# Essays in Microeconometrics

Tobias J. Klein

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften der Universität Mannheim

Dekan: Prof. Konrad Stahl, Ph.D.
Referent: Prof. Konrad Stahl, Ph.D.
Korreferent: Prof. Dr. Enno Mammen
Tag der mündlichen Prüfung: 28. Juli 2006

# ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors Enno Mammen and Konrad Stahl for their support and numerous insightful discussions. Furthermore, I am grateful to Andrew Chesher for his support and helpful suggestions during my stay at the University College London in the academic year 2003/4.

Chapters 2 through 4 of this dissertation have been written over the last four years. I appreciate the valuable comments on early versions of my work received in internal workshops at the University of Mannheim and University College London as well as in contributed sessions on various conferences and a seminar presentation. Versions of Chapter 2 were presented at the 2004 ENTER Jamboree in Barcelona, the 2004 Spring Meeting of Young Economists, and the 2004 Econometric Society Meeting in Madrid. Chapter 3 was presented at the 2005 IAB conference "The Empirical Evaluation of Labour Market Programmes" in Nuremberg and the 2006 North American Winter Meeting of the Econometric Society in Boston. Chapter 4 was presented at the 2006 SFB Conference in Caputh, at the 2006 International IO Conference in Boston and at the Wissenschaftszentrum Berlin in a seminar.

I would like to thank my colleagues, as well as Josh Angrist, Heski Bar-Isaac, Erich Battistin, Richard Blundell, Pedro Carneiro, Jim Heckman, Stefan Hoderlein, Pierre Hoonhout, Johannes Koenen, Jürgen Maurer, Cristina Santos, Melanie Schienle, Axel Ockenfels, Klaus Schmidt, Henry Schneider, and Edward Vytlacil for comments. I would also like to express my gratitude to two anonymous referees of the Junior Economits Peer-reviewed Discussion Paper Series (JEPS) for their detailed reports on Chapter 3. Moreover, I would like to thank Richard Blundell, Lorraine Dearden, as well as Leslie McGranahan for sharing their data which were used in Chapter 3.

I am especially grateful to my coauthors for Chapter 4, Christian Lambertz, Giancarlo Spagnolo, and Konrad Stahl. I have not only enjoyed working with them, but also benefitted greatly from this collaboration. The four of us acknowledge the excellent research assistance of Florian Hauber and Johannes Koenen. We also truly appreciate having had the opportunity to extensively work on the associated paper during the 2005 C.E.P.R. European Summer Symposium in Economic Theory in Gerzensee, Switzerland.

Furthermore, I thank the Department of Economics at the University College

London for its hospitality during the academic year 2003/4. Financial support by the University of Mannheim and the German Research Foundation (DFG) over the last four years as well as the European Commission Marie Curie program has made writing this thesis possible and is therefore gratefully acknowledged.

Last, but for sure not least, I would like to thank my girlfriend, my family, and my friends for supporting, encouraging, and backing me during the last four years. Their interest in my work has been a constant source of motivation.

TABLE OF CONTENTS

# 1. INTRODUCTION

Over the last decades individual field data has become more and more available to researchers. In three self-contained chapters, this thesis is concerned with the analysis of such data and some associated challenges.

An important aspect therein is what is referred to, e.g. by Fisher (1966) and Manski (1995), as the *identification problem in the social sciences*. It arises once the observed variables on the right hand side of a structural equation are correlated with the unobserved variables.[1] In econometrics this is usually referred to as the problem of *simultaneity* or *endogeneity*. Chapter 2 and 3 of this thesis address associated challenges, focussing on the *identification* of structural parameters of interest in the presence of endogeneity.

Koopmans (1945, p. 462) was among the first to study identification and notes that "economic data are not controlled values selected for the purpose of an experiment. Economic variables are produced by society." For example, in a standard bivariate regression model,

$$y_i = \alpha + \beta x_i + \varepsilon_i, \tag{1.1}$$

in which $y_i$ is individual $i$'s wage and $x_i$ the number of years of schooling, the latter could be endogenous because it depends on the idiosyncratic wage premium $\varepsilon_i$. In an alternative formalization there are additional so-called confounding variables such as ability which determine both wages and schooling choices.[2] In both cases, the returns to an additional year of schooling, $\beta$, are not identified: they can in general not be recovered from observations.

According to Koopmans (1945, p. 462) this "leads to the requirement that statistical methods of fitting take into account the formation of economic variables through a complete system of equations." Therefore, a "separation between problems of statistical inference... and problems of identification (Koopmans 1949, p. 132)" becomes necessary.

A formalization of the *notion of identification* can at least be traced back to Koopmans and Reiersol (1950) with important contributions by Hurwicz (1950a, 1950b), Koopmans, Rubin, and Leipnik (1950), and Wald (1950). They typically refer to an econometric model as a structure and say that two structures $S$ and $S'$ are observationally equivalent (indistinguishable) if, conditional on exogenous variables, the two distributions of endogenous variables generated by the structures $S$ and $S'$ are identical for all possible values of the exogenous variables. The structure $S$ is (uniquely) identifiable if there is no observationally equivalent structure $S'$. Finally, a parameter $\theta$ of a structure $S$ is identifiable if it is the same for all structures $S'$ observationally equivalent to $S$.

[1] We follow Goldberger (1972) and call an equation *structural* if it represents a causal link rather than a mere empirical association.

[2] See Fisher (1935, Ch. 7) and Yates (1937) for the notion of confounding.

Several approaches have been suggested in order to overcome the problem of endogeneity and achieve identification. For example, attempts to formulate systems of simultaneous equations were already made in papers by Sewall Wright on path coefficitents in the 1920's and 1930's (Wright 1934, e.g.). In this context, Koopmans (1945, p. 450, italics added) even emphasizes that "the formation of economic variables can *only* be described by a system of simultaneous equations."

A more reduced form approach is to exploit properties of so-called instrumental variables which, in the textbook formulation, are correlated with the observed right hand side variable, $x_i$ in our case, but uncorrelated with the unobserved error term, $\varepsilon_i$. It is well-known that such a source of exogenous variation can, under appropriate conditions, be used to uncover the effect of $x_i$ on $y_i$. As opposed to a system of structural equations, in which the *causal link* between the instrument and the endogenous variable is modelled, it is a reduced form approach because it only exploits the *correlation* between the endogenous variable and the instrument. The use of such instrumental variables can be traced back to Appendix B of Philip G. Wright (1928).[3]

Chapter 2 and 3 of this thesis try to contribute to the modern literature on identification and estimation using instrumental variables. Chapter 2 is concerned with the problem of evaluating the impact of a program on an outcome variable, e.g. the effect of on-the-job training on wages. The model we consider is completely nonparametric. In a stripped-down version, $x_i$ from above is a binary indicator for the treatment decision that takes on the value 1 if an individual participated in a program. The framework that is used in Chapter 2 is more general than (1.1) because it allows the effect of $x_i$ on $y_i$, the idiosyncratic gains from participation, to be stochastic. That is, (1.1) becomes

$$y_i = \alpha + \beta_i x_i + \varepsilon_i \tag{1.2}$$

with $x_i$, $\varepsilon_i$, and $\beta_i$ being scalar random variables.

We face a fundamental identification problem because not only the treatment decision, $x_i$, and the error term, $\varepsilon_i$, might depend on each other, but also the treatment decision and the idiosyncratic gains from participation, $\beta_i$. Imbens and Angrist (1994) were the first to exploit monotonicity of the treatment decision in the instruments in order to identify an average treatment effect parameter using instrumental variables. More recently, Heckman and Vytlacil (1999, 2000a, 2000b, 2005) suggested estimation of a variety of treatment effect parameters using a local version of their approach. However, identification hinges on the same monotonicity assumption that is fundamentally untestable. We investigate the sensitivity of respective estimates to reasonable departures from monotonicity that

[3] For a historic perspective see parts of Bowden and Turkington (1984), Goldberger (1972), Morgan (1990), Angrist and Krueger (2001), and Stock and Trebbi (2003).

are likely to be encountered in practice and relate it to properties of a structural parameter. One of our results is that the bias vanishes under a testable linearity condition. Our findings are illustrated in a Monte Carlo analysis.

Thereafter, in Chapter 3, we propose and implement an estimator for identifiable features of correlated random coefficient models with binary endogenous variables and nonadditive errors in the outcome equation. The framework is similar to (1.2) in nature but now $x_i$ is a vector consisting of one binary endogenous variable and $K > 0$ exogenous variables. It is still nonparametric in the direction of the unobservables but parametric in the direction of the exogenous covariates in order to reduce the dimensionality of the estimation problem. The estimator we propose is suitable, e.g., for estimation of the average returns to college education if they are heterogeneous across individuals and correlated with the schooling choice. The estimated features are of central interest to economists and are directly linked to the marginal and average treatment effect in policy evaluation. They are identified under assumptions weaker than typical exclusion restrictions used in the context of classical instrumental variables analysis. In our application for the U.K., we relate levels of expected wages to unobserved ability, measured ability, family background, type of secondary school, and the decision whether to attend college.

While Chapters 2 and 3 focus on recovering identifiable features of a structural equation, Chapter 4 follows a different approach. It is concerned with feedback mechanisms in electronic markets that allow partners to rate each other after a transaction. These mechanisms are considered crucial for the success of anonymous internet trading platforms. Rather than estimating a system of structural equations we document an asymmetry in the feedback behavior on eBay, propose an explanation based on the micro structure of the feedback mechanism and the time when feedbacks are given, and support this explanation by findings from a large data set. Our analysis implies that the informational content of feedback records is likely to be low. We argue that the reason for this is that agents leave feedbacks strategically. Negative feedbacks are given late, in the "last minute," or not given at all, most likely because of the fear of retaliative negative feedback. Conversely, positive feedbacks are given early in order to encourage reciprocation. Towards refining our insights into the observed pattern, we look separately at buyers and sellers, and relate the magnitude of the effects to the trading partners' experience.

Hence, the econometric approach that is taken in Chapter 4 is to describe individual feedback behavior using standard nonparametric methods in a first step. This descriptive evidence is paired in a second step with institutional details and economic reasoning in order to derive normative conclusions that aim at improving the feedback mechanism.

# BIBLIOGRAPHY

Angrist, J. D., and A. B. Krueger (2001): "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 15(4), 69–85.

Bowden, R. J., and D. A. Turkington (1984): *Instrumental Variables*. Cambridge University Press, Cambridge, MA.

Fisher, F. (1966): *The Identification Problem in Econometrics*. McGraw Hill.

Fisher, R. A. (1935): *The Design of Experiments*. Oliver & Boyd, Edinburgh.

Goldberger, A. S. (1972): "Structural Equation Methods in the Social Sciences," *Econometrica*, 40(6), 979–1001.

Heckman, J. J., and E. J. Vytlacil (1999): "Local Instrumental Variables and Latent Variables Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730–4734.

——— (2000a): "Local Instrumental Variables," in *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, ed. by C. Hsiao, K. Morimune, and J. Powell, pp. 1–46. Cambridge University Press, Cambridge.

——— (2000b): "Local Instrumental Variables," NBER Working Paper No. T0252.

——— (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

Hurwicz, L. (1950a): "Generalization of the Concept of Identification," in *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10. John Wiley and Sons, New York.

