Stefan Holst Bache
PhD Dissertation

CREATES
Center for Research in Econometric Analysis of Time Series

Quantile regression: Three econometric studies

Department of Economics and Business
Aarhus University • Denmark

Quantile regression: three econometric studies

By Stefan Holst Bache

A dissertation submitted to
Business and Social Sciences, Aarhus University,
in partial fulfilment of the requirements of
the PhD degree in
Economics and Management

CREATES
Center for Research in Econometric Analysis of Time Series

Contents

"If you want to know the end, look at the beginning"

- African proverb

Preface

This PhD dissertation is the written product of the work I have done during my 3-year PhD programme at Aarhus University in the period 2008–2011. It reflects some, but certainly not all, of the activities I have undertaken in my pursuit of the PhD title. Here, in the last section written (yet the first to be read), I will ponder briefly on some of these other activities, which have definitely brought me the completeness and diversity I believe a PhD programme should offer.

The programme offered by the School of Economics and Management[1] at Aarhus University (in my case, jointly with CREATES) has been rich and has provided me with wide-ranging professional and social experiences. First of all, there has been an extensive amount of coursework. Many of these courses have been specialized short-courses, where the topics and lecturers have been chosen in collaboration with the students. Here, I enjoyed co-arranging a course on quantile regression given by Roger Koenker, and one on empirical processes given by Jørgen Hoffmann-Jørgensen. Being able to invite such high-profiled researchers to give courses shows that quality is a high priority at the center and at the department.

Then there is teaching. On my first day at Aarhus University, I was called to a meeting where I was told that I was to be a teaching assistant (TA) in first-year mathematics and statistics. A bit of a "scoop" really, considering my research interests. These TA positions were followed by one in regression analysis, and then another in econometrics. I also designed and lectured a two-week intensive course in programming for econometrics with Johannes Tang Kristensen, a fellow PhD student. The importance of teaching as a part of the programme, I believe, cannot be understated: interacting with students, correcting assignments, identifying good and bad work, and not least—being required to know the material inside and out, are all very essential components. I guess my point here is this: not all PhD students I know got the chance to teach material relevant to their research; and even though I found some of the teaching obligations rather hard work, in retrospect I think that I got a pretty good deal.

During the three years I have attended a lot of seminars and conferences, and of course the customary dinners; many of them outside Denmark. Such events have given me the opportunity to meet and mingle with many people in my own and related fields, and also to present my work and get feedback. This has not only been

[1] After the recent merger with Aarhus School of Business, the department's new name is Department of Economics and Business.

very pleasant socially, but also professionally rewarding as such events build network and serve as a forum for discussing and sharing ideas. One such event, which deserves special mention, is the Econometric Game, held every year in Amsterdam. 25 universities (at least the two years I competed) from around the world meet, with teams of 5 students, for a 3-day intensive competition in applied econometrics. In 2010, our team managed to get Aarhus University placed on the podium: 3rd place.

Finally, I would like to mention my research stay at University of Illinois at Urbana-Champaign (UIUC). In a Danish PhD programme, it is customary to arrange a stay of one semester or more at a foreign university. I was fortunate that I could arrange to visit Professor Roger Koenker at UIUC in the period September–December 2010. This was the perfect change of research environment for me, because many people here work with quantile based methods and because of the very pleasant and interesting group of people I had contact with, not least Professor Roger Koenker himself.

As will be clear from the above, in the course of my PhD studies many people have contributed to—and have had great influence on—my programme. Needless to say, I am thankful to all these people. Further, I owe a special debt of gratitude to Professor Christian M. Dahl, my primary advisor and the person who got me interested in econometrics in the first place (when lecturing Financial Econometrics at University of Southern Denmark). In the best possible way he has kept me busy and always on the right track. I would also like to thank my secondary advisor, Professor Niels Haldrup, who is also the director of CREATES. He has put a lot of energy and effort into the PhD programme, with great success, and I feel lucky to have been a part of this research center. My dissertation is far from typical for CREATES (as there are not a great many t subscripts to be found in my work), but I have really enjoyed all the time series talks and events. I also extend my gratitude to Professor Roger Koenker, who made my stay at UIUC both pleasant and productive, and who continues to serve as a great inspiration through his work. In my efforts to finish my third paper I ran into computing capacity problems as the particular server at Statistics Denmark could not offer the required resources. Henning Bunzel saved the day (in fact much more than a day) by arranging access to the same data on one of his servers. For this I am also thankful. Finally, I would like to thank my fellow PhD students (there are quite a few of them) without whom the many hours spent on the university premises would have been less inspiring and fun; my family, who is always there for me provide me with loads of love and wisdom (and linguistic support). I dedicate this dissertation to Marianne, Malthe, and our yet-to-be-born Philipa, who had to bear with me when caught up in work. I know it has not always been easy. Thank you.

August, 2011.
Stefan Holst Bache.

Updated preface

The pre-defense took place on November 2nd, 2011 at Aarhus University. I am grateful to the members of the assessment committee, Bent Jesper Christensen (chairman, Aarhus University), Søren Leth-Petersen (Copenhagen University), and Shakeeb Khan (Duke University), for carefully reading my dissertation and for taking the time and energy to give many useful and constructive comments and suggestions. Their input has most definitely added value to this revised version of the dissertation and given food for thought.

November, 2011.
Stefan Holst Bache.

A note on structure

The usual structure of a dissertation, such as this one, is a collection of papers in a "submission-ready" format. Sometimes these papers have a clear-cut relation to one-another, and not much is needed said between chapters. Sometimes less so. The common topic here is "quantile regression," but other than that, the papers differ quite a bit, one being theoretical in nature, the two others empirical with very different applications. With the hope of bringing each paper into context, I provide a short preface for each paper.

For better typographical exposition, the papers have been recompiled into the format of the dissertation (B5). They may therefore appear slightly different than in their working paper format. The contents, however, have no significant changes. Since two of the papers are joint work, there may be variations in writing style and notation. I hope that I still succeed in obtaining some kind of unity.

English summary

This dissertation contains three papers with "quantile regression" as the common denominator. This particular regression technique emerged as a generalization of the least-absolute-deviations technique in 1978 when *Econometrica* published the seminal paper "Regression Quantiles" by Roger Koenker and Gilbert Bassett. Since then, its use and development have become widespread topics in both applied and theoretical work. Many classic regression techniques aim at modelling conditional expectations; several theoretical results and tricks exist for this purpose and the literature on specialized estimation procedures is vast. For some applications, however, it is sometimes important to direct attention towards conditional quantiles, rather than conditional expectations, particularly if tail behavior is essential. Quantile regression is one of the most prevalent methods for giving a more detailed description of conditional distributions, and with it we get interesting answers and intriguing challenges. Some statistical issues, e.g. unobserved heterogeneity and sample selection, still require research in possible solutions. I will now briefly summarize the three papers, each of which takes up some of these challenges.

The first paper, "Headlights on tobacco road to low birthweight" (with Christian Møller Dahl and Johannes Tang Kristensen; forthcoming in *Empirical Economics*), is about the utilization of a "short panel" (where individuals are observed only a few times) and quantile regression in order to investigate the effect of prenatal smoking on the conditional birthweight distribution. Low birthweight outcomes are associated with considerable social and economic costs, and therefore the possible determinants of low birthweight are of great interest. Smoking, which has received considerable attention in this matter, is interesting from an economic perspective, in part due to the possibility of influencing smoking habits through policy making. It is widely believed that maternal smoking reduces birthweight; however, the crucial difficulty in estimating such effects is the unobserved heterogeneity among mothers and the fact that estimation of conditional mean effects seems potentially inappropriate. We provide a unified view on the estimation of relationships between prenatal smoking and birthweight outcomes with quantile regression approaches for panel data and emphasize their differences. Our findings provide strong evidence that smoking during pregnancy has adverse consequences for birthweight outcomes. The documented connection between babies' birthweight and their overall health, along with the costs associated with low birthweight, makes this a very important result. The effect appears to increase the further one moves to the left in the birthweight distribution, especially when measured relative to birthweight at the corresponding quantiles.

In my second paper, "Minimax regression quantiles and non-diagonally weighted quantile regression", motivated by the need for more ways to deal with correlated data, I set out to find a new estimation approach for quantile regression. I provide an alternative estimation framework and large-sample (asymptotic) properties for the ba-

sic estimator which turn out to be equivalent to those of the classic quantile regression estimator. This framework is based on estimating functions to which one can obtain consistent roots by finding the minimax of a certain deviance function. The new and flexible estimating framework for quantile regression allows for both linear and non-linear specifications and has potential for generalization. The proposed minimax approach, and extensions of the basic estimator, could therefore prove to have important implications. This is demonstrated in a proposal of a generalized estimator for non-diagonally weighted quantile regression. A simulation example is presented where this new estimator has improved efficiency.

In the third paper, "Does entrepreneurship really (not) pay?" (with Christian Møller Dahl and Ulrich Kaiser), we investigate patterns in the distribution of income in wage- and self-employment, using Danish register data. A variety of quantile-based methods are used to show relationships between experience in self-employment, labor-market experience and income at various locations of the conditional distribution. We find that the timing of income and career income stream is very important to model explicitly because income patterns are very different in wage-employment and self-employment. To this end we develop a quantile regression model of career composition in terms of earned experience and occupational choice where the response variable is (discounted) average income. The results from estimating the model strongly indicate that early stages of the firm are critical and associated with very low incomes. Self-employment at other stages of the experience profiles seems to be competitive with wage-employment, and our findings also suggest great benefits to the entrepreneur from experience in wage-employment. We also find support for the jack-of-all-trades theory, which suggests that broad experience is essential to the successful entrepreneur.

Danish summary (Dansk resumé)

Denne afhandling indeholder tre studier hvor *kvantil-regression* er fællesnævner. Denne regressionsteknik fandt vej til den statistiske/økonometriske værktøjskasse i 1978, hvor *Econometrica* udgav artiklen "Regression Quantiles" af Roger Koenker og Gilbert Bassett. Siden da er teorien blevet videreudviklet og har fundet anvendelse i mange empiriske studier. Klassiske regressionsteknikker finder ofte anvendelse i analyser af betingede middelværdier. Literaturen her er omfangsrig, og der er mange værktøjer og specialiseringsmuligheder til rådighed for analytikeren. I nogle tilfælde kan det dog have speciel interesse at rette analysen mod betingede kvantiler snarere end betingede middelværdier, hvis fokus fx er på én eller begge haler i den betingede fordeling. Til at give en detaljeret beskrivelse af den betingede fordeling er kvantil-regression én metode, måske den mest udbredte, og med denne åbner der sig en spændende verden med både interessante svar og nye udfordrende spørgsmål. En del af de statistiske øvelser som er løst og studeret nøje for de mere klassiske teknikker, eksempelvis uobserverede karakteristika eller endogenitet, kræver stadigvæk forskning i velegnede løsninger, når det gælder kvantil-regression. Jeg vil her give et kort resumé af de tre studier der indgår i denne afhandling, og som hver især tager nogle af disse udfordringer op.

Det første studie, "Headlights on tobacco road to low birtweight" (med Christian Møller Dahl og Johannes Tang Kristensen; optaget til publicering i *Empirical Economics*), belyser spørgsmålet om effekten af rygning på fødselsvægt, og hvordan man kan benytte et datasæt med et antal individer der hver er observeret få gange (et såkaldt "kort panel") i samspil med kvantil-regression. Lav fødselsvægt er blevet forbundet med både store direkte økonomiske omkostninger og socio-økonomiske omkostninger som fx dårligere uddannelsesforløb og ringere evne til at klare sig godt på arbejdsmarkedet som følge af kognitive effekter. Fødselsvægt betragtes som et generelt mål for et barns sundhed. Det er derfor ikke overraskende at der er fokus på de mulige årsager til lav fødselsvægt, og én af dem der har påkaldt sig særlig opmærksomhed er rygning under graviditet. Ud fra et økonomisk synspunkt er rygning interessant da adfærd her muligvis kan påvirkes politisk. Som eksempel kan nævnes tobaksafgifter og rygeforbudet i Danmark på offentlige steder. Det fremherskende synspunkt er at rygning under graviditet påvirker fødselsvægt negativt, men den præcise effekt er svær at måle. Det vanskelige er at analysere sammenhængen i venstre hale af fødselsvægtsfordelingen (der hvor problemet er værst) og at tage højde for at selve valget at ryge ikke er tilfældigt. Vi beskriver i vores studie flere tilgange til denne problematik og viser muligheder for at kontrollere for mødres (uobserverede) heterogenitet i anvendelsen af kvantil-regression. Vores resultater bekræfter den negative effekt af rygning på fødselsvægt og antyder at effekten er tiltagende i venstre hale af fødselsvægtsfordelingen, specielt hvis den måles relativt til faktisk fødselsvægt.

Motiveret af problemstillingerne med korrelerede observationer i anvendelsen af kvantil-regression foreslår jeg i mit andet studie, "Minimax regression quantiles and

non-diagonally weighted quantile regression", en ny tilgang til estimation af betingede kvantiler. Jeg viser at denne alternative metode (estimator) har samme egenskaber i store stikprøver (asymptotiske egenskaber) som den klassiske kvantil-regressions-estimator. Denne nye estimator bygger på passende estimationsfunktioner (*estimating functions*) for hvilke rødderne giver en konsistent løsning til kvantil-regressions-problemet. I praksis findes rødderne her ved af identificere et mini-maks punkt af en specielt konstrueret "afvigelses-funktion", som er designet til at have specifikke sammenhæng til de førnævnte estimationsfunktioner. Der er mulighed for at specificere både lineære og ikke-lineære kvantilfunktioner. Et stort bidrag fra studiet er efter min mening generaliseringen af denne metode til at kunne håndtere korrelerede observationer og ikke-diagonal vægtning i stil med litteraturen vedrørende de såkaldte "generalized estimating equations (GEE)" og mere generelt: udforskningen af alternative fremgangsmåder der kan vise sig at bidrage til løsninger på de mange udfordringer ved kvantilregression.

I det tredje og sidste studie, "Does entrepreneurship really (not) pay?" (med Christian Møller Dahl og Ulrich Kaiser), undersøger vi indkomstfordelingen for hhv. selvstændige og lønmodtagere i Danmark. Vi benytter et omfangsrigt register-datasæt fra Danmarks Statistik og bruger forskellige kvantil-baserede metoder til at belyse sammenhængen mellem erfaring som selvstændig, arbejdsmarkedserfaring og lønniveau, for forskellige kvantiler i den betingede fordeling. Vi finder frem til at indkomstmønsteret over tid er vigtigt at modellere eksplicit, da det er meget forskelligt for selvstændige og lønmodtagere. For at imødekomme denne problemstilling udvikler vi en model for karriere-sammensætning i form af optjent erfaring i de to typer af ansættelse på forskellige tidspunkter i karriereren. Den forklarede variabel er diskonteret gennemsnitlig indkomst. Vores resultater viser at tidlige stadier som selvstændig er problematiske, kritiske, og forbundet med lav indkomst. I senere stadier er indkomsten konkurrencedygtig med lønnen som lønmodtager for forskellige niveauer af erfaring. Vi finder også at indtægten som selvstændig er positivt associeret med arbejdsmarkedserfaring, og at det også er fordelagtigt at have en alsidig erfaring. Sidstnævnte effekt er kendt fra den såkaldte "jack-of-all-trades" teori, og vores resultater understøtter således dennes validitet.

Preface to paper #1

This paper was my first project and is forthcoming in *Empirical Economics*. Its development through several revisions has shaped much of my perception of the quantile regression problem. It was submitted in an earlier version and a referee offered, amongst others, the following comment: "Equation (xx) is not a very attractive way of introducing quantile regression..." Even before receiving the referee report, we had come to the same conclusion and had already decided to take action. What was wrong? I'll try to make a long story short: a typical application of regression starts by assuming a model, say

$$Y = m(X;\beta) + \varepsilon.$$

In other words, a parameterized functional relationship is assumed "up to some residual" or random component. Many applications of quantile regression, including an early version of our paper, start with such a model or *data generating process*, and then argue how quantile regressions for different indices alter the parameter vector β when ε is heteroskedastic. Not only is this confusing in terms of interpretation, it can also be more restrictive than really needed. The "framework" we adopted to introduce the estimators in this paper is currently my preferred way of thinking of the objects "regression quantiles": forget for a moment the data generating process, forget the ε-way. Think in terms of uniform random variable, $U \in (0,1)$, and assume existence of a conditional quantile function (inverse distribution function) which satisfies

$$Y = q(U,X) = F_Y^{-1}(U|X).$$

Now, think about how to approximate the function $q(\cdot)$ and how good such an approximation is. It can be shown (references are provided in the paper) that for a linear approximation, $q(U,X;\beta(U)) = X'\beta(U)$, the quantile regression estimator provides the best linear approximation to the true conditional quantile function (in a certain weighted minimum-squared-error sense).

The interpretation is straight-forward. At $U = \tau$, the estimated coefficient $\widehat{\beta}(\tau)$ is a "best" approximation of an effect which can be linearly attributed to X at this location of the distribution, regardless of data generation and any ε.

This paper presents a few panel data methods within this framework, and gives an analysis of the effect of maternal smoking on birthweight outcomes. The estimators have least-squares analogues but a "philosophical" word of caution is: be very careful not simply to adopt the classic way of thinking. One might even think of our "OLS" columns as "dangerous" as they could encourage a *too* direct comparison. Perhaps if, from early days, a symbol other than β was used to denote regression quantiles some confusion could have been avoided.

Paper #1

Headlights on tobacco road to low birthweight

Evidence from a battery of quantile regression estimators and a heterogeneous panel

2011

Stefan Holst Bache
Aarhus University, Department of Economics and Business and Creates
Christian Møller Dahl
University of Southern Denmark, Department of Economics and Creates
Johannes Tang Kristensen
Aarhus University, Department of Economics and business and Creates

Keywords:
Quantile regression, low birthweight, panel data, unobserved heterogeneity

Abstract. Low birthweight outcomes are associated with considerable social and economic costs, and therefore the possible determinants of low birthweight are of great interest. One such determinant which has received considerable attention is maternal smoking. From an economic perspective this is in part due to the possibility that smoking habits can be influenced through policy conduct. It is widely believed that maternal smoking reduces birthweight; however, the crucial difficulty in estimating such effects is the unobserved heterogeneity among mothers and the fact that estimation of conditional mean effects seems potentially inappropriate. We provide a unified view on the estimation of relationships between prenatal smoking and birthweight outcomes with quantile regression approaches for panel data and emphasize their differences. This paper contributes to the literature in three ways: i) we focus not only on one technique, but provide evidence from several approaches and highlight a variety of statistical issues; ii) the performance of the methods are thoroughly tested in a simulated environment, and recommendations are given on their appropriate use; iii) our results are based on a detailed data set, which includes many relevant control variables for socio-economic, wealth and personal characteristics.

1 Introduction and Motivation

The potential adverse health consequences of low birthweight outcomes, along with the considerable economic burden they are believed to impose on society, have attracted much attention by researchers in both medical and economic literature. The use of birthweight as a proxy for the general health condition of infants is commonplace, as it has been linked to a vast array of health related complications, both short- and long-term.

The most severe event, perinatal mortality, has been found to be more likely in the event of a low birthweight outcome. Several studies find statistical evidence of this linkage, see e.g. Bernstein et al. (2000), Almond et al. (2005), and Black et al. (2007). Furthermore, it is believed that low birthweight may lead to complications such as epilepsy, mental retardation, blindness, and deafness. For a review and references, see Hack et al. (1995). While many of these complications are directly observable, some studies also consider less obvious socio-economic implications of low birthweight, a very popular topic being school performance. Kirkegaard et al. (2006) find a graded relationship between birthweight and school performance. In a follow-up study with 5,319 Danish children aged 9–11, they conclude that the risk of reading, spelling and arithmetic disabilities is greater with low birthweight children. Similarly, Corman and Chaikind (1998) find that repeating a grade, or special class attendance, is more likely among low birthweight children. This may suggest that even future earnings and labour market outcomes may be affected by birthweight. According to Black et al. (2007), this is indeed the case.

The strong evidence that low birthweight has adverse effects has naturally led to substantial efforts towards identifying the determinants of these undesirable outcomes. One such determinant which has received much attention in the literature is maternal smoking habits during pregnancy. Statistical efforts suggest a strong correlation between low birthweight and maternal smoking, see e.g. Bernstein et al. (1978) and Permutt and Hebel (1989). Other studies document the adverse effect of smoking on some of the above-mentioned complications directly, e.g. Wisborg et al. (2000), Wisborg et al. (2001), and Linnet et al. (2006). Medical research gives several reasons why cigarette smoking may affect birthweight. An explanation that seems to stand out is that the foetus may suffer from chronic hypoxic stress as a consequence of smoking, see DiFranza et al. (2004) and Hofhuisi et al. (2003). An interesting observation is that smoking does not seem to have a significant adverse effect on all birth outcomes. Wang et al. (2002) conclude that the association between maternal cigarette smoking and reduced birthweight is modified by maternal genetic susceptibility.

From an economic perspective, interest lies not with the individual as such, but rather with society as a whole. Maternal smoking habits are thus an especially interesting determinant since it is believed to be modifiable through policy conduct, e.g. by regulating taxes on tobacco products or by introducing smoking prohibitions in public areas. While medical research gives much attention to why smoking causes low birthweight, the above has led economists to focus primarily on the extent of this effect, and the associated costs. This perspective has the advantage of allowing analysts to disregard the specific medical links between maternal smoking and low birthweight, when using appropriate methods.

In an attempt to estimate the direct costs associated with low birthweight, Almond et al. (2005) use data from hospitals in New York and New Jersey to find that the costs peak at $150,000 (in year 2000 dollars) for newborns weighing 800 grams. In contrast, an infant weighing 2,000 grams has an estimated cost of $15,000. The soaring costs at the low end of the birthweight distribution highlight an important point. Traditional least squares regression essentially models the conditional mean, and may be uninformative about tail behavior. One way to overcome this problem is to use a quantile regression approach, which can provide estimation results across the entire distribution. This is done by Abrevaya (2001) and Koenker and Hallock (2001), who find justification for the quantile approach since regression estimates vary throughout the distribution. It is, however, troublesome to consider the estimated effects as causal, because the analyses do not account for unobserved heterogeneity. Not only is the susceptibility of smoking effects among mothers different, as noted above, but there are undoubtedly many other individual characteristics which cannot be accounted for.

Econometric panel data models allow controlling for (time invariant) unobserved individual heterogeneity. However, their extension to a quantile regression framework is still somewhat limited. In a recent paper, Abrevaya and Dahl (2008) consider the extension of the "correlated random effects" model by Chamberlain (1984) to a quantile regression framework, and estimate the effects of various birth inputs on birthweight, using data from Arizona and Washington. Their results indicate that the negative effects of smoking, albeit present, are significantly lower in magnitude across all quantiles than the corresponding cross-sectional estimates.

This paper in part extends the results of Abrevaya and Dahl, using Danish data, which in itself is novel: no previous study has applied such techniques to data with this origin. The advantage of our data, relative to those used in existing literature, lies in the richness and availability of variables. Quite naturally, however, there are fewer observations due to a small geographical area. Finally, the Danish Civil Registration System allows perfect linkage of the data. Based on the idea of the above mentioned study, we consider a new correlated random effects specification for quantile regression which, at the cost of a more restricted specification, allows for the use of an unbalanced dataset and benefits from a more parsimonious amount of regressors. Finally, we consider fixed effects approaches to quantile regression, in particular we examine a model specification by Koenker (2004) and suggest a simple a two-stage fixed effects approach.

Before we delve into our birthweight application in Section 3, we take a tour in the realm of quantile regression for panel data. Our treatment offers a unified framework in which the different approaches can be discussed and compared appropriately. We are not aware of a similar discussion, and believe it to be novel and relevant more generally. Simulations will serve to illustrate features and test performance of the discussed estimation methods.

2 Econometric setup

Panel data and quantile regression - a prologue

A chief difficulty in examining the causal effect of prenatal smoking, and other relevant observable variables, on birthweight outcomes is the possible existence of influential but unobservable determinants. The identification and measurement of all such determinants is an impossible task, and it can thus be necessary to control for such unobserved effects. When repeated measurements for each individual are available, analysts will try to utilize this panel structure of the data to either filter out, or in some other way deal with e.g. time-invariant unobserved characteristics. There is a very well developed machinery with a variety of estimation procedures available for linear least-squares models, and hence the issues are here easily mitigated. Often, and indeed in the present analysis, the conditional mean of the response is not of primary interest, but rather our attention is directed towards the conditional quantiles. Not surprisingly, combining the power of panel data methods with that of quantile regression methods is not a topic of little interest. While there are methods available to do so, the topic is still relatively undeveloped, and there are some very important subtleties that often do not receive sufficient attention. In particular, one needs to be very careful in defining the quantities or parameters of interest for reasons that we will try to make clear. The main purpose of this section is to discuss some relevant procedures for panel data quantile regression. Simulations will serve both to evaluate the appropriateness of the methods and to investigate their performance. The discussion is of broader relevance than to the current analysis and we hope it will help others in choosing the best approach for their particular application.

There are many ways of introducing quantile regression. One can take a structural approach, assume a data-generating process (DGP), and describe how the error term may change the coefficients at various quantiles. It is often hard to specifically link a DPG to its quantile function. Therefore, another common approach is to think of an approximation to the quantile function directly, and not emphasize how data is generated. Angrist et al. (2006) show that, in the linear case, the quantile regression estimator in a certain sense gives the best linear approximation to the true conditional quantile function. Related is the Skorohod representation where the response variable is generated by a (quantile) function that depends, amongst other things, on a rank-variable (or quantile index).

We shall start the discussion with the following definition of the conditional quantile function. Let Y denote the response variable of interest with distribution function F_Y and let X be a vector of covariates on which we condition, and denote by $\tau \in (0,1)$ the quantile index. We define

$$(1) \qquad Q_Y(\tau \mid X) \equiv \inf\{y : F_Y(y \mid X) \geqslant \tau\}.$$

Then, in Skorohod representation,

$$(2) \quad Y = Q_Y(U \mid X), \quad U \mid X \sim \text{uniform}(0,1).$$

It is well-known that the function is a solution to the minimization problem

$$(3) \quad Q_Y(\tau \mid X) \in \underset{q_\tau(X)}{\arg\min}\, \mathrm{E}\left[\rho_\tau(Y - q_\tau(X))\right],$$

where $\rho_\tau(u) = (\tau - 1\{u < 0\})u$, and the minimization is over all measurable functions. It is also a solution to the estimating function $\mathrm{E}[1\{Y \leqslant q_\tau(X)\}] = \tau$, or equivalently $\mathrm{E}[\tau - 1\{Y \leqslant q_\tau(X)\}] = 0$. For an alternative to (3) for solving a parameterized version of the latter estimating equation, see Bache (2010).

Our discussion is focussed on the linear-in-parameters subset of functions over which (3) is minimized, i.e. $q(X; \beta(\tau)) = X^\mathrm{T}\beta(\tau)$, and we shall investigate some possible ways of incorporating information from repeated measurements to alleviate identification issues that arise due to characteristics not included in the model. The methods we discuss can be categorized into a fixed-effects and a correlated-random-effects framework. The terminology is carried over from their least-squares analogies, and we will not philosophize about the appropriateness of it. We will, however, now argue why these two branches distinguish themselves even more from each other in a quantile regression setting.

Consider the following much simplified setup. We imagine a population of individuals, where each can be either of two types, say $c = 0$ or $c = 1$. Types are attributes in the sense that they are constant and not under the individuals' control (e.g. one could think about genetic traits). The analyst has no information about types, i.e. for all purposes they are unobservable.

The question of interest is how a treatment x (e.g. smoking during pregnancy) affects the distribution of an outcome variable (e.g. birthweight). Figure 1 shows three distributions that may be of interest.

Now, if type is considered a part of the "noise" or "error term" in the model[2], which it may well be, say to a politician who wants to start a campaign against prenatal smoking, then the unconditional density is the relevant one. On the other hand, even though information in the data is not sufficient to reveal type, a doctor for example may have "inside" information on type, and would want a model that conditions on c. In this case, the conditional distributions are of interest. The main point here is that the marginal effects of changes in X on the quantile function conditioning both X and c are *not* the same as if conditioning only on X.

Fixed-effects approaches often incorporate estimates of c as a way of conditioning on it, e.g. as the estimator by Koenker (2004). We will also suggest a simple 2-stage

[2]In a treatment of quantile regression, "unexplained ranking mechanism" may be better terminology.

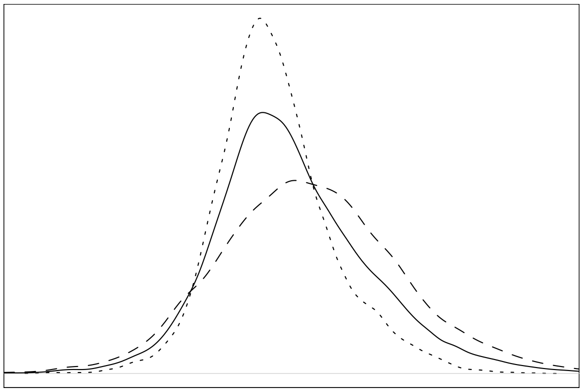


Figure 1: Three densities of Y: one conditional on $c = 1$ (dashed), one conditional on $c = 0$ (dotted), and one unconditional of c (solid). All densities are here unconditional of x.

plug-in method as a computationally simpler alternative. An inherent issue with the fixed-effects methods is the incidental parameter problem that arises when the number of repeated measurements is small and fixed.

The correlated random effects model, suggested first for regression quantiles by Abrevaya and Dahl (2008), henceforth the AD model, tries to control for dependence between X and c, which can bias the estimates, if ignored, when a "random assignment" interpretation is wanted. The idea is that one can generate one or more "sufficient covariates" from the repeated observations which carry information that can correct for the bias. For the present application, we will also consider an alternative specification that can be seen as a restricted version of the AD model, but which will allow us to use an unbalanced panel, and which is more parsimonious in terms of number of regressors.

We formally define the four estimators in section 2, but first we present an illustrative example of the point raised in this section about conditioning on unobserved time-invariant variables.

An illustrative simulation experiment

Consider again the simple setup with a two-type population. The model that generated the data depicted in Figure 1 is generated as

$$(4) \qquad y_{mb} = x_{mb} + c_m + (1 + x_{mb} + c_m)\varepsilon_{mb}, \qquad b = 1,2;\ m = 1,\ldots,M,$$

where one half of the population has $c_m = 1$, the other $c_m = 0$. The variable x_{mb} is a binary treatment and equals 1 with probability $0.5 + 0.2c_m$. The disturbance ε_{mb} is a standard normal random variable. The subscripts m and b denote "mother" and "birth" respectively, and are chosen in the light of the birthweight application.[3] Note that the unobserved type affects both location and scale of the response distribution. We also define the counterfactual variables $\tilde{x}_{mb}$ and $\tilde{y}_{mb}$, where $\tilde{x}_{mb}$ equals 1 with probability 0.5 and $\tilde{y}_{mb}$ is determined by (4) with $\tilde{x}_{mb}$ in place of x_{mb}. These then represent a counterfactual world, where type has no impact on treatment probability. We will consider three "targets" for the estimate of a coefficient for the effect of being treated (setting $x = 1$):

$$\Delta_\tau := Q_Y(\tau \mid x_{mb} = 1) - Q_Y(\tau \mid x_{mb} = 0) \tag{5}$$

$$\widetilde{\Delta}_\tau := Q_{\tilde{Y}}(\tau \mid \tilde{x}_{mb} = 1) - Q_{\tilde{Y}}(\tau \mid \tilde{x}_{mb} = 0) \tag{6}$$

$$\Delta_{\tau|c} := 1 + Q_\varepsilon(\tau) = 1 + Q_N(\tau), \tag{7}$$

where $Q_N(\tau)$ is the τth quantile of a standard normal random variable. Equations (5) and (6) are the treatment's effect on the distribution of y_{mb} and $\tilde{y}_{mb}$, respectively, unconditionally on c_m (i.e. c is here part of the error). Equation (7) is the effect on the distribution of y_{mb} conditional on c_m, i.e. $Q_Y(\tau|c_m, x_{mb} = 1) - Q_Y(\tau|c_m, x_{mb} = 0)$.

Let $\widehat{\beta}(\tau)$ be the estimate from a quantile regression of y_{mb} on x_{mb} and $\tilde{\beta}(\tau)$ the one from a quantile regression of y_{mb} on x_{mb} and $\bar{x}_{m\circ}$, where the latter variable is an average over the b dimension. Further, let $\check{\beta}(\tau)$ be the "oracle" estimate, where knowledge of type is assumed, from a regression of y_{mb} on x_{mb} and c_m. Table 1 shows the results from a simulation experiment, comparing these estimators with the three targets defined above for a single quantile index, $\tau = 1/5$.

The results in this example show quite clearly that the estimators identify different quantities. A very noteworthy observation is that the estimator $\widetilde{\beta}(\tau)$, a "correlated random effects" type estimator that we will define below, does *not* try to estimate c_m, and does not suffer from an incidental parameter problem. Instead, it uses information constructed from all observations for each individual to correct for correlation between c_m and x_{mb} to get a "random assignment" interpretation. The fixed effects estimator, on the other hand, relies on some kind of estimate of c_m (above it is simply assumed known), and will most likely not perform as well as we have just seen when the number of waves is small. The estimator $\widehat{\beta}(\tau)$ is not really of interest, but it shows what happens with the usual quantile regression estimator when the treatment is endogenous.

[3] In our application birth parity is what defines the waves, and does not represent time as such. In the examples, however, one can think of b as representing time.

	$\widehat{\beta}(0.2)$	$\widetilde{\beta}(0.2)$	$\check{\beta}(0.2)$
Δ_τ	**−0.0040** (**0.1188**)	−0.0228 (0.1619)	−0.0385 (0.1295)
$\widetilde{\Delta}_\tau$	0.0338 (0.1234)	**0.0070** (**0.1604**)	−0.0087 (0.1239)
$\Delta_{\tau\|c}$	0.0452 (0.1270)	0.0184 (0.1613)	**0.0026** (**0.1236**)

(a) $M = 999$

	$\widehat{\beta}(0.2)$	$\widetilde{\beta}(0.2)$	$\check{\beta}(0.2)$
Δ_τ	**0.0010** (**0.0380**)	−0.0289 (0.0597)	−0.0404 (0.0567)
$\widetilde{\Delta}_\tau$	0.0308 (0.0489)	**0.0009** (**0.0522**)	−0.0106 (0.0411)
$\Delta_{\tau\|c}$	0.0422 (0.0567)	0.0123 (0.0536)	**0.0008** (**0.0398**)

(b) $M = 9{,}999$

Table 1: Bias and root mean squared error (in parentheses) for the three estimators $\widehat{\beta}(\tau)$, $\widetilde{\beta}(\tau)$, and $\check{\beta}(\tau)$ against the three targets Δ_τ, $\widetilde{\Delta}_\tau$, and $\Delta_{\tau|c}$. The simulation setup is $\tau = 1/5$, "MC iterations"= 999, number of waves $B = 2$, and number of individuals $M = 999$ (panel a) and $M = 9{,}999$ (panel b). It should be noted that $\Delta_{\tau|c}$ can be calculated analytically, whereas $\widetilde{\Delta}_\tau$ and Δ_τ themselves are obtained by simulation.

The estimators defined

Throughout, we take Y_{mb} to be the random response variable and X_{mb} to be the corresponding covariate vector. We denote by $Z_{mb} \subset X_{mb}$ a subset that has time-varying and possibly endogenous covariates. The subscripts m and b ("mother" and "birth") are chosen to reflect the dimensions in our birthweight study. Lower case letters will denote outcomes in the sample with $m = 1,\dots,M$ and $b = 1,\dots,B_m$. The pairs $\{Y_{mb}, X_{mb}\}$ are assumed to be independent and identically distributed (IID), and so whenever no confusion arises subscripts m and b are omitted from notation.

For the purpose of this presentation of the models, we take the view that we can approximate the conditional quantile function reasonably well by a linear-in-parameters specification, and will not argue a specific DGP.[4] The following assumption states the standard QR problem in terms of the framework we shall work with to describe the panel data approaches.

Assumption (A.QR): Linear quantile approximation representation. The class of functions over which (3) *is minimized is linear, such that*

(8) $$q_\tau(X) = q(X,\tau) = X^{\mathrm{T}}\beta(\tau) \quad \textit{and}$$

(9) $$Y = q_\tau(X,U),$$

where U is a rank variable with $U|X \sim$ uniform(0,1) (independently of X). For any two possible ranks u_1 and u_2 it is assumed that

(10) $$u_1 < u_2 \Leftrightarrow q(X,u_1) < q(X,u_2) \Leftrightarrow Y_1|X < Y_2|X.$$

[4]In our simulations, of course, we need to assume some data-generating mechanism.

