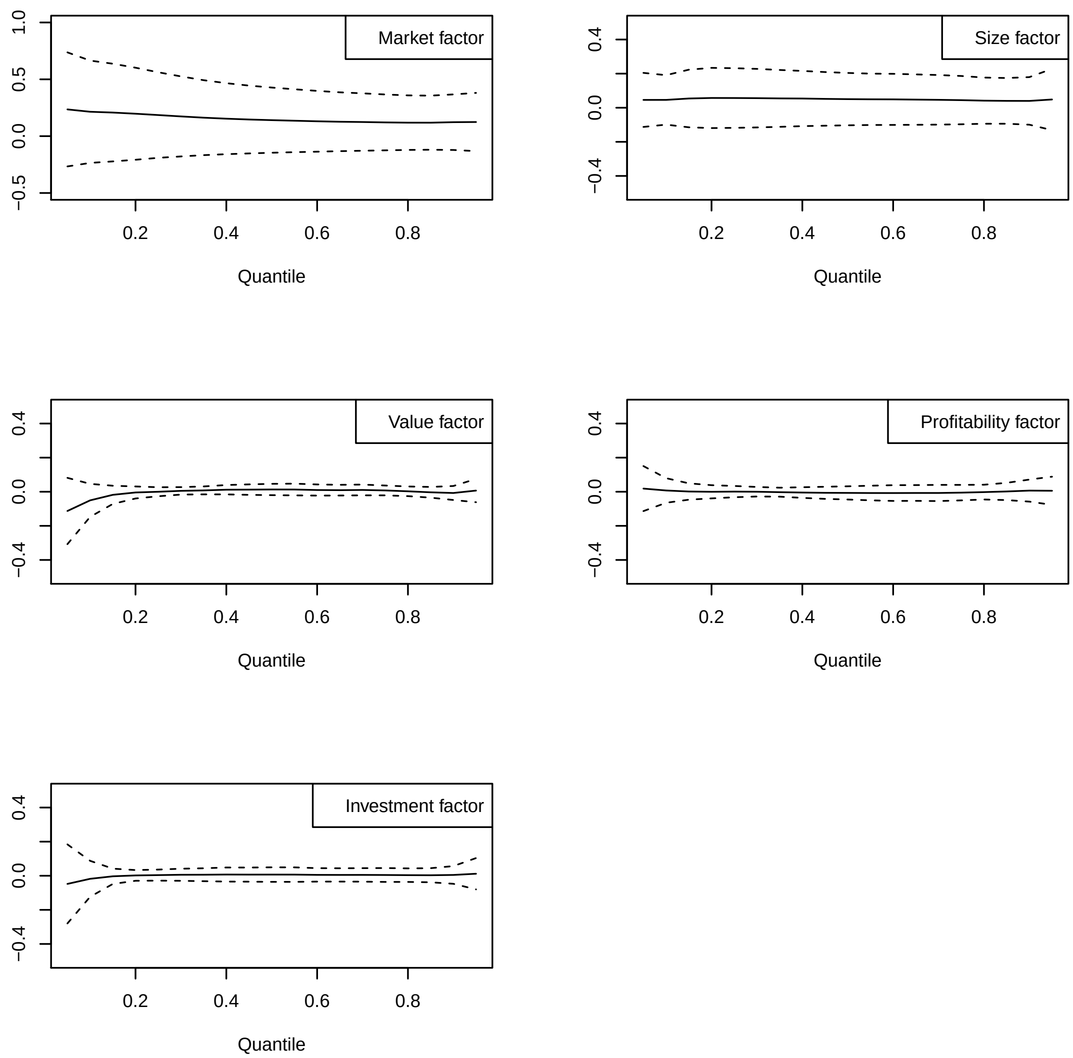
Figure B.1, B.2 and B.3 plot the decomposition results across quantiles and with 95% confidence intervals. It is obvious to observe that all are decomposition results are not statistically significant.