——— (1950b): "Prediction and Least Squares," in *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10. John Wiley and Sons, New York.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

KOOPMANS, T. C. (1945): "Statistical Estimation of Simultaneous Economic Relations," *Journal of the American Statistical Association*, 40(232, Part 1), 448–466.

——— (1949): "Identification Problems in Economic Model Construction," *Econometrica*, 17(2), 125–144.

KOOPMANS, T. C., AND O. REIERSOL (1950): "The Identification of Structural Characteristics," *Annals of Mathematical Statistics*, 21(2), 165–181.

KOOPMANS, T. C., H. RUBIN, AND R. B. LEIPNIK (1950): "Measuring the Equation System of Dynamic Economies," in *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10. John Wiley and Sons, New York.

MANSKI, C. F. (1995): *Identification Problems in the Social Sciences*. Harvard University Press, Cambridge, MA.

MORGAN, M. S. (1990): *The History of Econometric Ideas*. Cambridge University Press, Cambridge, MA.

STOCK, J. H., AND F. TREBBI (2003): "Who Invented Instrumental Variable Regression?," *Journal of Economic Perspectives*, 17(3), 177–194.

WALD, A. (1950): "Note on the Identification of Economic Relations," in *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10. John Wiley and Sons, New York.

WRIGHT, P. G. (1928): *The Tariff on Animal and Vegetable Oil*. MacMillan, New York.

WRIGHT, S. (1934): "The Method of Path Coefficients," *Annals of Mathematical Statistics*, 5(3), 161–215.

YATES, F. (1937): "The Design and Analysis of Factorial Experiments," Discussion paper, Technical Communication no. 35 of the Commonwealth Bureau of Soils.

# 2. HETEROGENEOUS TREATMENT EFFECTS: INSTRUMENTAL VARIABLES WITHOUT MONOTONICITY?

## 2.1 Introduction

### 2.1.1 Monotonicity

A fundamental identification problem in program evaluation arises if the treatment decision depends on the idiosyncratic gain from participation even if we condition on observables. This selection into treatments on unobservables precludes the use of the usual econometric tools such as matching type estimators, conventional instrumental variables analysis, and standard simultaneous equations models because their respective estimates of treatment effect parameters are generally biased.

Imbens and Angrist (1994) were the first to exploit monotonicity of the treatment decision in instrumental variables in order to identify a local average treatment effect parameter (LATE analysis). These instrumental variables are assumed to be independent of the pair of potential outcomes conditional on covariates in the outcome equation. They have identifying power if, conditional on these covariates, they have an impact on the treatment probability. The monotonicity assumption is that a *hypothetical* change in the instruments either has no impact on a unit's treatment status, or changes its treatment status in the same direction as it does for all other units for which it has an impact.

More recently, Heckman and Vytlacil (1999, 2000a, 2000b, 2005; HV in the remainder) suggested estimation of a variety of treatment effect parameters using a local version of their approach called local instrumental variables analysis (LIV analysis), drawing on a stronger condition on the support of the propensity score, the treatment probability conditional on the instruments.

Both approaches, LATE and LIV analysis, are in principle able to cope with unobserved dependence between the treatment decision and the outcome. They are intuitive, elegant, and easy to implement. Their generality consists of the fact that neither a parametric specification of the joint distribution of unobservables and observables, nor peculiarities of the data set or the economic question of interest are of need. However, identification in both approaches hinges on the same monotonicity assumption. In general, estimates of treatment effects will be biased if it does not hold.

A violation of the monotonicity assumption is nicely motivated in Example 2 of Imbens and Angrist (1994). When we think of two officials screening applicants for a social program, we would expect that for every set of characteristics of the applicants (the covariates in the outcome equation) the admission rate differs between the two officials. When it is unlikely that the identity of the official affects the outcome of participation or nonparticipation in the program, then, conditional on the characteristics of the applicant, this identity qualifies as an instrument. Suppose the admission rate for official A was higher than for official B. Then, in this setup monotonicity holds whenever *any* applicant who would have been accepted by official B *is* accepted by official A. Imbens and Angrist (1994) note that "this is unlikely to hold if admission is based on a number of criteria." In this case, monotonicity is violated.

In this paper, we aim at quantifying the degree of violation of the monotonicity assumption in order to investigate the consequences of a violation when we unjustifiably rely on this assumption. In particular, we investigate the sensitivity of both LIV and LATE estimates to reasonable departures from monotonicity that are likely to be encountered in practice. We focus on the bias in estimates of the marginal and average as well as the local average treatment effect. Importantly, highly sensitive estimates would question the suitability of LIV and LATE analysis for applied work as the monotonicity assumption is fundamentally untestable since it is identifying.[1]

It turns out that this can be done in a unifying approach because LATE and LIV analysis are similar with respect to the identification strategy and only differ with respect to the requirements on the support of the propensity score. In particular, conditional on covariates in the outcome equation, LIV analysis requires derivatives of the expected outcomes with respect to the propensity score at infinitely many values of the propensity score to be identified whereas LATE analysis is based on a finite set of level estimates.

### *2.1.2 Local Departures From Monotonicity*

The approach in this paper is to study the impact of *local departures* from monotonicity on our estimates. Taylor series approximations to the bias terms are derived. This is in the tradition of local specification error analysis suggested by Kiefer and Skoog (1984).[2] It has also been successfully applied by Chesher (1991), Chesher and Schluter (2002) and Battistin and Chesher (2004) in the context of measurement error. Lately, Chesher and Santos Silva (2002) studied the impact of uncontrolled taste variation in discrete choice models by modelling local departures from a multinomial logit model.

The virtue of this approach is that it allows us to keep in touch with the original structure. At the same time, we are able to explore what the sensitivity of LATE and LIV estimates depends on when monotonicity is in fact violated. In our case,

[1] HV propose a joint test for monotonicity and the existence of instruments. By itself, monotonicity is fundamentally untestable.

[2] Angrist, Imbens, and Rubin (1996) take a different approach and relate the bias in conventional IV estimates to the proportion of non-compliers, i.e. the units for which monotonicity is violated, and the treatment effect heterogeneity. We feel that our approach is fruitful because it allows us to express the treatment effect of non-compliers in terms of, up to an approximation error, possibly identifiable quantities.

the original structure consists of selection models of the form

$$D = \mathbb{I}\{\nu(Z) \geq \tilde{V}\}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function, $\nu : \mathbb{R}^k \rightarrow \mathbb{R}$ is a function of a $k$-vector of instrumental variables, and $\tilde{V}$ is an unobserved individual threshold, which is continuously distributed with distribution function $F_{\tilde{V}}$. Moreover, $Z$ and $\tilde{V}$ are assumed to be independently distributed. From now on, we keep conditioning on covariates in the outcome equation implicit.

These selection models imply monotonicity. To see this let $w$ and $z$ be two values of $Z$ with $\nu(w) < \nu(z)$.[3] Then, given $\tilde{V}$, $D$ can never change from 1 to 0 if $Z$ changes from $w$ to $z$. This is the original monotonicity assumption of Imbens and Angrist (1994). Vytlacil (2002) shows that such a selection model does not impose any additional restrictions on the data generating process. Therefore, without loss of generality, from now on, we represent the original LATE assumptions in this form.

Central to our generalization

$$D = \mathbb{I}\{\mu(\nu(Z), \sigma U) \geq \tilde{V}\},$$

$\mu : \mathbb{R}^2 \rightarrow \mathbb{R}$, $\mathbb{E}[\mu(\nu(Z), \sigma U)|Z] = \nu(Z)$, is an additional scalar random component $\sigma U$ which gives rise to additional individual heterogeneity and, under appropriate conditions, directly leads to a violation of the monotonicity assumption whenever $\sigma U$ is non-degenerate.[4] For degenerate $\sigma U$, however, the model is constructed so that it is equivalent to the canonical selection model. In this framework, a local departure from monotonicity is given by a change from $\sigma = 0$ to $\sigma > 0$.

The following example illustrates this point.

**Example 1 (Random Coefficients):** Consider the index selection model

$$D = \mathbb{I}\{Z\tilde{\gamma} \geq \tilde{V}\}.$$

The canonical index selection model would postulate that $\tilde{\gamma} = \gamma$, a $k$-vector of parameters. In a probit model, e.g., the additional assumption is made that $\tilde{V}$ is standard normally distributed. Then, given $\tilde{V}$, if $Z\gamma$ changes from $w\gamma$ to $z\gamma > w\gamma$, $D$ can only change from 0 to 1, remain 0, or remain 1, but can never change from 1 to 0. Now, let $\tilde{\gamma}$ be a vector of random coefficients $\tilde{\gamma} = \gamma + \sigma U \cdot \iota$,

[3] For any random variable $A$ and any vector of random variables $B$ we denote realizations thereof by lowercase letters, the c.d.f. of $A$ evaluated at $A = a$ by $F_A(a)$, the conditional c.d.f. of $A$ given $B = b$ evaluated at $A = a$ by $F_{A|B=b}(a)$, and the respective p.d.f.'s by $f_A(a)$ and $f_{A|B=b}(a)$.

[4] Conversely, Vytlacil (forthcoming) states conditions under which monotonicity still holds. These conditions will not be assumed to hold in this paper.

$\sigma \geq 0$, $\iota$ being a $k$-vector of ones, and $U$ non-degenerate and independent of $Z$. This is an example for the generalized selection model motivated above. If $\sigma > 0$, given $\tilde{V}$, $D$ is no longer monotone in $Z$ because now, under fairly general conditions on the distribution of $U$, there exist realizations $u$ and $u'$ of $U$ such that $w(\gamma + \sigma u \cdot \iota) > z(\gamma + \sigma u' \cdot \iota)$ while $w\gamma < z\gamma$. Consequently, monotonicity is violated. On the other hand, this model nests the canonical index selection model as the special case in which $\sigma = 0$. A local departure from monotonicity is hence given by an external change from $\sigma = 0$ to a small $\sigma > 0$. □

In this paper, we are interested in the bias of treatment effect parameter estimates that can be attributed to such a violation of the monotonicity assumption. We derive a second order approximation to respective bias terms in $\sigma$ about $\sigma = 0$ that can be used to assess the accuracy of LIV and LATE estimates without monotonicity. We show that the respective bias depends primarily on the dependence between the individual gains from participation in the program, $Y_1 - Y_0$, and the normalized selection threshold $V = F_{\tilde{V}}(\tilde{V})$ from the selection model, which is normalized to be uniformly distributed. Our results can be expressed in terms of a structural parameter, the so-called marginal treatment effect, $m(v) \equiv \mathbb{E}[Y_1 - Y_0 | V = v]$. It was introduced by Björklund and Moffitt (1987) and is the average treatment effect conditional on the selection threshold being equal to a certain value $v$. In many applications such as returns to schooling, we would expect this marginal treatment effect to be decreasing and possibly nonlinear in $v$ when we think of $v$ as being psychic or effort costs.

We show that under appropriate assumptions, the bias in estimates of the marginal treatment effect at $V = v$ is

$$\sigma^2/2 \cdot \frac{\partial^2 m(v)}{\partial v^2} + o(\sigma^2).$$

Here, $\partial^2 m(v)/\partial v^2$ can be replaced by an identified feature of the outcome equation without changing the order of the approximation error. Hence, a bias correction procedure is available if we have prior information about $\sigma$. In case no such prior information is available, a sensitivity analysis can still be undertaken by evaluating this expression at different values of $\sigma$. Moreover, according to our approximation, the bias is of order $o(\sigma^2)$ whenever $\partial^2 m(v)/\partial v^2$ is zero, independent of the presence of any unobserved heterogeneity $\sigma U$ in the population. Following up on this result, we show that if the marginal treatment effect is indeed linear in $V$, which is a testable condition on the data generating process, and if additional assumptions hold, estimates are unbiased. Arguably, this set of assumptions is more general than the assumption of constant treatment effects which underlies traditional instrumental variables analysis.