Example: Normal location-scale model. Let

(11) $q(X,\tau) = X^{\mathrm{T}}\beta(\tau) = X^{\mathrm{T}}(\beta + \gamma\Phi^{-1}(\tau)),$

where Φ is the standard-normal CDF, then we have the implied familiar location-scale DGP

(12) $Y = X^{\mathrm{T}}\beta + (X^{\mathrm{T}}\gamma)\Phi^{-1}(U).$

The above representation in (8)–(9) is often referred to as the Skorohod representation. If there are unobserved effects that affect both X and the ranking U, then we cannot represent the problem as above and use $\beta(\tau)$ as the quantity of interest, as illustrated in the example in the previous section. In the following, we will see how one can possibly get around this problem if such unobserved characteristics are time-invariant (or here, "birth-invariant").

Correlated random effects quantile regression

This presentation takes a slightly different approach than the one taken in Abrevaya and Dahl (2008), but the message remains the same. We now extend the model to include unobserved characteristics C_m that influence Y_{mb} either directly, through $Z_{mb} \subset X_{mb}$, or both. These characteristics are assumed to be time-invariant characteristics, so dependence with Z_{mb} is one-way. The random data pairs $\{C_m, Y_{m1}, X_{m1}, Y_{m2}, X_{m2}, \ldots\}$ are assumed to be IID. We wish to consider C_m a part of the unexplained ranking mechanism in the model, but at the same time control for endogenous effects propagated through Z_{mb}. To achieve this goal, we assume that repeated measurements of Z_{mb} allow for construction of sufficient covariate(s) S_m, and let the conditional quantile function of interest be $Q_{Y_{mb}}(\tau \mid X_{mb}, S_m)$. Sufficiency is to be understood as to allow for the following extension of (A.QR). Again, we will often omit subscripts to simplify notation.

Assumption (A.CRE): Correlated random effects representation. Consider a representation similar that of (A.QR), with

(13) $q(X,S,\tau) = X^{\mathrm{T}}\beta(\tau) + S^{\mathrm{T}}\pi(\tau)$ *and*

(14) $Y = q(X,S,U),$

for some variable S, constructable from repeated observations of Z, such that $U|X,S \sim$ $\mathrm{uniform}(0,1)$.

This allows us to think of a response process $\widetilde{Y}(U) \equiv Y - S^{\mathrm{T}}\pi(U)$ as being Y "corrected" at level U for effects of C through Z, and as having τth conditional quantiles

$X^{\mathrm{T}}\beta(\tau)$ for $U=\tau$. Put this way, it is emphasized that the correction gives $\beta(\tau)$ the interpretation of marginal effects in a "counterfactual world" where X is not determined by C, but where C is allowed to work directly on Y through the ranking U.

Under the assumptions in (A.CRE), the quantity of interest, $\beta(\tau)$, is identified from the data, and can be estimated by means of standard quantile regression of Y on X and S, with empirical criterion function

$$(15)\quad \big(\widehat{\pi}(\tau),\widehat{\beta}(\tau)\big)=\underset{\pi,\beta}{\arg\min}\sum_{m=1}^{M}\sum_{b=1}^{B_m}\rho_\tau(y_{mb}-s_m^{\mathrm{T}}\pi-x_{mb}^{\mathrm{T}}\beta).$$

The separability assumption that the effects from C through Z can be captured by S in a linear fashion may not be as restrictive as one would think at first, since it is allowed to vary with τ. It may therefore provide a good linear approximation at each quantile, and seems no more restrictive than believing the linear specification of q in the first-place. In general, note that the assumptions do not restrict the effect from C to a location-shift.

Example: Recall our previous introductory example. There we had a single endogenous binary treatment variable x_{mb} $(=z_{mb})$ and unobserved types c_m that affect both the probability of treatment, as well as location and scale of the outcome. Intuitively, the mean of x (over b) carries information about type, and we let $s_m=\bar{x}_{m\circ}\equiv\sum_b x_{mb}/B_m$. The simulations above confirmed that this was an acceptable choice and that the procedure was appropriate in this case.

More specifically, we mention here the following two specific CRE models, in terms of constructing S; the one proposed by Abrevaya and Dahl (2008), and another which can be seen as a special case of the first.

The "AD" model: Assume (A.CRE), that the data constitutes a balanced panel $B_m=B$, and that $S_m=(Z_{m1}^{\mathrm{T}},\dots,Z_{mB}^{\mathrm{T}})^{\mathrm{T}}$ are sufficient covariates.

The "CREM" model: Assume (A.CRE), and that $S_m=\bar{Z}_{m\circ}$, i.e. b-means of observed outcomes of Z_{mb}, are sufficient covariates.

The second, where the "M" in the acronym stands for *mean*, is the one deployed in the previous example. It can be seen as a special case of the AD model, in which S is essentially a weighed average, in the sense that the effect from C through Z is assumed to be the same for each b. In addition to being more parsimonious, in terms of number of regressors, this restriction allows the use of an unbalanced panel. It is possible to extend the AD model, using dummy variables, to allow for some kinds of unbalanced panels (Fitzenberger et al., 2010).[5]

[5]We thank the editor in chief for pointing this out to us. Our results did not seem to depend on the choice of S, so we did not explore this option explicitly.

For the linear least-square analogues of the AD and CREM models, by Chamberlain (1984) and Mundlak (1978) respectively, one has a linear DGP, $y_{mb} = x_{mb}^{\mathrm{T}}\beta + c_m + \varepsilon_{mb}$, and one thinks of projecting c_m onto observables as

$$(16) \quad c_m = x_{m1}^{\mathrm{T}}\pi_1 + \cdots + x_{mB}^{\mathrm{T}}\pi_B, +\eta_m \quad \text{respectively}$$

$$(17) \quad c_m = \bar{x}_{m\circ}^{\mathrm{T}}\pi + \eta_m,$$

for the two models. For quantile regression, using a DGP as a starting point seems restrictive as there is not necessarily a well-defined link between a DGP and a linear quantile representation. Furthermore, it would require a discussion of the critical joint distribution of the error terms in the y_{mb} and c_m equations. For further detail on—and a slightly different presentation of—the CRE approach we refer to the aforementioned article by Abrevaya and Dahl.

Fixed effects estimators

These estimators, in general, condition on the unobserved time-invariant characteristics C using some kind of estimate of them. Loosely speaking, the approaches in a way model $Q_Y(\tau \mid X, C)$ by $Q_Y(\tau \mid X, \widehat{C})$. As we saw in the previous simulation example, if one knows C, this approach is obviously perfectly valid, yet for a target quantity slightly different from that of the CRE approaches. Therefore, if one assumes that good estimates of C are available, these can be used in following representation.

Assumption (A.FE): Quantile regression conditional on fixed effects. The class of functions over which (3) *is minimized is linear in X and C such that*

$$(18) \quad q(X, C, \tau) = X^{\mathrm{T}}\beta(\tau) + C\delta(\tau) \quad \textit{and}$$

$$(19) \quad Y = q(X, C, U),$$

where $U|X, C \sim \mathrm{uniform}(0,1)$ and $q(\cdot)$ satisfies an ordering property equivalent to that in (10) *of (A.QR).*

Essentially, the assumption here is that C is the only source of endogeneity and that it enters linearly on equal grounds with X. Conditioning on C will give $\beta(\tau)$ the same interpretation as if C was just another covariate and not a part of the unexplained ranking U.

However, in many cases the number of waves, B_m, is small (fixed) while M is large. In such a setting, good estimates of M individual effects are hard to come by, a situation referred to as the "incidental parameter problem". The estimates of C are usually not of particular interest, but it is not straight-forward to infer the consequences of M-inconsistent estimates of C on the estimates of $\beta(\tau)$.

Koenker (2004) suggested an interesting method to overcome some of the difficulties of the FE approach, in which estimation of C is intrinsic (i.e. it is a one-step estimator). The idea is to (i) estimate the model for several τs simultaneously, restricting the effect of C at each τ to be the same; and (ii) penalize estimates of the fixed effects to shrink them towards zero. Both (i) and (ii) in some sense reduce the dimensionality added by introducing estimation of fixed effects. Again, our presentation of the model is slightly alternative, compared to its original, such that it fits well into our framework.

"PFE(k): Penalized FE QR model". Define M parameters, $\alpha_m = C_m\delta(\tau)$ *(for all* τ*). Assume that*

$$q(X_{mb}, C_m, \tau) = X_{mb}^{\mathrm{T}}\beta(\tau) + \alpha_m \quad \textit{and} \tag{20}$$
$$Y = q(X_{mb}, C_m, U_{mb}), \tag{21}$$

Let $\tau_1, \ldots, \tau_k$ *be* k *distinct quantile indices. Further define* $w_1, \ldots, w_k$ *to be weights that define the relative impact of each of these indices on estimation. The parameters in* (20) *are estimated from the following empirical criterion function:*

$$\left(\widehat{\beta}(\tau_1), \ldots, \widehat{\beta}(\tau_k), \widehat{\alpha}_1, \ldots, \widehat{\alpha}_M\right) = \underset{\beta_1,\ldots,\beta_k,\alpha_1,\ldots,\alpha_M}{\arg\min} \sum_{j=1}^{k}\sum_{m=1}^{M}\sum_{b=1}^{B} w_j \rho_{\tau_j}(y_{mb} - x_{mb}^{\mathrm{T}}\beta_j - \alpha_m) + \lambda \sum_{m=1}^{M} |\alpha_m|. \tag{22}$$

Let $M \to \infty$*,* $B \to \infty$ *and* $M^a/B \to 0$ *for some* $a > 0$*. Then under some regularity conditions,* $\widehat{\beta}(\tau_j)$*,* $j = 1, \ldots, k$*, are consistent and converge to Gaussian random vectors. See (Koenker, 2004, Theorem 1) for the details.*

It is pretty clear from this representation that the price one pays for the gained sparseness is that unobserved *fixed effects are only allowed to affect* Y *by location shifts.* Further, often B is fixed, and consistency is not guarantied by the theorem. The parameter λ is a tuning/calibration parameter that allows one to control the impact of the penalty, and how to choose it optimally is an open research question. When $\lambda \to 0$, one has a (weighted) dummy variable regression, and when $\lambda \to \infty$ the penalty sets all FE terms to zero, effectively leaving a pure (weighted) cross-section regression. As with other penalty methods, one might consider a "post-penalty" estimation of the model where terms shrunken to zero are omitted, since non-zero terms have been affected by the penalty. On the other hand, the penalty also helps "controlling" the many FE terms by not letting them take unreasonably large values.

A simpler approach, which does not restrict C to have the same impact across quantiles in return for sparseness, and which does not need calibration, is the following two-step fixed effects estimator estimator, which resembles recent work by Arulampalam et al. (2007).

"2SFE: 2-step FE QR model." Assume (A.FE). In a first step, obtain estimates $\tilde{C}_m$ of C_m from a least-squares within groups estimation. In the second step, $\tilde{C}_m$ are used in place of C in (18), from which estimates of $\delta(\tau)$ and $\beta(\tau)$ are obtained.

An important implicit assumption here is that a linear approximation is appropriate for both the conditional expectation and the conditional quantiles. To hope for any asymptotic justification, one would need $B \to \infty$, which we do not consider here, as it is not relevant for our application. It has come to our knowledge that Canay (2010) has work on the asymptotic aspect of this estimator. For our purpose, performance is compared to Koenker's method and the CRE estimators in the following simulation section.

Additional simulation evidence

Indeed, the number of estimation methods under consideration allows for many interesting simulation setups. We have no desire of letting a simulation section eclipse the main part of this section or the application in question. Therefore, we have chosen a simple setup that highlights some important features and issues of the procedures discussed. In short, our findings are:

- The CRE methods do not suffer from an incidental parameters problem and perform well even when omitted effects have scale effects on the response.
- The CREM and AD models have very similar levels of performance.[6]
- FE methods generally have difficulty for short panels, and the size of the bias appears to depend critically on both τ and the actual setup. However:
  - estimating more quantiles simultaneously can be advantageous, and
  - post-estimating the "selected model" where penalized FE terms are removed can improve performance in some cases.
- The FE methods perform worse when c_m has a scale effect (i.e. effects varying over quantiles). For the PFE model, this is not surprising given its specification.
- PFE estimates are generally bounded by cross-sectional estimates and a dummy-variable quantile regression, as the theory suggests.
- Being a FE estimator, 2SFE is also cursed by incidental parameters. It benefits solely from simple estimation and from the fact that it needs no calibration.

[6] We have tried specifications where the dependence between $x_{1,mb}$ and c_m depends on b to accommodate AD. This, however, had little effect and CREM performed equally well.

Our simulation setup, which easily encompasses all of the estimation procedures, has the following data generating mechanism:

$$
\begin{aligned}
(23)\quad y_{mb} &= 10 - 3x_{1,mb} + x_{2,mb} + c_m + \eta_{mb} \\
\eta_{mb} &= (1 + x_{1,mb} + \gamma c_m)\varepsilon_{mb} \\
\varepsilon_{mb} &\sim N(0,1)
\end{aligned}
$$

where c_m equals 0 with probability 1/2 and is distributed as standard normal otherwise; the binary variable $x_{1,mb}$ equals 1 with a probability, p_m, that depends on c_m in the following way:

$$
(24)\quad p_m = \begin{cases} 0.25 & \text{if } c_m > 0.2 \\ 0.75 & \text{if } c_m < -0.2 \\ 0.50 & \text{otherwise.} \end{cases}
$$

The idea is that the probability of treatment is only affected if individuals distinguish themselves sufficiently. We let $x_{2,mb}$ be a sum of five uniform variables on $(-0.5, 0.5)$, which then lies in $(-2.5, 2.5)$. We can think of this setup as a simplified (and scaled) simulation of our birthweight application where we have an overall intercept, from which some members of the population distance themselves (when $c_m \neq 0$). Smoking, $x_{1,mb}$, has a negative direct effect, and it has a scale effect. It is correlated with c_m which makes it necessary to control for if we want to have some notion of random assignment interpretation. The variable x_2 plays the part of "other" observables in the model. The parameter $\gamma \in \{0,1\}$ lets us control whether c_m has a scale effect in addition to its location effect, which would cause the effect from c_m to vary across quantiles.

As previously discussed, we have target quantities for the coefficient estimates on $x_{1,mb}$ that differ for CRE and FE approaches. Estimators in the latter category have the target $\beta_{fe}(\tau) = Q_N(\tau) - 3$, while those in the former have $\beta_{cre}(\tau) = Q_{c+(2+\gamma c)\varepsilon}(\tau) - Q_{c+(1+\gamma c)\varepsilon}(\tau) - 3$. These quantities are derived in a similar fashion as those in equations (6) and (7) but for the model in (23).[7]

We report results for the following estimators: (i) A pure dummy variable quantile regression, (ii) PFE(1) and PFE(3); (iii) a post-penalty estimation of the latter, where penalized FE terms are removed; (iv) 2SFE; and (v) the two CRE methods. We also consider a cross-sectional estimate against both targets to evaluate the bias when ignoring c_m completely. All results can be read in Tables 8 and 9 in Appendix A. Here, in the main text, we present a few selected results in Table 3. To illustrate the asymmetric performance, particularly of the FE estimators, we consider $\tau \in \{1/4, 1/2, 3/4\}$. The corresponding targets are presented in Table 2.

[7] Specifically, we have $\beta_{fe}(\tau) = Q_Y(\tau|c, x_2, x_1 = 1) - Q_Y(\tau|c, x_2, x_1 = 0) = [10 - 3 + x_2 + c + (2 + \gamma c)Q_N(\tau)] - [10 + x_2 + c + (1 + \gamma c)Q_N(\tau)] = Q_N(\tau) - 3$, and that $\beta_{cre}(\tau) = Q_Y(\tau|x_2, x_1 = 1) - Q_Y(\tau|x_2, x_1 = 0) = Q_{c+\eta}(\tau|x_1 = 1) - Q_{c+\eta}(\tau|x_1 = 0) - 3 = Q_{c+(2+\gamma c)\varepsilon}(\tau) - Q_{c+(1+\gamma c)\varepsilon}(\tau) - 3$.

	$\tau = 1/4$	$\tau = 1/2$	$\tau = 3/4$
FE	-3.6745	-3	-2.3255
CRE, $\gamma = 0$	-3.6299	-3	-2.3701
CRE, $\gamma = 1$	-3.6114	-3.0253	-2.3797

Table 2: Coefficient targets for the estimators.

		$B = 2$		$B = 3$		$B = 5$	
	τ	$M = 499$	$M = 999$	$M = 499$	$M = 999$	$M = 499$	$M = 999$
Cross-section FE target	0.25	−0.3380	−0.3402	−0.3412	−0.3443	−0.3435	−0.3405
		(0.3713)	(0.3561)	(0.3618)	(0.3538)	(0.3560)	(0.3467)
	0.50	−0.3724	−0.3740	−0.3755	−0.3718	−0.3718	−0.3722
		(0.3968)	(0.3867)	(0.3920)	(0.3808)	(0.3813)	(0.3770)
	0.75	−0.4321	−0.4344	−0.4310	−0.4282	−0.4186	−0.4245
		(0.4598)	(0.4468)	(0.4489)	(0.4373)	(0.4301)	(0.4303)
Dummy regression	0.25	0.6717	0.6626	0.3531	0.3474	0.1438	0.1431
		(0.6980)	(0.6770)	(0.3813)	(0.3619)	(0.1763)	(0.1595)
	0.50	−0.0028	−0.0119	−0.0047	−0.0055	−0.0035	−0.0053
		(0.1898)	(0.1395)	(0.1272)	(0.0904)	(0.0865)	(0.0645)
	0.75	−0.6773	−0.6864	−0.3596	−0.3598	−0.1438	−0.1475
		(0.7033)	(0.7003)	(0.3877)	(0.3736)	(0.1752)	(0.1645)
PFE(3) Post est.	0.25	0.0877	0.0807	0.1105	0.1081	0.0614	0.0600
		(0.1967)	(0.1425)	(0.1682)	(0.1421)	(0.1157)	(0.0905)
	0.50	−0.0224	−0.0321	−0.0474	−0.0430	−0.0253	−0.0269
		(0.1821)	(0.1357)	(0.1264)	(0.0951)	(0.0875)	(0.0676)
	0.75	−0.2531	−0.2582	−0.2052	−0.2002	−0.1040	−0.1086
		(0.3030)	(0.2844)	(0.2476)	(0.2207)	(0.1431)	(0.1279)
2SFE	0.25	0.3921	0.3844	0.2598	0.2551	0.1563	0.1553
		(0.4285)	(0.4014)	(0.2882)	(0.2707)	(0.1844)	(0.1687)
	0.50	−0.0020	−0.0110	−0.0208	−0.0210	−0.0259	−0.0267
		(0.1640)	(0.1169)	(0.1162)	(0.0844)	(0.0879)	(0.0664)
	0.75	−0.4036	−0.4080	−0.3004	−0.2981	−0.1961	−0.2005
		(0.4355)	(0.4249)	(0.3262)	(0.3101)	(0.2155)	(0.2104)
CREM	0.25	−0.0164	−0.0253	−0.0225	−0.0247	−0.0210	−0.0185
		(0.2063)	(0.1416)	(0.1439)	(0.1029)	(0.1020)	(0.0739)
	0.50	−0.0004	−0.0092	−0.0103	−0.0073	−0.0062	−0.0071
		(0.1749)	(0.1237)	(0.1290)	(0.0916)	(0.0904)	(0.0638)
	0.75	−0.0018	−0.0052	0.0048	0.0057	0.0144	0.0093
		(0.2017)	(0.1398)	(0.1431)	(0.1030)	(0.1045)	(0.0745)

Table 3: Bias and root mean squared error (rmse) for simulation of (23) with $\gamma = 0$, i.e. no scale effect of the individual effects.

A few things should be mentioned here about the results. First, the choice of penalty parameter λ in practice is an unresolved problem. For this simulation we have chosen one that approximately sets 30% of the FE terms to zero, i.e. "some but not quite enough". To get the same relative impact, this needs to be calibrated as B varies for a given M, but not vice versa. Second, the reported root mean squared error results (rmse) cannot be compared across FE and CRE methods, since targets and variance of the "total error" terms are different.

The cross section quantile regression suffers a serious bias away from both targets. Note that it does converge to some quantity in the sense that, as the sample grows, rmse and bias are more or less equal. The dummy variable regression performs very well for the median but not the other quartiles. Penalizing the FE terms and estimating all quantiles simultaneously offers improvements in the tails and post-estimating the "selected model" seems to add a little to this improvement. As expected from the theory, as B grows these models perform better. The 2SFE model performs poorly, yet better than the dummy regression, both for $\gamma = 0$ and $\gamma = 1$. If one specifies Koenker's PFE method well, 2SFE is outperformed by it.

The CRE methods perform very well, and only the total sample size seems to matter (i.e. no incidental parameters curse). They seem to be close to unbiased, with diminishing variance. Table 9 in the Appendix shows that the FE methods cannot handle a scale effect of c_m, and that they become even more biased. This is not the case for the CRE methods, which still have high performance.

In conclusion, pooled cross-sectional estimation is rarely a good idea when there are omitted individual effects correlated with included variables if a "random assignment" interpretation is intended. It also appears to be the case that one can do much better than a pure dummy regression (at least in the tails). The FE methods suffer from the incidental parameters curse and high sensitivity to the data generating mechanism. It is possible to improve estimates in a FE setting by calibrating Koenker's approach, but it is hard in practice to determine whether it is done optimally.

The CRE methods in general have good performance, and do not seem to have trouble with either small B or scale effects of the individual effects. We need to stress again that they estimate something different from the FE methods, and these findings do not imply that one should discard the latter. It all depends on what the target of interest is.

In any case, with a short panel it appears that the CRE results are more reliable. From an economic policy perspective the CRE target is perhaps also more sensible in our birthweight application, as the politician is interested not in the individual as such, but in society as a whole. Why then condition on individual effects? It is part of the unexplained ranking mechanism. In this light, we will put more emphasis on the CRE results in our application. However, we will report a selection of FE results as well.

3 Data description

We now return from our methodological excursion to put our birthweight application back in the spotlight. The data which are used throughout the analyses are in part obtained from Aarhus University Hospital, Skejby, in Denmark. In the Aarhus region this hospital is the only one with a maternity ward, and thus the data in fact represent a broad population group, i.e. all economic and social classes. Furthermore, the data are enriched with socio-economic characteristics of the mothers. These additional data have been made available by Statistics Denmark and are linked by means of the Danish Civil Registration System.

The methods we have discussed above require a panel of mothers with two or more registered births. Only singleton births are included since multiple births babies (e.g. twins) tend to be lighter. Moreover, stillbirths are excluded, and thus the population of interest are singleton live births. For the variables of interest the data offer an unbalanced panel consisting of 16,602 births and 7,900 mothers. Except for the AD model, the estimation strategies discussed allow for an unbalanced panel, and this will be the primary data set. However, in the interest of comparison with this model, estimations have also been conducted on the basis of a balanced subset of the data which have been constructed with the first two births by each mother. This includes a total of 12,670 births. The descriptive statistics for the (unbalanced) dataset are given in Table 5. The sample ranges from the year 1992 to 2005. The included variables and their role in the analysis are the topic of the remainder of this section.

The choice of birthweight as the dependent variable gives rise to an important question: should gestational age be included as an explanatory variable? There is no doubt that gestational age is correlated with birthweight. However, in the present analysis our interest lies in the total effect of maternal smoking on birthweight, including any effects propagated through gestational age. Therefore it is not necessary to include gestational age as an explanatory variable, it might even be inappropriate as it could have undesirable effects due to multicollinearity. In this context it should also be emphasised that on the matter of not including gestational age as an explanatory variable, we follow recent leading econometric studies on birthweight, see e.g. Abrevaya (2006, 2001), Abrevaya and Dahl (2008), Chernozhukov and Fernández-Val (2011), and Koenker and Hallock (2001).

The primary regressors of interest regard the mothers' smoking habits. These are represented by two separate variables: whether or not mothers smoked at the time they became pregnant (*smoked before*), and whether or not they smoked during the pregnancy (*smoked during*). Both variables are binary, as the data unfortunately do not offer details on smoked quantities. The analysis therefore cannot account for the size of the treatment, which of course may be a drawback, since it seems reasonable to believe that quantity could be important. For identification of the separate effects of the two smoke variables, it is necessary that there are mothers who actually start or stop smoking when becoming pregnant. 2,019 of 16,602 births are given by mothers who stop smoking at the time of pregnancy (12.66%). On the other hand only 10 births are given by mothers who start smoking at the time of pregnancy (0.06%). Thus the change of behavior is therefore largely one-way.

The included variables can be roughly categorised into six categories. First, there are variables relating to the behaviour of the mothers. Already mentioned are the two smoke variables. Further, we include a variable, *drink*, which indicate whether or not the mother has consumed alcohol during the pregnancy.[8] Drinking habits are also believed to be harmful to the foetus and warnings are often explicitly printed on alcoholic containers. Related to the behavioural category is also the extent to which the mother actively tried to become pregnant. Another possibility is that the pregnancy was either unwanted or accidental. To control for this, we include a dummy variable for use of *birth control pills* within four months before becoming pregnant.

A second category that seems obviously related to the health of the baby, and thus possibly birthweight, is the general health or physical ability of the mother. An important variable in this category is occurrence of pregnancy *complications*, which we represent with an aggregated dummy variable which covers things such as premature contractions, bleedings, excessive vomiting, infections, and intrauterine growth restriction. Especially this last example is important, and may be caused by factors such as high blood pressure, heart disease, malnutrition, and substance abuse. A potential problem is that tobacco smoke may also be the source of this complication, and in effect leave us with an issue of separability of effects. This will be discussed further in the next section, where we present our results. The remaining variables in this category are the number of *doctor visits* and *prenatal visits* during pregnancy, whether the mother has had diabetes at one or more of the registered pregnancies, and finally whether artificial insemination was required.

We also wish to control for effects related to wealth status, and thus include yearly after-tax *income* in 1,000 DKK, yearly *unemployment benefits* in 1,000 DKK and finally *home size* measured in square meters. Here, the values are those registered for the year of pregnancy. This category may be related to e.g. the ability to ensure good surroundings and a proper diet etc.

It has also previously been found that there is a linkage between birthweight outcomes and socio-economic factors such as marital status and level of education, see e.g. Abrevaya and Dahl (2008). We therefore include variables that indicate if the mother was *married* during the pregnancy period, the mother's *education* (a categorical variable summarised in Table 4, and included as dummy variables), and whether the mother was a *student* when pregnant. The latter variable may indicate how freely time can be organised and may proxy for how stressful workdays are.

[8]Here, alcohol is defined as consumption of more than 1 Danish standard drink (12 grams of pure alcohol) per week.

The final two categories concern characteristics of the mother and child respectively. They include *height, weight*, and *age* of the mother (the latter two are also included in squares), and dummy variables for birth parity and the sex of the child.

Category	Description
0	No education.
1	Primary school (9 years compulsory, 1 year optional).
2	Secondary pre-university high school (3 years), or technical college, craftsmen, etc.(2–5 years).
3	College: short-cycle higher education programme (1–2 years).
4	College: medium-cycle higher education programme (3–4 years).
5	3-year academic (Bachelor) degree.
6	5-year academic (Master) degree.
7	PhD degree and above.

Table 4: Description of *education* categories.

A natural concern is whether or not there are relevant seasonality effects which should be controlled for. Buckles and Hungerman (2008) discuss whether the time of year affects birthweight and conclude that this is the case. They attribute such effects to a strong correlation with socio-economic characteristics, which are well represented in our data. Dehejia and Lleras-Muney (2004) investigate effects of unemployment rates on babies' health, and suggest that high unemployment is positively correlated with healthy babies. This is just one example of general year-specific phenomena which may have effects which are desirable to control for. In our analyses we do this by including birth-year dummy variables.

The overall choice of variables is in part motivated by previous studies such as Abrevaya and Dahl (2008) and Koenker and Hallock (2001). Some results are therefore comparable, and may confirm previous findings. To analyse the effect of maternal smoking, we use data on smoking both during and before pregnancy, allowing for smoke to have a causal effect in different ways. This approach differs from previous studies and relates to the discussion of whether "last-minute" intervention could be effective.

	1st child		2nd child		3rd child		4th child	
Variable	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
Birthweight	3503.82	(534.17)	3650.23	(531.20)	3674.22	(554.12)	3624.56	(547.75)
Smoked during	0.14		0.13		0.16		0.21	
Smoked before	0.30		0.24		0.25		0.28	
Drink	0.04		0.03		0.04		0.03	
Birth control pills	0.26		0.16		0.14		0.11	
Complications	0.22		0.25		0.28		0.26	
Doctor visits	3.07		3.00		2.94		2.80	
Prenatal visits	5.17		4.74		4.56		4.47	
Test tube baby	0.02		0.01		0.00		0.01	
Diabetes	0.01		0.01		0.02		0.03	
Income	107.04	(40.65)	135.62	(105.72)	149.39	(67.37)	151.94	(50.48)
Unemployment benefits	6.43	(17.41)	6.97	(18.57)	6.29	(17.98)	4.37	(14.52)
Home size	93.64	(45.20)	112.74	(44.26)	125.59	(43.71)	131.47	(44.26)
Married	0.43		0.65		0.76		0.72	
Student	0.23		0.14		0.09		0.08	
Education cat. 0	0.01		0.01		0.01		0.01	
Education cat. 1	0.13		0.12		0.17		0.29	
Education cat. 2	0.49		0.42		0.37		0.32	
Education cat. 3	0.05		0.06		0.04		0.03	
Education cat. 4	0.18		0.23		0.25		0.21	
Education cat. 5	0.05		0.04		0.03		0.02	
Education cat. 6	0.09		0.12		0.12		0.10	
Education cat. 7	0.00		0.01		0.01		0.01	
Height	168.56	(6.04)	168.57	(6.07)	168.20	(6.03)	167.36	(5.95)
Weight	63.92	(10.83)	65.24	(11.89)	65.79	(12.37)	65.71	(12.40)
Age	27.57	(3.75)	30.34	(3.82)	32.47	(3.83)	33.99	(4.18)
Male child	0.51		0.50		0.51		0.52	

	Birthweight quantiles			
Quantile	1st child	2nd child	3rd child	4th child
10%	2880	3030	3030	3002
25%	3200	3330	3350	3300
50%	3500	3650	3660	3650
75%	3850	4000	4020	3990
90%	4150	4300	4350	4288
Observations	6642	7416	2181	363

Table 5: Descriptive statistics for the Aarhus Birth Cohort.

4 Results

We consider our main estimation results to be the CRE estimates; especially those for the unbalanced panel and the CREM specification. From an economic policy perspective, the interpretation of the CRE estimates seems most appropriate: to the politician unobserved individual effects should be part of the unexplained ranking or distribution mechanism, yet for estimation purposes some notion of random assignment is called for to take into account the dependence between the unobserved and included covariates. Also, the CRE estimators are not cursed by incidental parameters. Their specifications give no reason why they should be, and indeed our simulations confirmed their good performance, also for short panels.

The FE estimators, on the other hand, are questionable in such a setting. We give some insight into some bounds for some available calibration options and argue that, even though the estimators are cursed, they seem to lead to conclusions that are difficult to refute.

Investigating the alleged negative effects of smoking behavior on birthweight outcomes is a prime objective in this analysis. Therefore we initiate the presentation of our results with a detailed discussion of the matter, using evidence from our battery of estimators, after which we elaborate on some of our other findings. For a good discussion on the interpretation of quantile treatment effects, see for example Firpo (2007).

Table 6 summarizes the results for the *smoked during* variable as estimated by a variety of methods. These surely have one thing in common: a statement that prenatal smoking do not have negative direct effects on birthweight would be hard to justify, given the evidence from any of the estimators.

Consider first the results for the unbalanced data set.[9] Overall, the cross-section estimates provide the largest estimates (in absolute value) of the smoking effect. This holds, not only for the methods and calibrations shown here, but for all the many variations we have tried. The CREM "correction" estimates, S(CREM), absorbs an increasing part of the large effect alleged by a pure cross-section estimator, the further we move to the right in the birthweight distribution. This indicates that in the left tail, where we then have the largest adverse effect smoke during pregnancy, dependence between smoke and unobserved individual effects does not interfere much. Supposedly, there is more such "joint dependence" with birthweight at the larger quantiles. This supports a conclusion that smoking is more severe where it hurts the most: where babies are already prone to be low achievers when it comes to birthweight. In fact, all estimators except for 2SFE, lead to the conclusion that the adverse effect *relative* to birthweight is increasing to the left in the distribution when comparing point estimates to birthweight quantiles reported in Table 5. The 2SFE estimator here predicts a constant relative effect.

The PFE estimator can generally be calibrated to give results that lie between those of a pure dummy-variable estimator and a pure cross-section estimator. We will shortly discuss this in a little more detail. In our application, a pure dummy-variable regression is numerically infeasible. Letting $\lambda \to 0$ to approach the dummy-variable estimates, we learn that the effect of smoking decreases in absolute value. No value of λ, however, leads to effects of smoking as small as claimed by the 2SFE estimator. The value of λ is non-negligible, and we are not aware of a practical rule for choos-

[9]For all our estimations we have used a blocked pairwise subsampling bootstrap, as deemed appropriate by Abrevaya and Dahl (2008). The idea is that when sampling a mother, all her births are included to deal with the dependence in the observations. In practice we sample mothers (with replacement) until the desired sample size is obtained in number of births. For unbalanced panels the sub-sample size would vary if it was specified by a number of mothers. For robustness, we have compared various sub-sample sizes, and using larger samples did not alter the reported standard errors noticeably.

Estimates for *smoked during*		Quantile Regressions				
		10%	25%	50%	75%	90%
Unbalanced data	CS	-188.063 ***	-181.629 ***	-169.165 ***	-177.479 ***	-200.441 ***
		(28.681)	(22.116)	(18.526)	(21.528)	(27.491)
	CREM	-190.485 ***	-112.107 ***	-75.991 ***	-90.337 ***	-2.081
		(49.047)	(35.800)	(27.952)	(34.573)	(49.101)
	S(CREM)	1.511	-81.205 *	-118.446 ***	-118.538 ***	-224.515 ***
		(60.811)	(47.264)	(36.936)	(44.050)	(57.950)
	PFE(5), $\lambda = 0.8$	-161.989 ***	-163.087 ***	-155.812 ***	-148.444 ***	-167.575 ***
		(26.249)	(18.478)	(16.592)	(19.996)	(24.541)
	PFE(5), $\lambda = 0.8$, p.e.	-162.821 ***	-156.776 ***	-146.783 ***	-186.265 ***	-183.513 ***
		(32.061)	(24.873)	(24.113)	(25.319)	(32.148)
	2SFE	-70.869 ***	-83.486 ***	-99.452 ***	-104.189 ***	-108.669 ***
		(26.305)	(21.007)	(19.840)	(20.379)	(24.320)
Balanced data	CREM	-247.002 ***	-154.444 ***	-88.071 ***	-110.895 ***	3.708
		(54.908)	(39.487)	(32.845)	(41.679)	(59.003)
	S(CREM)	76.550	-24.508	-98.878 **	-81.969	-214.324 ***
		(67.040)	(48.564)	(39.680)	(50.903)	(69.351)
	AD	-231.362 ***	-174.003 ***	-57.873	-126.478 ***	12.694
		(58.602)	(39.997)	(35.973)	(44.777)	(61.797)
	S_1(AD)	34.427	17.522	-52.118 *	19.774	-32.024
		(48.437)	(33.175)	(30.156)	(36.824)	(51.533)
	S_2(AD)	37.590	-21.545	-77.352 **	-88.525 **	-184.486 ***
		(48.435)	(37.411)	(34.246)	(41.459)	(54.086)

Asterisks denote the significance level (double-sided). *: 10%, **: 5%, ***: 1%.
Bootstrapped standard errors are given in parentheses. The bootstrap was done using a sample size of 3,000 births and 499 iterations.

Table 6: Results for *smoked during* from a selection of estimators. The $S(\cdot)$-results are for the added CRE variables constructed from repeated measurements of *smoke during* (see the last part of Section 2). For the PFE estimates, λ refers to the penalty parameter, and "p.e." is where the model is re-estimated without penalty and zero-FE-terms.

ing it appropriately. However, our simulations confirm that the performance of the PFE estimator can be calibrated to outperform the 2SFE estimator. Since the latter acts like a lower bound in our case, there is strong evidence that there are significant direct adverse effects from smoke during pregnancy, also from a fixed effects perspective. However, there is much uncertainty about how adverse. The choice $\lambda = 0.8$, for which the results are reported in the table, penalizes to a degree where 30% of the mothers share intercept, and the remaining 70% are sufficiently different to get their own. Re-estimating the selected model has a more pronounced effect in the right tail. An argument for re-estimation is that the penalty affects the non-zero FE terms, whereas an argument against re-estimation is to preserve some degree of control over the size of estimated individual effects.