Figure B.1: Aggregate decomposition effects



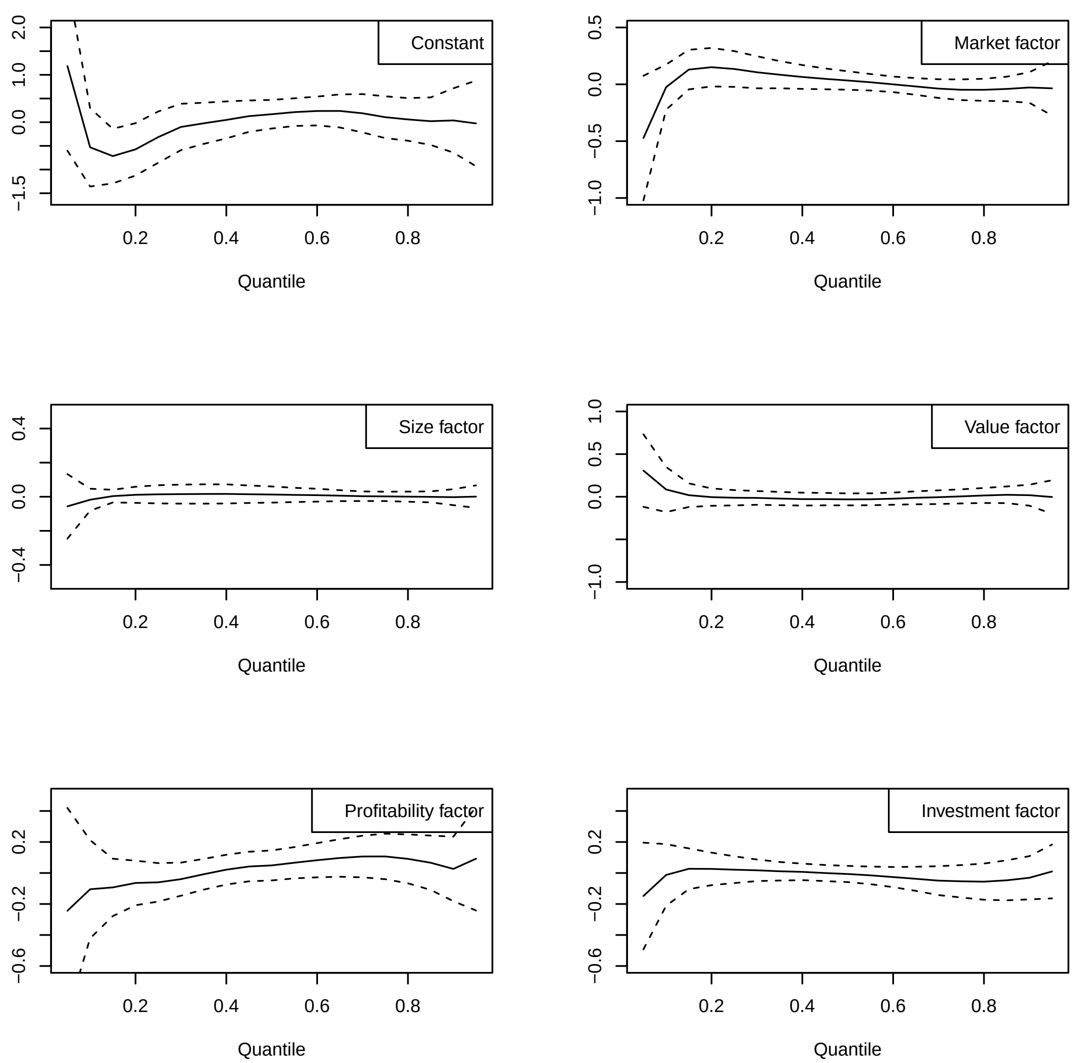
Notes: The figure plots the aggregate decomposition results across quantiles and with 95% confidence intervals. The left panel plots the aggregate composition effects across quantiles with 95% confidence intervals. The right panel plots the aggregate structure effects across quantiles with 95% confidence intervals.

Figure B.2: Detailed composition effects



Notes: The figure plots the detailed composition effects across quantiles and with 95% confidence intervals. Each panel plots a detailed decomposition effect related to the factor denoted in the topright of the panel.

Figure B.3: Detailed structure effects

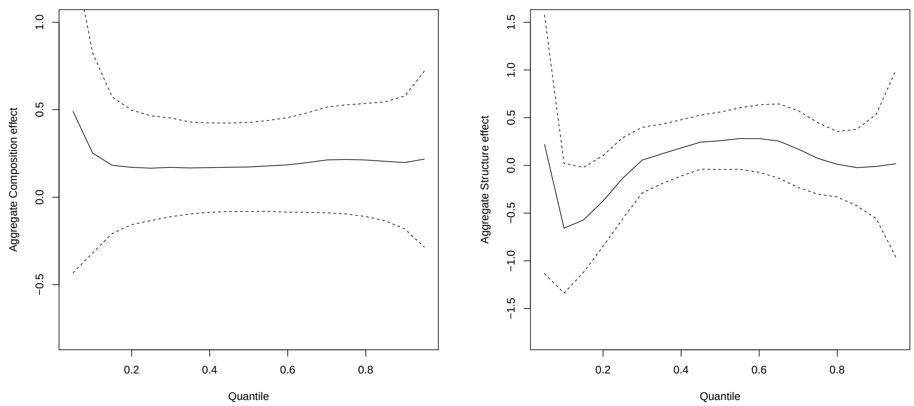


Notes: The figure plots the detailed structure effects across quantiles and with 95% confidence intervals. Each panel plots a detailed structure effect related to the factor denoted in the topright of the panel.