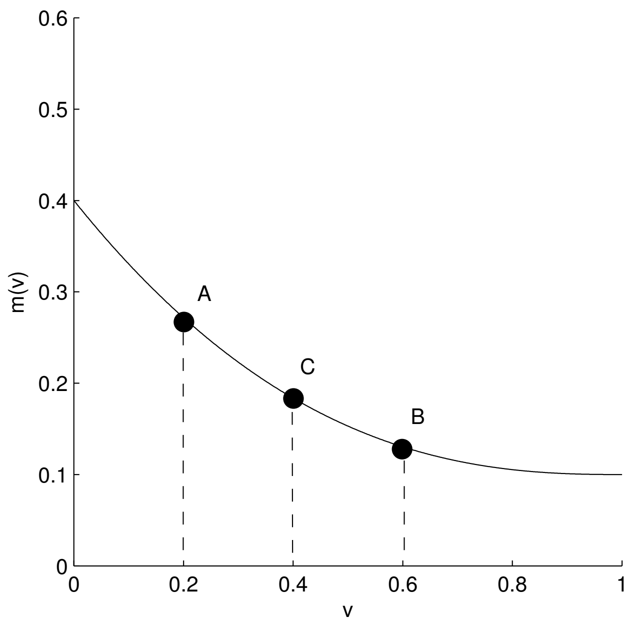


*Fig. 2.1:* Intuition behind the main result.

Even though the respective support conditions needed for identification differ LIV estimates of average treatment effect parameters as well as LATE estimates can be expressed in terms of the marginal treatment effect. Therefore, our findings translate directly into respective expressions for the bias in estimates of average treatment effect parameters in the context of LIV analysis and the local average treatment effect in the context of LATE analysis.

Example 2 is meant to give the intuition behind the main result.

**Example 2 (Intuition, Example 1 continued):** For the ease of the exposition let $F_{\tilde{V}}$ be known. According to our normalization $V = F_{\tilde{V}}(\tilde{V})$ is uniformly distributed, an equivalent representation of the index selection model in Example 1 is

$$D = \mathbb{I}\{F_{\tilde{V}}(Z\tilde{\gamma}) \geq V\},$$

and $\Pr(D = 1|Z) = \Pr\left(V \leq F_{\tilde{V}}(Z\tilde{\gamma})|Z\right) = \mathbb{E}\left[F_{\tilde{V}}(Z\tilde{\gamma})\middle|Z\right]$. Notice that, since only the left hand side of this equation is identified from observations, only $\mathbb{E}\left[F_{\tilde{V}}(Z\tilde{\gamma})\middle|Z\right]$ is known and in general, $F_{\tilde{V}}(Z\tilde{\gamma})$ is not. Suppose we wanted to identify $m(0.4)$. In Section 2.2 we will show that $m(0.4)$ can be estimated from observations for which $F_{\tilde{V}}(Z\tilde{\gamma})$ takes on values in a neighborhood around 0.4. Under monotonicity, $\sigma = 0$ and hence $F_{\tilde{V}}(Z\tilde{\gamma})$ is known because it is equal to $F_{\tilde{V}}(Z\gamma)$ which in turn is equal to $\Pr(D = 1|Z)$, recalling that $F_{\tilde{V}}$ is known. Consequently, $m(0.4)$ is identified since now, we can select observations for which $F_{\tilde{V}}(Z\tilde{\gamma}) = F_{\tilde{V}}(Z\gamma) = 0.4$. Now suppose $\sigma$ and $U$ are such that with respective probability one half either $F_{\tilde{V}}(Z\tilde{\gamma}) = 0.2$ or $F_{\tilde{V}}(Z\tilde{\gamma}) = 0.6$ whenever $\mathbb{E}\left[F_{\tilde{V}}(Z\tilde{\gamma})\middle|Z\right] = 0.4$. Then, monotonicity is violated and from observations with $\mathbb{E}\left[F_{\tilde{V}}(Z\tilde{\gamma})\middle|Z\right]$ taking on values around 0.4 we would estimate $0.5m(0.2) + 0.5m(0.6)$, the convex combination of points $A$ and $B$ in Figure 2.1. This corresponds to point $C$ only if the marginal treatment effect is linear in $V$. Only then, our estimate of $m(0.4)$ would still be unbiased even if monotonicity failed to hold. □

Section 2.2 lays out the formal framework. Section 2.3 contains the main theoretical results. We illustrate these findings and assess the accuracy of our approximation to the bias term in a Monte Carlo study which is carried out in Section 2.4. Section 2.5 concludes.

## 2.2 Formal Framework and Identification

We adopt the usual convention in program evaluation and say that if a unit is not treated, we observe an indicator variable $D$ being equal to zero and a realization of $Y_0$ and if it is treated, we observe $D$ being equal to one and a realization of $Y_1$. Usually, $Y_0$ and $Y_1$ are referred to as potential outcomes. They are real valued scalar random variables. We write $Y \equiv (1 - D)Y_0 + DY_1$. Our analysis can be thought of as being conditional on exogenous covariates as, e.g., in Vytlacil (2002).

As we have argued in the introduction we focus on the class of models in which identifying power is derived from monotonicity of the treatment decision in the instruments. This is expressed in terms of the selection model

$$D = \mathbb{1}\{\nu(Z) \geq \tilde{V}\} \tag{2.1}$$

with $\nu : \mathbb{R}^k \to \mathbb{R}$ being a function of the $k$-vector of instrumental variables. Stochastic restrictions and regularity conditions are given below.

Assumption 1 (Existence of Instruments): *$Z$ is independent of* $(Y_0, Y_1, \tilde{V})$.

Assumption 2 (Regularity Conditions I): *(i) $Y_0$ and $Y_1$ have finite first moments and (ii) the distribution of $\tilde{V}$ is absolutely continuous with respect to Lebesgue measure.*

Assumption 1 presumes that the instruments can be excluded from the outcome equation. This is a considerably weaker condition than conditional independence in matching, for example. Assumption 2 (i) will ensure that all structural parameters of interest are well defined. By Assumption 2 (ii) we can, w.l.o.g., normalize $\tilde{V}$ to be uniformly distributed.

Under this normalization, (2.1) becomes

$$D = \mathbb{1}\{P \geq V\} \tag{2.2}$$

where $P$ is shorthand notation for the propensity score $P(Z) \equiv \Pr(D = 1|Z)$, the probability to observe an individual to be treated conditional on the vector of instruments. Under Assumption 1 this probability is identified from observations.

### *2.2.1 Structural Parameters of Interest*

The effect of a deviation from monotonicity on a variety of structural parameters of interest can be expressed in terms of the marginal treatment effect

$$m(v) \equiv \mathbb{E}[Y_1 - Y_0|V = v]. \tag{2.3}$$

The marginal treatment effect by itself is of economic interest in many applications.[5] In this paper, we also focus on the bias in estimates of the population average treatment effect

$$\Delta^{\text{ATE}} \equiv \mathbb{E}[Y_1 - Y_0] = \int_0^1 m(v)\, dv \tag{2.4}$$

and the local average treatment effect

$$\Delta^{\text{LATE}}(v_l, v_h) \equiv \mathbb{E}[Y_1 - Y_0|v_l \leq V \leq v_h] = \frac{1}{v_h - v_l} \int_{v_l}^{v_h} m(v)\, dv, \tag{2.5}$$

where $v_l < v_h$.[6] However, our results extend easily to other average treatment effect parameters of interest such as the population average treatment effect on the treated, for example.

Note that by Assumption 2 all parameters that are considered here exist.

[5] See HV for a detailed discussion. For empirical studies of the returns to college education see Björklund and Moffitt (1987), Carneiro, Heckman, and Vytlacil (2005), Carneiro and Lee (2005), and Klein (2006). In this context, $V$ has the interpretation of unobserved ability which both has an impact on the decision of whether to attend college and the return from doing so. The dependence of this return on unobserved ability is of central interest to policy makers.

[6] Angrist, Graddy, and Imbens (2000) derive the marginal treatment effect as the limit form of the local average treatment effect and show that, conversely, the local average treatment effect is an average marginal treatment effect, though not the population average.

### *2.2.2 Identification of Structural Parameters*

In this subsection we briefly review and connect the identification results by HV and Imbens and Angrist (1994). We first turn to the former.

We call $p$ a limit point of the support of $P$ if $P$ has a continuous density in a neighborhood around $p$ which is bounded away from zero. Note that at $P = p$ derivatives of differentiable functions of $P$ are identified from observations.

Given Assumption 1 and 2 the marginal treatment effect is identified at $V = p$ if $p$ is a limit point of the support of $P$. To see this write

$$\mathbb{E}[Y|P = p] = \mathbb{E}[Y_0] + \int_0^p m(v)\, dv, \tag{2.6}$$

where the integral is equal to

$$p \cdot \mathbb{E}[Y_1 - Y_0|D = 1, P = p] = p \cdot \mathbb{E}[Y_1 - Y_0|V \leq p] = p \cdot \int_0^p m(v)/p\, dv$$

noting that the density of $V$ conditional on $V \leq p$ is $1/p$. Note also that this representation heavily draws on the monotonicity in the selection model.

$\mathbb{E}[Y|P = p]$ is differentiable with respect to $p$ since, by Assumption 2 (i), $m$ is integrable with respect to $V$. Differentiating both sides of (2.6) with respect to the propensity score yields

$$\frac{\partial \mathbb{E}[Y|P = p]}{\partial p} = m(p) \tag{2.7}$$

by Leibnitz' rule. The left hand side is identified from observations at limit points $p$ so that the marginal treatment effect is identified at $V = p$, the desired result.

The local average treatment effect can be identified from observations under weaker support conditions. Specifically, let $w$ and $z$ be two points of support of $P$ with $p_l < p_h$. Imbens and Angrist (1994) show that under Assumption 1 and 2 and the monotonicity in the selection model

$$\frac{\mathbb{E}[Y|P = p_h] - \mathbb{E}[Y|P = p_l]}{p_h - p_l} = \Delta^{\text{LATE}}(p_l, p_h). \tag{2.8}$$

Taking limits for $p_l \to p_h$ shows that (2.8) directly corresponds to (2.7).

HV emphasize that the local average treatment effect can always be expressed as a function of the marginal treatment effect as well. They show that

$$\Delta^{\text{LATE}}(p_l, p_h) = \frac{1}{p_h - p_l} \int_{p_l}^{p_h} m(v)\, dv.$$

Finally, note that therefore, there are at least two alternative ways for identifying the average treatment effect in this framework. First, if all $p \in (0, 1)$ are limit points of the support of $P$ the marginal treatment effect is identified at all values of $V$ in the open interval $(0, 1)$ and the average treatment effect can be obtained via (2.4) noting that the probability of $V$ being either 0 or 1 is equal to zero and first moments are finite. This is the approach suggested by LIV analysis. Alternatively, if 0 and 1 are in the support of $P$ we can make use of the fact that the average treatment effect is equal to $\Delta^{\text{LATE}}(0, 1)$, which is identified in this case. This is the approach suggested by LATE analysis.