In the lower part of Table 6, we provide CRE results from the smaller balanced panel. We include this mainly to show that the AD and CREM specifications give very similar estimates, justifying the more simple CREM specification of S which then allows for the inclusion of more observations given by the unbalanced panel. The point estimates for the balanced panel indicate a slightly larger effect, compared to the unbalanced one, even though we cannot deem them statistically different. One explanation, however, could be that mothers who give birth to unhealthy babies (in terms of low birthweight) choose not to get a third or fourth child, this resulting in some kind of sample selection issue. This would again point to the unbalanced panel as the more appropriate.

A general point that we need to emphasize is that even at the first decile births are not categorised as low birthweight (often defined as 2,500 grams), cf. Table 5. Unfortunately, it is not possible to obtain reasonable results for lower quantiles due to the very few extreme observations. However, it does not seem reasonable to expect the adverse effects of smoking to diminish as we move into the extreme left of the distribution. In fact, the trend in the CRE models suggests exactly the opposite. To illustrate the trend visually we have plotted point CRE estimates for a whole range of quantiles in Figure 2 (left panel). For comparison we show some FE estimates in the right panel. As PFE(k) estimations are problematic for such a “grid”, we use a PFE(1) specification. The right panel, where we also include the cross-section estimates, also serves to illustrate how estimates move as a function of λ.

To add a little insight into what happens when varying λ, we consider the PFE(5) point estimates for *smoked during* and the ratio of FE terms shrunken to zero as a function of λ. We present this sensitivity analysis in Figure 3. The top panel shows how the effect decreases at all quantiles as we penalize more. The trend stops at $\lambda \approx 1.6$ where the bottom panel shows that almost all FE terms are set to zero. We also show the penalized ratio for PFE(1) . It seems that FE terms are (fully) affected by the penalty in chunks. This feature is most pronounced for PFE(1), and it appears that this effect is smoothed out for PFE(k) as k increases.

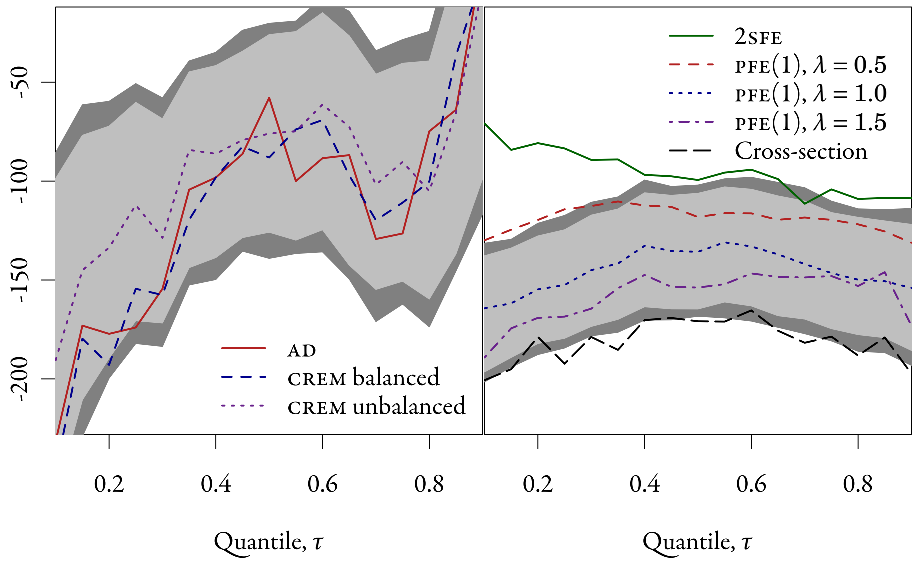


Figure 2: *Smoke during* estimates plotted for a grid of quantile indices and estimators. The light gray areas show the bootstrapped 90% confidence intervals for the unbalanced CREM model (left) and the PFE(1) model with $\lambda = 1$ (right), and the dark gray areas extend these to 95% confidence intervals.

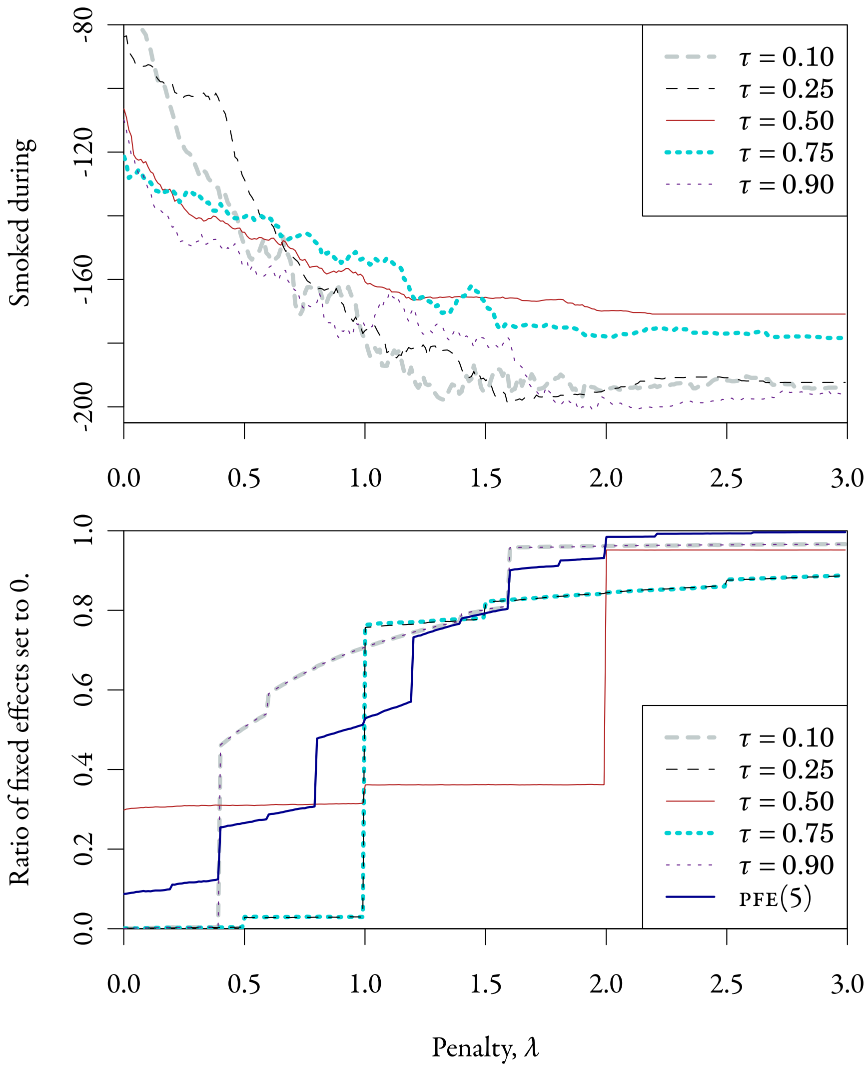


Figure 3: The top panel shows the PFE(5) estimates for *smoked during* as the penalty parameter λ varies. The bottom panel shows the ratio of FE-terms shrunken to zero because of the penalty as a function of λ.

Our analysis, of course, also includes data on whether the mothers *smoked before* their pregnancies. Interestingly, however, this seems to be unimportant when it comes to birthweight: no estimators find noteworthy significance. We therefore do not present a similar discussion for this variable, although some results can be found in Appendix B. We find the absence of statistical significance a very important and quite comforting result. It suggests that smoking behavior prior to pregnancy is not crucial for birthweight outcome, and it gives future mothers who are smoking a very good argument to quit in time. Intervention policy can therefore prove profitable, as reducing the number of low birthweight babies reduces both monetary as well as socio-economic costs.

Maternal smoking is also one of the prime interests in the study by Abrevaya and Dahl (2008), which is based on American data, more specifically natality data from Washington and Arizona. In their study they only have a smoke variable comparable to our *smoked during*, but for this variable they also find significant negative effects. However, their findings suggest somewhat less severe effects, in the range around -80 to -60 grams. Whether this can in fact be attributed to actual differences or is a consequence of measurement error is hard to say, but the latter could perhaps be attributed to smoking being less of a taboo in Denmark, which could lead to a lower degree of misreporting. Furthermore, they find that cross sectional estimates exaggerate the effect of smoking even more than in this case.

In addition to the variables pertaining to maternal smoking our analysis also contains a number of other interesting variables. A set of results for the CREM model using the unbalanced dataset is given in Table 7. Again, we will focus on these results and only consider those from the other models on a few occasions. A complete set of results for all the models is available as a separate appendix, and the most relevant results are listed in Appendix B. For the CRE-specifications we use all variables that vary from birth to birth to construct S_m, except education and year dummy-variables. The variables *height* and *diabetes* do not vary in our sample and are therefore not used, either for S_m or in the fixed effects model.

One behavioural aspect which receives great attention, especially related to pregnancies, and which is the subject of much societal debate, is drinking habits. This is also a topic where policy is conducted in the attempt to affect people's behaviour, e.g. taxes and age restrictions on the purchase of alcohol. In the light of the strong belief that alcohol intake during pregnancy has negative health implications on the foetus, it seems puzzling that these estimations show no significance in the CREM model and no or very little in the other models considered. There could be many reasons why no significance shows up in our results, one of which could be measurement or reporting errors. Even so, drinking may be the cause of many other health implications which are not related to birthweight.

In the behaviour category we also have the variable *birth control pills*. As mentioned earlier this can be thought to proxy for whether the pregnancy was planned

	Quantile Regressions					
	10%	25%	50%	75%	90%	OLS
Smoked during	-190.485 ***	-112.107 ***	-75.991 ***	-90.337 ***	-2.081	-94.897 ***
	(49.047)	(35.800)	(27.952)	(34.573)	(49.101)	(23.618)
Birth control pills	-27.853	-52.478 ***	-28.928 **	-21.657	-19.116	-33.813 ***
	(25.097)	(18.130)	(14.118)	(16.678)	(22.708)	(11.650)
Complications	-122.386 ***	-67.361 ***	-46.073 ***	-29.562 **	-11.029	-65.723 ***
	(24.632)	(16.595)	(12.271)	(14.727)	(21.520)	(10.636)
Prenatal visits	109.582 ***	90.162 ***	77.477 ***	80.245 ***	73.484 ***	64.011 ***
	(11.583)	(5.906)	(4.583)	(5.080)	(6.969)	(11.670)
Test tube baby	-2.683	64.214	30.674	-57.277	-251.771 ***	-34.417
	(108.153)	(70.387)	(61.737)	(63.434)	(86.622)	(46.240)
Diabetes	181.554 *	217.704 ***	280.896 ***	315.398 ***	373.631 ***	282.529 ***
	(94.004)	(70.214)	(55.456)	(62.333)	(88.341)	(54.962)
Student	20.743	4.311	1.137	10.260	51.933 *	17.787
	(31.961)	(22.674)	(21.060)	(22.319)	(29.430)	(15.434)
Height	6.662 ***	8.423 ***	10.489 ***	11.419 ***	11.183 ***	10.269 ***
	(1.525)	(1.088)	(0.975)	(1.074)	(1.320)	(0.906)
Weight	5.722	2.412	-8.622	2.040	23.161 **	7.950
	(14.546)	(9.823)	(8.403)	(8.182)	(11.050)	(6.514)
Weight^2	-0.026	0.008	0.063	-0.010	-0.127 *	-0.036
	(0.099)	(0.066)	(0.058)	(0.054)	(0.071)	(0.043)
Age	-40.769	-24.258	-19.409	-29.653 *	-0.204	-14.971
	(30.152)	(20.280)	(16.161)	(17.768)	(26.393)	(14.840)
Age^2	0.867 *	0.516	0.402	0.465	0.065	0.333
	(0.467)	(0.319)	(0.256)	(0.285)	(0.420)	(0.236)
Second child	177.725 ***	162.143 ***	150.112 ***	175.701 ***	156.739 ***	162.162 ***
	(21.562)	(14.632)	(12.540)	(14.267)	(19.457)	(11.515)
Third child	213.334 ***	191.121 ***	190.076 ***	252.706 ***	231.066 ***	209.153 ***
	(33.522)	(24.012)	(22.312)	(24.301)	(31.575)	(20.485)
Fourth child	172.238 ***	182.380 ***	181.915 ***	233.246 ***	202.214 ***	191.631 ***
	(56.800)	(41.156)	(37.683)	(40.316)	(51.240)	(34.217)
Male child	121.921 ***	128.415 ***	137.299 ***	162.926 ***	186.615 ***	145.904 ***
	(16.220)	(12.039)	(10.115)	(12.611)	(16.705)	(8.042)

Asterisks denote the significance level (double-sided). *: 10%, **: 5%, ***: 1%.
Bootstrapped standard errors are given in parentheses. The bootstrap was done using a sample size of 3,000 births and 499 iterations.

Table 7: Results for the CREM estimation using the unbalanced data set. Insignificant variables are not reported here but in Appendix B. These are: *smoked before, drink, doctor visits, income, unemployment benefits, home size, married,* and all education categories. Dummy variables for birth year are mostly significant but removed from this table in the interest of space. Results for the constructed CRE variables (i.e. those in S_m) can also be found in Appendix B. The OLS estimates are from a Mundlak regression with the projection being the same as S_m.

or not. We do see some signs of significance for this variable, but the extent of this varies between models. However, in general the point estimates are negative as would be expected if it indeed acts as a proxy for unwanted pregnancies.

In the health category of variables we have three variables which show clearly significant effects. *Prenatal visits* are of special interest since it is a preventive measure intended to ensure good health of the foetus, and therefore its effect is of great interest to policy makers. The main problem with *prenatal visits*, however, is that there may be two reasons for consulting a midwife, either as a routine/precautionary measure or because of complications. It is not possible to directly distinguish between these two effects of the variable. Therefore the estimations also include *complications*, which in part controls for this, thus leaving us with the preventive effect. We see that *prenatal visits* are significant, indicating a positive preventive effect. Further, *complications* have a significant negative effect as would be expected. However, this variable is problematic, as indicated in the last section, since it includes cases of intrauterine growth restrictions, which may be one of the channels through which smoking reduces birthweight. It is not possible to separate out this part of the variable, and consequently as a robustness check regressions have been run without *complications*. The resulting outcome had only minor changes in the point estimates and did not alter any conclusions. On this basis it is concluded that the prevalence of intrauterine growth restrictions does not constitute a problem for the interpretation of the results, in particular those for *smoked during*.

The last significant variable in this category is *diabetes*, which has a positive effect on the right tail of the birthweight distribution. This is in accordance with the medical literature, where diabetes is commonly accepted as a birthweight-increasing factor. Finally, both *doctor visits* and *test tube baby* show only little significance in the CREM model. In a few cases, they show moderate significance in the fixed effects models. The former can be thought of as a general measure of the mother's health. However, it is hard to say how good a proxy it really is, since it represents, not only birth-related health, but also general illness or even hypochondria. The fact that the latter is mostly insignificant need not say anything about causality, but may be due to a very small number of test tube babies in the sample.

The variables in the wealth and socio-economic categories are in general all insignificant. We do, however, see two exceptions. First, the variable *student* is mostly significant in the fixed effects models, which is in contrast to the CRE models. Second, the education variables do show moderate signs of significance in some of the fixed effects specifications, but there is not any general consensus on the significance between the models. That these categories are largely unimportant is not particularly unexpected when considering the welfare system in Denmark, where the social benefits available in general (and to mothers in particular) are quite generous. This is in contrast to the results from e.g. Abrevaya and Dahl (2008). They find, for instance, that marital status is highly significant. A reason for this difference could be that

America has a substantial social gap compared to Denmark. This will undoubtedly have consequences for unmarried mothers in America, who do not have the same social benefits as offered in Denmark. Another, perhaps more subtle reason could be the extent to which marriage can proxy for unobserved characteristics or ability of women. The choice of why and when to get married may be culturally dependent, which is supported by the descriptive statistics. There is quite a difference in proportions of pregnancies in and out of wedlock in their American data and our Danish data, which suggests that it is more uncommon to have children out of wedlock in America. When combined, these arguments may be used to explain why the American data suggest that marriage has a positive effect and no such evidence is found in the Danish data.

Abrevaya and Dahl also find that education has a significant effect, while we find little evidence of such an effect. This could very well be due to the costs associated with education in America. This is in contrast to Denmark where education is free. The variable may therefore proxy for wealth status which, as argued before, seems irrelevant in Denmark.

The mother's characteristics are largely insignificant in the "main" terms of the CREM specification (those in X_{mb}). The augmented CRE terms, however, do show some significance (those in S_m). This could be interpreted as a "part" of the heterogeneity, e.g. for *weight*: the overall stature of the mother can be more important than birth-specific fluctuations. *Height*, is highly significant. However, no extra CRE variable is constructed for height because it is birth-invariant. The fact that the *height* and *weight* variables are significant, for one part of the specification or the other, seems very natural. What is puzzling, though, is that *age* does not appear to have much significance. Abrevaya and Dahl (2008) find a significant effect of *age*, and the literature suggests that there is an optimal age, see e.g. Royer (2004).

For these variables the fixed effects models differ considerably. First, because the variable *height* is birth invariant in our sample it cannot be included in these models and should instead be captured by the fixed effect. Second, *weight* is in most cases significant. This is to be expected as it cannot be captured by the fixed effects, but will most likely affect birthweight. Finally, *age* does show moderate signs of significance in some of the fixed effects specifications, but not to a degree where we are confident enough to draw any firm conclusions.

Regarding child characteristics, we find that the parity variables *second, third* and *fourth child,* and *male child* are significant and positive across all quantiles. This is to be expected since it is generally acknowledged that the birthweight of male children is higher on average, and that birthweight increases with parity of the mother. This confirms the results of previous studies.

5 Concluding remarks

In this paper we have found strong evidence that smoking during pregnancy has adverse consequences for birthweight outcomes. The documented connection between babies' birthweight and their overall health, along with the costs associated with low birthweight, makes this a very important result. The effect appears to worsen the further one moves to the left in the birthweight distribution, especially when measured relative to birthweight at the corresponding quantiles.

The significant effect of smoking has been documented before, but we add to these results in several ways. The richness of the applied data set allowed us to control for many potentially important characteristics which were not included in previous studies. Furthermore, we use several estimators and provide a detailed discussion of their differences in interpretation and performance. Given the results from this battery of estimators, the adverse effect of smoking on birthweight seems irrefutable, regardless of estimation approach and which of the two discussed interpretations is desired for the estimated coefficients.

As icing on the cake, our analysis used information on smoking behavior prior to pregnancy, allowing for a separation of effects. Only smoking during pregnancy has a pronounced significant effect, a result speaking for intervention campaigns as a worthwhile activity.

6 Acknowledgements

We are grateful to Jason Abrevaya, Niels Haldrup, Marianne Simonsen, Chiara Orsini, Ulrich Kaiser, Tine B. Henriksen, and Kirsten Wisborg for helpful comments. Lise V. Hansen offered invaluable assistance by providing access to the data from Aarhus University Hospital, Skejby. The terms of the data-sharing agreements do not allow release of these data. However, codes for estimation in R are available from the corresponding author upon request. We would also also like to thank Roger Koenker, who wrote the `quantreg` package for R that was used as starting point for the quantile estimations in this paper. The authors gratefully acknowledge financial support from the Danish Social Sciences Research Council (grant 275-06-0105) and the Center for Research in Econometric Analysis of Time Series, CREATES, funded by The Danish National Research Foundation.

A Simulation results

	τ	$B = 2$, $M = 499$	$B = 2$, $M = 999$	$B = 3$, $M = 499$	$B = 3$, $M = 999$	$B = 5$, $M = 499$	$B = 5$, $M = 999$
Cross-section FE target	0.25	−0.3380	−0.3402	−0.3412	−0.3443	−0.3435	−0.3405
		(0.3713)	(0.3561)	(0.3618)	(0.3538)	(0.3560)	(0.3467)
	0.50	−0.3724	−0.3740	−0.3755	−0.3718	−0.3718	−0.3722
		(0.3968)	(0.3867)	(0.3920)	(0.3808)	(0.3813)	(0.3770)
	0.75	−0.4321	−0.4344	−0.4310	−0.4282	−0.4186	−0.4245
		(0.4598)	(0.4468)	(0.4489)	(0.4373)	(0.4301)	(0.4303)
Dummy regression	0.25	0.6717	0.6626	0.3531	0.3474	0.1438	0.1431
		(0.6980)	(0.6770)	(0.3813)	(0.3619)	(0.1763)	(0.1595)
	0.50	−0.0028	−0.0119	−0.0047	−0.0055	−0.0035	−0.0053
		(0.1898)	(0.1395)	(0.1272)	(0.0904)	(0.0865)	(0.0645)
	0.75	−0.6773	−0.6864	−0.3596	−0.3598	−0.1438	−0.1475
		(0.7033)	(0.7003)	(0.3877)	(0.3736)	(0.1752)	(0.1645)
PFE(3)	0.25	−0.0873	−0.0958	−0.0493	−0.0513	−0.0927	−0.0911
		(0.1872)	(0.1486)	(0.1283)	(0.0981)	(0.1322)	(0.1129)
	0.50	−0.1391	−0.1425	−0.1534	−0.1530	−0.1426	−0.1444
		(0.2089)	(0.1818)	(0.1887)	(0.1722)	(0.1641)	(0.1560)
	0.75	−0.3664	−0.3694	−0.2452	−0.2396	−0.1828	−0.1889
		(0.3977)	(0.3852)	(0.2753)	(0.2547)	(0.2062)	(0.2009)
PFE(3) Post est.	0.25	0.0877	0.0807	0.1105	0.1081	0.0614	0.0600
		(0.1967)	(0.1425)	(0.1682)	(0.1421)	(0.1157)	(0.0905)
	0.50	−0.0224	−0.0321	−0.0474	−0.0430	−0.0253	−0.0269
		(0.1821)	(0.1357)	(0.1264)	(0.0951)	(0.0875)	(0.0676)
	0.75	−0.2531	−0.2582	−0.2052	−0.2002	−0.1040	−0.1086
		(0.3030)	(0.2844)	(0.2476)	(0.2207)	(0.1431)	(0.1279)
PFE(1)	0.25	−0.1993	−0.2071	−0.1275	−0.1322	−0.1027	−0.1006
		(0.2582)	(0.2368)	(0.1725)	(0.1542)	(0.1395)	(0.1207)
	0.50	−0.2239	−0.2315	−0.1827	−0.1818	−0.1331	−0.1344
		(0.2615)	(0.2520)	(0.2144)	(0.1989)	(0.1570)	(0.1477)
	0.75	−0.3690	−0.3749	−0.2996	−0.2943	−0.2865	−0.2934
		(0.3980)	(0.3902)	(0.3247)	(0.3066)	(0.3009)	(0.3008)
2SFE	0.25	0.3921	0.3844	0.2598	0.2551	0.1563	0.1553
		(0.4285)	(0.4014)	(0.2882)	(0.2707)	(0.1844)	(0.1687)
	0.50	−0.0020	−0.0110	−0.0208	−0.0210	−0.0259	−0.0267
		(0.1640)	(0.1169)	(0.1162)	(0.0844)	(0.0879)	(0.0664)
	0.75	−0.4036	−0.4080	−0.3004	−0.2981	−0.1961	−0.2005
		(0.4355)	(0.4249)	(0.3262)	(0.3101)	(0.2155)	(0.2104)
Cross-section CRE target	0.25	−0.3825	−0.3848	−0.3858	−0.3889	−0.3881	−0.3851
		(0.4123)	(0.3989)	(0.4042)	(0.3973)	(0.3992)	(0.3906)
	0.50	−0.3724	−0.3740	−0.3755	−0.3718	−0.3718	−0.3722
		(0.3968)	(0.3867)	(0.3920)	(0.3808)	(0.3813)	(0.3770)
	0.75	−0.3875	−0.3898	−0.3864	−0.3836	−0.3740	−0.3799
		(0.4182)	(0.4036)	(0.4063)	(0.3937)	(0.3869)	(0.3864)
CREM	0.25	−0.0164	−0.0253	−0.0225	−0.0247	−0.0210	−0.0185
		(0.2063)	(0.1416)	(0.1439)	(0.1029)	(0.1020)	(0.0739)
	0.50	−0.0004	−0.0092	−0.0103	−0.0073	−0.0062	−0.0071
		(0.1749)	(0.1237)	(0.1290)	(0.0916)	(0.0904)	(0.0638)
	0.75	−0.0018	−0.0052	0.0048	0.0057	0.0144	0.0093
		(0.2017)	(0.1398)	(0.1431)	(0.1030)	(0.1045)	(0.0745)
AD	0.25	−0.0126	−0.0247	−0.0217	−0.0227	−0.0191	−0.0181
		(0.2028)	(0.1399)	(0.1432)	(0.1029)	(0.1021)	(0.0728)
	0.50	0.0005	−0.0093	−0.0086	−0.0077	−0.0066	−0.0071
		(0.1723)	(0.1237)	(0.1287)	(0.0919)	(0.0905)	(0.0635)
	0.75	−0.0066	−0.0060	0.0028	0.0062	0.0127	0.0095
		(0.2018)	(0.1391)	(0.1424)	(0.1026)	(0.1046)	(0.0749)

Table 8: Bias and root mean squared error (rmse) for simulation of (23) with $\gamma = 0$, i.e. no scale effect of the individual effects.

	τ	$B = 2$		$B = 3$		$B = 5$	
		$M = 499$	$M = 999$	$M = 499$	$M = 999$	$M = 499$	$M = 999$
Cross-section FE target	0.25	−0.1453	−0.1508	−0.1465	−0.1493	−0.1482	−0.1487
		(0.1853)	(0.1688)	(0.1739)	(0.1641)	(0.1671)	(0.1591)
	0.50	−0.4118	−0.4218	−0.4112	−0.4131	−0.4159	−0.4170
		(0.4326)	(0.4307)	(0.4252)	(0.4205)	(0.4240)	(0.4213)
	0.75	−0.5625	−0.5662	−0.5608	−0.5582	−0.5664	−0.5645
		(0.5855)	(0.5777)	(0.5774)	(0.5665)	(0.5753)	(0.5698)
Dummy regression	0.25	0.6746	0.6619	0.4037	0.4078	0.2317	0.2344
		(0.6973)	(0.6714)	(0.4241)	(0.4179)	(0.2482)	(0.2430)
	0.50	−0.0000	−0.0127	−0.0027	0.0003	−0.0020	−0.0005
		(0.1766)	(0.1134)	(0.1126)	(0.0826)	(0.0761)	(0.0573)
	0.75	−0.6744	−0.6871	−0.4076	−0.4071	−0.2327	−0.2313
		(0.6971)	(0.6963)	(0.4270)	(0.4179)	(0.2478)	(0.2395)
PFE(3)	0.25	0.1404	0.1339	0.2401	0.2387	0.1974	0.2002
		(0.1979)	(0.1605)	(0.2655)	(0.2514)	(0.2160)	(0.2099)
	0.50	−0.1861	−0.1970	−0.1769	−0.1756	−0.1575	−0.1571
		(0.2367)	(0.2179)	(0.2049)	(0.1910)	(0.1745)	(0.1667)
	0.75	−0.5255	−0.5303	−0.4710	−0.4708	−0.4405	−0.4393
		(0.5514)	(0.5426)	(0.4889)	(0.4803)	(0.4517)	(0.4452)
PFE(3) Post est.	0.25	0.3908	0.3829	0.3568	0.3556	0.3281	0.3289
		(0.4276)	(0.3999)	(0.3795)	(0.3667)	(0.3424)	(0.3363)
	0.50	−0.0235	−0.0364	−0.0378	−0.0369	−0.0209	−0.0185
		(0.1717)	(0.1140)	(0.1137)	(0.0866)	(0.0778)	(0.0593)
	0.75	−0.5240	−0.5309	−0.4238	−0.4220	−0.3684	−0.3682
		(0.5515)	(0.5447)	(0.4442)	(0.4331)	(0.3797)	(0.3745)
PFE(1)	0.25	−0.0206	−0.0255	0.0462	0.0446	0.0653	0.0668
		(0.1284)	(0.0848)	(0.1107)	(0.0834)	(0.1002)	(0.0874)
	0.50	−0.2401	−0.2503	−0.1690	−0.1690	−0.1137	−0.1118
		(0.2737)	(0.2633)	(0.1992)	(0.1855)	(0.1363)	(0.1257)
	0.75	−0.5009	−0.5108	−0.4130	−0.4143	−0.4239	−0.4217
		(0.5255)	(0.5229)	(0.4299)	(0.4235)	(0.4341)	(0.4272)
2SFE	0.25	0.5075	0.5054	0.3986	0.3962	0.2813	0.2839
		(0.5331)	(0.5167)	(0.4157)	(0.4048)	(0.2940)	(0.2902)
	0.50	−0.0004	−0.0091	−0.0226	−0.0220	−0.0273	−0.0270
		(0.1585)	(0.1052)	(0.1112)	(0.0823)	(0.0799)	(0.0611)
	0.75	−0.5913	−0.5983	−0.5150	−0.5165	−0.4201	−0.4197
		(0.6162)	(0.6104)	(0.5314)	(0.5256)	(0.4294)	(0.4250)
Cross-section CRE target	0.25	−0.2084	−0.2139	−0.2095	−0.2124	−0.2113	−0.2118
		(0.2380)	(0.2269)	(0.2296)	(0.2230)	(0.2249)	(0.2192)
	0.50	−0.3865	−0.3965	−0.3859	−0.3878	−0.3906	−0.3918
		(0.4087)	(0.4060)	(0.4008)	(0.3957)	(0.3992)	(0.3963)
	0.75	−0.5083	−0.5120	−0.5066	−0.5040	−0.5122	−0.5104
		(0.5336)	(0.5247)	(0.5250)	(0.5132)	(0.5220)	(0.5162)
CREM	0.25	−0.0061	−0.0116	−0.0107	−0.0126	−0.0256	−0.0243
		(0.1600)	(0.1084)	(0.1138)	(0.0802)	(0.0853)	(0.0649)
	0.50	−0.0217	−0.0310	−0.0118	−0.0095	−0.0066	−0.0081
		(0.1890)	(0.1255)	(0.1254)	(0.0945)	(0.0911)	(0.0657)
	0.75	−0.0002	−0.0084	−0.0031	−0.0048	−0.0144	−0.0103
		(0.2146)	(0.1518)	(0.1538)	(0.1107)	(0.1075)	(0.0784)
AD	0.25	−0.0076	−0.0137	−0.0129	−0.0132	−0.0232	−0.0231
		(0.1586)	(0.1056)	(0.1161)	(0.0803)	(0.0842)	(0.0646)
	0.50	−0.0201	−0.0271	−0.0115	−0.0094	−0.0059	−0.0089
		(0.1889)	(0.1243)	(0.1274)	(0.0939)	(0.0901)	(0.0659)
	0.75	0.0010	−0.0063	−0.0017	−0.0053	−0.0149	−0.0103
		(0.2167)	(0.1513)	(0.1491)	(0.1102)	(0.1072)	(0.0792)

Table 9: Bias and root mean squared error (rmse) for simulation of (23) with $\gamma = 1$, i.e. the individual effects have scale effects.

B Empirical results

	Quantile Regressions					
	10%	25%	50%	75%	90%	OLS
Smoked during	-231.362 ***	-174.003 ***	-57.873	-126.478 ***	12.694	-109.920 ***
	(58.602)	(39.997)	(35.973)	(44.777)	(61.797)	(27.174)
Smoked before	-14.954	-9.730	-14.592	-21.266	-32.916	-17.915
	(45.387)	(32.215)	(26.704)	(32.912)	(47.865)	(21.165)
Drink	2.451	-14.800	-34.876	-70.408 *	-59.946	-56.370 *
	(62.183)	(46.010)	(37.541)	(40.433)	(54.935)	(30.156)
Birth control pills	-34.450	-53.682 ***	-30.988 *	-19.569	-22.692	-33.379 ***
	(27.588)	(18.999)	(17.313)	(19.016)	(27.585)	(12.648)
Complications	-100.794 ***	-61.137 ***	-26.939 *	-23.624	2.355	-54.648 ***
	(30.312)	(20.413)	(14.485)	(17.466)	(24.221)	(13.054)
Doctor visits	6.119	18.112 *	16.192 **	11.760	8.599	15.857 *
	(12.870)	(9.804)	(8.155)	(8.255)	(11.528)	(8.399)
Prenatal visits	92.698 ***	84.246 ***	74.827 ***	74.228 ***	79.509 ***	55.543 ***
	(15.312)	(8.311)	(5.941)	(5.990)	(7.938)	(13.689)
Test tube baby	20.611	-27.565	-0.812	-40.657	-181.511 *	-54.088
	(127.855)	(79.473)	(65.856)	(73.092)	(97.172)	(50.084)
Diabetes	103.043	150.006 *	256.696 ***	285.550 ***	282.415 ***	221.760 ***
	(121.416)	(84.986)	(70.130)	(63.294)	(95.084)	(68.298)
Income	-0.080	-0.016	-0.122	-0.048	-0.100	-0.114
	(0.365)	(0.270)	(0.221)	(0.255)	(0.347)	(0.178)
Unemployment benefits	-0.430	-0.112	0.034	-0.220	-0.190	-0.328
	(0.612)	(0.457)	(0.373)	(0.428)	(0.596)	(0.298)
Home size	-0.255	-0.316	0.138	0.115	0.039	-0.114
	(0.275)	(0.219)	(0.204)	(0.255)	(0.306)	(0.154)
Married	-2.424	-2.771	-3.197	13.586	-29.501	9.930
	(32.435)	(24.068)	(18.997)	(22.819)	(30.852)	(15.214)
Student	1.707	5.730	9.500	27.206	38.908	11.019
	(38.078)	(25.659)	(21.703)	(25.454)	(35.496)	(16.583)
Height	6.879 ***	9.110 ***	10.965 ***	11.956 ***	11.858 ***	10.711 ***
	(1.676)	(1.242)	(1.095)	(1.185)	(1.496)	(1.009)
Weight	14.741	6.959	8.628	11.938	18.380	14.603 *
	(19.519)	(13.179)	(11.240)	(11.493)	(14.762)	(8.772)
Weight^2	-0.088	-0.016	-0.051	-0.080	-0.092	-0.085
	(0.134)	(0.091)	(0.077)	(0.078)	(0.099)	(0.060)
Age	-20.490	-14.207	-44.707 **	-55.060 **	-59.251	-51.252 ***
	(44.525)	(29.026)	(21.477)	(24.239)	(36.313)	(18.965)
Age^2	0.446	0.272	0.801 **	0.940 **	1.034 *	0.943 ***
	(0.689)	(0.463)	(0.348)	(0.390)	(0.588)	(0.303)
Second child	172.737 ***	176.777 ***	161.944 ***	167.637 ***	173.776 ***	161.824 ***
	(38.773)	(25.237)	(20.194)	(22.451)	(32.241)	(18.413)
Male child	122.307 ***	143.390 ***	142.821 ***	174.666 ***	204.769 ***	152.488 ***
	(19.951)	(14.244)	(12.124)	(13.834)	(18.670)	(9.888)
Education Cat. 1	89.874	-88.905	-50.289	14.492	47.121	-42.019
	(117.650)	(82.602)	(59.506)	(65.153)	(79.828)	(57.275)
Education Cat. 2	127.210	-45.128	10.293	77.617	140.297 *	30.576
	(112.123)	(80.818)	(58.588)	(63.556)	(78.042)	(56.087)
Education Cat. 3	143.775	-16.251	-9.094	68.587	129.169	30.857
	(119.275)	(82.667)	(62.246)	(68.082)	(84.486)	(59.790)
Education Cat. 4	142.865	-34.228	19.312	110.734 *	141.894 *	43.511
	(113.361)	(80.779)	(59.365)	(65.097)	(80.660)	(58.183)
Education Cat. 5	179.570	-1.531	14.288	91.657	168.319 *	54.905
	(119.177)	(81.216)	(61.821)	(67.131)	(86.391)	(60.299)
Education Cat. 6	168.811	-1.022	30.895	91.665	156.424 *	61.238
	(115.427)	(83.092)	(60.580)	(65.951)	(83.129)	(58.816)
Education Cat. 7	214.014	23.381	38.629	154.758	123.545	68.702
	(160.280)	(120.056)	(95.431)	(104.422)	(110.882)	(89.514)

Asterisks denote the significance level (double-sided). *: 10%, **: 5%, ***: 1%.
Bootstrapped standard errors are given in parentheses. The bootstrap was done using a sample size of 3,000 births and 499 iterations.

Table 10: Estimation results from the AD model using the balanced dataset. Main variables.