B.2 Robustness Results With Confidence intervals

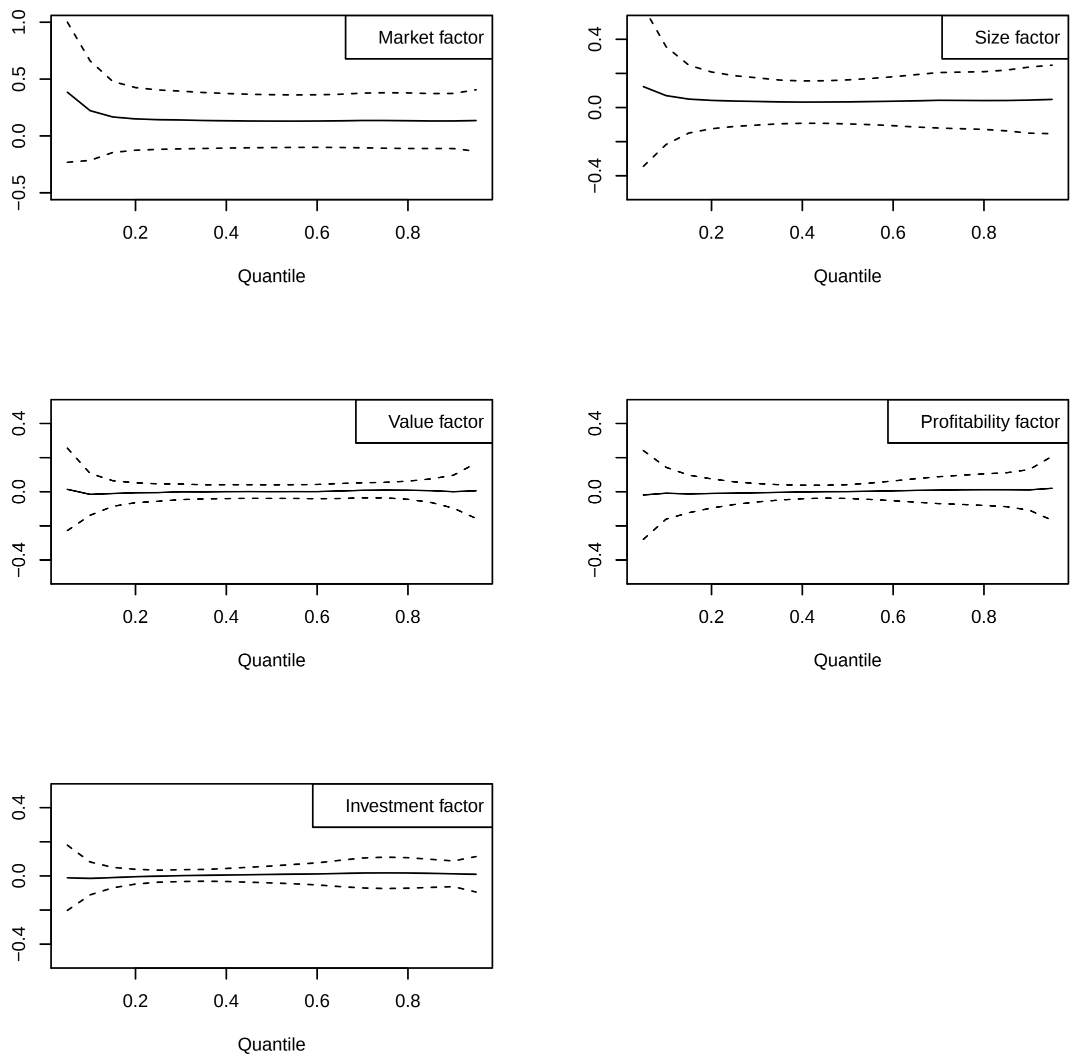
Figure B.4 to B.6 show the results of the robustness checks with 95% confidence intervals. Obviously, all the results are statistically nonsignificant too.