A similar argument holds for the local average treatment effect. Both approaches differ with respect to the requirements on the support of the propensity score but share their dependence on monotonicity. Therefore, we shall refer to respective estimates as monotonicity estimates.

## 2.3 The Impact of a Deviation From the Monotonicity Assumption

In this section we study the impact of local departures from monotonicity on monotonicity estimates of structural parameters that can be expressed in terms of the marginal treatment effect. First, local departures will be motivated and formalized in terms of a generalization of the selection model. Then, we study the impact of such local departures on monotonicity estimates of the marginal treatment effect. Finally, we use these results to derive expressions for the bias in monotonicity estimates of the average and local average treatment effect.

### 2.3.1 Generalized Selection Model

Instead of (2.2) let

$$D = \mathbb{I}\{Q(P, \sigma U) \geq V\}, \tag{2.9}$$

$Q : \mathbb{R}^2 \to \mathbb{R}$, with

$$\mathbb{E}[Q(P, \sigma U)|P] = P$$

and $\sigma \geq 0$.[7]

We shall refer to $Q(P, \sigma U)$ as the individual shape function in the remainder and assume that it is a nontrivial function of $\sigma U$, where $U$ is a continuously distributed non-degenerate scalar random variable. $\sigma U$ reflects, e.g., differences in

[7] We stick to our normalization that $V$ is uniformly distributed on the unit interval. We consider it most instructive to think of $V$ as being precisely the same random variable as before. Of course, (2.2) is a normalized version of (2.1) and (2.9) is a normalized version of the model $D = \mathbb{I}\{\mu(\nu(Z), \sigma U) \geq \tilde{V}\}$ from the introduction. An example of such a model would be the model from Example 1 with normally distributed $\tilde{V}$, a random coefficient probit model.

preferences or risk attitudes which let, through the individual shape function, the impact of observables $Z$ on $D$ differ across individuals. This additional randomness causes monotonicity to fail. Besides, $U$ is assumed to be purely random.

Assumption 3 (Random Noise): *$U$ is independent of $Z$ and $(Y_0, Y_1, V)$ and has probability density function $f_U$.*

If $\sigma > 0$ the individual shape function is a finer measure of the treatment probability than $P$. In particular, by $V$ being uniformly distributed independent of $(Z, U)$,

$$Q(P, \sigma U) \equiv \Pr(D = 1|P, \sigma U).$$

### *2.3.2 Impact of Local Departures from Montonicity*

In this subsection we study the impact of local departures from monotonicity on estimates of structural parameters that are based on the monotonicity assumption. These impacts can be expressed in terms of the marginal treatment effect. The generalized selection model that was developed above is central to this analysis. In particular, monotonicity holds if $\sigma = 0$ and a local departure from monotonicity is given by a change from $\sigma = 0$ to a small $\sigma > 0$.

We derive an approximation to the bias term by performing a second order Taylor series expansion in $\sigma$ about $\sigma = 0$. This yields an approximation to the biased estimate of the marginal treatment effect at $V = p$ which we compare to an approximation to the true value.

In particular, if $\sigma = 0$, the representation for $\mathbb{E}[Y|P = p]$ in (2.6) is still valid because the canonical selection model in (2.9) is equivalent to the generalized selection model in (2.2). Therefore, if $\sigma = 0$ the marginal treatment effect is identified at $V = p$. If $\sigma > 0$, identification of the marginal treatment effect fails because the representation in (2.6) is no longer valid. A representation in which we replace the propensity score by the individual shape function is not feasible since the value of the individual shape function is not identified from observations because (2.9) is not invertible in $V$.

The approximation will be derived under the following two assumptions. As for notation, partial derivatives of a function $f(a)$ with respect to its argument evaluated at $a = 0$ are denoted by $\partial f(0)/\partial a$. Second and third partial derivatives as well as cross derivatives are denoted accordingly.

Assumption 4 (Regularity Conditions II): *(i) $Q(p, \sigma u)$ and $\partial Q(p, \sigma u)/\partial p$ are continuously differentiable in $\sigma u$ up to the third order around $\sigma u = 0$ and (ii) $m(v)$ is three times continuously differentiable.*

ASSUMPTION 5 (Pure Noise): *We have (i)* $\frac{\partial^2 Q(p,0)}{\partial(\sigma u)\, \partial p} = 0$ *and (ii)* $\frac{\partial Q(p,0)}{\partial(\sigma u)} = 1$.

Assumption 4 is a regularity condition. The first part of Assumption 5 postulates that the impact of infinitesimal changes in $\sigma U$ on the value of the individual shape function at $\sigma U = 0$ do not depend on the level of the propensity score. The second part is a normalization. Furthermore, we normalize $U$ so that $\mathbb{E}[U] = 0$ and $\mathbb{E}[U^2] = 1$. Then, $\sigma$ is the standard deviation of the random component $\sigma U$ and the first order approximation to the standard deviation of the variance of $Q(P, \sigma U)$ conditional on $P$. The main result is summarized in the following theorem and its two corollaries.

THEOREM 1: *Let $v$ be a limit point of the support of $P$. Then, under Assumptions 1-5 the bias in estimates of $m(v)$ is given by*

$$B^{MTE*} = \sigma^2/2 \cdot \frac{\partial^2 m(v)}{\partial v^2} + o(\sigma^2).$$

*Proof.* Appendix. □

COROLLARY 1.1: *Let Assumptions 1-5 hold. The bias in estimates of the average treatment effect is equal to*

$$B^{ATE*} = \sigma^2/2 \cdot \left( \left.\frac{\partial m(v)}{\partial v}\right|_{v=1} - \left.\frac{\partial m(v)}{\partial v}\right|_{v=0} \right) + o(\sigma^2)$$

*provided that the support of $P$ is equal to the open unit interval or contains the points 0 and 1. The bias in estimates of the local average treatment effect between $v_l$ and $v_h$ is equal to*

$$B^{LATE*} = \sigma^2/2 \cdot \left( \left.\frac{\partial m(v)}{\partial v}\right|_{v=v_h} - \left.\frac{\partial m(v)}{\partial v}\right|_{v=v_l} \right) + o(\sigma^2)$$

*provided that the support of $P$ is equal to the open interval $(v_l, v_h)$ or contains the points $v_l$ and $v_h$.*

*Proof.* Appendix. □

COROLLARY 1.2: *Under Assumptions 1-5, if $\sigma^2$ is known, a bias correction procedure in which we substitute biased estimates for $B^{MTE*}$, $B^{ATE*}$, and $B^{LATE*}$ is feasible in the sense that the order of the approximation error remains unchanged.*

Corollary 1.2 follows from the bias term being a multiple of $\sigma^2$. This allows us to use biased estimates of the marginal treatment effect and its derivatives to estimate the bias term without changing the order of the approximation error. If $\sigma^2$ is unknown, a sensitivity analysis can still be undertaken in which we calculate the approximation to the bias term for different values of $\sigma$.

A particularly interesting result is that the bias is of order $o(\sigma^2)$ if the marginal treatment effect is linear in $V = v$, i.e. if $\partial^2 m(v)/\partial v^2 = 0$. This result is independent of $\sigma$. This is a testable condition on the data generating process because it implies that $\mathbb{E}[Y|P = p]$ is quadratic in $p$.

In fact, this observation yields identifying conditions not relying on monotonicity which are weaker than the ones used in classical instrumental variables analysis. We summarize this finding in a theorem.[8]

Theorem 2: *Let the marginal treatment effect be linear in $v$ so that*

$$\mathbb{E}[Y|Q(P, \sigma U) = q] = \alpha + \beta q + \gamma q^2$$

*for some constants $\alpha, \beta, \gamma$. Moreover, let Assumptions 1-2 hold and let the variance of $Q(P, \sigma U)$ conditional on $P$ be equal to $\tilde{\sigma}$. Then, the marginal, average, and local average treatment effect are identified if the the support of the propensity score contains at least three points.*

*Proof.* We have

$$\begin{aligned}
\mathbb{E}[Y|P = p] &= \alpha + \beta\mathbb{E}[Q(P, \sigma U)|P = p| + \gamma\mathbb{E}[Q(P, \sigma U)^2|P = p| \\
&= \alpha + \beta p + \gamma \left(\tilde{\sigma} + p^2\right) \\
&= \tilde{\alpha} + \beta p + \gamma p^2,
\end{aligned}$$

where $\tilde{\alpha} = \alpha + \gamma\tilde{\sigma}$. Thus, $\beta$ and $\gamma$ are identified from observations. Consequently, the marginal treatment effect,

$$m(v) = \beta + 2\gamma v,$$

is identified at all $v$, and therefore, the average and local average treatment effect are identified as well. □

To summarize, our analysis has shown that the curvature of the marginal treatment effect in $V$ determines the magnitude of the bias when monotonicity does not hold. As a rule of thumb, we have that the less curved the marginal treatment

[8] As before, we can think of the exposition here as being conditional on exogenous covariates.

effect is in the value of the selection threshold $V$, the less biased monotonicity estimates are when monotonicity does not hold.

The relationship between the curvature of the marginal treatment effect and the bias of the monotonicity estimator as well as the accuracy of our approximation are illustrated in the following Monte Carlo study.

## 2.4 Monte Carlo

We simulated data in order to characterize the bias that arises from a violation of the monotonicity assumption as well as the accuracy of our analytical approximation to the bias term. For $R = 1.000$ repetitions and values of a curvature parameter $\rho$ we generated $N = 2.000$ data points. Since biases in estimates of average treatment effect parameters are functions of biases in estimates of the marginal treatment effect we focus on the respective mean bias in estimates of $m(0.5)$ as a function of the curvature $\rho$.

Specifically, we drew values of $P$ and $V$ from a uniform distribution, respectively, and let $Q = P + \sigma U$, where $\sigma = 0.2$ and being $U$ drawn from a uniform distribution with support $[-1, 1]$.[9] Next, we calculated $D = \mathbb{1}\{Q \geq V\}$. Notice that by construction, monotonicity of the treatment decision in $Q$ holds whereas monotonicity in $P$ is violated. In the spirit of the empirical results in Carneiro, Heckman, and Vytlacil (2005) we let

$$Y = 2.2 + 0.5Q - \frac{(1-Q)^{\rho+1}}{\rho+1} + \varepsilon,$$

where $\varepsilon$ was drawn from a normal distribution with mean 0 and variance 0.04. In their application, the treatment decision is whether to attend college or not. For $\rho = 2$ our simulations yield data similar to theirs. The marginal treatment effect at $V = q$ is given by the derivative with respect to $Q$, evaluated at $Q = q$:

$$m(q) = 0.5 + 1.5(1-q)^{\rho}.$$

It is decreasing in $q$. The second and third derivative are equal to $\partial m(q)/\partial q = -1.5\rho(1-q)^{\rho-1}$ and $\partial^2 m(q)/\partial q^2 = 1.5\rho(\rho-1)(1-q)^{\rho-2}$, respectively. Observe that for $\rho = 1$ the marginal treatment effect is linear in $q$ whereas for $\rho > 1$ it is a convex function of $q$. For low values of $q$ , e.g. $q = 0.3$ this function is the more convex the higher $\rho$.