	Quantile Regressions					
	10%	25%	50%	75%	90%	OLS
Smoked during (i)	34.427	17.522	-52.118 *	19.774	-32.024	-17.594
	(48.437)	(33.175)	(30.156)	(36.824)	(51.533)	(26.501)
Smoked during (ii)	37.590	-21.545	-77.352 **	-88.525 **	-184.486 ***	-71.036 **
	(48.435)	(37.411)	(34.246)	(41.459)	(54.086)	(31.282)
Smoked before (i)	21.684	28.449	12.392	-6.865	18.485	18.441
	(33.333)	(26.032)	(22.050)	(26.626)	(38.512)	(18.975)
Smoked before (ii)	-4.789	16.983	12.363	52.183 *	54.067	23.716
	(39.965)	(30.875)	(27.373)	(30.820)	(42.059)	(23.153)
Drink (i)	8.472	-15.287	-9.150	17.372	-7.084	14.925
	(45.042)	(43.505)	(37.213)	(42.717)	(52.395)	(30.150)
Drink (ii)	-43.062	17.824	54.152	12.678	-6.692	34.730
	(55.961)	(45.819)	(35.595)	(39.593)	(59.906)	(35.509)
Birth control pills (i)	-7.967	12.434	-12.738	-17.245	-26.740	-13.656
	(25.764)	(18.860)	(16.818)	(18.882)	(25.430)	(13.801)
Birth control pills (ii)	34.883	50.037 **	18.906	20.941	32.899	37.296 **
	(27.015)	(20.325)	(17.742)	(18.587)	(28.468)	(15.719)
Complications (i)	-65.146 **	-26.441	-22.255	-17.894	-11.197	-37.698 **
	(26.855)	(18.963)	(14.585)	(18.625)	(24.055)	(15.105)
Complications (ii)	-88.984 ***	-55.124 ***	-31.961 **	-31.331 *	-42.418 *	-57.531 ***
	(27.224)	(20.279)	(15.756)	(16.976)	(24.115)	(14.989)
Doctor visits (i)	4.390	-2.316	1.404	-1.955	6.506	9.751
	(12.360)	(9.992)	(8.851)	(9.915)	(13.147)	(8.532)
Doctor visits (ii)	1.283	-3.983	-1.185	-4.714	-8.715	-3.847
	(12.543)	(9.229)	(8.435)	(8.106)	(11.210)	(7.503)
Prenatal visits (i)	29.114 **	30.386 ***	36.961 ***	38.947 ***	30.897 ***	27.070 **
	(12.883)	(9.344)	(6.827)	(6.081)	(7.713)	(11.495)
Prenatal visits (ii)	0.370	-0.392	2.953	3.607	-0.300	0.848
	(9.028)	(5.291)	(4.562)	(5.091)	(5.969)	(6.655)
Test tube baby (i)	46.405	32.526	44.364	17.930	91.791	62.780
	(108.364)	(65.722)	(53.476)	(72.293)	(110.727)	(55.740)
Test tube baby (ii)	-41.714	20.727	-30.053	78.943	147.260 *	38.015
	(101.538)	(82.967)	(77.418)	(84.050)	(82.467)	(62.333)
Income (i)	0.211	0.154	-0.276	-0.390	-0.390	-0.143
	(0.328)	(0.268)	(0.243)	(0.282)	(0.365)	(0.213)
Income (ii)	0.110	0.005	0.082	0.047	-0.001	0.083
	(0.326)	(0.245)	(0.194)	(0.198)	(0.266)	(0.167)
Unemployment benefits (i)	-0.305	-0.689	-0.689 *	-0.578	-0.780	-0.540
	(0.555)	(0.430)	(0.388)	(0.449)	(0.541)	(0.337)
Unemployment benefits (ii)	0.342	0.439	0.542	0.847 **	0.719	0.727 **
	(0.571)	(0.479)	(0.380)	(0.388)	(0.565)	(0.328)
Home size (i)	0.268	0.376 *	-0.071	0.208	0.183	0.192
	(0.225)	(0.207)	(0.192)	(0.228)	(0.252)	(0.163)
Home size (ii)	0.197	0.322	0.086	0.140	0.059	0.220
	(0.269)	(0.207)	(0.171)	(0.192)	(0.271)	(0.143)
Married (i)	-35.504	-21.928	-2.066	-6.983	17.387	-22.931
	(23.669)	(19.285)	(16.372)	(19.568)	(25.276)	(15.421)
Married (ii)	10.370	-0.958	-12.085	-20.238	10.430	-3.058
	(26.864)	(21.047)	(16.674)	(19.218)	(27.180)	(14.911)
Student (i)	48.297	38.346 *	23.560	4.320	-2.628	23.562
	(30.253)	(22.501)	(20.822)	(25.053)	(35.786)	(18.565)
Student (ii)	21.763	16.683	-7.744	-12.135	-3.007	6.474
	(34.996)	(26.086)	(22.569)	(23.948)	(31.045)	(19.320)
Weight (i)	-2.419	-6.812	-7.124	-22.844 **	-25.814 *	-11.814
	(16.352)	(10.952)	(9.852)	(11.132)	(13.401)	(8.255)
Weight (ii)	11.216	22.988 **	21.388 **	26.031 ***	23.605 **	18.239 **
	(14.437)	(10.603)	(9.404)	(9.048)	(11.045)	(7.457)
$Weight^2$ (i)	-0.010	0.004	0.018	0.140 *	0.153 *	0.050
	(0.112)	(0.075)	(0.069)	(0.078)	(0.092)	(0.057)
$Weight^2$ (ii)	-0.029	-0.109	-0.079	-0.113 *	-0.110	-0.063
	(0.100)	(0.073)	(0.065)	(0.062)	(0.072)	(0.052)
Age (i)	34.032	-0.010	-17.079	-19.052	-20.056	-0.984
	(57.090)	(43.502)	(32.998)	(38.446)	(50.820)	(31.604)
Age (ii)	26.236	30.125	77.768 **	58.312	80.949	65.563 *
	(63.144)	(44.987)	(34.394)	(38.459)	(55.363)	(33.800)
Age^2 (i)	-0.679	-0.107	0.193	0.326	0.362	-0.051
	(0.998)	(0.764)	(0.593)	(0.679)	(0.920)	(0.560)
Age^2 (ii)	-0.514	-0.461	-1.216 **	-0.947	-1.311	-1.091 **
	(1.038)	(0.725)	(0.561)	(0.636)	(0.927)	(0.551)
Male child (i)	-18.425	-32.615 **	-25.624 *	-33.839 **	-13.955	-30.625 ***
	(19.717)	(14.770)	(13.300)	(14.540)	(19.145)	(11.199)
Male child (ii)	11.833	-5.623	-9.206	-14.778	-51.292 ***	-5.882
	(19.609)	(15.003)	(12.137)	(14.254)	(18.668)	(11.647)

Asterisks denote the significance level (double-sided). *: 10%, **: 5%, ***: 1%.
Bootstrapped standard errors are given in parentheses. The bootstrap was done using a sample size of 3,000 births and 499 iterations.

Table 11: Estimation results from the AD model using the balanced dataset. CRE added variables.

	Quantile Regressions					
	10%	25%	50%	75%	90%	OLS
Smoked during	-190.485 ***	-112.107 ***	-75.991 ***	-90.337 ***	-2.081	-94.897 ***
	(49.047)	(35.800)	(27.952)	(34.573)	(49.101)	(23.618)
Smoked before	-6.101	-14.429	-30.103	-36.834	-47.618	-21.863
	(35.629)	(27.941)	(24.240)	(26.244)	(37.643)	(19.926)
Drink	11.578	-47.183	-33.420	-45.828	-13.504	-42.091
	(51.397)	(37.665)	(34.045)	(36.649)	(49.262)	(26.156)
Birth control pills	-27.853	-52.478 ***	-28.928 **	-21.657	-19.116	-33.813 ***
	(25.097)	(18.130)	(14.118)	(16.678)	(22.708)	(11.650)
Complications	-122.386 ***	-67.361 ***	-46.073 ***	-29.562 **	-11.029	-65.723 ***
	(24.632)	(16.595)	(12.271)	(14.727)	(21.520)	(10.636)
Doctor visits	5.378	10.838	15.874 **	11.317	-0.099	12.845 *
	(10.409)	(8.081)	(6.916)	(7.094)	(10.166)	(6.765)
Prenatal visits	109.582 ***	90.162 ***	77.477 ***	80.245 ***	73.484 ***	64.011 ***
	(11.583)	(5.906)	(4.583)	(5.080)	(6.969)	(11.670)
Test tube baby	-2.683	64.214	30.674	-57.277	-251.771 ***	-34.417
	(108.153)	(70.387)	(61.737)	(63.434)	(86.622)	(46.240)
Diabetes	181.554 *	217.704 ***	280.896 ***	315.398 ***	373.631 ***	282.529 ***
	(94.004)	(70.214)	(55.456)	(62.333)	(88.341)	(54.962)
Income	-0.126	-0.126	-0.088	-0.127	-0.120	-0.108
	(0.276)	(0.196)	(0.183)	(0.203)	(0.284)	(0.135)
Unemployment benefits	-0.303	-0.185	-0.135	-0.468	0.128	-0.410 *
	(0.510)	(0.358)	(0.311)	(0.350)	(0.484)	(0.239)
Home size	-0.219	0.016	-0.010	0.042	-0.079	-0.072
	(0.268)	(0.191)	(0.180)	(0.225)	(0.294)	(0.143)
Married	-5.950	-1.556	-2.479	-6.801	-19.225	-4.313
	(27.684)	(19.025)	(16.058)	(19.336)	(27.507)	(13.429)
Student	20.743	4.311	1.137	10.260	51.933 *	17.787
	(31.961)	(22.674)	(21.060)	(22.319)	(29.430)	(15.434)
Height	6.662 ***	8.423 ***	10.489 ***	11.419 ***	11.183 ***	10.269 ***
	(1.525)	(1.088)	(0.975)	(1.074)	(1.320)	(0.906)
Weight	5.722	2.412	-8.622	2.040	23.161 **	7.950
	(14.546)	(9.823)	(8.403)	(8.182)	(11.050)	(6.514)
$Weight^2$	-0.026	0.008	0.063	-0.010	-0.127 *	-0.036
	(0.099)	(0.066)	(0.058)	(0.054)	(0.071)	(0.043)
Age	-40.769	-24.258	-19.409	-29.653 *	-0.204	-14.971
	(30.152)	(20.280)	(16.161)	(17.768)	(26.393)	(14.840)
Age^2	0.867 *	0.516	0.402	0.465	0.065	0.333
	(0.467)	(0.319)	(0.256)	(0.285)	(0.420)	(0.236)
Second child	177.725 ***	162.143 ***	150.112 ***	175.701 ***	156.739 ***	162.162 ***
	(21.562)	(14.632)	(12.540)	(14.267)	(19.457)	(11.515)
Third child	213.334 ***	191.121 ***	190.076 ***	252.706 ***	231.066 ***	209.153 ***
	(33.522)	(24.012)	(22.312)	(24.301)	(31.575)	(20.485)
Fourth child	172.238 ***	182.380 ***	181.915 ***	233.246 ***	202.214 ***	191.631 ***
	(56.800)	(41.156)	(37.683)	(40.316)	(51.240)	(34.217)
Male child	121.921 ***	128.415 ***	137.299 ***	162.926 ***	186.615 ***	145.904 ***
	(16.220)	(12.039)	(10.115)	(12.611)	(16.705)	(8.042)
Education Cat. 1	-17.762	-76.618	-81.724	-39.997	-9.637	-62.830
	(90.017)	(63.440)	(54.685)	(51.733)	(62.760)	(44.738)
Education Cat. 2	40.745	-39.634	-34.390	17.789	63.845	-5.916
	(86.174)	(60.040)	(51.815)	(50.007)	(59.730)	(44.813)
Education Cat. 3	84.970	-8.360	-37.973	-6.862	31.144	-7.380
	(92.000)	(63.567)	(56.402)	(55.421)	(64.993)	(46.993)
Education Cat. 4	57.207	-29.640	-18.659	30.217	85.484	9.992
	(85.169)	(59.404)	(52.030)	(50.103)	(61.687)	(44.595)
Education Cat. 5	67.216	-30.498	-59.437	9.144	77.633	-2.762
	(93.874)	(65.271)	(55.834)	(59.343)	(69.266)	(49.610)
Education Cat. 6	80.965	1.085	-12.183	19.262	66.071	22.282
	(86.395)	(62.012)	(55.002)	(53.307)	(67.144)	(46.642)
Education Cat. 7	56.158	-22.984	-48.886	3.971	40.766	-18.372
	(157.415)	(96.129)	(87.963)	(96.698)	(147.423)	(82.569)

Asterisks denote the significance level (double-sided). *: 10%, **: 5%, ***: 1%.
Bootstrapped standard errors are given in parentheses. The bootstrap was done using a sample size of 3,000 births and 499 iterations.

Table 12: Estimation results from the CREM model using the unbalanced dataset. Main variables.

	Quantile Regressions					
	10%	25%	50%	75%	90%	OLS
Smoked during	1.511	-81.205 *	-118.446 ***	-118.538 ***	-224.515 ***	-123.482 ***
	(60.811)	(47.264)	(36.936)	(44.050)	(57.950)	(30.579)
Smoked before	-2.960	36.036	31.171	63.577 *	88.842 *	45.941 *
	(43.834)	(34.982)	(30.794)	(33.820)	(47.633)	(26.418)
Drink	-41.244	12.411	37.354	-0.066	-56.913	18.728
	(67.546)	(58.164)	(45.985)	(52.983)	(70.383)	(37.602)
Birth control pills	20.963	42.465	3.423	-5.526	-5.043	13.155
	(34.027)	(26.099)	(22.278)	(25.759)	(34.068)	(18.917)
Complications	-135.696 ***	-75.278 ***	-38.337 *	-44.836 *	-34.767	-79.684 ***
	(35.896)	(24.149)	(20.843)	(24.062)	(31.528)	(17.880)
Doctor visits	10.269	1.704	-3.187	-1.286	0.357	5.962
	(16.818)	(11.226)	(10.423)	(12.417)	(17.171)	(8.744)
Prenatal visits	8.670	18.729 **	31.123 ***	28.191 ***	19.459 **	20.307 ***
	(8.377)	(8.150)	(7.073)	(7.784)	(9.857)	(6.213)
Test tube baby	9.859	-63.569	-5.916	101.877	312.990 **	53.572
	(136.485)	(97.028)	(80.492)	(97.485)	(129.928)	(69.107)
Income	0.327	0.248	0.106	0.100	0.018	0.134
	(0.315)	(0.252)	(0.225)	(0.255)	(0.354)	(0.175)
Unemployment benefits	0.503	0.260	0.591	0.938 *	0.165	0.791 *
	(0.711)	(0.542)	(0.498)	(0.556)	(0.743)	(0.404)
Home size	0.504	0.262	0.253	0.223	0.272	0.322 *
	(0.317)	(0.250)	(0.220)	(0.276)	(0.342)	(0.193)
Married	2.626	-11.628	-5.422	13.333	36.962	1.738
	(32.593)	(22.949)	(19.948)	(22.826)	(31.406)	(16.984)
Student	42.309	50.671	58.122 **	36.587	-10.418	35.052
	(40.198)	(31.235)	(27.134)	(29.995)	(40.104)	(22.354)
Weight	18.993	25.524 **	33.761 ***	18.262 **	-0.876	16.575 **
	(14.521)	(10.053)	(9.554)	(8.925)	(12.387)	(7.507)
Weight^2	-0.106	-0.158 **	-0.188 ***	-0.078	0.035	-0.084 *
	(0.098)	(0.067)	(0.067)	(0.060)	(0.080)	(0.050)
Age	52.479	33.363	20.231	7.749	5.752	16.427
	(39.830)	(27.214)	(21.470)	(24.502)	(34.454)	(20.514)
Age^2	-1.115 *	-0.702	-0.460	-0.112	-0.163	-0.382
	(0.644)	(0.438)	(0.352)	(0.398)	(0.550)	(0.332)
Male child	-15.270	-30.872	-34.649 *	-41.648 **	-46.620 *	-30.403 **
	(25.973)	(20.708)	(18.213)	(19.345)	(24.349)	(15.446)

Asterisks denote the significance level (double-sided). *: 10%, **: 5%, ***: 1%.
Bootstrapped standard errors are given in parentheses. The bootstrap was done using a sample size of 3,000 births and 499 iterations.

Table 13: Estimation results from the CREM model using the unbalanced dataset. CRE added variables.

	Quantile Regressions					
	10%	25%	50%	75%	90%	OLS
Smoked during	-70.869 ***	-83.486 ***	-99.452 ***	-104.189 ***	-108.669 ***	-94.360 ***
	(26.305)	(21.007)	(19.840)	(20.379)	(24.320)	(19.320)
Smoked before	-55.565 ***	-41.376 ***	-19.328	-10.835	-3.529	-24.233 *
	(18.267)	(15.541)	(14.587)	(15.951)	(18.783)	(14.206)
Drink	-24.374	-22.919	-44.797 *	-51.251 **	-59.697 *	-39.255 *
	(29.747)	(24.710)	(25.253)	(25.508)	(30.727)	(23.552)
Birth control pills	-36.884 **	-31.089 **	-37.452 ***	-30.736 **	-27.382 *	-34.084 ***
	(14.850)	(12.219)	(11.802)	(12.597)	(14.772)	(11.521)
Complications	-96.524 ***	-74.428 ***	-54.155 ***	-31.219 ***	-22.950 *	-64.428 ***
	(16.128)	(12.741)	(11.166)	(11.377)	(13.732)	(11.104)
Doctor visits	14.708 *	10.948	12.629 *	13.281 **	9.717	12.974 *
	(8.338)	(7.224)	(6.623)	(6.666)	(6.830)	(7.323)
Prenatal visits	76.715 ***	70.372 ***	69.121 ***	69.656 ***	68.376 ***	62.657 ***
	(14.315)	(10.600)	(9.265)	(8.646)	(8.266)	(13.404)
Test tube baby	-45.804	-53.412	-18.727	-50.580	-39.651	-35.884
	(56.989)	(47.253)	(39.209)	(40.931)	(48.205)	(40.673)
Income	-0.094	-0.095	-0.114	-0.136	-0.135	-0.109
	(0.147)	(0.124)	(0.119)	(0.127)	(0.152)	(0.116)
Unemployment benefits	-0.311	-0.259	-0.411	-0.231	-0.247	-0.365
	(0.347)	(0.275)	(0.258)	(0.270)	(0.326)	(0.255)
Home size	-0.043	-0.128	-0.098	-0.007	-0.120	-0.086
	(0.145)	(0.120)	(0.115)	(0.122)	(0.138)	(0.112)
Married	-1.042	-1.205	-1.521	-4.776	-4.244	-1.310
	(13.302)	(11.668)	(10.787)	(11.433)	(13.451)	(10.709)
Student	37.874 **	14.307	5.707	3.321	-16.208	5.924
	(17.172)	(14.461)	(13.833)	(14.664)	(17.734)	(13.347)
Weight	6.290	9.372 **	7.402 **	5.478	10.136 **	7.808 **
	(4.425)	(3.942)	(3.468)	(3.594)	(4.093)	(3.539)
Weight^2	-0.036	-0.052 *	-0.033	-0.015	-0.042	-0.036
	(0.031)	(0.027)	(0.024)	(0.025)	(0.028)	(0.024)
Age	-9.353	-20.214	-24.367 **	-28.692 **	-40.657 ***	-23.597 *
	(16.801)	(13.553)	(12.277)	(12.582)	(14.517)	(12.242)
Age^2	0.009	0.177	0.254	0.350 *	0.541 **	0.245
	(0.278)	(0.228)	(0.206)	(0.210)	(0.241)	(0.204)
Second child	153.467 ***	156.642 ***	159.453 ***	148.996 ***	150.982 ***	154.368 ***
	(13.608)	(11.138)	(10.753)	(10.909)	(13.908)	(10.496)
Third child	173.195 ***	184.148 ***	196.344 ***	194.057 ***	201.286 ***	187.332 ***
	(23.305)	(19.337)	(18.227)	(19.007)	(23.195)	(18.328)
Fourth child	183.072 ***	207.923 ***	206.771 ***	203.407 ***	208.761 ***	194.235 ***
	(45.266)	(39.430)	(35.272)	(36.674)	(44.848)	(35.431)
Male child	144.927 ***	140.211 ***	142.491 ***	149.684 ***	143.478 ***	145.187 ***
	(11.433)	(9.381)	(8.734)	(9.158)	(11.089)	(8.548)
Education Cat. 1	-38.800	-25.320	27.994	50.535	104.376 **	21.033
	(57.321)	(33.010)	(27.240)	(35.568)	(50.233)	(27.652)
Education Cat. 2	-49.741	-29.418	33.218	57.239 *	110.142 **	22.576
	(54.799)	(32.347)	(27.109)	(34.744)	(48.344)	(26.771)
Education Cat. 3	-57.021	-29.995	31.191	44.736	105.960 **	18.574
	(58.305)	(35.817)	(29.071)	(37.073)	(51.821)	(28.921)
Education Cat. 4	-48.247	-21.351	33.079	52.249	112.424 **	22.250
	(56.191)	(33.309)	(27.153)	(34.958)	(48.894)	(26.871)
Education Cat. 5	-56.763	-17.026	37.185	60.919	113.416 **	25.409
	(57.618)	(34.794)	(29.122)	(37.192)	(53.932)	(28.824)
Education Cat. 6	-41.767	-26.221	36.781	50.065	104.212 **	21.113
	(57.214)	(33.233)	(27.344)	(35.580)	(50.290)	(27.318)
Education Cat. 7	-109.022	-46.016	29.793	34.105	112.564	7.597
	(90.562)	(61.026)	(48.152)	(65.498)	(99.115)	(48.731)

Asterisks denote the significance level (double-sided). *: 10%, **: 5%, ***: 1%.
Bootstrapped standard errors are given in parentheses. The bootstrap was done using a sample size of 3,000 births and 499 iterations.

Table 14: Estimation results from the 2SFE model using the unbalanced dataset.

	Quantile Regressions				
	10%	25%	50%	75%	90%
Smoked during	-162.837 ***	-156.775 ***	-146.782 ***	-186.265 ***	-183.502 ***
	(34.303)	(25.604)	(23.264)	(24.305)	(32.937)
Smoked before	-24.508	-14.232	-5.846	12.113	29.442
	(26.155)	(20.022)	(17.802)	(19.141)	(27.079)
Drink	-41.675	-33.005	-49.502 *	-52.400 *	-43.625
	(38.308)	(29.008)	(28.939)	(29.333)	(38.988)
Birth control pills	-22.839	-22.815	-24.609 *	-7.255	-26.680
	(19.812)	(13.959)	(13.109)	(13.852)	(19.669)
Complications	-151.500 ***	-93.982 ***	-62.792 ***	-24.545 *	-8.647
	(20.297)	(13.777)	(12.187)	(13.502)	(19.534)
Doctor visits	7.704	8.782	14.639 **	13.681 **	8.451
	(9.085)	(7.003)	(6.103)	(6.370)	(8.415)
Prenatal visits	101.149 ***	88.274 ***	82.826 ***	82.325 ***	82.236 ***
	(12.888)	(7.595)	(5.724)	(5.359)	(6.441)
Test tube baby	43.098	21.891	33.384	5.572	-35.821
	(68.260)	(49.740)	(38.826)	(44.940)	(61.586)
Income	0.010	-0.015	-0.043	-0.070	-0.103
	(0.183)	(0.142)	(0.136)	(0.153)	(0.193)
Unemployment benefits	-0.268	-0.277	-0.142	0.142	-0.357
	(0.411)	(0.303)	(0.259)	(0.305)	(0.390)
Home size	0.094	0.157	0.274 **	0.312 **	0.318 *
	(0.163)	(0.142)	(0.136)	(0.137)	(0.178)
Married	-7.534	2.276	-1.178	-3.597	-13.208
	(17.240)	(13.761)	(13.302)	(14.068)	(18.967)
Student	86.441 ***	63.563 ***	45.732 ***	24.006	33.635
	(22.890)	(16.876)	(17.352)	(17.373)	(25.137)
Weight	31.928 ***	27.255 ***	27.080 ***	23.050 ***	26.062 ***
	(5.952)	(5.399)	(4.906)	(4.881)	(5.529)
Weight^2	-0.170 ***	-0.132 ***	-0.127 ***	-0.094 ***	-0.106 ***
	(0.042)	(0.038)	(0.035)	(0.034)	(0.039)
Age	-4.217	-29.836 *	-30.644 **	-32.730 **	-44.297 **
	(19.625)	(15.671)	(13.528)	(14.332)	(19.300)
Age^2	0.058	0.485 *	0.526 **	0.545 **	0.733 **
	(0.325)	(0.259)	(0.222)	(0.235)	(0.318)
Second child	186.303 ***	169.990 ***	161.051 ***	160.180 ***	168.794 ***
	(18.279)	(13.763)	(12.066)	(14.487)	(19.541)
Third child	167.977 ***	202.422 ***	202.170 ***	220.322 ***	230.686 ***
	(30.166)	(24.527)	(22.574)	(24.571)	(31.334)
Fourth child	191.796 ***	183.375 ***	204.031 ***	228.417 ***	191.609 ***
	(55.278)	(46.668)	(40.472)	(42.210)	(54.352)
Male child	146.002 ***	140.324 ***	143.942 ***	152.610 ***	152.521 ***
	(14.798)	(10.770)	(9.261)	(10.380)	(14.746)
Education Cat. 1	-171.625 *	-69.412	33.134	46.669	-78.829
	(90.106)	(73.051)	(69.667)	(69.450)	(80.900)
Education Cat. 2	-74.140	27.718	128.365 *	148.536 **	41.025
	(87.029)	(69.838)	(66.251)	(67.651)	(80.363)
Education Cat. 3	-31.520	49.709	162.331 **	142.743 *	21.934
	(94.834)	(78.860)	(75.283)	(76.471)	(91.134)
Education Cat. 4	-72.511	47.528	154.641 **	173.598 **	63.779
	(89.993)	(71.321)	(66.672)	(68.443)	(83.084)
Education Cat. 5	-80.900	-1.067	115.933	152.616 **	85.175
	(97.066)	(79.928)	(75.752)	(76.968)	(90.828)
Education Cat. 6	-39.452	71.311	165.983 **	162.931 **	36.387
	(93.207)	(73.391)	(69.194)	(70.773)	(87.325)
Education Cat. 7	-111.444	7.248	183.595 *	172.339	46.874
	(174.631)	(128.867)	(106.956)	(116.731)	(182.347)

Asterisks denote the significance level (double-sided). *: 10%, **: 5%, ***: 1%.
Bootstrapped standard errors are given in parentheses. The bootstrap was done using a sample size of 3,000 births and 499 iterations.

Table 15: Estimation results from the post estimated PFE(5) model with $\lambda = 0.8$ and using the unbalanced dataset.

7 References

Abrevaya, J. (2001). The Effects of Demographics and Maternal Behavior on the Distribution of Birth Outcomes. *Empirical Economics 26*(1), 247–257.

Abrevaya, J. (2006). Estimating the Effect of Smoking on Birth Outcomes Using a Matched Panel Data Approach. *Journal of Applied Econometrics 21*(4), 489–519.

Abrevaya, J. and C. Dahl (2008). The effects of birth inputs on birthweight. *Journal of Business and Economic Statistics 26*(4), 379–397.

Almond, D., K. Y. Chay, and D. S. Lee (2005). The Costs Of Low Birth Weight. *The Quarterly Journal of Economics 120*(3), 1031–1083.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006). Quantile regression under misspecification, with an application to the us wage structure. *Econometrica 74*(2), 539–563.

Arulampalam, W., R. A. Naylor, and J. Smith (2007, November). Am I Missing Something? The Effects of Absence From Class on Student Performance. Warwick Economic Research Paper 820, University of Warwick, Department of Economics.

Bache, S. H. (2010, August). Minimax regression quantiles. CREATES Research Paper 2010-54, CREATES, Aarhus University.

Bernstein, I. M., J. D. Horbar, G. J. Badger, A. Ohlsson, and A. Golan (2000). Morbidity and Mortality Among Very-Low-Birth-Weight Neonates with Intrauterine Growth Restriction. *American Journal of Obstetrics and Gynecology 182*(1), 196–206.

Bernstein, I. M., J. A. Mongeon, G. J. Badger, L. Solomon, S. H. Heil, and S. T. Higgins (1978). Maternal Smoking and Its Association With Birth Weight. *Obstetrics & Gynecology 106*(5), 986–991.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2007). From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes. *Quarterly Journal of Economics 122*(1), 409–439.

Buckles, K. and D. Hungerman (2008, December). Season of birth and later outcomes: Old questions, new answers. NBER Working Paper 14573, National Bureau of Economic Research.

Canay, I. (2010). A simple approach to quantile regression for panel data. Working paper, Northwestern University.

Chamberlain, G. (1984). Panel Data. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2, pp. 1247–1318. Elsevier Science B. V.

Chernozhukov, V. and I. Fernández-Val (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *Review of Economic Studies 78*(2), 559–589.

Corman, H. and S. Chaikind (1998). The Effect of Low Birthweight on the School Performance and Behavior of School-Aged Children. *Economics of Educations Review 17*(3), 307–316.

Dehejia, R. and A. Lleras-Muney (2004). Booms, busts, and babies' health. *Quarterly Journal of Economics 119*(3), 1091–1130.

DiFranza, J. R., C. A. Aligne, and M. Weitzman (2004). Prenatal and Postnatal Environmental Tobacco Smoke Exposure and Children's Health. *Pediatrics 113*(4), 1007–1015.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica 75*(1), 259–276.

Fitzenberger, B., K. Kohn, and Q. Wang (2010). The erosion of union membership in germany: determinants, densities, decompositions. *Journal of Population Economics 24*(1), 141–165.

Hack, M., N. K. Klein, and H. G. Taylor (1995). Long-term developmental outcomes of low birth weight infants. *The Future of Children 5*(1), 176–196.

Hofhuisi, W., J. C. de Jongste, and P. J. F. M. Merkus (2003). Adverse Health Effects of Prenatal and Postnatal Tobacco Smoke Exposure on Children. *Archives of Disease in Childhood 88*(12), 1086–1090.

Kirkegaard, I., C. Obel, M. Hedegaard, and T. B. Henriksen (2006). Gestational Age and Birth Weight in Relation to School Performance of 10-Year-Old Children: A Follow-up Study of Children Born After 32 Completed Weeks. *Pediatrics 118*(4), 1600–1606.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis 91*(1), 74–89.

Koenker, R. and K. F. Hallock (2001). Quantile Regression. *The Journal of Economic Perspectives 15*(4), 143–156.

Linnet, K. M., C. Obel, E. Bonde, P. Hove, Thomsen, N. J. Secher, K. Wisborg, and T. B. Henriksen (2006). Cigarette Smoking During Pregnancy and Hyperactive-Distractible Preschooler's: A follow-up Study. *Acta Paediatrica 95*(6), 694–700.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica 46*(1), 69–85.

Permutt, T. and J. R. Hebel (1989). Simultaneous-Equation Estimation in a Clinical Trial of the Effect of Smoking on Birth Weight. *Biometrics 45*(2), 619–622.

Royer, H. (2004, June). What All Women (and Some Men) Want to Know: Does Maternal Age Affect Infant Health? Working Paper 68, Center of Labor Economics, University of California, Berkely.

Wang, X., B. Zuckerman, C. Pearson, G. Kaufman, C. Chen, G. Wang, T. Niu, P. H. Wise, H. Bauchner, and X. Xu (2002). Maternal Cigarette Smoking, Metabolic Gene Polymorphism, and Infant Birth Weight. *The Journal of the American Medical Association 287*(2), 195–202.

Wisborg, K., U. Kesmodel, T. B. Henriksen, S. F. Olsen, and N. J. Secher (2000). A Prospective Study of Smoking During Pregnancy and SIDS. *Archives of Disease in Childhood 83*(3), 203–206.

Wisborg, K., U. Kesmodel, T. B. Henriksen, S. F. Olsen, and N. J. Secher (2001). Exposure to Tobacco Smoke in Utero and the Risk of Stillbirth and Death in the First Year of Life. *American Journal of Epidemiology 154*(4), 322–327.

Preface to paper #2

After spending some time trying to slightly demystify the topic of panel data in applications of quantile regression, I was still intrigued by the idea of a general way to deal with correlated observations and non-diagonal weighting in spirit similar to the literature on generalized estimating equations (GEE).

The subgradient of the classic quantile regression criterion function gives very intuitive *estimating functions*: they clearly show that what is sought is a fitted quantile-specific function satisfying the condition that the conditional probability of not exceeding its value is τ—the quantile index being considered. In this paper I start with the most basic estimating function for quantile regression and investigate optimality (in terms of the functions, rather than the final estimator) and the connection to the classic criterion function. I develop a minimax estimating framework which provides an alternative way to obtain solutions to the estimating equations. I generalize the basic estimator such that correlation can modelled directly by non-diagonal weighting in the spirit of the literature on generalized estimating equations.

The main contribution of the paper, I think, is its initiative to seek alternative ways to approach estimation and the proposed generalization of the basic estimator. Indeed, for conditional expectation modelling, we have a handful of frameworks available for dealing with different types of problems, e.g. OLS, GMM, and maximum likelihood. It is interesting that estimating function theory provides a synthesis of these methodologies (and even finite sample justification for maximum likelihood), and one wonders if the same could be true for conditional quantile modelling.

At the time of writing, I have made a prototype R-script for computation of the linear model available on my website in hope of aiding numerical implementation. (`http://www.stefanbache.dk/minimax-prototype-code.zip`)

Paper #2

Minimax regression quantiles and non-diagonally weighted quantile regression

2011

Stefan Holst Bache
Aarhus University,
Department of Economics and Business and CREATES

Keywords:
Estimating function, minimax estimation, non-linear quantile regression, quantile regression, weighted quantile regression

Abstract. A new quantile regression framework is developed, and consistency as well as asymptotic normality is established for the basic estimator. These asymptotic properties coincide with those of the classic estimator for quantile regression. Estimators in this framework are based on estimating functions to which one can obtain consistent roots by finding the minimax of a certain deviance function. This function is conveniently constructed to have certain properties and relations to the estimating functions. The new and flexible estimating framework for quantile regression allows for both linear and non-linear specifications and has potential for generalization. The proposed minimax approach, and extensions of the basic estimator, could therefore prove to have important implications. This is demonstrated in a proposal of a generalized estimator for non-diagonally weighted quantile regression. A simulation example is presented where this new estimator has improved efficiency. It is argued that modified gradient-type algorithms can be used to compute solutions to the optimization problem, and one such is outlined to aid practical implementation.

1 Introduction

The regression quantiles methodology, as introduced some thirty years ago by Koenker and Bassett (1978), has become a well established and popular empirical tool amongst researchers and practitioners. It compliments its more mature cousin, that of least squares, in providing a sometimes more complete picture of distributional relationships between variables of interest. Here, the phrasing "more mature" is meant to express that some modelling challenges are better understood and have elegant solutions in the least squares methodology, while they still keep researchers busy with searching for quantile regression analogues. Examples present themselves when faced with selection issues, dependent data, unobserved heterogeneity, etc. Much work has been done in dealing with these, and other common problems, yet there still seems to be work to do. The rapidly expanding literature on quantile regression methods is indeed an indicator of this. This paper offers a new and alternative approach to quantile regression. This may then serve as a new building block for such investigations. The alternative approach to regression quantiles taken here, is based on an artificially constructed deviance criterion function, for which the minimax point is a consistent solution to appropriate estimating equations. The minimax approach allows for linear and non-linear regression functions which satisfy certain regularity restrictions and identification assumptions. One potential drawback of such minimax estimators is that the powerful linear programming algorithms, in which one usually puts ones faith to provide numerical estimates of the linear quantile regression model, no longer apply. It does however turn out that one can often use modified gradient-type optimization methods to compute solutions to the minimax problem, and for model specifications where linear programming does not apply the minimax approach may provide numerical benefits.

In what follows, the basic estimator is derived along with the minimax framework. Its asymptotic properties are presented, and the computational aspect is considered. Finally, a generalized estimator which models potentially correlated data is presented and a simulation experiment shows that it can provide efficiency gains.

2 A brief notational overview

Let $(\Omega, \mathscr{F}, P)$ denote the underlying probability space. Modelling conditional quantiles, allows us to investigate relations between $Y(\omega) \in \mathbb{R}$, a real continuous random variable, and $X(\omega) \in \mathscr{X}$, a k-dimensional random vector. Define $S = \mathbb{R} \times \mathscr{X}$, the sample space with distribution law π. By y and x we will refer to the first and the last k elements, respectively, of a point $s \in S$, i.e. points in $\mathbb{R}$ and $\mathscr{X}$. The observed sample, $\{s_i\} = \{y_i, x_i\}$, with $i = 1, \ldots, n$, is assumed to be independent and identically distributed according to π. Denote by Θ_0 the parameter set, an analytic subset of the compact metric space (Θ, d). We set out to find a criterion function $D : S \times \Theta_0 \to \mathbb{R}$ with the help of which we can estimate the parameters, $\theta_0 \in \Theta_0$, in a model for the quantiles of $Y(\omega)$, conditional on $X(\omega) = x$. Henceforth, dependence on ω is dropped from the notation. Let $\mu : \mathscr{X} \times \Theta_0$ be such a known and fixed model for a given quantile index $\tau \in (0,1)$. The dependence on τ is also omitted from the notation. For convenience, μ_i^θ and μ_x^θ are sometimes used to abbreviate $\mu(x_i, \theta)$ and $\mu(x, \theta)$. Finally, unless otherwise specified, denote by E the expectation operator with respect to π.