Figure B.4: Aggregate decomposition effects



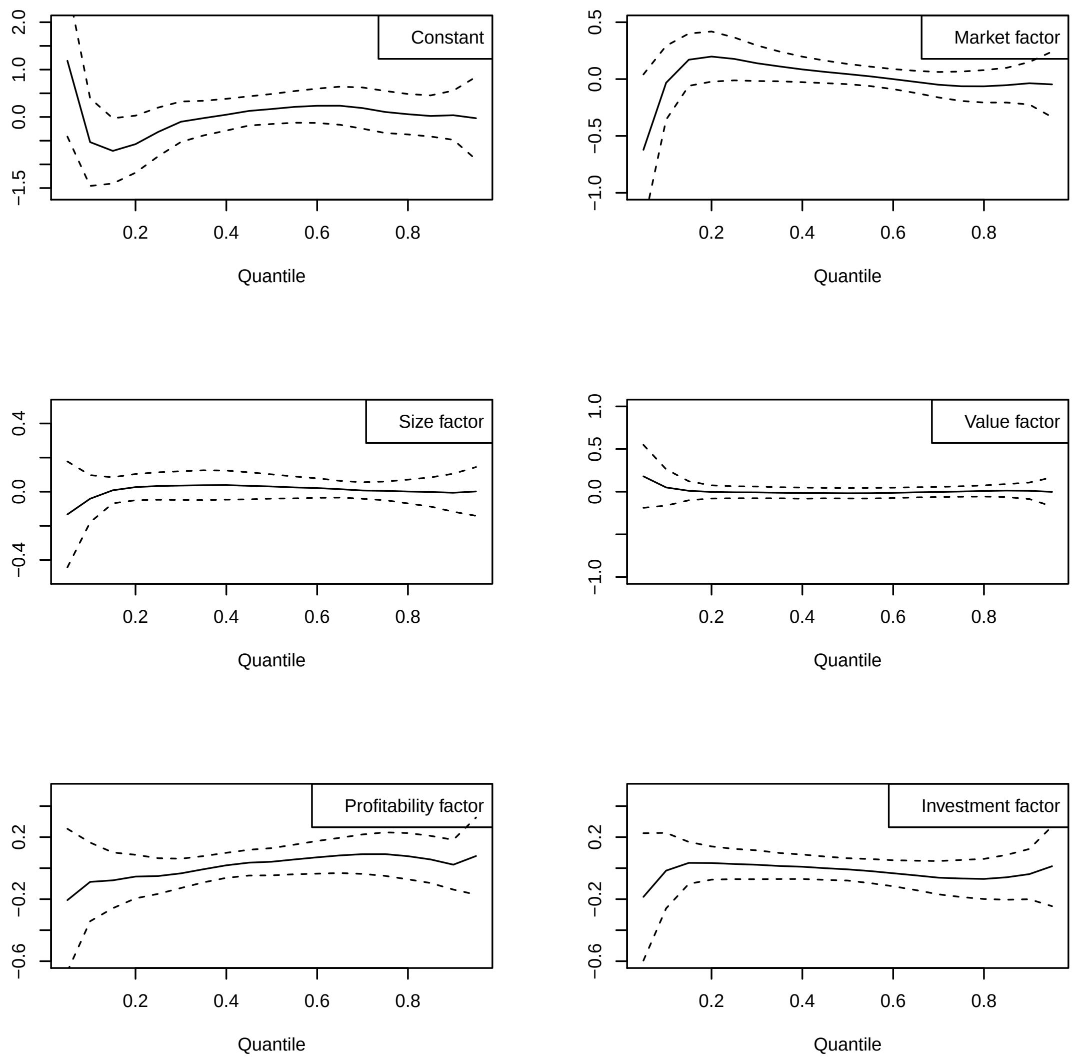
Notes: The figure plots the aggregate decomposition results across quantiles and with 95% confidence intervals. The left panel plots the aggregate composition effects across quantiles with 95% confidence intervals. The right panel plots the aggregate structure effects across quantiles with 95% confidence intervals.

Figure B.5: Detailed composition effects



Notes: The figure plots the detailed composition effects across quantiles and with 95% confidence intervals. Each panel plots a detailed decomposition effect related to the factor denoted in the topright of the panel.

Figure B.6: Detailed structure effects



Notes: The figure plots the detailed structure effects across quantiles and with 95% confidence intervals. Each panel plots a detailed structure effect related to the factor denoted in the topright of the panel.