[9] The values of the individual shape function can be interpreted as approximating a probability measure. Simulating them in a way such that they lie between zero and one turned out to impose a number of additional restrictions on the relationship between $P$ and $Q$. Hence, we decided to pursue this *ad hoc* but clear way to generate them.

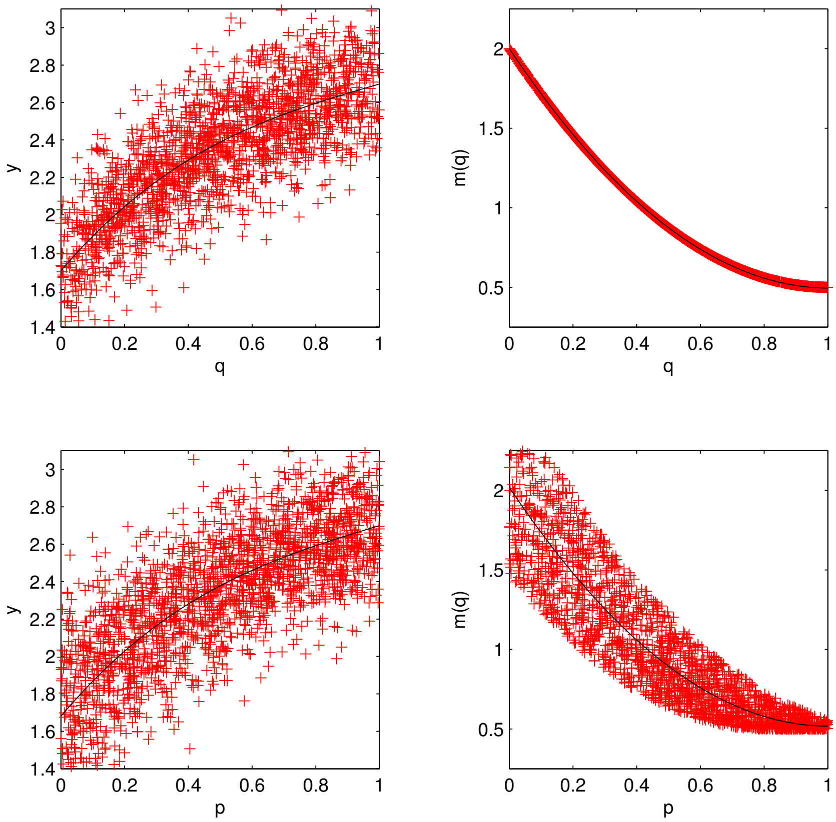


*Fig. 2.2:* One draw of generated data for $\rho = 2$.

For this Monte Carlo, estimates of the marginal treatment effect as well as its first and second derivative were obtained by fitting a third order polynomial to the data.

Figure 2.2 shows one draw of generated data for $\rho = 2$. Solid lines are estimated means over all repetitions. On the left, $y$ is plotted against $p$ and $q$. On the right, the marginal treatment effect, $m(q)$, is plotted against $p$ and $q$. Obviously, when we plot $m(q)$ against $q$ we get the marginal treatment effect itself. However, plotting $m(q)$ against perturbed values $p$ of $q$ yields a distribution of marginal treatment effects for every $p$.

Figure 2.3 shows the distribution of estimates of $m(0.5)$ that are based on $P$ and $Q$. Monotonicity of $D$ with respect to $Q$ holds by construction, whereas monotonicity of $D$ with respect to $P$ is violated. The figure shows that estimates based on $P$ are in general upward biased and more dispersed.

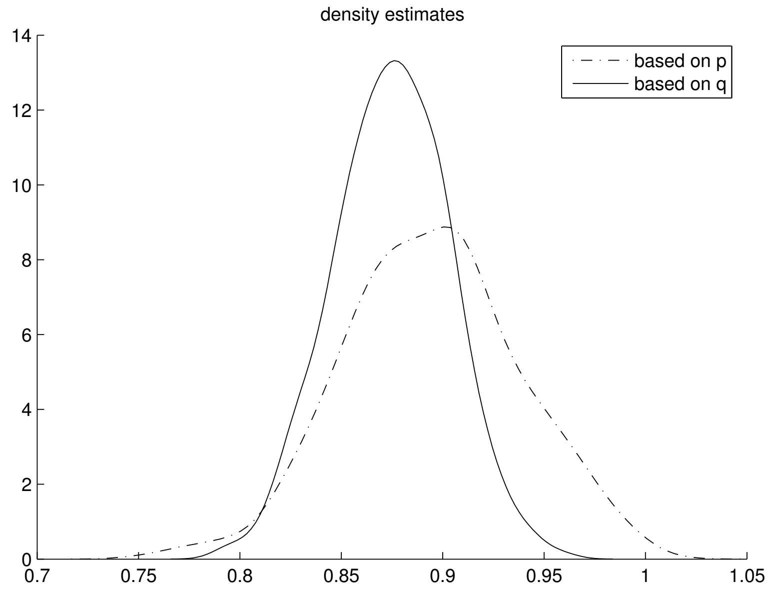


*Fig. 2.3:* Distribution of estimates of $m(0.5)$ based on $P$ and $Q$. $\rho = 2$. The true value is 0.875.

In Section 2.3 we have shown that the bias which arises from a violation of the monotonicity assumption is the higher the more convex the marginal treatment effect is in $v$. Next, in Figure 2.4, we plot the dependence of the mean bias in estimates of the marginal treatment effect at $V = 0.5$ against $\rho$. Additionally, we plot the analytical approximation to the bias term, $\hat{\sigma}^2/2 \cdot \partial^2 m(0.5)/\partial v^2$, and the estimated approximation, $\hat{\sigma}^2/2 \cdot \widehat{\partial^2 m(0.5)}/\partial v^2$, where $\hat{\sigma}^2$ is the sample variance of $\sigma U$ and $\widehat{\partial^2 m(0.5)}/\partial v^2$ is the estimate of the second derivative of the marginal treatment effect at $V = 0.5$. It shows that here, the approximation is reasonably accurate.

In general, our finding that the bias is the lower the more linear the marginal treatment effect is in $v$ seems to be confirmed in this Monte Carlo.

## 2.5 Concluding Remarks

In this paper, we have investigated the sensitivity of estimates of treatment effect parameters that hinge on monotonicity to violations of this assumption. An analytical approximation to the bias term has been derived. In general, we find that estimates are the more sensitive to violations of monotonicity the more curvature the marginal treatment effect exhibits in $V$. This analytical result was illustrated

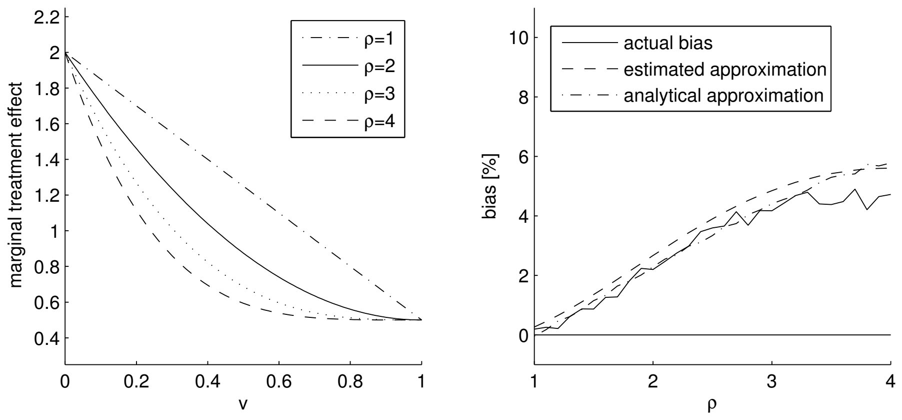


*Fig. 2.4:* The marginal treatment effect for different values of $\rho$ on the left and the bias as well as the accuracy of the approximation as a function of $\rho$, for $V = 0.5$, on the right.

in a Monte Carlo study.

Moreover, the bias can be estimated from the data up to a parameter $\sigma$ without changing the order of the approximation error. Therefore, our results have practical relevance which we summarize in the following three points. First, a bias correction procedure is available if there is prior knowledge on the variance of the individual shape function conditional on the propensity score. Second, there is an easy to implement way for applied researchers to qualitatively assess how sensitive their respective results are to likely deviations from monotonicity. In particular, a sensitivity analysis can be carried out by calculation the bias for different values of $\sigma$. Finally, our analysis reveals that whenever the marginal treatment effect is linear in $V$, the bias is of order $o(\sigma^2)$.

It is well known that traditional instrumental variables analysis hinges on the assumption of constant marginal treatment effects. This constitutes a special case of the marginal treatment effect being linear in $V$. In Theorem 2, we have shown that the assumption of a linear marginal treatment effect together with a conditional variance restriction is a generalization of traditional instrumental variables analysis which does *neither* rely on monotonicity as opposed to LIV analysis in its original formulation, *nor* does it require a strong conditional independence assumption to hold.

## *Appendix: Proofs*

Observe that by the regularity conditions in Assumption 2 all structural parameters considered here exist. Moreover, observe that a second order Taylor series expansion of the expressions considered below can be performed by the differentiability of the individual shape function and the marginal treatment effect that was introduced in Assumption 4.

We prove Theorem 1 using Lemma 1 and Lemma 2.

Lemma 1: *Under Assumptions 1-4,*

$$
\begin{aligned}
&\frac{\partial}{\partial p}\mathbb{E}[Y|P=p] \\
&\quad= \frac{\partial}{\partial Q(p,0)}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)] \\
&\qquad+\sigma^2/2\cdot\frac{\partial^3}{\partial Q(p,0)^3}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\left(\frac{\partial Q(p,0)}{\partial(\sigma u)}\right)^2 \\
&\qquad+\sigma^2/2\cdot\frac{\partial^2}{\partial Q(p,0)^2}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2} \\
&\qquad+\sigma^2\cdot\frac{\partial^2}{\partial Q(p,0)^2}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\frac{\partial Q(p,0)}{\partial(\sigma u)}\cdot\frac{\partial^2 Q(p,0)}{\partial(\sigma u)\,\partial p}+o(\sigma^2).
\end{aligned}
$$

*Proof.* The proof is in 5 steps.