3 Estimating functions for quantile regression

First, consider the task of computing a sample τ-quantile. For the moment we let $F(y) = P(Y \leqslant y)$ denote the continuously differentiable distribution function of Y. Given a sample, the minimizer, $\hat{q}$, of the asymmetric loss-function $\sum_{i=1}^n \rho_\tau(y_i - q)$, with $\rho_\tau(u) = u(\tau - 1\{u \leqslant 0\})$, is well-known to be a consistent estimator of the τth quantile of Y, i.e. $Q_Y(\tau) \equiv F^{-1}(\tau) \equiv \inf\{y : F(y) \geqslant \tau\}$. Heuristically, the argument roughly goes that

$$\frac{\mathrm{d}}{\mathrm{d}q}\left[\tau\int_q^\infty (Y-q)\mathrm{d}F - (1-\tau)\int_{-\infty}^q (q-Y)\mathrm{d}F\right] = F(q) - \tau,$$

so the population minimum is obtained at $F(q) = \tau$. Koenker and Basset's quite clever and fruitful idea, then, was to extend this to a regression setting, letting $q \equiv \mu(x,\theta)$, with x being covariates of y and $\theta \in \Theta_0$ being the unknown parameter vector of interest. The function μ is assumed to be known, and in their original paper it is linear in the parameters. To some it is perhaps not intuitive, without the above argument, that the resulting regression curve implied by $\hat{\theta} = \hat{\theta}(\tau)$, the minimizer of the regression version of the asymmetric loss-function, splits the regression residuals $\{\epsilon\}_{i=1}^n$ such that a fraction, τ, of these are non-positive. This property, however, can in some sense be taken as a defining property of the quantile regression curve, and estimating functions reflecting this is therefore a natural starting point. We will now explore this in more detail for quantile regression models where the conditional quantile function for Y, or its approximation, is known and of the form $Q_Y(\tau|x) = \mu(x,\theta_0)$. Here, $\theta_0 = \theta_0(\tau) \in \Theta_0$ denotes the true parameter; or, as μ is most likely viewed as an approximation of the true conditional quantile function, population solution may be a more appropriate term.

Let $b(\theta) = \tau - 1\{Y \leqslant \mu(X,\theta)\}$ be a re-centered parameter dependent Bernoulli variable, and write $b(s,\theta) = \tau - 1\{y \leqslant \mu(x,\theta)\}$. Define

$$H(\theta) = \int_S b(s,\theta)\pi(\mathrm{d}s) = \mathrm{E}\left[\tau - 1\{Y \leqslant \mu(X,\theta)\}\right].$$

We shall make two identifying assumptions:

Assumption 1. *$\mu(\theta) \neq \mu(\vartheta)$ $\forall\vartheta \neq \theta$ and $H(\theta_0) = 0$ (note, thus that $H(\theta) \neq 0$ $\forall\theta \neq \theta_0$).*

Hence, as argued, we have the natural starting point in this unbiased estimating function. The empirical counterpart, where dependence on the observed sample is suppressed the from notation, is

$$H_n(\theta) = \frac{1}{n}\sum_{i=1}^n b_i(\theta) \equiv \frac{1}{n}\sum_{i=1}^n b(s_i,\theta) = \frac{1}{n}\sum_{i=1}^n \left[\tau - 1\{y_i \leqslant \mu(x_i,\theta)\}\right].$$

The term "estimating function", which is regularly used for $nH_n(\theta)$ and not $H(\theta)$, is here to be understood in the sense of Godambe (1960) and the vast amount of literature following this work. A natural question is whether this estimating function is optimal. It is typical to consider optimality within a class of empirical estimating functions where each summand is weighted, i.e. $\mathcal{G} := \{G_n : G_n(\theta) = \frac{1}{n}\sum_{i=1}^n a_i(\theta)b_i(\theta)\}$. The following proposition states the theoretically optimal, but unfortunately quite impractical, estimating function within this class, i.e. it defines the optimal $a_i(\theta)$.

PROPOSITION 1. *Let $\mathcal{G}$ and $b_i(\theta)$ be defined as above. Assume that the distribution function of Y, conditional on $X = x$, is continuously differentiable. Then the jointly optimal estimating function $G_n^* \in \mathcal{G}$ is given by*

$$G_n^*(\theta) = \frac{1}{n}\sum_{i=1}^n a_i^*(\theta)b_i(\theta), \quad a_i^*(\theta) = -\frac{f_i(\mu_i^\theta)}{\tau(1-\tau)}\dot{\mu}_i^\theta.$$

Here, f_i is the density of Y given that $X = x_i$; $\dot{\mu}_i^\theta$ is the vector of first-order derivatives of μ_i^θ with respect to θ; and $\tau(1-\tau) = \mathrm{var}(b(\theta_0))$.

Note: The optimality statement is to be understood in the Godambe sense, such that for real-valued estimating functions optimality of G_n^* means that

$$\frac{\mathrm{E}\{(G_n^*)^2\}}{\{\mathrm{E}\partial G_n^*/\partial\theta\}^2} \leqslant \frac{\mathrm{E}\{(G_n')^2\}}{\{\mathrm{E}\partial G_n'/\partial\theta\}^2} \quad \forall G_n' \in \mathcal{G},$$

and for vector-valued estimating functions that $\Sigma_{G_n'} - \Sigma_{G_n^*}$ is non-negative definite, where Σ_{G_n} is the variance-covariance matrix of $\{\mathrm{E}(\partial G_n/\partial\theta)\}^{-1}G_n$. With non-differentiable functions, as the G_ns in the present case, one exchanges differential and expectation operators in the above expressions, cf Godambe and Thompson (1984). Also, this is a finite-sample result, rather than one of asymptotic nature.

Proof. A classic result, which we do not prove here, is that

(1) $\quad a_i^*(\theta) = \mathrm{E}\{b_i^2(\theta_0)\}^{-1}\mathrm{E}\{\partial b_i(\theta)/\partial\theta\},$

where $\{x_i\}$ is treated as a fixed sequence. For references, see e.g. Godambe (1985), Godambe (1987), and Ferreira (1982). To make the expression in (1) well-defined, we shall exchange differential and expectation operators, due to non-differentiability of $b_i(\theta)$. Write

$$\partial\mathrm{E}\{b_i(\theta)\}/\partial\theta = \partial[\tau - \mathrm{P}(Y \leqslant \mu(X,\theta)|X = x_i)]/\partial\theta = -f_i(\mu_i^\theta)\dot{\mu}_i^\theta.$$

Clearly, $\mathrm{E}\{b_i^2(\theta_0)\} = \tau(1-\tau)$, and the result follows. □

This result is similar to the one offered by Jung (1996) for quasi likelihood estimation of the median, and an unpublished one by Godambe (2001) for median estimating functions. The presence of the density function, evaluated at the quantile, in the optimal estimating function is problematic since it is usually unknown, and distributional assumptions are mostly unwanted. One could consider an approach where the density is estimated either non-parametrically or by imposing some assumptions on the shape. The potential of using some form of prior information on the density also seems intriguing. These potentials are outside the scope of this paper. Another important part of this proposition is the variance term, which in the basic setting is a constant. A generalization of the estimating functions which models covariance explicitly, a concept introduced by Liang and Zeger (1986a,b), is an obvious approach to dealing with different forms of inter-panel correlation. We will return to this idea in Section 7 where we build this feature into the minimax framework to be considered shortly. First, we shall disregard the constant variance term and the density weights and focus on the estimating function $G_n(\theta) = \frac{1}{n}\sum_{i=1}^n \dot{\mu}_i^\theta[\tau - 1\{y_i \leqslant \mu(x_i,\theta)\}]$, which equated to zero now gives k equations in k unknowns.

A note on the vector-estimating function G_n is that it equals the vector of first-order derivatives of the classical objective function for quantile regression, where they exist. The minimizer of the latter therefore, heuristically speaking, provides consistent roots of G_n. We shall next turn to the main result of this paper, namely another consistent solution: the minimax of an artificially constructed deviance function.

4 The minimax criterion

The minimax approach presented here is inspired by the results for a variety of estimating functions provided by Li (1996, 1997). However, the estimating function $G_n(\theta)$ does not fall into the category of functions considered there, but here it will be shown that a similar deviance function can be constructed, and that consistency is preserved. The proof much parallels Li's, but requires some extension and is presented in full for completeness.

Let $G(s,\theta) = \dot{\mu}_x^\theta[\tau - 1\{y \leqslant \mu(x,\theta)\}]$. From the discussion of estimating functions, we have by assumption that only the parameter of interest, θ_0, will satisfy the vector-valued estimating equation $G(\theta_0) = 0$, where

$$G(\theta) = \int_S G(s,\theta)\pi(\mathrm{d}s) = \int_S \dot{\mu}_x^\theta[\tau - 1\{y \leqslant \mu(x,\theta)\}]\pi(\mathrm{d}s),$$

where, again, $\dot{\mu}$ is the vector of first-order derivatives with respect to θ. The empirical counterpart, which was discussed above, is

$$G_n(\theta) = \frac{1}{n}\sum_{i=1}^n \dot{\mu}_i^\theta(\tau - 1\{y_i \leqslant \mu(x_i,\theta)\}).$$

Note, that even though we here take outset in the population quantities, Proposition 1 made optimality statements based on a finite sample version of the estimating function.

Solving $G_n(\theta) = 0$ is obviously not particularly practical and often there is not an exact solution. However, for reasons that will soon become clear, the claim is that we can make use of the following criterion function from $S \times \Theta \times \Theta$ into $\mathbb{R}$:

$$D(s,\theta,\vartheta) = [\mu(x,\vartheta) - \mu(x,\theta)][\tau - 1\{y \leqslant \bar{\mu}(\theta,\vartheta)\}],$$

where $\bar{\mu}(\theta,\vartheta) = [\mu(x,\theta) + \mu(x,\vartheta)]/2$. Now, integrating out s with respect to π gives us the so-called information function:

$$D(\theta,\vartheta) = \int_S [\mu(x,\vartheta) - \mu(x,\theta)][\tau - 1\{y \leqslant \bar{\mu}(\theta,\vartheta)\}]\pi(\mathrm{d}s).$$

Further, we will restrict attention to Θ_0, so we define

$$D(\theta,\vartheta) \equiv \begin{cases} \infty, & \text{if } \theta \in \Theta \backslash \Theta_0 \\ -\infty, & \text{if } \theta \in \Theta_0, \vartheta \in \Theta \backslash \Theta_0. \end{cases}$$

Make note of the following properties of the deviance function D:

(p1) negativity at θ_0: $D(\theta_0,\vartheta) < 0 \;\forall \vartheta \neq \theta_0$,

(p2) trivial roots: $D(\theta,\theta) = 0$,

(p3) antisymmetry: $D(\theta,\vartheta) = -D(\vartheta,\theta)$.

We now impose an additional, but quite weak identifying assumption, which is satisfied trivially by including e.g. a continuous intercept parameter in the model.

ASSUMPTION 2. *Whenever $\theta \neq \theta_0$ there exists a point $\vartheta \in \Theta_0$ which squeezes $\mu(x,\vartheta)$ in between $\mu(x,\theta)$ and $2\mu(x,\theta_0) - \mu(x,\theta)$, i.e. we can find ϑ such that*

$$\begin{aligned} &\mu(x,\theta) > \mu(x,\vartheta) > 2\mu(x,\theta_0) - \mu(x,\theta) \quad \textit{if } \mu(x,\theta) > \mu(x,\theta_0), \\ &\mu(x,\theta) < \mu(x,\vartheta) < 2\mu(x,\theta_0) - \mu(x,\theta) \quad \textit{if } \mu(x,\theta) < \mu(x,\theta_0). \end{aligned}$$

This assumption allows us to add the following property to the list above:

(p4) positivity for some $\theta \in \Theta_0 \backslash \theta_0$: whenever $\theta \neq \theta_0$ one can choose ϑ such that $D(\theta,\vartheta) > 0$.

Property (p1) follows from the definition of $\mu(x,\theta_0)$ and τ; (p2)–(p3) are evident; and to realise (p4), fix x and choose ϑ according to Assumption 2. This will make $D(\theta,\vartheta)$, conditional on x, positive. By the law of iterated expectations, (p4) then follows. These properties imply that

$$\sup_{\vartheta \in \Theta_0} D(\theta_0,\vartheta) = \inf_{\theta \in \Theta_0} \sup_{\vartheta \in \Theta_0} D(\theta,\vartheta), \qquad \theta_0 = \operatorname*{arginf}_{\theta \in \Theta_0} \sup_{\vartheta \in \Theta_0} D(\theta,\vartheta).$$

Also, we have that the minimax point is achieved at one of the trivial roots, so

$$\inf_{\theta\in\Theta_0}\sup_{\vartheta\in\Theta_0} D(\theta,\vartheta)=0.$$

A final important property is the one that connects $D(\theta,\vartheta)$ to the estimating function $G(\theta)$:

$$\text{(p5)}\quad \left.\frac{\partial D(\theta,\vartheta)}{\partial\vartheta}\right|_{\vartheta=\theta}=G(\theta).$$

The empirical criterion function is given by

$$D_n(\theta,\vartheta)=\frac{1}{n}\sum_{i=1}^{n}[\mu(x_i,\vartheta)-\mu(x_i,\theta)][\tau-1\{y_i\leqslant\bar{\mu}(x_i,\theta,\vartheta)\}],$$

and on the basis of the aforementioned properties, it is natural to estimate θ_0 by the minimax estimator

$$\widehat{\theta}_n\equiv\operatorname*{arginf}_{\theta\in\Theta_0}\sup_{\vartheta\in\Theta_0} D_n(\theta,\vartheta).$$

As Li (1996) also points out in his general discussion, the properties of such deviance functions are closely related to properties of the likelihood-ratio, which can also be stated as a minimax problem. With these constructed deviance functions, however, one gets similar properties without assuming the existence of a likelihood function. Some applications, including the present, do not have a likelihood function as a natural starting point. More generally, estimating functions need not be the score or derivative of any function. Further, we need not require differentiability of D_n, and it is therefore consistent even when the estimating function, when viewed as derivative of D_n, does not exist. Consistency of the basic minimax estimator is given in Theorem 1 below. The asymptotic distribution is given in Theorem 2. Other estimators can be constructed within such a framework by carefully constructing the deviance function. In Section 7 an extension that allows for non-diagonal weighting of the observations is presented.

Theorem 1. *Under assumptions 1 and 2, and those in Section 2, any minimax point*

$$(2)\quad \widehat{\theta}_n=\operatorname*{arginf}_{\theta\in\Theta_0}\sup_{\vartheta\in\Theta_0} D_n(\theta,\vartheta)$$

is consistent, i.e. $\widehat{\theta}_n\xrightarrow{p}\theta_0$.

Proof. Let $D(\theta) = D(\theta, \theta_0) = \mathrm{E}D_n(\theta, \theta_0)$ and $D_n^\circ(\theta) = D_n(\theta, \theta_0) - D(\theta)$. First, we shall consider the uniform convergence properties of $D_n^\circ(\theta)$, which are essential to the proof. To this end we will utilize a few results from the theory of empirical processes. Let $\mathscr{C}$ be a collection of subsets of the sample space S. This collection is said to *pick out* a certain subset, say Z, of a finite set $S^{(n)} = \{s_1, \ldots, s_n\} \subset S$ if it can be written as $Z = S^{(n)} \cap C$, for some $C \in \mathscr{C}$. If $\mathscr{C}$ picks out all of the possible 2^n subsets, then it is said to shatter $S^{(n)}$. Let $V(\mathscr{C})$ be the smallest n such that no set of size n is shattered by C. If $V(\mathscr{C})$ is finite, then $\mathscr{C}$ is called a Vapnik-Červonenkis class (or VC-class). A class of real-valued measurable functions on S is said to be a VC-subgraph class if the collection of subgraphs of these functions forms a VC-class of sets in $S \times \mathbb{R}$. We shall find that the relevant class of functions for D_n is a VC-subgraph class. First, the class of indicator functions $\mathscr{I} \equiv \{1\{y \leqslant (\mu(\theta, x) + \mu(\vartheta, x))/2\} : s \in S, (\theta, \vartheta) \in \Theta_0 \times \Theta_0\}$ is a classic example of a VC-subgraph class. Let $\tilde{\mu}(\theta, \vartheta, x) = \mu(\vartheta, x) - \mu(\theta, x)$. Now, τ and $\tilde{\mu}$ are fixed functions from $\Theta_0 \times \Theta_0 \times S$ into $\mathbb{R}$, and thus by Lemma 2.6.18 (iv)-(v) of Van der Vaart and Wellner (1996), the classes $-\mathscr{I} \equiv \{-\imath : \imath \in \mathscr{I}\}$ and $\tau - \mathscr{I} \equiv \{\tau - \imath : \imath \in \mathscr{I}\}$ are VC-subgraph classes. Finally, by (v) in the same lemma $\mathscr{D} \equiv \{D_n : s \in S, (\theta, \vartheta) \in \Theta_0 \times \Theta_0\}$ is a VC-subgraph class. This property implies that it is uniformly Glivenko-Cantelli in π (sometimes called a GC-π class), i.e. it is a sufficient condition for the following to hold:

$$(3) \qquad \sup_{D_n \in \mathscr{D}} |D_n - D| \to 0 \text{ a.s.}, \qquad \sup_{\theta \in \Theta_0} |D_n^\circ(\theta)| \to 0 \text{ a.s.}$$

Now, we will derive the result that for any $\varepsilon > 0$

$$(4) \qquad \lim_{n \to \infty} P\left\{ \inf_{\theta \in \Theta_0} \sup_{\vartheta \in \Theta_0} D_n(\theta, \vartheta) < \varepsilon \right\} = 1.$$

Then we show that this is contradicted if the assertion of the theorem is false. Let $T \subseteq \Theta_0$ and note that

$$(5) \qquad \inf_{\theta \in T} D_n(\theta, \theta_0) \geqslant \inf_{\theta \in T} D_n^\circ(\theta) + \inf_{\theta \in T} D(\theta).$$

Since $-\sup_{\theta \in T} |D_n^\circ(\theta)| \leqslant \inf_{\theta \in T} D_n^\circ(\theta) \leqslant \sup_{\theta \in T} |D_n^\circ(\theta)|$, then by (3) the first term on the right-hand side of (5) converges in probability to 0. So, for any $\varepsilon > 0$ we now have that

$$(6) \qquad \lim_{n \to \infty} P\left\{ \inf_{\theta \in T} D_n(\theta, \theta_0) \geqslant \inf_{\theta \in T} D_n^\circ(\theta) + \inf_{\theta \in T} D(\theta), \quad \left|\inf_{\theta \in T} D_n^\circ(\theta)\right| < \varepsilon \right\} = 1.$$

If $T = \Theta_0$, then $\inf_{\theta \in T} D(\theta) = 0$ and (6) implies

$$(7) \qquad \lim_{n \to \infty} P\left\{ \sup_{\vartheta \in \Theta_0} D_n(\theta_0, \vartheta) < \varepsilon \right\} = \lim_{n \to \infty} P\left\{ \inf_{\theta \in \Theta_0} D_n(\theta, \theta_0) > -\varepsilon \right\} = 1.$$

The first equality is due to the anti-symmetry of D_n. This confirms the validity of (4).

Let O be a small open ball centered at θ_0. Now, if the theorem was false we would have $\limsup_{n\to\infty} P\left\{\widehat{\theta}_n \notin O\right\} > 0$. We shall use this to contradict (4). Let $T = \Theta_0 \backslash O$ in (6), which then implies

$$\lim_{n\to\infty} P\left\{\inf_{\theta\notin O} D_n(\theta,\theta_0) > \delta\right\} = 1,$$

since we can take $\varepsilon = \delta$ and $\inf_{\theta\in T} D(\theta) = 2\delta > 0$ by Assumption 2 and the properties of D and μ. This, in turn, implies

$$\lim_{n\to\infty} P\left\{\inf_{\theta\notin O}\sup_{\vartheta\in\Theta_0} D_n(\theta,\vartheta) > \delta\right\} = 1. \tag{8}$$

Now, write

$$\begin{aligned}
&\limsup_{n\to\infty} P\left\{\inf_{\theta\in\Theta_0}\sup_{\vartheta\in\Theta_0} D_n(\theta,\vartheta) \geqslant \delta\right\} \geqslant \\
&\limsup_{n\to\infty} P\left\{\inf_{\theta\in\Theta_0}\sup_{\vartheta\in\Theta_0} D_n(\theta,\vartheta) = \inf_{\theta\notin O}\sup_{\vartheta\in\Theta_0} D_n(\theta,\vartheta),\ \inf_{\theta\notin O}\sup_{\vartheta\in\Theta_0} D_n(\theta,\vartheta) > \delta\right\}.
\end{aligned} \tag{9}$$

By (8) the right-hand side of (9) reduces to

$$\begin{aligned}
&\limsup_{n\to\infty} P\left\{\inf_{\theta\in\Theta_0}\sup_{\vartheta\in\Theta_0} D_n(\theta,\vartheta) = \inf_{\theta\notin O}\sup_{\vartheta\in\Theta_0} D_n(\theta,\vartheta)\right\} \geqslant \\
&\limsup_{n\to\infty} P\left\{\widehat{\theta}_n \notin O\right\} > 0.
\end{aligned}$$

This is in contradiction to (4) and proves the theorem. □

THEOREM 2. *Let f_i and F_i denote the density and distribution functions of Y given $X = x_i$. Write μ_i for $\mu(x_i,\theta_0)$ and μ_x for $\mu(x,\theta_0)$. Assume that*

(i) $0 < f_{Y|X=x} < M\ \forall x \in \mathscr{X}$, *for some* $M \in \mathbb{R}_+$.

(ii) $V \equiv \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \dot{\mu}_i\dot{\mu}_i^{\mathrm{T}} = \int_{\mathscr{X}} \dot{\mu}_x\dot{\mu}_x^{\mathrm{T}}\pi(\mathrm{d}x)$ *has full rank.*

(iii) $\int_{\mathscr{X}} \left\|\dot{\mu}(X,\theta)\right\|^2 \pi(\mathrm{d}x) < \infty \quad \forall\theta \in \Theta_0$.

(iv) $\theta_0 \in \operatorname{interior}(\Theta_0)$.

Then, $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \tau(1-\tau)\Gamma^{-1}V\Gamma^{-1})$, *where*

$$\Gamma = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n f_i(F_i^{-1}(\tau))\dot{\mu}_i\dot{\mu}_i^{\mathrm{T}}.$$

Proof. We show that the conditions for Theorem 3.3 of Pakes and Pollard (1989) are satisfied. We will refer to these as (p-i)–(p-v) to avoid confusion with the assumptions of this theorem. Let $D_L(\theta)$ denote the vector of left partial derivatives of D_n with respect to ϑ, evaluated at $\vartheta = \theta$. Since $\widehat{\theta}_n$ is the optimizer of D_n, the summands of $D_L(\widehat{\theta}_n)$ for which $y_i \neq \mu(x_i, \widehat{\theta}_n)$ cancel out. Thus, the only important contributions to $G_n(\widehat{\theta}_n)$ occur at the kinks $y_i = \mu_i(x_i, \widehat{\theta}_n)$. We can therefore bound the absolute value of the jth coordinate of $G_n(\widehat{\theta}_n)$ by

$$(10)\quad |G_{n,j}(\widehat{\theta}_n)| \leqslant \max_{1 \leqslant i \leqslant n} \{\|\dot{\mu}(x_i, \widehat{\theta}_n)\|\} \frac{1}{n} \sum_{i=1}^{n} 1\{y_i = \mu(x_i, \widehat{\theta}_n)\}(\tau - 1).$$

By assumption (iii), $\max_{1 \leqslant i \leqslant n}\{\|\dot{\mu}(x_i, \widehat{\theta}_n)\|\} \in o_p(n^{1/2})$, and since Y is continuously distributed, the normalized sum is $O_p(n^{-1})$. Therefore we have that $|G_{n,j}(\widehat{\theta}_n)| \in o_p(n^{-1/2})$, and thus also $\|G_n(\widehat{\theta}_n)\| \in o_p(n^{-1/2})$. This argument is similar to one made by Honoré (1992) and shows that (p-i) is satisfied.

Next, write $G(\theta)$ as

$$(11)\quad G(\theta) = \int_S \dot{\mu}_x^{\theta}[\tau - 1\{y \leqslant \mu_x^{\theta}\}]\pi(\mathrm{d}s) = \int_{\mathscr{X}} \dot{\mu}_x^{\theta}[\tau - F_{Y|X=x}(\mu_x^{\theta})]\pi_x(\mathrm{d}x).$$

Assumption (iii) allows us to differentiate under the integral sign (Billingsley, 1995), so:

$$(12)\quad \Gamma \equiv \frac{\partial G}{\partial \theta}\bigg|_{\theta=\theta_0} = -\int_{\mathscr{X}} f_{Y|X=x}(\mu_x)\dot{\mu}_x\dot{\mu}_x^{\mathrm{T}}\pi_x(\mathrm{d}x) = -\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_i(\mu_i)\dot{\mu}_i\dot{\mu}_i^{\mathrm{T}},$$

which by (i) and (ii) exists and is finite. This concludes on (p-ii).

Regarding the condition (p-iii), we have that the function class $\mathscr{I} \equiv \{1\{y_i \leqslant \mu_i^{\theta}\} : s_i \in S, \theta \in \Theta_0\}$ is universally P-Donsker, and so are $-\mathscr{I}$ and $\{\tau - \mu_i^{\theta}\imath : \imath \in \mathscr{I}\}$. The property is retained under addition, cf Van der Vaart and Wellner (1996, Theorem 2.10.6 and example 2.10.7). Therefore, for all positive sequences $\{\delta_n\}$ with $\delta_n \in o(1)$,

$$(13)\quad \sup_{\|\theta-\theta_0\| \leqslant \delta_n} \frac{\|G_n(\theta) - G(\theta) - G_n(\theta_0)\|}{n^{-1/2} + \|G_n(\theta)\| + \|G(\theta)\|} \in o_p(1),$$

cf Chen et al. (2003, Section 4).

Finally, for condition (p-iv), the assumption of independent and identically distributed observations allows an application of the standard central limit theorem:

$$(14)\quad G_n(\theta_0) \xrightarrow{\mathrm{d}} \mathscr{N}(0, J), \quad \text{where}$$
$$J = \tau(1-\tau)\int_{\mathscr{X}} \dot{\mu}_x\dot{\mu}_x^{\mathrm{T}}\pi(\mathrm{d}x)$$
$$= \tau(1-\tau)V.$$

Condition (p-v) is simply assumption (iv). Thus, all conditions for Theorem 3.3 of Pakes and Pollard (1989) are satisfied, and the result of this theorem follows. □

Example 1. As an illustration, consider the linear model. Let $\mu(x,\theta) = x^{\mathrm{T}}\theta$. Then,

$$D_n(\theta,\vartheta) = \frac{1}{n}\sum_{i=1}^{n} x_i^{\mathrm{T}}[\vartheta-\theta][\tau - 1\{y_i \leqslant (x_i^{\mathrm{T}}[\vartheta+\theta])/2\}].$$

The limiting distribution is then

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \tau(1-\tau)\Gamma_l^{-1} H_l \Gamma_l^{-1})$$
$$\Gamma_l = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_i(F_i^{-1}(\tau)) x_i x_i^{\mathrm{T}}$$
$$H_l = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}.$$

It is seen that the asymptotic distribution of the minimax estimator is equivalent to that of the classic quantile regression estimator (Koenker, 2005, Section 3.2.3). As in that case, it is recomended to use e.g. bootstrap methods for the minimax estimator due to the presence of the density function in the asymptotic variance. Examples are the "Bayesian bootstrap", see e.g. Hahn (1997) or Bose and Chatterjee (2003).

5 Behavior of the deviance function

The preceding section presented the deviance criterion function and showed the consistency and asymptotic normality of the minimax estimator $\widehat{\theta}_n$. We now inspect the function a little more closely to get an idea of the behaviour it displays. At first, it appears intractable, since it is piecewise linear with discontinuities. It does, however, reveal some properties that can be exploited for numerical aspects, something we will return to in the next section. For ease of graphical exposition, consider the one-dimensional case of estimating a sample quantile. Then we have $\Theta_0 \subset \mathbb{R}$ and no covariates. Figure 1 shows $D_n(\theta,\vartheta)$ as a function of ϑ for three different fixed values of θ, given the sample points $\{-5,-4,\ldots,5\}$, with $\tau = 1/2$. It is clearly seen that the further away θ is from the sample median, the easier it becomes to choose ϑ in order to get a large, positive function value. As θ approaches the sample median, the function flattens, and the supremum function value decreases. When θ equals the sample median, here $\theta = 0$, the best one can do in terms of choosing ϑ, is to set $\vartheta = \theta$ and get a function value of 0. Also, note that when $\theta < \theta_0$, the maximum function value is achieved by setting $\vartheta > \theta$, and vice versa. This gives an indication of the direction in which one should direct θ to find, in this case, the sample median.

An alternative geometric interpretation is as follows. Again, let us stick with the simplest possible case of finding a sample quantile. Let $\Delta(\vartheta) = \#\{i : y_i \leqslant (\theta+\vartheta)/2\}$

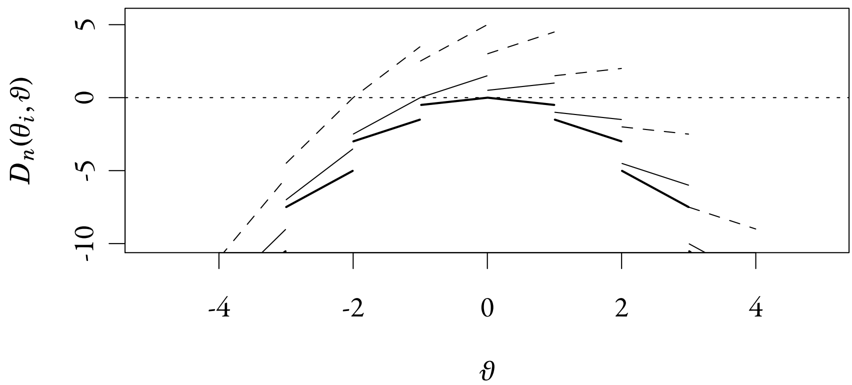


Figure 1: The deviance function $D_n(\theta, \vartheta)$ as a function of ϑ for three different θs. The dashed line represents $\theta_1 = -2$; the solid is for $\theta_2 = -1$; and when $\theta_3 = 0$, the sample median, we get the solid bold graph.

be the number of observations with a value less than or equal to $\bar{\mu}_i = (\theta + \vartheta)/2$, viewed as a function of ϑ. Then, if the sample is ordered with $y_1 \leqslant y_2 \leqslant \cdots \leqslant y_n$, and we fix $\theta = \theta^*$, we can write

$$\begin{aligned} nD_n(\vartheta) = nD_n(\theta^*; \vartheta) &= [\vartheta - \theta^*] \sum_{i=1}^{\Delta(\vartheta)} (\tau - 1) + [\vartheta - \theta^*] \sum_{i=\Delta(\vartheta)+1}^{n} \tau \\ &= [\vartheta - \theta^*][n\tau - \Delta(\vartheta)]. \end{aligned}$$

Given a θ, then ϑ is chosen to maximize the area of a rectangle where one side is increasing in ϑ and the other is decreasing. The job is then to chose a θ for which no ϑ can make this area positive. In higher dimensions, it is a little more complicated. As the $[\mu_i^{\vartheta} - \mu_i^{\theta}]$ term no longer can be taken outside the sums, we have many rectangles for which the sides are determined by a common parameter vector. The parameter θ must then be chosen such that no ϑ can make the sum of the areas positive, where areas can be negative. This does not as such get us closer to how one would go about computing this value of θ. The next section gives a suggestion on how to do this in practice.

6 Computing minimax regression quantiles

At first, the deviance function appears to be problematic: it is discontinuous and not everywhere differentiable. Furthermore, it is a function of both the model $\mu(\cdot,\theta)$ and the "deviant" model $\mu(\cdot,\vartheta)$. It is therefore natural to question its tractability in practice. It does, however, have some properties that allow for the implementation of relatively simple algorithms in the pursuit of the optimizer $\hat{\theta}_n$.

Importantly, D_n behaves similarly to partial derivatives for small model deviance: at $\theta = \vartheta$ it evaluates to 0, and when $D_n(\theta,\vartheta) > 0$ (which by property p1 happens for some ϑ when θ is not the optimizer) θ should be moved in the direction towards ϑ. If for some θ we have $D_n < 0$ for all values of ϑ, then θ is the optimizer. This behavior can be seen in Figure 1. Whenever μ is smooth enough one will find a good direction by examining D_n for ϑ in a small neighborhood of θ. Also, $\sup_\vartheta D_n(\theta,\vartheta)$ is much smoother than D_n itself and increases as θ moves away from optimum.

This suggests that one can use modified gradient type methods to solve the optimization problem. It turns out, for example, that Shor's r-algorithm (Shor et al., 1985) is quite successful here. More specifically, consider the "pseudo-gradient"

$$\delta D_n(\theta) = (\delta_1 D_n(\theta),\ldots,\delta_k D_n(\theta))^{\mathrm{T}}$$

with entries defined as follows. Let $\epsilon > 0$ be a small number and ϵ^j be a k-vector with the value ϵ in the jth entry and 0 elsewhere. Then, with

$$D^-_{nj}(\theta) \equiv D_n(\theta,\theta-\epsilon^j)/\epsilon$$
$$D^+_{nj}(\theta) \equiv D_n(\theta,\theta+\epsilon^j)/\epsilon$$

let

$$(15)\quad \delta_j D_n(\theta) \equiv \begin{cases} 0 & \text{if } D^+_{nj}(\theta) < 0 \text{ and } D^-_{nj}(\theta) < 0 \\ D^+_{nj}(\theta) & \text{otherwise} \end{cases}$$

Thus, $\delta D_n(\theta)$ examines, for each entry in θ, the effect of small model deviance in each direction, and chooses the direction accordingly.

The modified r-algorithm is then as follows:

1. Initialize $B_1 = I_{k\times k}$ (identity matrix)
2. Fix $\gamma \in (0,1)$.
3. $\theta_{l+1} = \theta_l + \lambda_l B_l B_l^{\mathrm{T}} \delta D_n(\theta_l)$
   where the step length λ_l is found to maximize $D_n(\theta_l,\theta_{l+1})$ (e.g. by a line-search)

4. $r_{l+1} = B_l^{\mathrm{T}}[\delta D_n(\theta_{l+1}) - \delta D_n(\theta_l)]$ and normalized to have length 1
5. $B_{l+1} = B_l[I_{k\times k} - \gamma r_{l+1} r_{l+1}^{\mathrm{T}}]$
6. Repeat 3–5 until $\delta D_n(\theta_l) = 0$ (note that the tolerance, ϵ, influences when this occurs).

The matrix B_l is called the space dilation matrix and has the interpretation of changing the coordinate system. Shor's r-algorithm is designed to optimize non-smooth functions, and its reported success is considerable even though few theoretical convergence results exist (Burke et al., 2008). In simulations it appears that the parameter γ is not critical, although it affects the number of iterations required.

It might happen that $|\delta_j D_n(\theta_{l+1}) - \delta_j D_n(\theta_l)|$ is too close to 0 for all j and makes r_{l+1} numerically undefined. Often this happens very close to the optimum, as the step-size λ_l could have been increased (this is also found up to a tolerance), and usually one can stop. Alternatively, re-initialize $B_{l+1} = I_{k\times k}$ and continue to improve precision.

7 A generalized minimax estimator

The basic estimator in the minimax framework offers little improvement over the classical quantile regression estimator: it has the same asymptotic properties and, indeed, finite sample solutions often coincide. The real incentive for the minimax framework is its potential for generalization. There are some intriguing challenges with quantile regression and panel data, see for example a survey by Bache et al. (2011). Here we consider how to allow for non-diagonal weighting as an alternative way to accommodate correlated data.

Let data, $\{s_{it}\} = \{y_{it}, x_{it}\}$ $(i = 1,\dots,n; t = 1, t\dots,T)$, be characterized by two dimensions where observation T-pairs are IID over i and have common correlation structure over t. Denote by $B_i(\theta)$ the vector $(b(s_{i1},\theta),\dots,b(s_{iT},\theta))^{\mathrm{T}}$; and by $M_i(\theta)$ the vector $(\mu(x_{i1},\theta),\dots,\mu(x_{iT},\theta))^{\mathrm{T}}$. Let $\dot{M}_i(\theta)$ be the $T\times k$ matrix of derivatives. Now, we generalize the constant variance term $V(b(\theta_0)) = \tau(1-\tau)$ to $W(\theta,\alpha) = V(B_i(\theta_0))^{1/2} R(\theta,\alpha) V(B_i(\theta_0))^{1/2}$, where $V(B_i(\theta_0))$ is a diagonal variance matrix with entries $V(b(\theta_0))$ which in this simple case all equal $\tau(1-\tau)$ and can be disregarded. The matrix $R(\theta,\alpha)$ is called the "working correlation matrix", and is specified up to the parameter vector α which is determined for each θ (so R is essentially viewed as a function of θ alone, but α may be estimated for each θ). We simplify notation and write $R(\theta) \equiv R(\theta,\alpha(\theta))$ and $\hat{R}(\theta) = R(\theta,\hat{\alpha}(\theta))$.

The generalized estimating function we consider is

$$\widetilde{G}(s,\theta) = \dot{M}(s,\theta) W^{-1}(\theta) B(s,\theta)$$

and the population equation is then

$$\widetilde{G}(\theta_0) = \int_S \widetilde{G}(s,\theta_0) = \dot{M}(s,\theta_0)^{\mathrm{T}} W^{-1}(\theta_0) B(s,\theta_0)\pi(\mathrm{d}s) = 0.$$

Note that the estimating function is unbiased for all values of α in $W(\theta,\alpha(\theta))$ a property called E-ancillarity. This very important property means that the correlation structure does not need to be correctly specified for consistency, a remarkable result in the literature of generalized estimating equations.

The empirical estimating equation is

$$\widetilde{G}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \dot{M}_i(\theta)^{\mathrm{T}} \hat{W}^{-1}(\theta) B_i(\theta) = 0.$$

We can then construct a minimax criterion function as

$$\widetilde{D}(\theta,\vartheta) = \int_S [M(s,\vartheta) - M(s,\theta)]^{\mathrm{T}} \hat{W}^{-1}(\theta,\vartheta) B(s,\theta,\vartheta)\pi(\mathrm{d}s)$$

with

$$\begin{aligned} B(s,\theta,\vartheta) &= (\tau - 1\{y_1 \leqslant \bar{\mu}(x_1,\theta,\vartheta)\},\ldots,\tau - 1\{y_T \leqslant \bar{\mu}(x_T,\theta,\vartheta)\})^{\mathrm{T}} \\ W(\theta,\vartheta) &= \mathrm{E}[B(s,\theta,\vartheta)B(s,\theta,\vartheta)^{\mathrm{T}}]. \end{aligned}$$

Note that the relevant correlation structure is that of the Bs and not the ys themselves.

The empirical generalized minimax criterion is

$$\begin{aligned} \widetilde{D}_n(\theta,\vartheta) &= \frac{1}{n}\sum_{i=1}^{n} [M_i(\vartheta) - M_i(\theta)]^{\mathrm{T}} \hat{W}_i^{-1}(\theta,\vartheta) B_i(\theta,\vartheta) \\ &\propto \frac{1}{n}\sum_{i=1}^{n} [M_i(\vartheta) - M_i(\theta)]^{\mathrm{T}} \hat{R}_i^{-1}(\theta,\vartheta) B_i(\theta,\vartheta). \end{aligned}$$

Theorem 3. *Assume that that $\hat{R}(\theta)$ is positive definite, along with appropriate analogues of assumptions 1 and 2. Then any minimax point*

$$(16)\quad \widehat{\theta}_n = \underset{\theta\in\Theta_0}{\arg\inf}\ \sup_{\vartheta\in\Theta_0} \widetilde{D}_n(\theta,\vartheta)$$

is consistent, i.e. $\widehat{\theta}_n \xrightarrow{p} \theta_0$.

Proof. Since $\hat{R}$ is a symmetric positive definite matrix, we can use a spectral decomposition to write $\hat{R}^{-1}(\theta) = U(\theta)L(\theta)U(\theta)^{\mathrm{T}}$, where L is a diagonal matrix with entries $1/\hat{e}_j$; $\hat{e}_j$ are the eigen values of $\hat{R}$; and U is a matrix of eigen vectors. Since the eigen values are strictly positive we have that R^{-1} is bounded. Each $\hat{e}_j = \hat{e}_j(\theta)$

are fixed functions of θ and $0 < 1/\hat{e}_j < \infty$. Furthermore the set of matrices $U(\theta)$ is compact and totally bounded (it is the set of rotations on the unit sphere). This together implies that the VC-subgraph property is retained and so $\widetilde{D}_n$ converges uniformly to $\widetilde{D}$. To see that $\widetilde{D}$ has the essential property (p1), realize that we sum terms of the form $m_t r_{tt'} b_{t'}$, where $m_t = \mu(x_{\cdot t}, \vartheta) - \mu(x_{\cdot t}, \theta_0)$, $b_t = \mathrm{E}b(s_{\cdot t}, \theta_0, \vartheta)$, and $r_{tt'} = \mathrm{cor}(b_t, b_{t'})$. If $m_t > 0$ and $r_{tt'} > 0$ then $b_{t'} < 0$. On the other hand if $m_t > 0$ and $r_{tt'} < 0$ then $b_{t'} > 0$. Argue similarly for $m_t < 0$. The products are therefore always negative when $\theta = \theta_0$ and $\vartheta \neq \theta_0$. Now, the result follows with arguments analogous to those in the proof of Theorem 1. □

Example 2. We round off the discussion with a small simulation exercise to illustrate the generalized minimax quantile regression estimator. Let $n = 99$, $T = 4$, and

$$\begin{aligned}
v_{it} &\sim \mathcal{N}(0,1)\\
\varepsilon_{i1} &= v_{it}\\
\varepsilon_{it} &= 0.8\varepsilon_{i,t-1} + v_{it}, \quad (t = 2,\dots,T)\\
u_{it} &\sim \mathcal{U}(0,2)\\
\bar{x}_i &\sim \mathcal{U}(0,5)\\
x_{it} &= \bar{x}_i + u_{it}\\
y_{it} &= 1 + 2x_{it} + (1 + x_{it})\varepsilon_{it}
\end{aligned}$$

We now use both classic quantile regression and the new estimator with $\tau = 1/2$ to estimate $\theta_0 = (1,2)$. R is specified in the simplest form as the sample-correlation matrix:

$$(17) \quad R(\alpha(\theta)) = \begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{12} & 1 & \alpha_{23} & \alpha_{24} \\ \alpha_{13} & \alpha_{23} & 1 & \alpha_{34} \\ \alpha_{14} & \alpha_{24} & \alpha_{34} & 1 \end{pmatrix}.$$

From a simulation with 500 replications, Table 1 reports a few comforting statistics from the estimated distributions. It is seen that in this example both range, interquartile range, and root mean squared error (RMSE) are smaller for the covariance-weighted generalized minimax estimator, in particular for $\hat{\theta}_1$. Figure 2 visualizes the simulated distributions for θ_1.

While we use a very simple specification for R above, one can use more complex correlation structures and relax the assumption of a balanced panel.

	Classic QR		(Generalized) Minimax QR	
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$
Minimum	-1.124	1.100	-0.9690	1.152
1st quartile.	0.471	1.816	0.5094	1.850
Median	0.944	2.032	0.9872	2.019
3rd quartile.	1.532	2.222	1.3990	2.198
Maximum	3.333	2.971	3.0840	2.813
RMSE	0.792	0.303	0.6733	0.265

Table 1: Statistics from the simulated distributions of the classic quantile regression estimator and the generalized minimax quantile regression estimator.

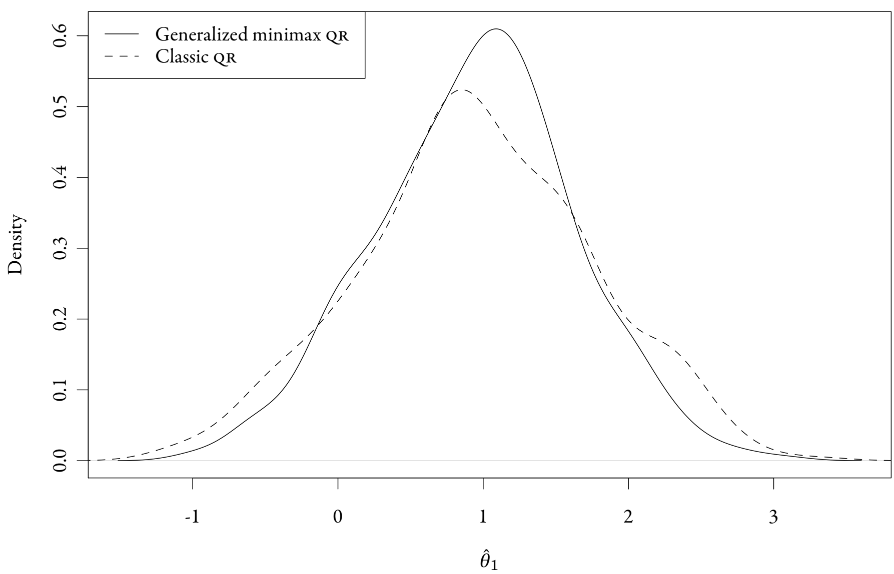


Figure 2: Simulated distribution of $\hat{\theta}_1$ for the classic quantile regression estimator (dashed) and the (generalized) minimax estimator (solid).

8 Concluding remarks

The minimax framework developed in this paper offers a new way to perceive regression quantiles: the estimating functions described provide an intuitive description of the quantity of interest as the roots of these functions. They do not, however, give a practical way to obtain this quantity as they may not have exact roots in finite samples. A deviance function is constructed to give consistent roots to these functions and it has restated the problem as that of a certain type of extremum estimators. This is interesting in itself, as it provides a quantile regression framework which incorporates a wide class of regression functions, and new numerical possibilities when model specification does not admit linear programming algorithms. This framework, as a building block, may prove useful in alleviating some problems related to quantile regression; a quite important example was provided with non-diagonal weighting in the spirit of generalized estimating equations. In this example, the generalized estimator had improved efficiency compared to a quantile regression estimator which ignores correlation.

9 Acknowledgements

The Author is grateful to Jørgen Hoffmann-Jørgensen, Roger Koenker, Michael Jansson, Christian M. Dahl, and Anders Bredahl Kock for helpful comments and suggestions. Financial support by the Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation, is gratefully acknowledged.

10 References

Bache, S. H., C. M. Dahl, and J. T. Kristensen (2011). Headlights on tobacco road to low birthweight: evidence from a battery of quantile regression estimators. *Empirical Economics, forthcoming*.

Billingsley, P. (1995, April). *Probability and Measure* (3 ed.). Wiley-Interscience.

Bose, A. and S. Chatterjee (2003). Generalized bootstrap for estimators of minimizers of convex functions. *Journal of statistical planning and inference 117*(2), 225–239.

Burke, J., A. Lewis, and M. Overton (2008). The speed of shor's r-algorithm. *IMA journal of numerical analysis 28*(4), 711–720.

Chen, X., O. Linton, and I. Van Keilegom (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica 71*(5), 1591–1608.

Ferreira, P. (1982). Multiparametric estimating equations. *Annals of the Institute of Statistical Mathematics 34*(1), 423–431.

Godambe, V. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics 31*(4), 1208–1211.

Godambe, V. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika 72*(2), 419–428.

Godambe, V. (1987). The foundations of finite sample estimation in stochastic processes - ii. *Proc. First World Congress of Bernoulli Society 2*, 49–59.

Godambe, V. (2001). Estimation of median: Quasi-likelihood and optimum estimating functions. *Working Paper, University of Waterloo, Statistics and Actuarial Sciences Working Paper Series*.

Godambe, V. and M. Thompson (1984). Robust estimation through estimating equations. *Biometrika 71*, 115–125.

Hahn, J. (1997). Bayesian bootstrap of the quantile regression estimator: a large sample study. *international economic review 38*(4), 795–808.

Honoré, B. (1992). Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica 60*(3), 533–565.

Jung, S. (1996). Quasi-Likelihood for Median Regression Models. *Journal of the American Statistical Association 91*(433), 251–267.

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

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Li, B. (1996). A minimax approach to consistency and efficiency for estimating equations. *The Annals of Statistics 24*, 1283–1297.

Li, B. (1997). On the consistency of generalized estimating equations. *Lecture Notes-Monograph Series 32*, 115–136.

Liang, K. and S. Zeger (1986a). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics 42*(1), 121–130.

Liang, K. and S. Zeger (1986b). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Shor, N. Z., K. C. Kiwiel, and A. Ruszcaynski (1985). *Minimization methods for non-differentiable functions*. New York, NY, USA: Springer-Verlag New York, Inc.

Van der Vaart, A. and J. Wellner (1996). *Weak convergence and empirical processes*. Springer Verlag.

Preface to paper #3

It is gratifying to experience interest in ones work, and we have been fortunate that our work on correlated random effects methods for quantile regression (paper #1) has made other researchers interested in the potential for implementing these in other projects.

One such project, initiated by professor Ulrich Kaiser, was about examining the returns to self-employment in Denmark, utilizing rich register data from Statistics Denmark. However, after giving a lot of thought to the problem we started to suspect that we would start by taking a different approach.

The main problem is the inherent "multivariate nature" of income types in that characterization of the income streams is very heterogeneous. Also the accumulation of experience in different types of employment seems to be very important. The dynamics of this evolution, switching in and out of self-employment, and the very different types of income seems to opt for a descriptively very thourough analysis.

We came up with what is probably the main contribution of this paper, a quantile regression model for career composition. It is not a causal model; rather it is descriptive in nature. It exploits the fact that one can argue that total career income (or average income) is the (or *an*) important income measure, and characterizes which income levels are observed in different occupations at various quantile indices of this (weighted) aggregated income. Here, the panel problem is in some sense transformed into a cross-sectional problem. The model indirectly shows when and why quantile regression estimates (in terms of the framework discussed in the preface to paper #1) cannot be interpreted as causal.

Although not causal, our model does show how the "current" environment turned out to favor various occupations.

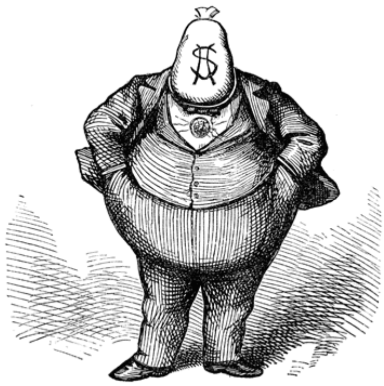
Paper #3

Does entrepreneurship really (not) pay?

2011

Stefan Holst Bache
Aarhus University, Department of Economics and Business and CREATES
Christian Møller Dahl
University of Southern Denmark, Department of Economics and CREATES
Ulrich Kaiser
University of Zurich, Department of Business Administration

Keywords:
Entrepreneurship, self-employment, quantile regression, career composition



Abstract. We investigate patterns in the distribution of income in wage- and self-employment on an aggregate level to identify important aspects for analyses of potential wage differentials. We employ a high-quality Danish register data set which allows us to construct good measures of longer-term income and experience. A variety of quantile-based methods are used to show relationships between experience in self-employment, labor-market experience and income at various locations of the conditional distribution. We find that the timing of income and career income stream is very important to model explicitly because income patterns are very different in wage-employment and self-employment. To this end we develop a quantile regression model of career composition in terms of earned experience and occupational choice where the response variable is (discounted) average income. The results from estimating the model strongly indicate that early stages of the firm are critical and associated with very low income and should be targeted for in-depth analysis of likely causes. Self-employment at other stages of the experience profiles seems to be competitive with wage-employment, and our findings also suggest great benefits to the entrepreneur from experience in wage-employment. We also find support for the jack-of-all-trades theory, which suggests that broad experience benefits the entrepreneur.

1 A good question with many (possible) answers

With his title, Hamilton (2000) posed the simple question "Does entrepreneurship pay?". Tergiman (2011) answers and entitles her study by a reordering of words: "Entrepreneurship does pay." Her conclusion challenges that of Hamilton's study, which concluded that pecuniary reasoning alone does not offer strong motivation for self-employment. The question asked by Hamilton is still the source of active debate. Why? Well, because policy makers all over the world tend to regard self-employment as a route out of poverty and social disadvantage as well as an engine of innovation, job creation and product-market competition (Blanchflower, 2000). Also, many economists have regarded self-employment as a major determinant of innovative activity ever since Schumpeter (1942). Finally, the social value of self-employment is undisputed, as pointed out by Baumol (1990) as well as Murphy et al. (1991). Van Praag and Versloot (2007) provide a survey of the more recent literature on the societal value of entrepreneurship. With self-employment being such a precious benefactor to society and the economy, it is somewhat surprising that comparatively little research has focused on the private incentives for self-employment. What's in it for the entrepreneur? From an economic policy perspective, answers to this question are quite valuable as they can indicate what political devices can be used to encourage self-employment and innovation.

Amongst existing studies on the topic there seem to be conflicting results. Some find evidence that in spite of a higher risk in self-employment, mean (and median) income is actually *below* the wages of the dependently employed, at least in the US, as found in the reviews by Van der Sluis et al. (2007) and Van Praag and Versloot (2007). This may suggest that the private value of self-employment is below the social value of self-employment. The existence of wage gaps between wage-employed and self-employed individuals leads many authors to conclude that there are substantial non-pecuniary (or "hedonic") values in connection with self-employment, see e.g. also Moskowitz and Vissing-Jørgensen (2002). Others find opposing conclusions. Tergiman (2011) tries to explain earnings differentials in a life-cycle model of occupational choice and concludes that entrepreneurship *does* pay both in early stages and late in the self-employed's life. le Maire and Scherning (2007) use a dynamic sample selection method to model income and its impact on the occupational choice. They find evidence that income level and variance are indeed important determinants, and that the level is more important to men, while the risk is more important to women. They also find mean earnings to be higher in self-employment for a certain income measure.

The opening line by le Maire and Scherning (2007) reads "Compared to wage work, self-employment is a fundamentally different occupation with respect to type and source of income." Indeed. It is also fundamentally different in many other aspects. Psychology literature on the entrepreneurial profile is vast, so much that Gartner (2007) coins it the "critical mess." Not only is it different to wage-work, it is highly diverse in itself. The question of whether entrepreneurship pays (the entrepreneur) is clearly complex, and the opposing answers may well be attributes of an entrepreneurial mess (economic as well as psychological). Here are a few questions on which the answer may depend critically:

- *what is an entrepreneur?* There are currently several definitions of entrepreneurship and entrepreneurial thinking in use. Who is the "main character"?

- *what is the relevant income?* How should the pecuniary benefits in self-employment be defined and measured, and how can these be compared to those of a dependently employed individual.

- *what is the relevant environment?* Rules, regulations, population characteristics, etc. determine existing opportunities and are key in opportunity creation for entrepreneurs.

We touch upon these questions in the next section where our data are presented. In addition to different outsets, in terms of the points mentioned above, a likely reason for the many opposing results is that it is very difficult to make causal analysis at a highly aggregated level while controlling for the very heterogeneous nature of the work force and selection mechanisms. In this paper it is our aim to obtain indicators at the aggregate level which can guide the direction of future research and analyses with a more specific micro-level focus. We do this in part by using high-quality data which allow for the construction of a purposeful definition of relevant income. We develop a simple model which takes the *income stream* into account, rather than *single observations*. It is a panel data model, but not in a traditional sense. The argument is that self-employed individuals often face a more erratic income stream than employees, in part because fluctuations directly affect the individual's income but also due to e.g. an expensive startup and initialization of a firm. The model is extended slightly to investigate whether a diverse experience can be beneficial to the entrepreneur, as predicted by the jack-of-all-trades theory by Lazear (2004). Before developing the model we present some benchmark regressions ala Hamilton (2000) and describe some aspects and trends found in the data.

2 A tour of the givens

In our study we use population data from parts of the so-called "Integrated Data Base for Labor Market Research (IDA)" provided by Statistics Denmark. Details about IDA are provided by Abowd and Kramarz (1999). It contains register data on all Danish residents. The data base provides detailed information on income and a wide range of other individual–specific characteristics, like educational background. It also provides information on the occupation from which the individuals earn their primary income.

Ida's information on annual income stems from the tax system and is supplemented by other registers maintained by Statistics Denmark. Our use of data from the tax system is hence like Gort and Lee (2007), and unlike Hamilton (2000), who uses self–reported income.[10] It is important to note that our data do not constitute a sample but the entire population of Danish residents observed between 1992 and 2002. Hence, we avoid certain sample selection issues. The period 1980–1991 is used as a "burn-in" period, in which we initiate the accumulation of experience.

The occupation of an individual in a given year is determined by Statistics Denmark according to the individual's primary labor market status established in the last week of November. A self–employed is defined by Statistics Denmark, and hence by us in this study, as an individual who owns a non–incorporated business, and either has employees or enjoys a wage–employment income below a fixed threshold, and is not registered as unemployed or non–employed (e.g. paternity leave).

The following restrictions are applied to our data to remedy a few potential issues. We (i) only consider males in order to avoid problems with maternity leave and discrimination issues; (ii) we restrict attention to individuals aged 31 to 59 years to make sure people have completed their educations and have not yet retired; (iii) we do not include public sector workers since wages in that sector are to a large extent seniority–based, and while the public sector could be a relevant alternative, it has no direct self-employment equivalent; (iv) we neither include individuals working in agriculture, since production is very different from manufacturing and services, nor workers whose sector of employment is unknown; (v) we exclude individuals from Greenland, individuals with an unknown place of residence, part–time workers, and individuals who did not work a full year. Finally, we only consider Danish citizens since the educational variables are poorly recorded for immigrants (Kaiser and Malchow-Møller, 2010) and there is evidence that they are marginalized into self-employment (le Maire and Scherning, 2007). The age restriction in (ii) is on the one hand restrictive: if age-profiles differ at either end it might distort measures of life-span income; On the other hand, dealing with e.g. systematic retirement differences, early retirement, etc., would complicate matters even more and is outside the scope of our analysis.

Our dependent variable is based on annual income, measured in the Danish currency (dkk). We use price indices obtained from Statistics Denmark to get income measure in year-2002 prices. While there exist estimates of hourly wages for wage-employed workers in ida, no such is available for self-employed and the annual measure is used for both types of occupation. We do, however, scale our income measure

[10]Income from self-employment is largely self-reported to the tax system, and misreporting constitutes tax-evasion.

	#	Description
Sector	1	Industry
	2	Energy
	3	Construction
	4	Trade
	5	Transportation
	6	Finance
Education	1	No qualifying education.
	2	craftsmen, skilled.
	3	Secondary pre-university high school (3 years), or technical college, etc. (2–-5 years).
	4	short- to medium-cycle higher education programme (1-–4 years).
	5	Long-cycle higher education programme (5+ years).

Table 1: Table of categorical variables. These are aggregations of detailed categorizations by Statistics Denmark.

to mimic an "hourly income" (also in DKK), according to the workload assumed in Danish employment contracts (which is set to 1820 hours/year). This is obviously sub-optimal, since some people work harder and longer than others. One can argue that many self-employed people in general work longer hours, and hence we face an upward bias for self-employed individuals. We only consider individuals' primary labor income, and thus exclude other secondary forms of income e.g. capital income, deductions, transfers, gifts, etc. For self-employment we measure income by net profit (revenues less expenses) from own business. This consists of the draw plus retained earnings. This is one of the three measures also used by Hamilton (2000). For wage-employed people, income is taken to be their net wage. We use income before tax to avoid issues of tax-optimizing income smoothing across years and tax brackets, various deductions, and other potentially distorting effects of the tax system.

Our main explanatory variables are (i) years of experience accumulated in the different occupational groupings. These are constructed from the primary occupations since 1980; (ii) business sector; and (iii) education. The latter two categorical variables are described in Table 1.

In addition to missing information on hours worked, there are a few potential problems with the data available. First, while income in wage-employment is mostly reported to the authorities by third-party (most importantly the employer), many self-employed individuals report their income themselves. Kleven et al. (2011) show that the tax-evasion rate is close to zero when income is reported by third-party, but substantial in comparison when income income is self-reported. However, they conclude that the overall tax-evasion rate in Denmark appears to very small in spite of the high marginal tax rate. Their study employs a tax-enforcement field experimental design and Danish tax data to reach this conclusion. Therefore, our income measure for self-employed individuals may be slightly downward biased. The Danish tax system, which is quite complicated with high margins compared to other industrialized countries, and the geography (and demography) are major characteristics of the country, ant thus of the relevant entrepreneurial environment. Our results, therefore, should not be applied uncritically to different environments. The methods, however, are easily implemented if data are available.

Second, using non-incorporated businesses as our definition of self-employment may exclude individuals whom it would be appropriate to classify as self-employed. In Denmark, "anpartselskab (ApS)" (loosely corresponding to a limited liability company) is an example of a commonly used company form where Statistics Denmark considers the owners as employees. In some instances this may be appropriate for our purpose, but certainly not in all. Indeed, more successful firms may become incorporated, in which case returns to self-employment are underestimated. On the other hand, from an expected-return-versus-risk point of view, one could argue that an incorporated firm has limited risk and is often jointly owned and managed. Optimally, incorporated firms should be treated as a separate category. Our particular subset of the IDA database does not allow us to remedy these problems.

Table 2 provides some descriptive statistics for the most recent years of our sample. It is immediately clear that the risk associated with income in self-employment is high compared to that in wage-employment, both when considering the standard deviation of mean income and the listed quantiles. Mean income in self-employment seems to be slightly higher when based on observations from a single year. If, however, one considers the distribution of individual mean-income obtained in the two occupations, wage-employment seems to have both higher expected income, and a much lower associated risk. In all cases, self-employment has a lower median income but a higher upper-quartile income. Another interesting observation is that, on average, individuals in self-employment also enjoy a good amount of experience in wage-employment, whereas the converse is not the case. Together, the observations made from the descriptive statistics motivate the model in Section 6.

		1999		2000		2001		2002		Within	
		SE	WE	SE	WE	SE	WE	SE	WE	SE	WE
	# Obs.	61942	490636	61004	502431	59798	508798	57813	504421	135774	816613
Income	mean	208.20	199.62	205.02	200.79	205.10	203.06	207.98	203.98	169.79	185.31
		(292.66)	(100.60)	(294.49)	(106.07)	(331.78)	(110.38)	(1114.72)	(107.81)	(253.78)	(86.76)
	10%	3.89	126.60	2.17	126.10	0.55	126.63	13.40	127.99	0.00	119.86
	25%	75.82	145.13	74.22	145.10	72.16	146.20	78.56	147.93	52.83	138.42
	50%	162.03	175.59	159.86	175.80	157.30	177.52	158.08	178.52	132.14	165.68
	75%	272.33	226.29	266.74	227.22	263.19	229.72	257.92	230.00	228.87	210.71
	90%	432.57	296.94	423.33	300.18	417.07	304.10	406.39	303.22	361.24	274.08
Experience	Age	46.18	43.71	46.19	43.80	46.27	43.93	46.29	43.99		
		(8.10)	(8.22)	(8.18)	(8.25)	(8.25)	(8.27)	(8.27)	(8.27)		
	SE	10.13	0.56	10.29	0.58	10.48	0.60	10.67	0.58		
		(6.08)	(2.05)	(6.37)	(2.13)	(6.62)	(2.19)	(6.85)	(2.17)		
	WE	7.33	16.90	7.96	17.58	8.54	18.28	9.23	18.99		
		(5.53)	(3.08)	(5.76)	(3.32)	(5.96)	(3.55)	(6.17)	(3.75)		
	UE	0.67	0.58	0.71	0.63	0.74	0.65	0.73	0.67		
		(1.42)	(1.26)	(1.46)	(1.33)	(1.50)	(1.37)	(1.47)	(1.39)		
	NE	0.57	0.58	0.62	0.66	0.67	0.73	0.66	0.79		
		(1.47)	(1.34)	(1.52)	(1.44)	(1.58)	(1.52)	(1.51)	(1.58)		
Education	# 1	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.26		
	# 2	0.55	0.52	0.55	0.51	0.55	0.51	0.55	0.51		
	# 3	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06		
	# 4	0.06	0.09	0.06	0.09	0.06	0.09	0.06	0.09		
	# 5	0.06	0.06	0.06	0.07	0.06	0.07	0.06	0.07		
Sector	# 1	0.10	0.33	0.10	0.33	0.10	0.33	0.09	0.33		
	# 2	0.00	0.02	0.00	0.02	0.00	0.02	0.01	0.02		
	# 3	0.22	0.12	0.22	0.13	0.23	0.12	0.23	0.12		
	# 4	0.34	0.21	0.33	0.21	0.32	0.21	0.31	0.21		
	# 5	0.12	0.14	0.11	0.14	0.11	0.14	0.11	0.14		
	# 6	0.23	0.18	0.23	0.19	0.24	0.19	0.24	0.20		

Table 2: Descriptive statistics. Means (and quantiles for income) and standard deviations (in parentheses). SE denotes self-employment, WE denotes wage-employment, UE denotes unemployment, and NE denotes non-employment. The "within" columns contain means and quantiles of individual mean-income.

3 In the footsteps: another regression

Hamilton (2000) used linear quantile regression to learn about the association between income in self-employment and wage-employment, respectively, and the level of experience at the three quartiles of the conditional income distribution. We will follow in Hamilton's footsteps, yet with a slightly higher level of flexibility.

First, recall the following formulation of the quantile regression problem. For a continuous random response variable of interest, W, with conditional distribution function $F_W(w|X)$ for conditioning set X, we define the τth conditional quantile of

W as

$$(1) \qquad Q_W(\tau|X) = F_W^{-1}(\tau|X) \equiv \underset{q \in \mathcal{Q}}{\arg\inf}\ \mathbb{E}[\rho_\tau(W - q(\tau, X))],$$

where $\mathcal{Q}$ is the set of all measurable functions, and $\rho_\tau(z) = z(\tau - 1\{z \leqslant 0\})$. We can now think of W in terms of its conditional quantile function and a (random) rank variable:

$$(2) \qquad W = Q_W(U|X), \quad U|X \sim \text{uniform}(0,1).$$

For linear parametric quantile regression, one then restricts $\mathcal{Q}$ to a linear subclass and lets $q(\tau, X) = X'\beta(\tau)$. Then by solving the empirical problem

$$(3) \qquad \min_\beta \frac{1}{N} \sum_{i=1}^{N} \rho_\tau(W - X'\beta)$$

one obtains estimates $\widehat{\beta}(\tau)$.

In many cases a linear specification is of course an approximation. Angrist et al. (2006) provide results showing that, in a certain sense, the estimator provides the best linear approximation to the true conditional quantile. We stress here that we do not think of W as generated by a linear data-generating process (DGP). A linear DGP need not correspond to a linear conditional quantile and vice versa. Actually, in some cases a linear quantile regression formulation is fully general, no matter how data are generated. To illustrate this, let $W_1 = F_1^{-1}(U_1)$ for uniform U and $W_2 = F_2^{-1}(U_2)$. Think of subscripts 1 and 2 to denote wage-employment and self-employment. Let W denote a variable which may belong to either category, i.e. a random draw from the population of self- and wage-employed individuals. The following regression is "fully general" (but not that exotic):

$$(4) \qquad Q_W(\tau|I^w, I^s) = I^w \alpha_1(\tau) + I^s \alpha_2(\tau).$$

Here I^w is an indicator variable for wage-employment and $I^s = 1 - I^w$ one for self-employment. By "fully general" is meant that $\widehat{\alpha}_1(\tau)$ is a sample quantile of W_1 and similarly $\widehat{\alpha}_2(\tau)$ is a sample quantile of W_2. The form of F_j is not important for consistency of these quantities (except for e.g. continuity of W to make the assumed relations valid). In this sense there are no parametric distributional assumptions.

When adding information to the conditioning set, say business sector, inclusion of an indicator for each sector category is not enough to keep this generality. The difference between the τth quantile for a wage- and self- employed in the trade sector need not be the same as that in the industry sector. One can include interaction terms, and the above generality is preserved.

Building on this idea of generality we estimate a "semi-saturated" model in which we have many categorical intercepts with an added linear approximation for experience and age variables. Let X^s denote XI^s and similarly with X^w for wage-employment. Then, for κ categorical combinations and p experience-related variables we specify the model as

$$\begin{aligned} Q_Y(\tau) &= I_1\alpha_1(\tau)+\cdots+I_K\alpha_\kappa(\tau) \\ &\quad + X_1^s\beta_1(\tau)+\cdots+X_p^s\beta_p(\tau)+X_1^w\beta_{p+1}(\tau)+\cdots+X_p^w\beta_{2p}(\tau). \end{aligned} \tag{5}$$

More specifically, we will include interactions for *occupational choice* (self- or wage employment), *year*, *sector*, and *education*. This gives a total of $\kappa = 660$ categorical intercepts. The experience-related variables included are the number of years in self-employment ($\mathcal{E}^s$), wage-employment ($\mathcal{E}^w$), unemployment ($\mathcal{E}^u$), non-employment ($\mathcal{E}^n$), age, and all of their squares. As a consequence of the added experience terms the $\alpha(\tau)$ estimates no longer correspond to sample quantiles. The model is again an approximation, yet a very flexible one in terms of a large number of intercepts. The $\beta(\tau)$ coefficients are restricted to be shared by all categories to get one common (aggregated) approximate measure of how experience is associated with income.

For inference, we use a weighted bootstrap scheme (sometimes called generalized bootstrap, see Bose and Chatterjee (2003), or Bayesian bootstrap, see Hahn (1997), who shows that the bootstrap distribution converges in probability for the quantile regression estimator). In each bootstrap iteration all observations for each individual are weighted by a unit exponential random variable (weighting individuals rather than observations is done to account for clustering and dependence in observations). In symbols, the procedure is as follows. The main regression solves

$$(\widehat{\alpha}(\tau),\widehat{\beta}(\tau)) = \operatorname*{arg\,min}_{\alpha,\beta}\sum_{i=1}^{N}\sum_{t\in T_i}\rho_\tau(W_{it}-1_\kappa'\alpha-X_{it}'\beta), \tag{6}$$

where T_i is the set of time periods where individual i is active in the labor market, $\alpha(\tau)=(\alpha_1(\tau),\ldots,\alpha_\kappa(\tau))'$, 1_κ is a κ-vector of ones, $\beta(\tau)=(\beta_1(\tau),\ldots,\beta_{2p}(\tau))'$, and $X_{it}=(X_{it1}^w,\ldots,X_{itp}^s,X_{it1}^w,\ldots,X_{itp}^s)'$. Each bootstrap iteration solves

$$(\alpha^*(\tau),\beta^*(\tau)) = \operatorname*{arg\,min}_{\alpha,\beta}\sum_{i=1}^{N}\sum_{t\in T_i}e_i\rho_\tau(W_{it}-1_\kappa'\alpha-X_{it}'\beta) \tag{7}$$

$$e_i \sim \exp(1). \tag{8}$$

and the empirical bootstrap distribution of θ^* is used for inference about $\widehat{\theta}$, where $\theta=(\alpha,\beta)$.

The results are presented in Table 3. It is found that for all the quantile indices considered, there is a positive association between experience in self-employment and income in self-employment, an effect which increases toward the right tail of this income distribution. This is visualized in Figure 1. This is not surprising: while experience is a measure of ability it may also proxy for maturity of the firm in that it take time to build up capital, clientele, etc. For wage-employment there is a positive association between experience (in wage-employment) and income for low-to-median earners, but a negative association for high-earners. This is illustrated in Figure 2. The negative association in the right tail may be due to several things. While some careers can be characterized by an experience-promotion cycle, other, perhaps "high-skill" jobs, are characterized by rare or specialized skill sets, which are not necessarily obtained through labor-market experience. Another explanation can be that new hires face different terms than existing employees, which may disfavor "settled employees". The negative association may also be misleading if the linear specification is a poor approximation, or because the regression disregards individuals' income stream and simply pools observations. Indeed this pattern is not shared by our later model specification, which deals with both issues. We have done robustness checks, by estimating the above regression for each year, each sector, and each education level. All regressions show the same patterns for the included variables.

The results also indicate that periods with no employment are negatively related to income at all quantiles with unemployment more adversely so than non-employment. Note that only active workers are included in the sample, so this result relates to previous periods with non- or unemployment.

One obvious criticism is that this model does not take into account the income flow or dynamics of earnings in the individuals' careers. For self-employed individuals in particular income flow is not necessarily steady; some entrepreneurs encounter initial periods of low or negative income before business is mature and produce acceptable or high payoffs. Others generally have "erratic" firm performance which, in contrast to dependent employment, directly affects income. We return to this in Section 5, and provide a first attempt to overcome some of the problems in Section 6.