*First step:* A second order Taylor series expansion of $Q(p,\sigma u)$ in $\sigma$ about $\sigma = 0$ yields

$$Q(p,\sigma u) = Q(p,0) + \sigma u\cdot\frac{\partial Q(p,0)}{\partial(\sigma u)} + (\sigma u)^2/2\cdot\frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2). \tag{2.10}$$

*Second step:* By the definition of the propensity score, Assumption 3, and the normalizations on $U$,

(2.11)

$$p = \mathbb{E}[Q(P,\sigma U)|P=p] = \mathbb{E}[Q(p,\sigma U)] = Q(p,0) + \sigma^2/2\cdot\frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2).$$

Combining this with (2.10) yields

$$Q(p,\sigma u) = p + \sigma u\cdot\frac{\partial Q(p,0)}{\partial(\sigma u)} + \sigma^2/2\cdot(u^2-1)\cdot\frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \tag{2.12}$$

and

$$\frac{\partial}{\partial p}Q(p,\sigma u) = 1 + \sigma u\cdot\frac{\partial^2 Q(p,0)}{\partial(\sigma u)\,\partial p} + \sigma^2/2\cdot(u^2-1)\frac{\partial^3 Q(p,0)}{\partial(\sigma u)^2\,\partial p} + o(\sigma^2). \tag{2.13}$$

*Third step:* A second order Taylor series expansion of $\partial\mathbb{E}[Y|Q(P,\sigma U) = Q(p,\sigma u)]/\partial Q(p,\sigma u)$ in $\sigma$ about $\sigma = 0$ yields

(2.14)
$$\begin{aligned}
&\frac{\partial}{\partial Q(p,\sigma u)}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,\sigma u)]\\
&= \frac{\partial}{\partial Q(p,0)}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)]\\
&\quad + \frac{\partial^2}{\partial Q(p,0)^2}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)] \cdot \frac{\partial Q(p,0)}{\partial(\sigma u)} \cdot \sigma u\\
&\quad + \frac{\partial^3}{\partial Q(p,0)^3}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)] \cdot \left(\frac{\partial Q(p,0)}{\partial(\sigma u)}\right)^2 \cdot (\sigma u)^2/2\\
&\quad + \frac{\partial^2}{\partial Q(p,0)^2}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)] \cdot \frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2} \cdot (\sigma u)^2/2 + o(\sigma^2).
\end{aligned}$$

*Fourth step:* We have

$$\begin{aligned}
&\frac{\partial}{\partial p}\mathbb{E}[Y|P = p]\\
&= \frac{\partial}{\partial p}\mathbb{E}[\mathbb{E}[Y|Q(P,\sigma U)]|P = p]\\
&= \frac{\partial}{\partial p}\int \mathbb{E}[Y|Q(P,\sigma U) = Q(p,\sigma u)]\; f_U(u)\; du\\
&= \int \frac{\partial}{\partial p}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,\sigma u)]\; f_U(u)\; du\\
&= \int \frac{\partial}{\partial Q(p,\sigma u)}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,\sigma u)] \cdot \frac{\partial}{\partial p}Q(p,\sigma u)\; f_U(u)\; du\\
&= \int \frac{\partial}{\partial Q(p,\sigma u)}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,\sigma u)] \cdot \left\{1 + \sigma u \cdot \frac{\partial^2 Q(p,0)}{\partial(\sigma u)\,\partial p}\right.\\
&\qquad \left. + \sigma^2/2 \cdot (u^2 - 1) \cdot \frac{\partial^3 Q(p,0)}{\partial(\sigma u)^2\,\partial p}\right\}\; f_U(u)\; du + o(\sigma^2),
\end{aligned}$$

where the first equality is by iterated expectations, the second follows from Assumption 3, the third from the integrand being finite, the fourth applies the chain rule, and the fifth uses (2.13).

Together with (2.14), this is

$$
\begin{aligned}
&\frac{\partial}{\partial p}\mathbb{E}[Y|P=p] \\
&= \int \Bigg\{ \frac{\partial}{\partial Q(p,0)}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)] \\
&\qquad + \frac{\partial^2}{\partial Q(p,0)^2}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\frac{\partial Q(p,0)}{\partial(\sigma u)}\cdot\sigma u \\
&\qquad + \frac{\partial^3}{\partial Q(p,0)^3}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\left(\frac{\partial Q(p,0)}{\partial(\sigma u)}\right)^2\cdot(\sigma u)^2/2 \\
&\qquad + \frac{\partial^2}{\partial Q(p,0)^2}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2}\cdot(\sigma u)^2/2 \Bigg\} \\
&\quad \times \left\{ 1+\sigma u\cdot\frac{\partial^2 Q(p,0)}{\partial(\sigma u)\,\partial p} + \sigma^2/2\cdot(u^2-1)\cdot\frac{\partial^3 Q(p,0)}{\partial(\sigma u)^2\,\partial p} \right\} \\
&\quad \times f_U(u)\,du + o(\sigma^2)
\end{aligned}
$$

and this in turn is

$$
\begin{aligned}
&= \int \Bigg\{ \frac{\partial}{\partial Q(p,0)}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)] \\
&\qquad + \frac{\partial^2}{\partial Q(p,0)^2}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\frac{\partial Q(p,0)}{\partial(\sigma u)}\cdot\sigma u \\
&\qquad + \frac{\partial^3}{\partial Q(p,0)^3}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\left(\frac{\partial Q(p,0)}{\partial(\sigma u)}\right)^2\cdot(\sigma u)^2/2 \\
&\qquad + \frac{\partial^2}{\partial Q(p,0)^2}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2}\cdot(\sigma u)^2/2 \\
&\qquad + \frac{\partial}{\partial Q(p,0)}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\sigma u\cdot\frac{\partial^2 Q(p,0)}{\partial(\sigma u)\,\partial p} \\
&\qquad + \frac{\partial^2}{\partial Q(p,0)^2}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\frac{\partial Q(p,0)}{\partial(\sigma u)}\cdot\sigma u\cdot\sigma u\cdot\frac{\partial^2 Q(p,0)}{\partial(\sigma u)\,\partial p} \\
&\qquad + \frac{\partial}{\partial Q(p,0)}\mathbb{E}[Y|Q(P,\sigma U)=Q(p,0)]\cdot\sigma^2/2\cdot(u^2-1)\cdot\frac{\partial^3 Q(p,0)}{\partial(\sigma u)^2\,\partial p} \Bigg\} \\
&\quad \times f_U(u)\,du + o(\sigma^2),
\end{aligned}
$$

where we already let multiples of $\sigma^2$ enter the remainder term. Using the normalizations on $U$, $\mathbb{E}[U] = 0$ and $\mathbb{E}[U^2] = 1$, this is

$$\begin{aligned}
&\frac{\partial}{\partial p}\mathbb{E}[Y|P = p] \\
&\quad = \frac{\partial}{\partial Q(p,0)}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)] \\
&\qquad + \sigma^2/2 \cdot \frac{\partial^3}{\partial Q(p,0)^3}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)] \cdot \left(\frac{\partial Q(p,0)}{\partial(\sigma u)}\right)^2 \\
&\qquad + \sigma^2/2 \cdot \frac{\partial^2}{\partial Q(p,0)^2}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)] \cdot \frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2} \\
&\qquad + \sigma^2 \cdot \frac{\partial^2}{\partial Q(p,0)^2}\mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)] \cdot \frac{\partial Q(p,0)}{\partial(\sigma u)} \cdot \frac{\partial^2 Q(p,0)}{\partial(\sigma u)\,\partial p} + o(\sigma^2).
\end{aligned}$$

□

LEMMA 2: *Under Assumptions 1-4,*

$$\begin{aligned}
&\frac{\partial}{\partial p}\mathbb{E}[Y|Q(P,\sigma U) = p] \\
&\quad = \frac{\partial\mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)]}{\partial Q(p,0)} \\
&\qquad + \frac{\partial^2\mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)]}{\partial Q(p,0)^2} \cdot \frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2} \cdot \sigma^2/2 + o(\sigma^2).
\end{aligned}$$

*Proof.* A second order Taylor series expansion of $\partial\mathbb{E}[Y|Q(P,\sigma U) = p]/\partial p$ in $\sigma$

about $\sigma = 0$, using (2.11), yields

$$
\begin{aligned}
&\frac{\partial \mathbb{E}[Y|Q(P,\sigma U) = p]}{\partial p} \\
&= \left\{ \frac{\partial \mathbb{E}\left[ Y \middle| Q(P,\sigma U) = Q(p,0) + \sigma^2/2 \cdot \frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \right]}{\partial \left( Q(p,0) + \sigma^2/2 \cdot \frac{\partial Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \right)} \right\} \Bigg|_{\sigma=0} \\
&\quad + \left\{ \frac{\partial^2 \mathbb{E}\left[ Y|Q(P,\sigma U) = Q(p,0) + \sigma^2/2 \cdot \frac{\partial Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \right]}{\partial \left( Q(p,0) + \sigma^2/2 \cdot \frac{\partial Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \right)^2} \right. \\
&\qquad \left. \times \frac{\partial \left( Q(p,0) + \sigma^2/2 \cdot \frac{\partial Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \right)}{\partial \sigma} \right\} \Bigg|_{\sigma=0} \cdot \sigma \\
&\quad + \left\{ \frac{\partial^3 \mathbb{E}\left[ Y|Q(P,\sigma U) = Q(p,0) + \sigma^2/2 \cdot \frac{\partial Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \right]}{\partial \left( Q(p,0) + \sigma^2/2 \cdot \frac{\partial Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \right)^3} \right. \\
&\qquad \left. \times \left( \frac{\partial \left( Q(p,0) + \sigma^2/2 \cdot \frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \right)}{\partial \sigma} \right)^2 \right\} \Bigg|_{\sigma=0} \cdot \sigma^2/2 \\
&\quad + \left\{ \frac{\partial^2 \mathbb{E}\left[ Y|Q(P,\sigma U) = Q(p,0) + \sigma^2/2 \cdot \frac{\partial Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \right]}{\partial \left( Q(p,0) + \sigma^2/2 \cdot \frac{\partial Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \right)^2} \right. \\
&\qquad \left. \times \frac{\partial^2 \left( Q(p,0) + \sigma^2/2 \cdot \frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2} + o(\sigma^2) \right)}{\partial \sigma^2} \right\} \Bigg|_{\sigma=0} \cdot \sigma^2/2 \\
&\quad + o(\sigma^2) \\
&= \frac{\partial \mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)]}{\partial Q(p,0)} \\
&\quad + \frac{\partial^2 \mathbb{E}[Y|Q(P,\sigma U) = Q(p,0)]}{\partial Q(p,0)^2} \cdot \frac{\partial^2 Q(p,0)}{\partial(\sigma u)^2} \cdot \sigma^2/2 + o(\sigma^2).
\end{aligned}
$$

□

*Proof of Theorem 1.* Follows directly from Lemma 1, Lemma 2, and Assumption 5. □

*Proof of Corollary 1.1.* The proof for the average treatment effect is similar to the proof for the local average treatment effect. We present the former. Since

$$\Delta^{\text{ATE}} \equiv \mathbb{E}[Y_1 - Y_0] = \int_0^1 m(v)\, dv$$

in order to obtain the bias in the average treatment effect, we have to integrate over the bias in estimates of $m(v)$ for values of $v$ from 0 to 1. This yields

$$\begin{aligned} B^{\text{ATE}*} &= \int_0^1 \sigma^2/2 \cdot \frac{\partial^2 m(v)}{\partial v^2}\, dv + o(\sigma^2) \\ &= \sigma^2/2 \cdot \left( \left.\frac{\partial m(v)}{\partial v}\right|_{v=1} - \left.\frac{\partial m(v)}{\partial v}\right|_{v=0} \right) + o(\sigma^2). \end{aligned}$$

Observe that this proof holds for both LATE and LIV analysis since the LATE estimator can be written as an average over LIV estimates. □

# BIBLIOGRAPHY

Angrist, J. D., K. Graddy, and G. W. Imbens (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67(3), 499–527.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Society*, 91(434), 444–455.

Battistin, E., and A. Chesher (2004): "The Impact of Measurement Error on Evaluation Methods Based on Strong Ignorability," Mimeograph.

Björklund, A., and R. Moffitt (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, 69(1), 42–49.

Carneiro, P., J. J. Heckman, and E. Vytlacil (2005): "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education," Mimeograph.

Carneiro, P., and S. Lee (2005): "Ability, Sorting and Wage Inequality," c*e*mmap Working Paper CWP 16/05, University College London.

Chesher, A. (1991): "The effect of measurement error," *Biometrika*, 78(3), 451–462.

Chesher, A., and J. M. C. Santos Silva (2002): "Taste Variation in Discrete Choice Models," *Review of Economic Studies*, 69(1), 147–168.

Chesher, A., and C. Schluter (2002): "Welfare Measurement and Measurement Error," *Review of Economic Studies*, 69(2), 357–378.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Kiefer, N., and G. R. Skoog (1984): "Local Asymptotic Specification Error Analysis," *Econometrica*, 52(4), 873–886.

KLEIN, T. J. (2006): "College Education and Wages in the U.K.: Estimating Conditional Average Structural Functions in Nonadditive Models with Binary Endogenous Variables," JEPS Working Paper 06-001.

VYTLACIL, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70(1), 331–341.

——— (forthcoming): "A Note on Additive Separability and Latent Index Models of Binary Choice: Representation Results," *Oxford Bulletin of Economics and Statistics*.

# 3. COLLEGE EDUCATION AND WAGES IN THE U.K.: ESTIMATING CONDITIONAL AVERAGE STRUCTURAL FUNCTIONS IN NONADDITIVE MODELS WITH BINARY ENDOGENOUS VARIABLES

## 3.1 Introduction

In many econometric applications, the characterization of the impact of binary variables on an outcome variable is of central interest. Examples are the impact of an additional year of schooling, or attending college, on wages, or the impact of participation in a labor market program on unemployment duration. Often, conditional on observable covariates, these effects are considered to be heterogeneous across individuals and possibly correlated with the binary choice variable. This is true if the choice is based upon knowledge of the outcome that is superior to what is observed in the data.

In this paper, we model the log of individual wages, which we denote by $Y$, by a correlated random coefficient model of the form

$$Y = X'\varphi(D, U, V) \tag{3.1}$$

$$D = \mathbb{1}\{P(Z) \geq V\}, \tag{3.2}$$

where $D$ is the binary endogenous variable. It is equal to one if the individual graduated from college. $X$ is a $K$-vector of observable covariates in the wage equation, (3.1). In our application, we exploit a uniquely rich birth cohort data set, the British National Child Development Survey (NCDS), and include in $X$, among other variables, the type of secondary school that was attended, the social class of the parents, as well as other family background variables and accurately measured ability test scores at the age of 7 and 11. $Z$ is a vector of covariates in the selection equation, (3.2), and includes the variables in $X$ as well as the father's interest in the education of the child for which we assume that it can be excluded from the wage equation. As we will see below, such an exclusion restriction is not necessary for identification in our model but yields additional identifying power. $X$ and $Z$ include a constant as their respective first elements. $U$ is a vector and represents "luck", and $V$ is a scalar entering both the wage and the selection equation. It represents unobservable costs, benefits, and most importantly, talent and unobserved ability which we suppose to have an impact on both wages and the decision whether to attend college. Modelling this link is of economic interest and important once we aim at estimating the impact of changes in $D$ or $X$ on $Y$ from non-experimental field data. As for stochastic restrictions, we assume that $(U, V)$ are jointly independent of $(X, Z)$ and that $U$ is independent of $V$. This implies that $P(Z)$ is identified from observations.

In this model, *ceteris paribus* effects of changes in observables, $D$ and $X$, on wages depend on unobservables $U$ and $V$. In general, the model is not identified.[1]

[1] The model in (3.1) and (3.2) is nonadditive in the unobservables. Moreover, the vector $X$, in principle, could include approximating functions in a way such that the number of approximating functions grows with the sample size. Then, along with Newey (1997), (3.1) could be interpreted

However, we will show that under suitable conditions the expected level of wages, $Y$, for a given $D$, $X$, and $V$,[2]

$$\mathbb{E}[Y|D = d, X = x, V = v] = x'\mathbb{E}[\varphi(d, U, v)]$$

is identified from observations. We will refer to this identifiable feature of the wage equation as the *conditional average structural function* (CASF) and to $\mathbb{E}[\varphi(d, U, v)]$ as the vector of conditional average *ceteris paribus* effects, understanding the notion of *ceteris paribus* as holding all other factors constant, including $V$, averaging only over $U$. We believe that not only average *ceteris paribus* effects, where we average over $V$ and $U$, are of interest but also their dependence on $V$. In Section 3.3, we define the parameters of interest more formally and link them to a variety of treatment effect parameters that are considered in the literature on program evaluation.

Heckman and Vytlacil (1999, 2000a, 2000b; HV in the remainder) establish nonparametric identification of the CASF under weaker stochastic restrictions. They assume that $(U, V)$ are jointly independent of $Z$ conditional on $X$ instead of our assumption that $(U, V)$ are jointly independent of $(X, Z)$. However, they require conditions on the support of $P(Z)$ to hold conditional on $X$, whereas we require them to hold only unconditionally.

To illustrate this point, nonparametric estimation of the CASF as suggested by the identification result of HV is not feasible in our application because not only the expected level of wages needs to be estimated conditional on $X = x$, $D = d$, and $P(Z) = p$, but also its partial derivative with respect to $p$ which requires continuous variation of $P(Z)$ conditional on $D$ and $X$, a requirement that is not met in our data because there is no such continuous variable in $Z$. This shows that there is a tradeoff between flexibility of the model—the model by HV is fully nonparametric—and data requirements.

Carneiro, Heckman, and Vytlacil (2005) and Carneiro and Lee (2005) propose what we shall refer to as the additive model. Write $X$ as $(1, X'_{-1})'$. Then, instead of (3.1), which can be written as

$$Y = \varphi_1(D, U, V) + X'_{-1}\varphi_{-1}(D, U, V),$$

they consider a wage equation of the form

$$Y = \mu(D, U, V) + X'_{-1}\gamma(D, U)$$

as a series approximation of a general nonseparable structural equation $Y = g(X, D, U, V)$. Together with (3.2) this is a triangular structure similar to the ones considered by Chesher (2003) and Imbens and Newey (2003). The key difference, however, is that here (3.2) is not invertible in $V$ and hence identification fails since $V$ enters as an argument. Chesher (2005) shows that in this case set identification may still be feasible.

[2] We will denote (vectors of) random variables by uppercase letters and their respective typical elements by lowercase letters.

and show that the CASF is identified under the same stochastic restrictions and support conditions that we use in this paper.

One limitation of their model is that they do not allow for the effect of $X$ on $Y$ to depend on $V$. This nonseparability is an important aspect of unobserved heterogeneity and is of economic interest in many applications with binary endogenous variables. Estimates for the more general model that is proposed in this paper indicate that these nonseparabilities are present in our data. For example, the expected effect of the parents' social class measured by the father's occupation depends on the level of unobserved ability, $V$. Moreover, we find that imposing separability results in biases which are significant.

We estimate the model by local linear smoothing. Our estimator is a *local instrumental variables estimator* built on the conventional two stage least squares instrumental variables (IV) estimator, except that we let the coefficients depend on the value of $P(Z)$. It turns out that it is substantially easier to implement than the estimator used for the additive model.[3]

Our results indicate that returns of attending college relative to obtaining just A-levels to be sizable. Moreover, we find evidence for heterogeneity of monetary returns. They are lower for individuals who actually attend college as compared to the returns for those who don't. This can be traced back to both observable and unobservable factors, and the interaction of the two. One finding is that returns are increasing in the father's years of education. Unlike other studies, we don't find clear cut evidence for sorting based on comparative or absolute advantage.

The paper is organized as follows. In Section 3.2, we embed our study into the literature. The full characterization of the econometric model, the identification result, and the proposed estimator are presented in Section 3.3. Section 3.4 contains the results from the empirical analysis. Section 3.5 concludes.

## *3.2 Related Results*

In this section, we briefly relate our model to the literature.[4] Furthermore, we discuss aspects of modelling unobserved heterogeneity, and most importantly ability.

In our application, we model two types of ability. The first consists of math and verbal ability test scores at the age of 7 and 11, which we include in our

[3] First applications are Carneiro, Heckman, and Vytlacil (2005), Carneiro and Lee (2005) and Heckman, Urzua, and Vytlacil (2004) where the model is estimated using a double residual regression involving several additional steps, see Robinson (1988).

[4] The question of how to estimate the returns to schooling and college education, which is closely related to the estimation of respective counterfactual wage levels, is one of the classical questions in econometrics. For two excellent surveys of the literature on the returns to schooling see Griliches (1977) and Card (2001). For an early survey on the returns to college education see Solmon and Taubman (1973).

set of covariates. The availability of this information is a key advantage of the NCDS since in many other data sets, e.g. the Family Expenditure Survey, the General Household Survey, or the Labor Force Survey, such precisely measured information is not available. Blundell, Dearden, and Sianesi (2005, BDS in the remainder) analyze the same data using OLS, IV, matching, and control function techniques. In Section 3.4, we compare our estimates of average returns to the ones of BDS.

The second type of ability is contained in $V$ which enters both the wage and the selection equation. In the statistics literature, $V$ is sometimes referred to as a confounding variable, see e.g. Fisher (1935, Ch. 7) and Yates (1937). For simplicity, we refert to $V$ as unobservable ability.

This is well in line with the economics literature, where the term "ability" is often used in different contexts and with different meanings. Griliches (1977, p. 7) defines it as "an unobserved latent variable that both drives people to get relatively more schooling and earn more income, given schooling, and perhaps also enables and motivates people to score better on various tests." Along those lines, Taubman and Wales (1972) and Taubman (1973) call it "mental ability" and Willis and Rosen (1979) use the expression "talent". On the other hand, Griliches (1977, p. 8) suggests that one could also interpret ability as "initial human capital". More broadly, Becker (1967) elaborates on whether there are several types of ability and Willis and Rosen (1979, p. S29) note that ability is potentially multi factoral.

In fact, in our model, $V$ is the projection of *all* unobservable factors that are common to the wage and the selection equation onto a scalar. Ashenfelter and Mooney (1968), Griliches and Mason (1972), Hansen, Weisbrod, and Scanlon (1970), Weisbrod and Karpoff (1968) and Leibowitz (1974) discuss and partly analyze the link between ability as well as other factors and earnings in more detail. Examples of such other factors include wealth, parent's income, status, social origin, motivation, quality of schooling, and idiosyncratic preferences. Here, we reach a natural limitation of our data since not all of those factors are observable. We proxy for some of these factors by including accurately measured family background variables, that are contained in the NCDS, in our set of covariates so that only the remaining variation is captured by $V$ if it is common to the wage and the selection equation, and $U$ if its only impact is on wages.

In general, econometric challenges arise from the fact that, *via* what we call unobserved ability, $V$, the return to schooling and college education is likely to be correlated with schooling and college choice once it results from optimizing behavior by economic agents who act on their knowledge of their ability. This gives rise to the classical selection problem in econometrics which could be overcome relatively easily if a perfect measure of ability was available, for example by including this measure into the set of regressors in the wage equation. Griliches (1977) discusses econometric consequences when an imperfect measure is used, i.e. when ability is measured with error. Along these lines, Chamberlain (1977) argues that it is instructive to think of unobserved ability as being a left-out variable.