We have not reported the estimates of the many intercepts. It is interesting to take a look at the "winners", i.e. which categories are associated with highest income at the quantile indices estimated when controlling for different types of work experience. Table 4 shows the top 3 significant intercepts for each of the considered quantile levels. It is interesting to note that for all of these, the top intercepts are categories with a high level of education. Also, trade and industry sectors seem to stand out as high-income categories across income levels.

	Variable / $\tau =$	Regression quantiles				
		0.10	0.25	0.50	0.75	0.90
Self-employment	$\mathscr{E}^s$	6.22	12.72	15.84	19.15	25.07
		(0.24)	(0.25)	(0.32)	(0.54)	(0.96)
	$(\mathscr{E}^s)^2$	-0.09	-0.27	-0.33	-0.39	-0.48
		(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
	$\mathscr{E}^w$	0.51^a	2.40	2.34	2.10	2.91
		(0.24)	(0.28)	(0.35)	(0.51)	(0.92)
	$(\mathscr{E}^w)^2$	0.03	-0.01^c	0.04	0.03^b	-0.02^c
		(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
	$\mathscr{E}^u$	-4.17	-12.60	-22.33	-35.74	-56.01
		(0.41)	(0.47)	(0.62)	(0.97)	(1.70)
	$(\mathscr{E}^u)^2$	0.48	1.32	2.16	3.39	5.37
		(0.05)	(0.06)	(0.08)	(0.13)	(0.23)
	$\mathscr{E}^n$	-7.70	-14.63	-17.78	-20.96	-24.53
		(0.31)	(0.45)	(0.63)	(0.92)	(1.57)
	$(\mathscr{E}^n)^2$	0.60	1.08	1.14	1.24	1.38
		(0.02)	(0.04)	(0.06)	(0.08)	(0.14)
	Age	1.06	2.05	4.61	8.73	13.28
		(0.34)	(0.39)	(0.47)	(0.67)	(1.36)
	Age^2	-0.02	-0.04	-0.08	-0.12	-0.18
		(0.00)	(0.00)	(0.01)	(0.01)	(0.02)
Wage-employment	$\mathscr{E}^s$	-6.21	-2.74	-2.29	-2.57	-2.85
		(0.15)	(0.09)	(0.10)	(0.17)	(0.32)
	$(\mathscr{E}^s)^2$	0.27	0.07	-0.02	-0.15	-0.36
		(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
	$\mathscr{E}^w$	8.76	4.62	2.20	-0.80	-5.15
		(0.12)	(0.10)	(0.10)	(0.17)	(0.36)
	$(\mathscr{E}^w)^2$	-0.24	-0.14	-0.11	-0.07	-0.02^a
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
	$\mathscr{E}^u$	-6.20	-8.34	-12.12	-18.76	-28.99
		(0.09)	(0.08)	(0.10)	(0.16)	(0.29)
	$(\mathscr{E}^u)^2$	0.44	0.61	0.87	1.36	2.11
		(0.01)	(0.01)	(0.01)	(0.02)	(0.04)
	$\mathscr{E}^n$	-3.29	-3.73	-5.50	-8.21	-12.02
		(0.13)	(0.12)	(0.12)	(0.18)	(0.31)
	$(\mathscr{E}^n)^2$	0.06	-0.02^c	-0.02^c	-0.03^c	-0.04^c
		(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
	Age	1.53	3.26	5.46	8.34	12.12
		(0.05)	(0.05)	(0.06)	(0.10)	(0.20)
	Age^2	-0.02	-0.04	-0.06	-0.09	-0.13
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 3: Coefficient estimates for the model in (6). All estimates, unless otherwise noted, are significantly different from zero at all conventional levels (p-value< 0). Standard errors are obtained by a weighted XY bootstrap with 199 bootstrap iterations.

Notes:

[a]: Significant at a 5% level.

[b]: Significant at a 10% level.

[c]: Insignificant at all conventional levels.

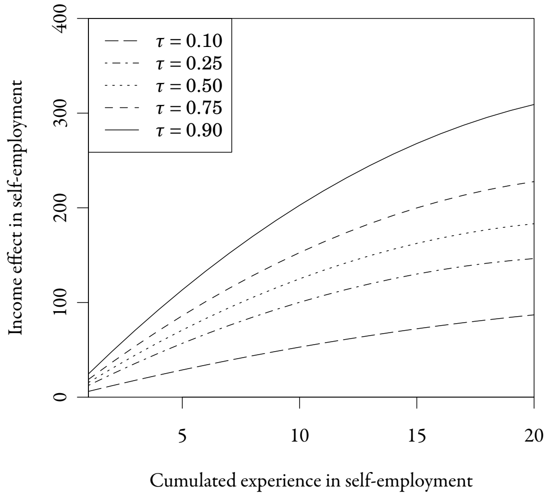


Figure 1: Relation between self-employment income and experience as self-employed for different quantile indices, τ, of the income-distribution.

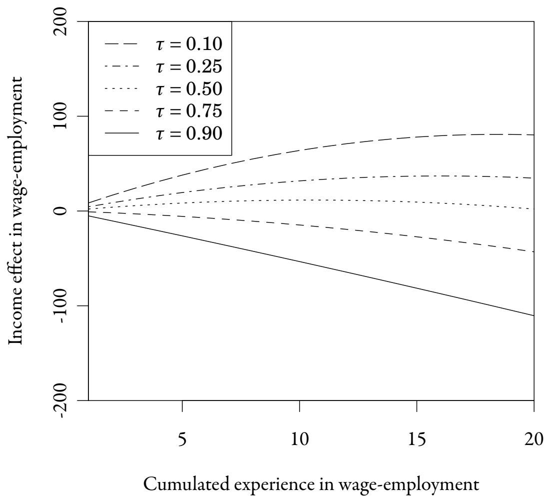


Figure 2: Relation between wage-employment income and experience as wage-employed for different quantile indices, τ, of the income-distribution.

Quantile	#	Sector	Education	Occupation	Estimate	pval
$\tau = 0.10$	1	2:energy	5:higher	wage	117.67	< 0.01
	2	1:industry	5:higher	wage	111.28	< 0.01
	3	6:finance	5:higher	wage	100.88	< 0.01
$\tau = 0.25$	1	2:energy	5:higher	wage	156.72	< 0.01
	2	1:industry	5:higher	wage	148.26	< 0.01
	3	4:trade	5:higher	wage	142.81	< 0.01
$\tau = 0.50$	1	4:trade	5:higher	wage	189.09	< 0.01
	2	2:energy	5:higher	wage	181.65	< 0.01
	3	1:industry	5:higher	wage	178.03	< 0.01
$\tau = 0.75$	1	4:trade	5:higher	self	268.52	< 0.01
	2	4:trade	5:higher	wage	260.76	< 0.01
	3	4:industry	5:higher	wage	234.20	< 0.01
$\tau = 0.90$	1	1:energy	5:higher	self	268.52	0.06
	2	4:trade	5:higher	self	260.76	< 0.01
	3	4:trade	5:higher	wage	234.20	< 0.01

Table 4: Top 3 sector/education/occupation combinations for different quantile levels in the model in (6). Only significant intercepts were compared, and for each sector/education/occupation the best year was chosen.

4 Comparing apples with pears: quantile equivalence

The income distributions available to the individual are very heterogeneous; so much in fact that a "fair" comparison and causal "modelling" is very difficult. First of all, the risk profiles are dissimilar, as is the effective income range. Also the timing of income flow distorts comparison: dependently employed individuals have a stable income "from day one", while a newly started business may need to build up reputation, clientele, and capital before producing a stable surplus. Even then, there is more uncertainty about yearly payoff.

Before considering such timing issues, we suggest a notion of quantile equivalence as a simple graphical tool related to QQ-plots for uncovering some interesting features of the data. Let $\mathscr{W}$ and $\mathscr{S}$ be two (income) distributions of interest. For example, these could be income distributions for wage-employed and self-employed people in the financial sector. Then for a given rank, $\tau_w \in (0,1)$ in $\mathscr{W}$, income is given by $\mathscr{W}^{-1}(\tau_w)$; similarly for τ_s in $\mathscr{S}$. We will call $\tilde{\tau}_s$ the equivalence quantile (of τ_w) when $\mathscr{S}^{-1}(\tilde{\tau}_s) = \mathscr{W}^{-1}(\tau_w)$. The quantile equivalence curve (QEC) $\tilde{\tau}_s(\tau_w) : (0,1) \to (0,1)$ provides a way to compare how well one has to do in relative terms to achieve the same level of income.

In wage-employment one will in many (but not all) cases face relatively little uncertainty about wage. A newly trained carpenter, say, or a university professor, will have a reasonable idea about what income he can expect, and possibly also how this evolves with experience/tenure. In a sense there is a market value of his labor/work.

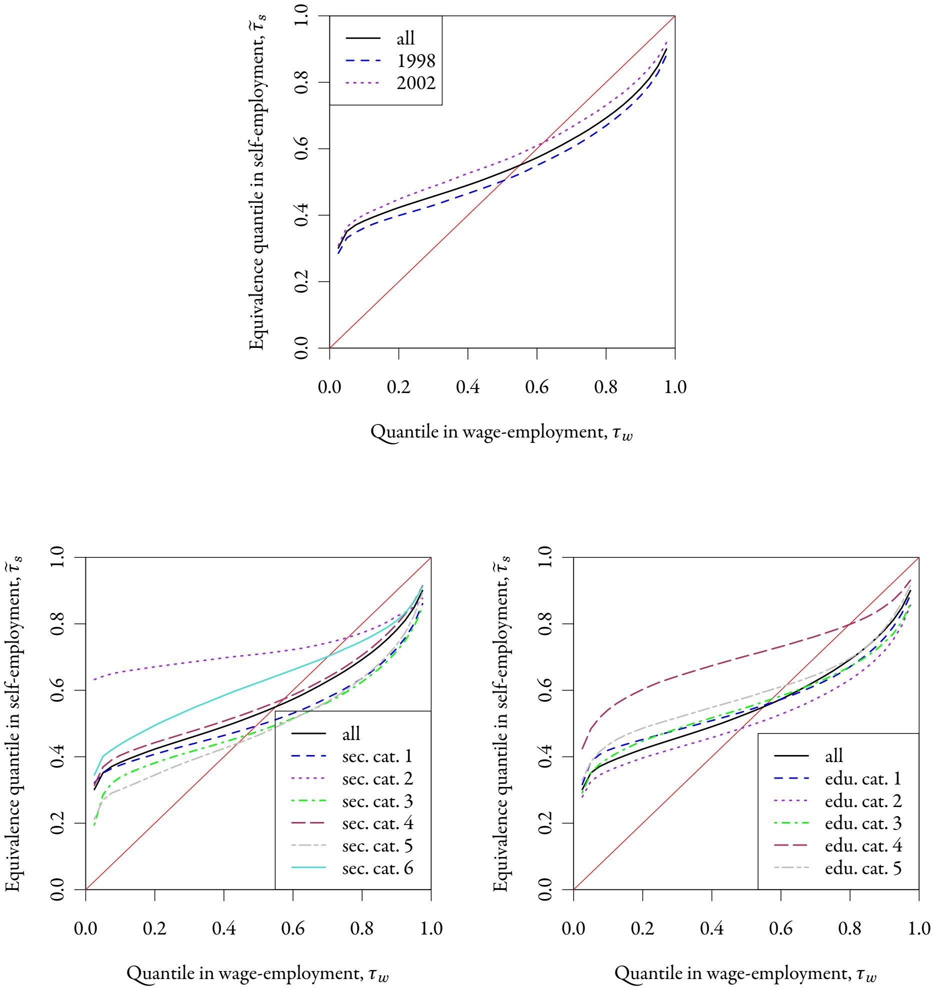


Figure 3: Quantile Equivalence Curves: The top panel shows how that the curve is similar across years. The remaining years in 1992–2002 lie in between the 1998- and 2002-curve with similar shape. The left panel shows a curve for each sector and the right panel shows a curve for each education level.

On the other hand, should he choose to start his own business, there is no such "market value". Using the idea of quantile equivalence, with information on one's rank in wage-employment, one can assess the "relative level of ability" required to earn at least the same in self-employment.

Figure 3 shows QECs, conditioning on year (top panel), sector (left panel) and education (right panel). The black solid curve is unconditional (i.e. all observations pooled) and is the same in all three panels. If the (conditional) distributions were equal, the QEC would be a 45 degree line. The figures show that wage-employment is favored in the left tail and self-employment is favored in the right tail. All the curves cross roughly at the median or to the right. Furthermore, the graph indicates that the income span is much smaller for the dependently employed. The energy sector appears extremely unfavorable to the self-employed. This could be because of a high capital requirement and barriers to entry. On the other hand, if one reaches the equivalence quantile, the curve is quite "flat". The finance sector also shows a different shape and seems to disfavor self-employment relatively to the remaining sectors. Another category that stands out is education category 4, short- to medium-cycle higher level. It is likely that small stores, internet shops, restaurants, etc. make up a lot of the businesses for this category. These are highly competitive markets and may explain this curve. Aside from these two examples, the shape and level of the curve is very similar across years, sectors and education levels.

We now turn to experience levels. Figure 4 shows how the curve changes with experience in self-employment and wage employment. We let 0–5 years of experience be categorized as low level of experience, 6–10 as medium, and above 10 as high. The left panels condition only in the distribution in which experience is earned, whereas the right panels condition in both distributions. The top-left panel in Figure 4 shows how the distribution in self-employment with certain levels of experience compare to the wage-income distribution (unconditional). One can think of a counterfactual world, in which one could "skip the first years" of building up reputation, capital, etc. and enter directly into medium or high level of experience. Having medium–high level of experience surely seems to make self-employment a more attractive alternative to wage-employment. The top-right panel, where both distributions are conditional on experience level in self-employment, indicates that once experienced in self-employment the monetary incentives to switch decline. We stress of course that this is not causal: on the contrary it probably indicates that those who did not make a satisfactory business switched back into wage-employment.

Moving attention to experience in wage-employment, we see that income for low-experienced employees compare more to that of the self-employed in the left tail with only a slight advantage. In the right tail, self-employment is much preferable. For medium experience and above, wage-employment is clearly preferable in the left tail and less unfavorable in the right.

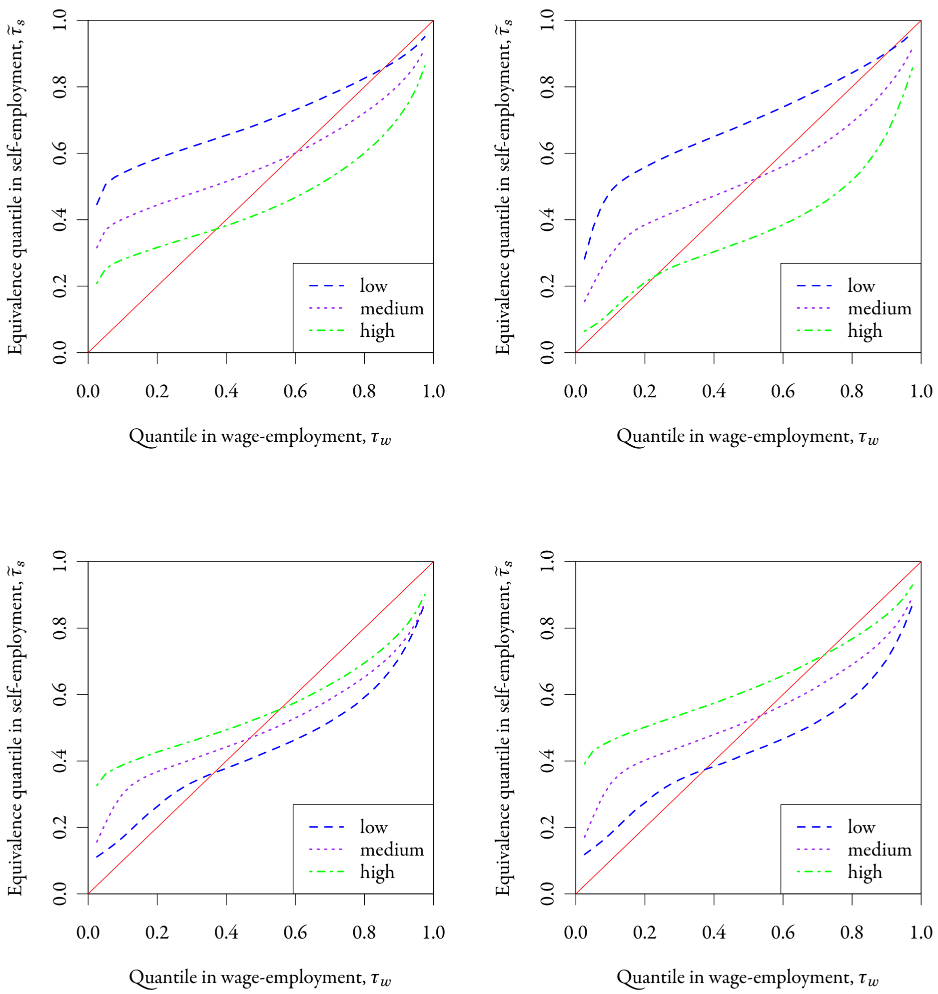


Figure 4: Quantile Equivalence Curves: The top panels condition on experience in self-employment and the bottom panels condition on experience in wage-employment. The left panels condition only in the distribution in which experience is earned, whereas the right panels condition in both distributions. Low level of experience correspond to 0–5 years, medium to 6–10 years, and high is above 10 years.

Figure 5 depict QECs where the vertical axis conditions on having some level of experience in self-employment, and the horizontal axis conditions on having earned that level of experience in wage-employment. The distributions compared this way are aligned slightly, in the sense that the QECs are closer to the 45 degree line than unconditionally, yet still favoring wage-employment in the left tail and self-employment in the right tail. Median incomes here appear to be close to equal for medium level of experience. For the inexperienced, wage-employment is almost always better paying, whereas a highly experienced entrepreneur easily competes with a highly experienced employee. A good question, then, is whether "good times" make up for the "bad".

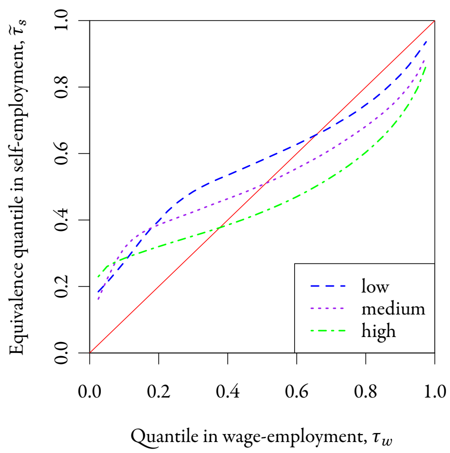


Figure 5: Quantile equivalence curvers: the vertical axis conditions on having some level of experience in self-employment, and the horizontal axis conditions on having earned that same level of experience in wage-employment

5 On the timing of returns to self-employment

One inherent problem in assessing the relationships between occupational choice and realized income levels is the possibly erratic nature of income stream in self-employment, at least by comparison with wage-employment. For many entrepreneurs, an initial period of low or negative outcome is encountered before a positive payoff is achieved. Perhaps one needs to build up a solid customer base or reach a certain level of capital before business is fruitful. Another type of entrepreneur starts businesses only to sell them at the right time and start new ones. Also, in self-employment the up- and down-periods generally affect income directly, in contrast to dependent employment. Therefore, a relevant question is whether periods of high income justify years of low income.

First, we present some noteworthy empirical numbers that seem to disfavor the self-employed. Table 5 shows the number of firms started each year by first-time entrepreneurs who were previously dependently employed; how many of these entrepreneurs have achieved at least one yearly income as high as the one they earned in wage-employment conditional on survival; and how many of these firms survive.

	Start	Init	Years after initializing first business.									
	year	#	1	2	3	4	5	6	7	8	9	10
Percentage conditional on survival	1993	913	0.266	0.552	0.692	0.765	0.807	0.837	0.850	0.867	0.882	0.893
	1994	938	0.248	0.538	0.654	0.736	0.784	0.813	0.840	0.869	0.874	
	1995	1027	0.263	0.553	0.690	0.778	0.814	0.835	0.855	0.863		
	1996	1071	0.276	0.567	0.710	0.776	0.808	0.834	0.857			
	1997	1211	0.265	0.587	0.679	0.745	0.793	0.830				
	1998	1407	0.330	0.593	0.692	0.757	0.782					
	1999	1451	0.298	0.527	0.632	0.698						
Survival ratio	1993	913	1.000	0.847	0.779	0.722	0.681	0.633	0.584	0.562	0.531	0.503
	1994	938	1.000	0.848	0.770	0.722	0.688	0.651	0.613	0.576	0.551	
	1995	1027	1.000	0.887	0.810	0.758	0.717	0.673	0.643	0.617		
	1996	1071	1.000	0.865	0.784	0.735	0.685	0.636	0.607			
	1997	1211	1.000	0.882	0.798	0.732	0.654	0.608				
	1998	1407	1.000	0.866	0.775	0.713	0.672					
	1999	1451	1.000	0.846	0.750	0.704						

Table 5: The top panel shows the ratio of individuals having achieved a yearly income of at least the income in their last year as dependently employed as a function of years after business initialization conditonal on firm survival. The bottom panel shows the ratio of businesses still alive. Individuals included were dependently employed the year before "start year", where their first business was initialized. The column "Init #" shows how many *first businesses* were started.

Although these are simple descriptive figures, they do indicate relative difficulty faced by the self-employed. Only 20-30% achieves an income comparable to their wage as employee in the first year of business. This figure increases to 50-60% after year 2 for those still in business, which is around 85%. After 6 years we pass 80%, with 60-70% still in business. The numbers seem to be in agreement for different starting years.

Having achieved just one equivalent income does not mean that income is satisfactory in general. What if we also require the periods of "high" income to cover the years with income lower than the alternative income in self-employment? Consider a dependently employed individual i and let R^w_{it} be his wage-income at time t. He now makes an occupational switch and initiates own business which in year $t+y$ gives an income of $R^s_{i,t+y}$. We assume that this individual could have earned the alternative income R^w_{it} (assumed constant) had he stayed in wage-employment. In other words, each year he pays $R^w_{i,t+y}$ to get $R^s_{i,t+y}$. After Y periods, define

$$\bar{r}_{i,t+Y} \equiv \frac{Y^{-1}\sum_{y=1}^{Y} \delta_{t+y} R^s_{t+y} - R^w_i}{R^w_i} \tag{9}$$

to be the "average returns to self-employment" (ARS), which shows whether the income in self-employment on average has covered the (lost) alternative income in wage employment. The deflators δ_t are used to take into account valuation differences across time. We take these to be price indices taken from Statistics Denmark. Beyond these, the measure assumes that timing of income is irrelevant. It also neglects possible wage-raises as a function of experience in wage-employment. However, the drawbacks make it a conservative measure in the sense that it lets doubt benefit the self-employed. Table 6 shows quantiles of the realized $\bar{r}_{i,t+Y}$, conditional on survival, for individuals who

- are wage-employed until (and including) 1992,
- are first-time entrepreneurs in 1993,
- observed (but not necessarily in self-employment) all years up until and including 2002,

where their 1992-wage is used as R^w_{it}. In their first year of business, 70% of the businesses have a negative ARS and even at the 90% quantile the "investment" of R^w_{it} has not been covered. Even as years pass, the numbers indicate that very few entrepreneurs are better off (in monetary terms) in self-employment. After 10 years, 70% of those still in self-employment have not yet archived an ARS of 1. These numbers strongly indicate that the composition of the occupational choice and the overall income stream should be accounted for when assessing the private value of entrepreneurship.

	Years after initializing first business.									
Quantile	1	2	3	4	5	6	7	8	9	10
90%	0.771	1.107	1.368	1.418	1.524	1.606	1.656	1.657	1.598	1.806
80%	0.232	0.509	0.699	0.812	0.951	1.038	1.086	1.072	1.070	1.072
70%	-0.100	0.186	0.340	0.475	0.577	0.635	0.673	0.702	0.735	0.729
60%	-0.363	-0.014	0.147	0.264	0.344	0.401	0.447	0.464	0.479	0.484
50%	-0.586	-0.218	-0.050	0.068	0.159	0.216	0.273	0.282	0.300	0.309
40%	-0.728	-0.385	-0.206	-0.102	-0.010	0.060	0.101	0.122	0.137	0.151
30%	-0.867	-0.507	-0.361	-0.256	-0.166	-0.114	-0.066	-0.061	-0.017	0.003
20%	-0.995	-0.670	-0.535	-0.431	-0.378	-0.299	-0.261	-0.264	-0.246	-0.220
10%	-1.068	-0.919	-0.763	-0.686	-0.632	-0.580	-0.553	-0.525	-0.514	-0.507

Table 6: Empirical quantiles of ARS, $\bar{r}_{i,1992+Y}$, $Y = 1, \ldots, 10$.

6 A quantile regression model for the composition of occupational choice

We have argued that the timing of income, and income stream over an individual's career is important, and some simple descriptive statistics in the previous section indeed seem to agree. We now offer a first attempt to model this in a quantile regression framework.

Our approach is ultimately descriptive: given current policy π (meant in the broad sense as the environment offered to the workers in the society), what are the income-distribution patterns in different occupations at different maturity levels of the career? How individuals distribute themselves into different occupations over time is considered a random process which depends on the policy π. Knowledge about income distribution can help the politician narrow down topics for further research and analysis to guide policy improvement. In the model, the policy is treated as given, or fixed, and is not modelled as such: we use it to make explicit that we treat the labor markets, and entrepreneurial opportunities, as heavily dependent on the environment, or *policy*. It also clarifies the "birds-eye" view taken in the model.

Let i and t index individuals and time. Each period, individuals participate in occupation $j \in \{w, s\}$, where they are rewarded according to their experience and their rank $U^j_{it} \in (0,1)$. The rank here represents ability, luck, and other characteristics not captured by experience. Thus ex-ante, individuals face a career determined by the process $\{U^j_{it}\}_t$, $t \in T_i$, which depends on π. The set T_i represents the time-periods in individual i's career. The income earned in period t by individual i with experience $\mathscr{E}_{it} = \{\mathscr{E}^w_{it}, \mathscr{E}^s_{it}\}$ and ranked U^j_{it} is given by

$$(10) \quad \phi^j_t(U^j_{it}, \mathscr{E}_{it}) = \phi^j_t(U^j_{it}, \mathscr{E}_{it}; \pi),$$

with ϕ^j_t monotonically increasing in U^j_{it} for fixed $\mathscr{E}_{it}$. The function ϕ^j_t is then a conditional quantile function, the inverse of the conditional income distribution function, and U^j_{it} is the quantile index.

We assume that the income relevant to the individual is total career income

$$(11) \quad W_i = \sum_{t \in T_i} \delta_t \phi^j_t(U^j_{it}, \mathscr{E}_{it}),$$

or equivalently the averaged measure

$$(12) \quad \bar{w}_i = W_i / \# T_i,$$

where $\# T_i$ is the cardinality of T_i. The "weights" δ_t are deflators to account for valuation differences over time. Using $\bar{w}_i$ allows us to model individuals with differing career lengths.

Assume values δ_t and functions ϕ^j such that

$$(13)\quad \delta_t \phi_t^j(u,e) = \phi^j(u,e) \quad \forall t,u,e$$

This assumption means that for appropriate choice of δ_t, the income functions themselves are time-invariant (but the input, of course, is not). Time dependence is thus assumed to work only through the ranks. We now write the income measure of interest as

$$(14)\quad \bar{w}_i = \frac{1}{\#T_i} \sum_{t\in T_i} \phi^j(U_{it}^j, \mathscr{E}_{it})$$

$$(15)\quad = \frac{1}{\#T_i^w + \#T_i^s} \left\{ \sum_{t\in T_i^w} \phi^w(U_{it}^w, \mathscr{E}_{it}) + \sum_{t\in T_i^s} \phi^s(U_{it}^s, \mathscr{E}_{it}) \right\},$$

where T_i^j are partitions of T_i corresponding to time periods spent in occupation j.

So far, the model is fairly flexible; the functions ϕ^j are left unspecified and both these income functions and the ranks are allowed to differ with j, the occupational choice. The two potentially critical points are (i) the assumption that the income functions are time invariant in real terms, and (ii) that the measure $\bar{w}_i$ renders the timing of income flow irrelevant beyond the corrections from δ_t. This timing, however, is not very comparable across occupational choice, or even within self-employment, and hence this "abstraction" seems worthwhile when interest is in comparison across occupation.

The random (aggregated) income, $\bar{w}_i$, can be characterized by a ranking mechanism and a quantile function. Define $\mathscr{U}_i : (0,1) \to \{U_{it}^j\}$, with $\mathscr{U}_i | \pi \sim \mathrm{uniform}(0,1)$, such that we can describe the random outcome $\bar{w}_i$ as

$$(16)\quad \bar{w}_i = \varphi(\mathscr{U}_i) = \varphi(\mathscr{U}_i; \mathscr{C}_i(\mathscr{U}_i)),$$

for some φ monotonically increasing in $\mathscr{U}_i$. We include the second equality here, to make explicit that we want to model the function in terms of career composition, $\mathscr{C}_i$, which from the model's perspective is determined by $\mathscr{U}_i$. The object $\mathscr{C}_i$ represents time spent in different occupations with certain levels of experience.

To link equations (15) and (16) in a way that allows sensible specification of $\mathscr{C}_i$ and estimate the value associated with different occupational choices for various levels of experience, we will use a discretization strategy. Assume that discrete levels $k \equiv k(\mathscr{E}_{it}) = \langle k^s, k^w \rangle \in \{l,m,h\} \times \{l,m,h\}$ of experience are sufficient to characterize the experience heterogeneity (l for low, m for medium, and h for high), in the sense that there exist functions ϕ^{jk} such that

$$(17)\quad \phi^j(U_{it}^j, \mathscr{E}_{it}) \approx \phi^{jk}(U_{it}^j).$$

Now, let the average income achieved by individual i in occupation j and experience category k be denoted by

$$(18) \quad \varphi^{jk}(\mathcal{U}_i) = \sum_{t \in T_i^{jk}} \frac{\phi^{jk}(U_{it}^j)}{\#(T_i^{jk})},$$

Then, by rearrangement we obtain

$$\bar{w}_i = \frac{1}{\#(T_i)} \sum_{t \in T_i} \phi^{jk}(U_{it}^j)$$

$$(19) \quad = \sum_j \sum_k \frac{\#(T_i^{jk})}{\#(T_i)} \varphi^{jk}(\mathcal{U}_i)$$

$$(20) \quad = \sum_j \sum_k X_i^{jk} \varphi^{jk}(\mathcal{U}_i) \equiv \varphi(\mathcal{U}_i; \mathcal{C}_i),$$

with $\mathcal{C}_i \equiv \{X_i^{jk}\}$. We observe $\bar{w}_i$ and the ratios $X_i^{jk} = X_i^{jk}(\mathcal{U}_i)$ that define the occupational composition, so by quantile regression we can estimate the value of the functions $\varphi^{jk}(\mathcal{U}_i)$ for different fixed levels of $\mathcal{U}_i$. Note that this does not specifically reveal information on the ranks U_{it}^j or the functions ϕ^{jk}. The transformation abstracts out the time dimension, and we are left with a cross-sectional type regression, where independence is assumed across individuals only.

The interpretation of estimates $\widehat{\varphi}^{jk}(\tau)$, for some level τ of $\mathcal{U}_i$, is as follows. At the level τ of the $\bar{w}_i$-income distribution, a period of labor in occupation j with experience level k is associated with a pay of $\widehat{\varphi}^{jk}(\tau)$ on average. It is important to stress that the model is descriptive rather than causal: one cannot fix some $\mathcal{C}_i$ and then predict $\bar{w}_i$ at each τ because $\mathcal{C}_i$ is itself a function of the quantile level. It is not a conditional model in the classic sense. With a slight abuse of language, one can call it a semi-parametric model (there are no parametric assumptions about the income functions), that smoothes out the erratic nature of the income stream. Which career patterns that are most likely at each level of $\mathcal{U}_i$ is not immediately clear from estimating the model, in the sense that it does not model the evolution of $X_{it}^j(\mathcal{U}_i)$, the weights put on each payoff function. Rather, these are fixed for a "snapshot" estimation of $\varphi^{jk}(\mathcal{U}_i)$.

In our empirical implementation of the model we use the same definitions of income as previously and use the price indices as δ_t. It might be argued that it would be appropriate to include general wage-level increase measures in δ_t, but we are unaware of a way to do this which is appropriate for both types of occupation. We can now construct, for each individual, the income $\bar{w}_i$. We use 5 and 10 as cut-off points for experience categories, i.e. 0–5 years is considered a low level of experience, 6–10 years is considered a medium level of experience, and more than 10 years is considered a high level of experience. From our experience variables we can then construct the regressors X_i^{jk} (there are 18 of them: 2 occupations with 3 levels of experience in each).

We will assume that entries and exists from the data set occur at random. This may not be entirely true as you may have e.g. early exits as a reaction to periods of high income, or people earning experience abroad (unaccounted for) to later re-enter the sample. Also, exits into unemployment or non-employment may be related to realized income. Such examples of "good fortune" and "poor luck" must then be attributed to the rank $\mathscr{U}$ and any such hidden effects are accounted for by the unspecified income functions. Also, as we do not observe full careers, we must assume that we observe "enough" to make the observed incomes overall representative. Another point is that we cannot directly account for firms' accumulated wealth and liquidation value. Our approach smoothes realized income over an 11-year period and if people smooth consumption our income measure is a potential "second-best".

We estimate the model for each sector individually but we ignore educational level. This has the effect that the estimates associate various income levels with certain occupational choices ignoring which education is more likely at the particular level. $\mathscr{U}_i$ can be thought of to work in part by assigning an educational level. One can think of sector being specified by π, the policy/environment being described, and education as being attributed to the random distribution $\mathscr{U}_i$.

We present the results for the trade sector in Table 7 and for the industry sector in Table 8. From presentational considerations we round off coefficients to nearest integer and only make note of less significant or insignificant estimates. Inference is again based on the weighted XY bootstrap with unit exponential weights and 299 bootstrap iterations. As trends in the results are similar and for brevity, we do not present the results for the remaining 4 sectors in the main text. We present more detailed results for all sectors in the appendix, numbers rounded to two digits along with standard errors.

We here find that income in wage-employment is positively related with experience across all estimated quantiles. This is in contrast to the findings of the model in Section 3, where there was a negative relation in the right tail of the distribution. In that model, the response distribution was that of single outcomes, whereas we now model an aggregated measure. There is no convincing evidence that experience in self-employment has a positive effect on income in wage-employment.

In self-employment our results suggest benefits from experience in both occupations. Income in self-employment most crucially seems to rely on experience in self-employment. At the median, wages in self-employment are only competitive for high levels of experience, or a medium level along with a medium or high level in wage-employment. Even at the 90% quantile wages for an individual with low level of experience in both occupations do not compete with any median-income wage-employee pay.

It seems that early stages in self-employment are critical. A policy maker who wants to encourage self-employment would probably focus analysis efforts on conditions faced by newly started firms or the obstacles faced during startup. Our results indicate that the pay in self-employment is competitive once a certain level of experience has been obtained. Of course, experience here most likely proxy also for maturity of the firm itself.

			Regression quantiles				
	$\mathcal{E}^s$	$\mathcal{E}^w$	0.1	0.25	0.5	0.75	0.9
Self-employment	l	l	-8	0	7	50	112
	m	l	0	17	82	168	280
	h	l	38	110	201	308	448
	l	m	-6	0	48	106	181
	m	m	12	71	144	235	385
	h	m	69	158	253	361	506
	l	h	-9	8	65	133	213
	m	h	35	132	223	338	500
	h	h	97	112^b	296	382	278^b
Wage-employment	l	l	33	68	107	145	213
	m	l	43	81	119	152	191
	h	l	14	58	109	146	196
	l	m	70	108	136	186	252
	m	m	75	114	140	176	236
	h	m	92	132	158	195	259
	l	h	122	141	170	222	294
	m	h	107	141	174	223	291
	h	h	107	161	169	239	350

Table 7: Point estimates of function values in model (20) for different quantile indices (values of $\mathcal{U}_i$) in the **trade** sector. All (non-zero) estimates are significantly different from zero with p-values < 0.01 unless otherwise noted. Inference is based on a weighed XY bootstrap with 299 bootstrap iterations.
Notes:
^{a}: Significant at a 5% level.
^{b}: Significant at a 10% level.
^{c}: Insignificant at all conventional levels.

	$\mathcal{E}^s$	$\mathcal{E}^w$	Regression quantiles 0.1	0.25	0.5	0.75	0.9
Self-employment	l	l	-5^b	0	9	54	104
	m	l	0	17	87	189	312
	h	l	40	109	201	316	460
	l	m	-3^b	5^a	56	117	199
	m	m	21	88	167	263	396
	h	m	84	175	267	394	466
	l	h	0	33	102	171	246
	m	h	58	148	230	343	499
	h	h	66^c	127^c	226^a	306^c	260^c
Wage-employment	l	l	36	77	117	152	216
	m	l	55	99	126	158	198
	h	l	25	78	118	146	183
	l	m	87	117	144	191	241
	m	m	87	121	144	176	229
	h	m	84	130	150	181	226
	l	h	124	140	165	204	261
	m	h	107	134	160	198	251
	h	h	130	147	168	201	259

Table 8: Point estimates of function values in model (20) for different quantile indices (values of $\mathcal{U}_i$) in the **industry** sector. All (non-zero) estimates are significantly different from zero with p-values < 0.01 unless otherwise noted. Inference is based on a weighed XY bootstrap with 299 bootstrap iterations.
Notes:
^{a}: Significant at a 5% level.
^{b}: Significant at a 10% level.
^{c}: Insignificant at all conventional levels.

7 Jack of all trades, expert of none?

Lazear (2004) provides a theory of entrepreneurship where it benefits the entrepreneur to be a "jack of all trades": while an individual who works for others should specialize in specific tasks, the entrepreneur's skill set must suffice for the variety of tasks involved in running a successful business. Lazear refers to this as a "weakest link" property of entrepreneurship. He uses data from Stanford Master of Business administration alumni to show that those ending up being entrepreneurs study a broader curriculum when they are in program.

Others have tested Lazear's theory, e.g Wagner (2003), who find empirical support using German micro data; Åstebro and Thompson (2011), who find that entrepreneurs have a more varied labor market experience, but that this varied experience is related to a lower household income; and Silva (2007), who argues that a spread of knowledge across fields does *not* increase the probability of becoming an entrepreneur.

We make a slight extension of our model and consider the impact of having obtained more than two years of experience in sectors other than the one under consideration. This is done by redefining the categorization k:

$$(21) \quad k = k(\mathcal{E}_{it}) = \langle k^s, k^w, k^{jack} \rangle$$
$$= \{l, m, h\} \times \{l, m, h\} \times \{jack, specialist\}$$

The model is estimated for a single sector, and a person is categorized as a "jack" if more than 2 years of his experience is obtained in a different sector, regardless of occupation, otherwise he is a "specialist".

Table 9 reports the results for the trade sector and Table 10 for the industry sector. While there is not a consistent *overall* pattern, there are some indications of benefits in self-employment to the jacks defined as above. For example, considering a medium to high level of experience in self-employment we see that estimates are higher for jacks than for specialists across all quantiles in both sectors presented. In wage employment, however, estimates are closer in value, perhaps with a slight overall advantage to specialists. Both observed patterns seem to provide justification for Lazear's claim.

				Regression quantiles				
	$\mathscr{E}^s$	$\mathscr{E}^w$	J/S	0.10	0.25	0.50	0.75	0.90
Self-employment	l	l	J	-9	0	6	48	106
	l	l	S	-5^a	0	10	59	129
	m	l	J	0	17	80	170	290
	m	l	S	0	23	90	164	263
	h	l	J	44	116	206	311	447
	h	l	S	1^c	67	157	287	464
	l	m	J	-8	2^c	54	121	204
	l	m	S	-6	0	44	99	169
	m	m	J	31	91	164	272	446
	m	m	S	3^b	48	119	197	309
	h	m	J	114	196	278	361	456
	h	m	S	30	120	209	332	526
	l	h	J	-5	17	79	154	260
	l	h	S	-12	3	58	123	194
	m	h	J	92	186	279	388	545
	m	h	S	14	93	176	282	443
	h	h	J	83^c	273^c	368^a	364^a	134^c
	h	h	S	97^c	127^c	218^a	315^a	218^c
Wage-employment	l	l	J	33	68	105	144	206
	l	l	S	27	61	111	157	233
	m	l	J	39	75	115	144	182
	m	l	S	49	88	127	166	227
	h	l	J	11	47	102	136	190
	h	l	S	26	70	120	154	199
	l	m	J	77	109	135	177	237
	l	m	S	64	106	139	196	262
	m	m	J	84	114	140	171	213
	m	m	S	64	110	137	178	239
	h	m	J	108	155	172	218	275
	h	m	S	82	118	150	189	255
	l	h	J	123	140	167	214	282
	l	h	S	122	142	175	230	304
	m	h	J	129	157	188	238	333
	m	h	S	97	133	168	217	287
	h	h	J	141	144	175	169^a	259^c
	h	h	S	105	152	162	245	350

Table 9: Point estimates of function values in model (20) with k defined as in (21) for different quantile indices (values of $\mathscr{U}_i$) in the i **trade** sector. The label "J" refers to "jack of all trades" and the label "S" refers to a "specialist" as explained in Section 7. All (non-zero) estimates are significantly different from zero with p-values < 0.01 unless otherwise noted. Inference is based on a weighed XY bootstrap with 299 bootstrap iterations.
Notes:
^{a}: Significant at a 5% level.
^{b}: Significant at a 10% level.
^{c}: Insignificant at all conventional levels.

	$\mathcal{E}^s$	$\mathcal{E}^w$	J/S	0.10	0.25	0.50	0.75	0.90
				Regression quantiles				
Self-employment	l	l	J	-4[a]	0	8[a]	52	104
	l	l	S	-17[c]	0	12[c]	48	73
	m	l	J	0	27	98	201	317
	m	l	S	0	6[c]	75	169	305
	h	l	J	44	113	204	319	452
	h	l	S	12[c]	88	180	307	487
	l	m	J	-7[a]	4[c]	53	120	199
	l	m	S	-2[c]	5[c]	57	114	196
	m	m	J	36	106	188	300	423
	m	m	S	2[c]	72	145	232	352
	h	m	J	127	202	300	415	590
	h	m	S	46	138	207	333	437
	l	h	J	0	37	109	183	256
	l	h	S	0	29	96	161	237
	m	h	J	88	180	283	396	544
	m	h	S	36	119	203	304	446
	h	h	J	168[c]	95[c]	205[c]	662[b]	1037[c]
	h	h	S	66[c]	99[c]	226[c]	226[c]	226[c]
Wage-employment	l	l	J	35	77	116	155	218
	l	l	S	41	77	121	163	252
	m	l	J	63	103	129	159	205
	m	l	S	51	95	126	156	198
	h	l	J	15	70	111	137	166
	h	l	S	43	90	128	160	205
	l	m	J	92	118	143	190	238
	l	m	S	79	116	147	195	247
	m	m	J	98	120	138	166	210
	m	m	S	77	121	147	180	238
	h	m	J	104	136	153	176	219
	h	m	S	76	125	148	182	228
	l	h	J	124	140	163	200	253
	l	h	S	124	141	168	211	271
	m	h	J	119	136	166	206	275
	m	h	S	100	133	158	197	248
	h	h	J	131	160	173	305	455
	h	h	S	124	142	168	187	259

Table 10: Point estimates of function values in model (20) with k defined as in (21) for different quantile indices (values of $\mathcal{U}_i$) in the i **industry** sector. The label "J" refers to "jack of all trades" and the label "S" refers to a "specialist" as explained in Section 7. All (non-zero) estimates are significantly different from zero with p-values < 0.01 unless otherwise noted. Inference is based on a weighed XY bootstrap with 299 bootstrap iterations.

Notes:

[a]: Significant at a 5% level.

[b]: Significant at a 10% level.

[c]: Insignificant at all conventional levels.

8 Concluding remarks

In this paper we have investigated some patterns in the income for self-employed individuals and the dependently employed using Danish population register data. We argue that it is important to model career income, rather than one-period income because of the different nature of income flow across occupations. We develop a simple model in a quantile regression framework which allows us to estimate how different levels of experience in different occupations are related to an aggregate measure of income at various levels of the distribution. Our model takes a "birds-eye" view rather than making any individual-level causal claims. We find that the early stages of self-employment appear to be critical, and argue that compensating for initial periods of low income seems difficult. Extending the model to allow for experience to stem from different business sectors allow us to comment on Lazear's jack-of-all-trades theory. The claim that a broad experience profile benefits the entrepreneur is indeed backed up by our estimation results.

In rounding off our discussion, it is safe to say: the answer to the ambiguous question posed in the title of this paper is itself ambiguous. We do find that median earnings in self-employment are competitive in the mature stages of the firm, but we question whether earnings then compensate for tough immature stages. We also find that a mixed career, both in terms of wage- and self-employment, and in terms of diverse experience, associates positively with entrepreneurial earnings.

A Estimation results not presented in main text

			Regression quantiles				
	$\mathscr{E}^s$	$\mathscr{E}^w$	0.1	0.25	0.5	0.75	0.9
Self-employment	l	l	-5.14	0.00	9.27	53.97	103.55
			(2.73)	(0.08)	(2.96)	(6.39)	(7.47)
	m	l	0.00	17.02	87.31	189.45	312.13
			(0.87)	(4.41)	(5.50)	(10.00)	(14.99)
	h	l	39.67	108.92	201.01	316.31	460.36
			(2.38)	(2.21)	(3.13)	(4.34)	(9.27)
	l	m	-3.37	5.01	55.79	117.39	199.28
			(1.83)	(2.26)	(3.88)	(5.17)	(7.56)
	m	m	20.96	88.09	167.09	262.83	396.41
			(3.77)	(4.75)	(5.36)	(9.70)	(17.11)
	h	m	83.91	175.50	267.43	394.03	465.80
			(16.43)	(13.04)	(13.91)	(18.80)	(40.94)
	l	h	0.00	32.87	102.19	170.76	246.22
			(0.36)	(3.03)	(2.90)	(3.65)	(7.08)
	m	h	58.03	148.21	229.78	343.24	499.07
			(11.17)	(10.10)	(10.69)	(13.59)	(22.86)
	h	h	65.61	126.91	225.59	306.47	259.65
			(68.60)	(82.72)	(111.32)	(251.13)	(354.60)
Self-employment	l	l	35.89	77.37	116.73	151.84	216.21
			(2.10)	(2.12)	(1.58)	(2.19)	(4.31)
	m	l	54.74	98.85	126.18	158.00	197.53
			(4.98)	(2.14)	(1.73)	(3.26)	(6.50)
	h	l	25.37	78.16	117.57	145.79	182.99
			(2.76)	(3.20)	(1.44)	(1.60)	(4.32)
	l	m	87.50	117.34	144.10	191.34	241.24
			(0.99)	(0.42)	(0.51)	(0.84)	(1.36)
	m	m	86.79	120.82	143.63	175.54	229.20
			(3.02)	(0.96)	(1.16)	(2.01)	(4.56)
	h	m	84.17	130.46	150.22	181.43	226.37
			(4.79)	(2.07)	(2.11)	(2.64)	(6.27)
	l	h	124.22	140.28	164.56	204.35	260.70
			(0.08)	(0.08)	(0.10)	(0.17)	(0.36)
	m	h	107.36	133.55	159.86	198.36	250.61
			(2.91)	(1.29)	(1.56)	(2.77)	(5.74)
	h	h	130.33	146.76	168.06	200.75	259.28
			(15.42)	(13.54)	(14.77)	(25.34)	(58.86)

Table 11: Point estimates of function values in model (20) for different quantile indices (values of $\mathscr{U}_i$) in the **industry** sector. Standard errors are shown in parentheses and are obtained by a weighted XY bootstrap with 299 bootstrap iterations.

			Regression quantiles				
	$\mathcal{E}^s$	$\mathcal{E}^w$	0.1	0.25	0.5	0.75	0.9
Self-employment	l	l	0.00	0.00	0.00	8.85	244.53
			(27.93)	(15.11)	(13.77)	(94.40)	(142.99)
	m	l	-15.84	0.00	1.09	64.44	261.93
			(12.28)	(8.97)	(15.43)	(49.67)	(76.96)
	h	l	26.37	91.99	187.40	323.14	528.00
			(18.88)	(9.73)	(14.55)	(24.66)	(38.20)
	l	m	0.00	0.00	0.00	47.99	153.70
			(0.00)	(0.00)	(4.23)	(28.12)	(62.69)
	m	m	-15.84	0.00	130.39	349.79	457.24
			(27.22)	(9.09)	(64.92)	(49.95)	(87.00)
	h	m	15.84	37.46	247.51	329.24	608.04
			(25.30)	(56.17)	(35.70)	(81.65)	(251.79)
	l	h	-0.88	0.00	0.00	11.26	163.24
			(1.67)	(0.00)	(0.00)	(11.18)	(29.12)
	m	h	-6.89	0.00	88.66	375.26	500.31
			(20.78)	(5.09)	(54.23)	(67.37)	(201.23)
	h	h	34.43	601.77	601.77	590.50	438.53
			(250.51)	(305.04)	(372.39)	(495.92)	(719.47)
Wage-employment	l	l	80.82	126.53	143.36	200.19	232.59
			(21.77)	(10.10)	(14.91)	(13.53)	(17.38)
	m	l	5.88	13.07	96.59	148.99	172.35
			(2.35)	(10.53)	(33.48)	(17.06)	(21.25)
	h	l	6.08	7.32	13.82	42.79	128.37
			(0.36)	(0.79)	(1.64)	(19.11)	(12.17)
	l	m	107.88	128.39	160.43	205.02	234.17
			(3.50)	(1.76)	(2.96)	(3.99)	(4.61)
	m	m	9.98	43.85	143.83	172.83	212.46
			(2.63)	(25.57)	(10.71)	(13.73)	(17.86)
	h	m	6.51	14.20	94.24	155.40	213.63
			(1.53)	(10.13)	(24.92)	(33.85)	(30.20)
	l	h	132.38	147.57	170.49	213.00	269.17
			(0.29)	(0.32)	(0.46)	(0.97)	(1.29)
	m	h	12.84	127.28	152.20	203.37	274.62
			(16.74)	(27.71)	(8.89)	(26.59)	(40.73)
	h	h	826.42	777.17	517.10	365.78	209.70
			(4.30)	(33.76)	(50.36)	(70.23)	(63.21)

Table 12: Point estimates of function values in model (20) for different quantile indices (values of $\mathcal{U}_i$) in the **energy** sector. Standard errors are shown in parentheses and are obtained by a weighted XY bootstrap with 299 bootstrap iterations.

	$\mathscr{E}^s$	$\mathscr{E}^w$	Regression quantiles 0.1	0.25	0.5	0.75	0.9
Self-employment	l	l	0.00	0.00	22.80	64.89	119.30
			(0.00)	(0.00)	(3.65)	(3.20)	(5.66)
	m	l	0.00	39.42	93.54	153.07	225.13
			(0.73)	(3.20)	(2.92)	(5.61)	(9.76)
	h	l	69.20	124.38	192.80	286.12	401.00
			(1.83)	(1.33)	(1.56)	(2.34)	(4.73)
	l	m	0.00	27.67	70.00	115.34	172.81
			(0.00)	(2.07)	(2.09)	(2.51)	(5.51)
	m	m	49.77	94.83	157.55	236.91	325.48
			(3.38)	(2.90)	(3.79)	(4.61)	(7.98)
	h	m	91.49	156.37	228.13	311.47	415.77
			(10.09)	(8.07)	(9.36)	(8.96)	(22.09)
	l	h	16.47	61.69	113.22	163.31	221.14
			(1.80)	(1.47)	(1.41)	(1.51)	(3.45)
	m	h	86.74	151.15	209.29	303.22	419.90
			(8.03)	(5.08)	(6.00)	(8.33)	(15.85)
	h	h	-31.28	7.03	33.39	124.82	139.70
			(32.97)	(22.97)	(41.43)	(64.67)	(69.66)
Wage-employment	l	l	33.88	63.70	109.51	150.42	198.78
			(1.77)	(2.63)	(2.14)	(3.01)	(4.49)
	m	l	50.04	87.94	125.50	160.86	203.76
			(5.11)	(3.66)	(2.62)	(4.33)	(5.13)
	h	l	44.76	78.60	116.05	144.43	185.63
			(3.18)	(2.57)	(2.05)	(3.01)	(4.79)
	l	m	62.69	108.46	134.49	172.39	221.90
			(1.50)	(0.80)	(0.70)	(1.10)	(1.85)
	m	m	79.04	116.03	138.68	168.13	213.12
			(3.28)	(1.88)	(1.36)	(2.84)	(5.56)
	h	m	91.30	127.29	151.60	178.51	228.63
			(7.63)	(2.84)	(2.68)	(5.17)	(9.56)
	l	h	122.25	136.91	157.06	187.05	228.54
			(0.12)	(0.10)	(0.15)	(0.25)	(0.38)
	m	h	108.90	133.17	159.38	190.62	243.43
			(3.20)	(1.56)	(1.94)	(3.48)	(7.27)
	h	h	181.78	165.05	209.17	246.88	272.86
			(40.60)	(23.03)	(33.83)	(65.38)	(103.41)

Table 13: Point estimates of function values in model (20) for different quantile indices (values of $\mathscr{U}_i$) in the **construction** sector. Standard errors are shown in parentheses and are obtained by a weighted XY bootstrap with 299 bootstrap iterations.

	$\mathcal{E}^s$	$\mathcal{E}^w$	Regression quantiles				
			0.1	0.25	0.5	0.75	0.9
Self-employment	l	l	-7.83	0.00	7.02	49.85	111.84
			(1.26)	(0.00)	(1.82)	(2.62)	(6.25)
	m	l	0.00	16.54	81.69	168.10	280.17
			(0.00)	(1.80)	(2.50)	(3.44)	(7.50)
	h	l	37.86	110.27	201.02	308.22	448.14
			(1.39)	(1.31)	(1.71)	(2.14)	(3.77)
	l	m	-6.23	0.00	47.63	106.20	181.12
			(1.04)	(0.24)	(1.56)	(2.35)	(4.15)
	m	m	12.47	71.41	144.21	235.36	384.75
			(2.65)	(2.37)	(3.44)	(4.94)	(13.11)
	h	m	69.40	158.36	252.94	361.43	506.05
			(7.06)	(6.64)	(8.24)	(11.52)	(24.63)
	l	h	-8.66	8.39	64.56	133.29	212.59
			(1.03)	(1.04)	(1.40)	(1.76)	(4.35)
	m	h	34.80	131.58	223.40	338.26	500.11
			(6.06)	(6.29)	(7.02)	(8.93)	(20.06)
	h	h	96.60	111.95	296.33	382.38	277.83
			(36.92)	(66.08)	(98.81)	(120.53)	(144.25)
Wage-employment	l	l	32.51	68.43	107.36	145.07	213.24
			(2.03)	(2.18)	(1.60)	(2.44)	(8.43)
	m	l	43.48	80.60	119.26	152.01	191.19
			(2.85)	(3.89)	(2.54)	(2.73)	(4.70)
	h	l	14.49	57.96	109.38	145.55	195.53
			(1.56)	(4.46)	(2.53)	(3.12)	(6.21)
	l	m	69.96	107.52	136.27	185.63	252.07
			(1.36)	(0.61)	(0.57)	(1.31)	(2.03)
	m	m	74.70	113.57	139.59	176.46	235.95
			(3.46)	(1.42)	(1.21)	(2.26)	(6.33)
	h	m	91.84	132.21	158.32	194.57	258.80
			(5.49)	(3.44)	(3.38)	(4.86)	(16.07)
	l	h	122.40	140.62	170.46	222.01	293.87
			(0.11)	(0.10)	(0.15)	(0.34)	(0.60)
	m	h	106.86	140.57	174.25	223.10	290.66
			(3.66)	(1.94)	(2.38)	(4.00)	(7.41)
	h	h	107.09	161.09	169.08	238.89	350.28
			(30.56)	(22.47)	(27.93)	(59.26)	(101.18)

Table 14: Point estimates of function values in model (20) for different quantile indices (values of $\mathcal{U}_i$) in the **trade** sector. Standard errors are shown in parentheses and are obtained by a weighted XY bootstrap with 299 bootstrap iterations.

	$\mathscr{E}^s$	$\mathscr{E}^w$	Regression quantiles 0.1	0.25	0.5	0.75	0.9
Self-employment	l	l	0.00	0.00	39.10	87.74	129.14
			(0.91)	(0.85)	(6.13)	(5.77)	(9.62)
	m	l	14.13	65.21	120.91	162.71	208.13
			(4.56)	(4.02)	(3.85)	(4.73)	(8.27)
	h	l	80.71	135.11	200.90	289.75	409.95
			(2.27)	(1.73)	(2.26)	(3.07)	(6.33)
	l	m	0.00	33.13	84.21	130.53	178.40
			(0.13)	(3.14)	(2.68)	(3.05)	(4.94)
	m	m	58.68	112.91	157.41	215.93	291.01
			(5.07)	(4.21)	(3.56)	(5.56)	(8.92)
	h	m	109.77	167.06	229.00	308.49	424.90
			(9.39)	(9.76)	(7.28)	(11.70)	(23.43)
	l	h	15.23	66.36	110.59	150.71	200.34
			(3.76)	(2.84)	(1.70)	(2.66)	(5.42)
	m	h	90.78	148.95	202.74	255.98	351.20
			(11.34)	(5.35)	(5.99)	(10.08)	(20.36)
	h	h	109.74	174.73	329.99	365.11	876.62
			(93.12)	(55.26)	(100.27)	(121.99)	(442.26)
Wage-employment	l	l	47.77	91.58	123.73	160.66	214.70
			(3.91)	(2.85)	(1.44)	(3.31)	(5.66)
	m	l	50.26	92.28	121.36	148.54	187.96
			(4.31)	(5.34)	(1.94)	(3.48)	(9.97)
	h	l	31.25	81.05	118.66	152.06	203.77
			(4.41)	(4.08)	(2.58)	(3.68)	(7.73)
	l	m	83.82	112.72	137.08	167.33	215.00
			(1.22)	(0.56)	(0.56)	(0.97)	(2.14)
	m	m	78.34	115.03	144.07	172.01	215.54
			(3.43)	(2.28)	(1.70)	(2.98)	(5.32)
	h	m	74.00	119.53	145.92	172.17	205.63
			(6.39)	(3.79)	(3.49)	(4.68)	(11.95)
	l	h	124.03	139.74	162.48	194.19	244.67
			(0.10)	(0.13)	(0.12)	(0.26)	(0.54)
	m	h	97.45	129.71	157.69	194.56	240.50
			(4.48)	(2.55)	(2.52)	(4.25)	(8.04)
	h	h	122.45	145.81	175.84	189.25	242.97
			(36.32)	(19.06)	(22.82)	(48.56)	(67.09)

Table 15: Point estimates of function values in model (20) for different quantile indices (values of $\mathscr{U}_i$) in the **transportation** sector. Standard errors are shown in parentheses and are obtained by a weighted XY bootstrap with 299 bootstrap iterations.

	$\mathcal{E}^s$	$\mathcal{E}^w$	0.1	0.25	0.5	0.75	0.9
			Regression quantiles				
Self-employment	l	l	0.00	0.00	26.43	77.24	147.83
			(0.50)	(0.00)	(2.68)	(4.04)	(7.99)
	m	l	0.00	17.62	83.47	170.23	288.81
			(0.00)	(2.67)	(4.47)	(5.62)	(11.43)
	h	l	30.23	109.03	223.96	381.93	606.61
			(2.25)	(2.57)	(2.97)	(5.10)	(11.88)
	l	m	0.00	14.87	67.66	141.86	244.81
			(0.00)	(1.56)	(1.81)	(2.61)	(7.40)
	m	m	12.66	68.90	145.58	237.52	362.96
			(2.52)	(2.79)	(3.68)	(5.84)	(12.04)
	h	m	39.85	132.52	260.25	436.95	749.21
			(7.05)	(9.00)	(11.16)	(22.42)	(44.96)
	l	h	0.00	26.42	95.49	181.40	299.12
			(0.00)	(1.44)	(1.83)	(2.41)	(5.91)
	m	h	28.54	114.80	217.80	351.74	509.32
			(4.11)	(6.38)	(6.85)	(10.44)	(28.53)
	h	h	96.02	120.85	175.84	469.64	859.60
			(41.18)	(70.70)	(118.95)	(167.52)	(351.23)
Wage-employment	l	l	40.16	87.97	133.59	189.58	249.72
			(2.14)	(2.78)	(2.35)	(3.28)	(4.32)
	m	l	24.32	82.67	137.62	206.13	312.45
			(3.59)	(5.70)	(5.76)	(7.28)	(13.88)
	h	l	7.86	25.24	100.12	162.10	241.41
			(0.40)	(2.42)	(4.26)	(4.61)	(10.63)
	l	m	92.07	128.06	179.40	229.21	287.31
			(1.26)	(0.72)	(0.96)	(1.13)	(2.09)
	m	m	70.31	127.57	172.29	245.22	313.16
			(4.91)	(2.35)	(3.22)	(4.76)	(7.04)
	h	m	41.64	130.45	177.23	265.95	336.62
			(9.81)	(4.65)	(4.82)	(11.94)	(19.94)
	l	h	134.04	164.58	212.37	272.47	345.77
			(0.19)	(0.23)	(0.30)	(0.37)	(0.65)
	m	h	103.45	145.58	199.93	272.32	359.49
			(4.53)	(3.08)	(3.86)	(6.11)	(11.62)
	h	h	121.72	141.89	229.05	330.44	645.11
			(35.75)	(28.34)	(64.56)	(134.20)	(286.61)

Table 16: Point estimates of function values in model (20) for different quantile indices (values of $\mathcal{U}_i$) in the **finance** sector. All (non-zero) estimates are significantly different from zero with p-values < 0.01 unless otherwise noted. Inference is based on a weighed XY bootstrap with 299 bootstrap iterations.

9 References

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006). Quantile regression under misspecification, with an application to the us wage structure. *Econometrica 74*(2), 539–563.

Baumol, W. (1990). Entrepreneurship: Productive, unproductive, and destructive. *The Journal of Political Economy 98*(5), 893–921.

Blanchflower, D. (2000). Self-employment in oecd countries. *Labour economics 7*(5), 471–505.

Bose, A. and S. Chatterjee (2003). Generalized bootstrap for estimators of minimizers of convex functions. *Journal of statistical planning and inference 117*(2), 225–239.

Gartner, W. (2007). Psychology, entrepreneurship, and the "critical mess". *The psychology of entrepreneurship*, 325–334.

Gort, M. and S. Lee (2007). The rewards to entrepreneurship. *Buffalo State University of New York, Working paper*.

Hahn, J. (1997). Bayesian bootstrap of the quantile regression estimator: a large sample study. *international economic review 38*(4), 795–808.

Hamilton, B. (2000). Does entrepreneurship pay? an empirical analysis of the returns of self-employment. *Journal of Political Economy*, 604–631.

Kaiser, U. and N. Malchow-Møller (2010). Is self-employment really a bad experience?: The effects of previous self-employment on subsequent wage-employment wages. *Journal of Business Venturing 26*(5), 572–588.

Kleven, H., M. Knudsen, C. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica 79*(3), 651–692.

Lazear, E. (2004). Balanced skills and entrepreneurship. *The American Economic Review 94*(2), 208–211.

le Maire, D. and B. Scherning (2007). Earnings, uncertainty, and the self-employment choice. *CEBR discussion paper, 2007-04*.

Moskowitz, T. and A. Vissing-Jørgensen (2002). The returns to entrepreneurial investment: A private equity premium puzzle? *The American Economic Review 92*(4), 745–778.

Murphy, K., A. Shleifer, and R. Vishny (1991). The allocation of talent: implications for growth. *The Quarterly Journal of Economics 106*(2), 503.

Schumpeter, J. (1942). *Socialism, capitalism and democracy*. Harper and Bros.

Silva, O. (2007). The jack-of-all-trades entrepreneur: Innate talent or acquired skill? *Economics Letters 97*(2), 118–123.

Tergiman, C. (2011). Entrepreneurship does pay. *New York University, Working paper*.

Van Praag, C. and P. Versloot (2007). What is the value of entrepreneurship? a review of recent research. *Small Business Economics 29*(4), 351–382.

Wagner, J. (2003). Testing lazear's jack-of-all-trades view of entrepreneurship with german micro data. *Applied Economics Letters 10*(11), 687–689.

Åstebro, T. and P. Thompson (2011). Entrepreneurs, jacks of all trades or hobos? *Research Policy 40*(5), 637–649.

DEPARTMENT OF ECONOMICS AND BUSINESS

AARHUS UNIVERSITY – BARTHOLINS ALLÉ 10 – BUILDING 1322
DK-8000 AARHUS C – TEL. +45 8715 0000 - www.econ.au.dk

PhD Theses:

1999-4	Philipp J.H. Schröder, Aspects of Transition in Central and Eastern Europe.
1999-5	Robert Rene Dogonowski, Aspects of Classical and Contemporary European Fiscal Policy Issues.
1999-6	Peter Raahauge, Dynamic Programming in Computational Economics.
1999-7	Torben Dall Schmidt, Social Insurance, Incentives and Economic Integration.
1999	Jørgen Vig Pedersen, An Asset-Based Explanation of Strategic Advantage.
1999	Bjarke Jensen, Five Essays on Contingent Claim Valuation.
1999	Ken Lamdahl Bechmann, Five Essays on Convertible Bonds and Capital Structure Theory.
1999	Birgitte Holt Andersen, Structural Analysis of the Earth Observation Industry.
2000-1	Jakob Roland Munch, Economic Integration and Industrial Location in Unionized Countries.
2000-2	Christian Møller Dahl, Essays on Nonlinear Econometric Time Series Modelling.
2000-3	Mette C. Deding, Aspects of Income Distributions in a Labour Market Perspective.
2000-4	Michael Jansson, Testing the Null Hypothesis of Cointegration.
2000-5	Svend Jespersen, Aspects of Economic Growth and the Distribution of Wealth.
2001-1	Michael Svarer, Application of Search Models.
2001-2	Morten Berg Jensen, Financial Models for Stocks, Interest Rates, and Options: Theory and Estimation.
2001-3	Niels C. Beier, Propagation of Nominal Shocks in Open Economies.
2001-4	Mette Verner, Causes and Consequences of Interrruptions in the Labour Market.
2001-5	Tobias Nybo Rasmussen, Dynamic Computable General Equilibrium Models: Essays on Environmental Regulation and Economic Growth.

2001-6	Søren Vester Sørensen, Three Essays on the Propagation of Monetary Shocks in Open Economies.
2001-7	Rasmus Højbjerg Jacobsen, Essays on Endogenous Policies under Labor Union Influence and their Implications.
2001-8	Peter Ejler Storgaard, Price Rigidity in Closed and Open Economies: Causes and Effects.
2001	Charlotte Strunk-Hansen, Studies in Financial Econometrics.
2002-1	Mette Rose Skaksen, Multinational Enterprises: Interactions with the Labor Market.
2002-2	Nikolaj Malchow-Møller, Dynamic Behaviour and Agricultural Households in Nicaragua.
2002-3	Boriss Siliverstovs, Multicointegration, Nonlinearity, and Forecasting.
2002-4	Søren Tang Sørensen, Aspects of Sequential Auctions and Industrial Agglomeration.
2002-5	Peter Myhre Lildholdt, Essays on Seasonality, Long Memory, and Volatility.
2002-6	Sean Hove, Three Essays on Mobility and Income Distribution Dynamics.
2002	Hanne Kargaard Thomsen, The Learning organization from a management point of view - Theoretical perspectives and empirical findings in four Danish service organizations.
2002	Johannes Liebach Lüneborg, Technology Acquisition, Structure, and Performance in The Nordic Banking Industry.
2003-1	Carter Bloch, Aspects of Economic Policy in Emerging Markets.
2003-2	Morten Ørregaard Nielsen, Multivariate Fractional Integration and Cointegration.
2003	Michael Knie-Andersen, Customer Relationship Management in the Financial Sector.
2004-1	Lars Stentoft, Least Squares Monte-Carlo and GARCH Methods for American Options.
2004-2	Brian Krogh Graversen, Employment Effects of Active Labour Market Programmes: Do the Programmes Help Welfare Benefit Recipients to Find Jobs?
2004-3	Dmitri Koulikov, Long Memory Models for Volatility and High Frequency Financial Data Econometrics.
2004-4	René Kirkegaard, Essays on Auction Theory.

2004-5	Christian Kjær, Essays on Bargaining and the Formation of Coalitions.
2005-1	Julia Chiriaeva, Credibility of Fixed Exchange Rate Arrangements.
2005-2	Morten Spange, Fiscal Stabilization Policies and Labour Market Rigidities.
2005-3	Bjarne Brendstrup, Essays on the Empirical Analysis of Auctions.
2005-4	Lars Skipper, Essays on Estimation of Causal Relationships in the Danish Labour Market.
2005-5	Ott Toomet, Marginalisation and Discouragement: Regional Aspects and the Impact of Benefits.
2005-6	Marianne Simonsen, Essays on Motherhood and Female Labour Supply.
2005	Hesham Morten Gabr, Strategic Groups: The Ghosts of Yesterday when it comes to Understanding Firm Performance within Industries?
2005	Malene Shin-Jensen, Essays on Term Structure Models, Interest Rate Derivatives and Credit Risk.
2006-1	Peter Sandholt Jensen, Essays on Growth Empirics and Economic Development.
2006-2	Allan Sørensen, Economic Integration, Ageing and Labour Market Outcomes
2006-3	Philipp Festerling, Essays on Competition Policy
2006-4	Carina Sponholtz, Essays on Empirical Corporate Finance
2006-5	Claus Thrane-Jensen, Capital Forms and the Entrepreneur – A contingency approach on new venture creation
2006-6	Thomas Busch, Econometric Modeling of Volatility and Price Behavior in Asset and Derivative Markets
2007-1	Jesper Bagger, Essays on Earnings Dynamics and Job Mobility
2007-2	Niels Stender, Essays on Marketing Engineering
2007-3	Mads Peter Pilkjær Harmsen, Three Essays in Behavioral and Experimental Economics
2007-4	Juanna Schrøter Joensen, Determinants and Consequences of Human Capital Investments
2007-5	Peter Tind Larsen, Essays on Capital Structure and Credit Risk

2008-1 Toke Lilhauge Hjortshøj, Essays on Empirical Corporate Finance – Managerial Incentives, Information Disclosure, and Bond Covenants

2008-2 Jie Zhu, Essays on Econometric Analysis of Price and Volatility Behavior in Asset Markets

2008-3 David Glavind Skovmand, Libor Market Models - Theory and Applications

2008-4 Martin Seneca, Aspects of Household Heterogeneity in New Keynesian Economics

2008-5 Agne Lauzadyte, Active Labour Market Policies and Labour Market Transitions in Denmark: an Analysis of Event History Data

2009-1 Christian Dahl Winther, Strategic timing of product introduction under heterogeneous demand

2009-2 Martin Møller Andreasen, DSGE Models and Term Structure Models with Macroeconomic Variables

2009-3 Frank Steen Nielsen, On the estimation of fractionally integrated processes

2009-4 Maria Knoth Humlum, Essays on Human Capital Accumulation and Educational Choices

2009-5 Yu Wang, Economic Analysis of Open Source Software

2009-6 Benjamin W. Blunck, Creating and Appropriating Value from Mergers and Acquisitions – A Merger Wave Perspective

2009-7 Rune Mølgaard, Essays on Dynamic Asset Allocation and Electricity Derivatives

2009-8 Jens Iversen, Financial Integration and Real Exchange Rates: Essays in International Macroeconomics

2009-9 Torben Sørensen, Essays on Human Capital and the Labour Market

2009-10 Jonas Staghøj, Essays on Unemployment and Active Labour Market Policies

2009-11 Rune Majlund Vejlin, Essays in Labor Economics

2009-12 Nisar Ahmad, Three Essays on Empirical Labour Economics

2010-1 Jesper Rosenberg Hansen, Essays on Application of Management Theory in Public Organizations - Changes due to New Public Management

2010-2 Torben Beedholm Rasmussen, Essays on Dynamic Interest Rate Models and Tests for Jumps in Asset Prices

2010-3 Anders Ryom Villadsen, Sources of Organizational Similarity and Divergence: Essays on Decision Making and Change in Danish Municipalities Using Organizational Institutionalism

2010-4 Eske Stig Hansen, Essays in Electricity Market Modeling

2010-5 Marías Halldór Gestsson, Essays on Macroeconomics and Economic Policy

2010-6 Niels Skipper, Essays on the Demand for Prescription Drugs

2010-7 Thomas Quistgaard Pedersen, Return predictability and dynamic asset allocation

2011-1 Malene Kallestrup Lamb, Health, Retirement and Mortality

2011-2 Lene Kromann, Essays on Productivity and Labour Supply

2011-3 Firew Bekele Woldeyes, Essays on Government Budgets in Developing Countries

2011-4 Anders Bredahl Kock, Forecasting and Oracle Efficient Econometrics

2011-5 Christian Bach, The Game of Risk

2011-6 Stefan Holst Bache, Quantile Regression: Three Econometric Studies