Early contributions discussing the selection problem in detail include Heckman (1978), Heckman and Robb (1985, 1986) and Willis and Rosen (1979). A variety of approaches to this challenging problem has been taken over the last four decades. Identifying assumptions include parametric assumptions, as well as conditional (mean) independence and monotonicity in order to identify mean returns. Also, quantile invariance has proved to be a powerful identifying assumption.[5]

Most of these approaches rely on the presence of IVs that can be excluded from the earnings equation. IVs that have been used are quarter of birth (Angrist and Krueger 1991) and parental interest in education (BDS) as well as, e.g., the level of tuition fees, distance to college, and parental education, see Card (2001) for details. Angrist and Krueger (2001) advocate the use of natural experiments such as institutional changes as instruments giving rise to variation exogenous to the earnings equation. In our application, we derive *additional* identifying power from the assumption that the father's interest in the education of the child can be excluded from the wage equation.

The approach we take in this paper has several key advantages. First, we do not restrict selection to be based solely on observables which underlies OLS regressions, classical IV estimation, the random coefficient model suggested by Heckman and Vytlacil (1998), and matching.[6] Second, we do not have to spec-

[5] For distributional assumptions see, e.g., Heckman (1978), and Aakvik, Heckman, and Vytlacil (2005). Conditional independence is assumed in Rosenbaum and Rubin (1983). Heckman and Vytlacil (1998) exploit additivity of the error term in a random coefficient framework. Imbens and Angrist (1994), Angrist, Graddy, and Imbens (2000), HV, Carneiro, Heckman, and Vytlacil (2005), Heckman and Vytlacil (2005), and Abadie, Angrist, and Imbens (2002) exploit monotonicity, which is implied by the selection model. In Section 3.3, it will become clear that this is what we do in this paper as well. Quantile invariance is relied on in Chernozhukov, Imbens, and Newey (2004) and Chernozhukov and Hansen (2005). It is well beyond the scope of this paper to review the literature. However, the reader is referred to, e.g. Blundell and Powell (2003) for the relationship between IV and control function estimators, HV as well as Heckman and Vytlacil (2005) for the relationship between estimators based on monotonicity and classical IV estimators and OLS, and BDS for a comparison of OLS, IV, matching and control function estimators.

[6] These models assume that conditional on observables, $D$ is independent of either the effect from changes in $D$, or the error term in the outcome equation, or both. Garen (1984), Heckman (1978), Newey, Powell, and Vella (1999) as well as Pinske (2000) and Blundell and Powell (2003) pursue a control function approach. Imbens and Newey (2003) generalize this approach. Newey and Powell (2003), Darolles, Florens, and Renault (2003), and Das (2005) investigate the case in which the error term is additive. Notice that in our case identification is complicated by the fact that the endogenous variable is binary so that a control function approach in which we include the first stage residual into the second stage is not feasible because the selection equation is not invertible in $V$. It will become clear in Section 3.3 that the estimation step in our approach boils down to estimation of the expected outcome conditional on $D$, $X$, and $P(Z)$. Identification of the

ify the joint distribution of unobservables which underlies parametric approaches. Third, our model is nonparametric in the dimension of the unobserved heterogeneity since the dependence of the random coefficients on $D$ and $V$ is not constrained by functional form assumptions. Forth, as we have already discussed in the introduction, data requirements are weaker than in the fully nonparametric setup of HV, and equal to the ones of the additive model of Carneiro, Heckman, and Vytlacil (2005). At the same time, our model is more general in the sense that we allow both the random coefficient and $D$ to depend on $V$.

## 3.3 Econometric Approach

This section contains the formal results underlying our analysis in Section 3.4. Our point of departure is the correlated random coefficients model that was given in (3.1) and (3.2). We restate it for convenience:

$$Y = X'\varphi(D, U, V) \tag{3.1}$$

$$D = \mathbb{1}\{P(Z) \geq V\}. \tag{3.2}$$

(3.1) is the wage equation and (3.2) is the selection equation. We impose the following stochastic restrictions.

Assumption 1 (Stochastic Restrictions): (i) $(U, V)$ *are jointly independent of* $(X, Z)$ *and* (ii) $U$ *is independent of* $V$.

This allows $Z$ to contain variables also included in $X$ and *vice versa*. Assumption 1(i) requires the unobservables $(U, V)$ to be jointly independent of the observables $(X, Z)$. This is considerably weaker than the IV type assumption that $D$ is independent of the unobservables in the outcome equation conditional on $Z$ and $X$. Assumption 1(ii) restricts the randomness in $Y$ through $U$ to be completely random so that $U$ represents luck, whereas $V$ can be thought of as a confounding factor.[7]

parameters of interest is achieved by exploiting the monotonicity implied by the selection model.

[7] Assumption 1(ii) is not restrictive. $\varphi(D, U, V)$ is a nonparametric function of the observable $D$ and unobservables $(U, V)$. Therefore, it can at most be identified up to normalizations on the joint distribution of unobservables. Assume that the joint distribution of unobservables is absolutely continuous with respect to Lebesgue measure. Then, the restrictions on the joint distribution of observables imposed by any joint distribution of $(\widetilde{U}, \widetilde{V})$ are the same as the ones imposed by the joint distribution of $(U, V)$, where $v = F_{\widetilde{V}|\widetilde{U}}(\widetilde{v})$ with $V$ being uniformly distributed independently of $U$. For example, we could have $U = \widetilde{U}$ or any positive monotone transformation thereof. See also Imbens and Newey (2003) for a related discussion.

Apart from the stochastic restrictions we assume that the following regularity conditions hold.

Assumption 2 (Regularity Conditions): *(i) All first moments exist and (ii) the distribution of V is absolutely continuous with respect to Lebesgue measure.*

Assumption 2(i) ensures that all parameters of interest are well defined. Assumption 2(ii) implies that $V$ is a continuous random variable. This allows us, w.l.o.g., to normalize $V$ from now on to be uniformly distributed on the unit interval, see, e.g., Vytlacil (2002) for details. From Assumption 1(i) it follows immediately that $P(Z)$ is identified from observations since it is equal to $\Pr(D = 1|Z)$. For simplicity, we will write $P$ for $P(Z)$ in the remainder, with typical element $p$.

### 3.3.1 Parameters of Interest

We have already argued in the introduction that the CASF,

$$\mathbb{E}[Y|D = d, X = x, V = v] = x'\mathbb{E}[\varphi(d, U, v)]$$

is of special interest in our application. The terminology we use was introduced by Blundell and Powell (2003) who suggest to focus on the average structural function, $\mathbb{E}[Y|D = d, X = x]$. Likewise, Imbens and Newey (2003) call it the average conditional response. Following Goldberger (1972), who calls an equation *structural* if it represents a causal link rather than a mere empirical association, we prefer to think of the wage equation as a structural equation. We believe that for a given $D$ and $X$ the *dependence* of the average structural function on $V$ is of central economic interest by itself and hence focus on the average *conditional* on $V$.

A second object of interest that is related to the CASF is the conditional average *ceteris paribus* effect of changes in $X_k$, e.g. the type of secondary school that was attended or the social class of the father, for a given $D$, $X_{-k}$, and $V$,[8]

$$\frac{\partial\mathbb{E}[Y|D = d, X = x, V = v]}{\partial x_k} = \mathbb{E}[\varphi_k(d, U, v)].$$

Moreover, we are interested in the expected *ceteris paribus* effect of changes in $D$ for a given $X$ and $V$,

$$\begin{aligned}\mathbb{E}[Y|D = 1, X = x, V = v] - \mathbb{E}[Y|D = 0, X = x, V = v]\\ = x'(\mathbb{E}[\varphi(1, U, v)] - \mathbb{E}[\varphi(0, U, v)]).\end{aligned}$$

[8] The $k$th element of a vector $x$ is denoted by $x_k$. The remaining elements are denoted by $x_{-k}$. For discrete $X_k$ the partial derivative is replaced by an appropriate difference.

This is Björklund and Moffit's (1987) marginal treatment effect. It is the expected effect of a college degree on wages for a given level of unobserved ability and for a given vector of covariates. The well-known average treatment effect, averaged over the population distribution of unobserved ability, for a given $X = x$ is given by

$$x' \int_0^1 (\mathbb{E}[\varphi(1, U, v)] - \mathbb{E}[\varphi(0, U, v)])dv, \tag{3.3}$$

recalling that we have normalized $V$ to be uniformly distributed.

### *3.3.2 Identification*

In this subsection, we show identification of the CASF at a given $D$, $X$, and $V$ under Assumption 1 and 2. The estimator we implement, which is built on local linear smoothing, is proposed thereafter.

Because of the multiplicative structure of the wage equation, identification of the CASF at $D = d$, $V = v$, and any $X = x$ is equivalent to identification of the conditional average *ceteris paribus* effects. The average structural function as well as average *ceteris paribus* effects are identified at $D = d$ if the CASF is identified at all $V$ in the open unit interval, recalling that we have normalized its distribution to be uniformly distributed and that the endpoints have probability measure zero. Finally, if the (conditional) average structural function is identified at both $D = 0$ and $D = 1$, the average (marginal) treatment effect is as well.

From the model in (3.1), it follows that

$$\mathbb{E}[Y|D = 1, P = p, X = x] = x'\mathbb{E}[\varphi(1, U, V)|D = 1, P = p, X = x] \tag{3.4}$$

which is equal to

$$x'\mathbb{E}[\varphi(1, U, V)|P \geq V, P = p, X = x]$$

by the selection model in (3.2). But this is

$$x'\mathbb{E}[\varphi(1, U, V)|X = x, p \geq V].$$

By Assumption 1(i) we get that this is equal to

$$\mathbb{E}[x'\varphi(1, U, V)|p \geq V] = x'\mathbb{E}[\varphi(1, U, V)|p \geq V] =: x'\beta(1, p).$$

Note that $\mathbb{E}[\varphi(1, U, V)|p \geq V]$ is a function of $p$ which we will denote by $\beta(1, p)$ in the remainder. Since the left hand side of (3.4) is identified from observations at points of support $X = x$ and $P = p$, $\beta(1, p)$ is identified if we observe at least $K$ linearly independent values of $X$ for $D = 1$ (rank condition). $\beta(0, p)$ is defined accordingly and a similar result holds for $D = 0$.

Starting from this, we show that the CASF is identified. We state the result in a theorem which resembles Lemma 1 from Carneiro and Lee (2005). Following HV, they show nonparametric identification under weaker stochastic restrictions than the ones in Assumption 1, at the price of stronger support conditions that need to hold for their result. Only when they estimate the model they impose the restrictions in Assumption 1. We show the proof for two reasons. First, strictly speaking, our identification result is not implied by their Lemma 1, even though their proof is similar to ours. Second, our rank condition differs from theirs.

We call $p$ a limit point of the support of $P$, if $P$ has a continuous density in a neighborhood around $p$ which is bounded away from zero. Note that at $P = p$ derivatives of differentiable functions of $P$ are identified from observations.

Theorem 3 (Identification): *Assume that $\beta(0, p)$ and $\beta(1, p)$ are continuously differentiable with respect to $p$ and that we observe at least $K$ linearly independent realizations of $X$ for every $D$ and $P = p$ (rank condition). Then, under Assumptions 1 and 2 the CASF is identified at $V = p$, where $p$ is a limit point of the support of $P$, and given by*

$$x'\mathbb{E}[\varphi(0, U, p)] = x'\left(\beta(0, p) - (1 - p)\frac{\partial\beta(0, p)}{\partial p}\right)$$
$$x'\mathbb{E}[\varphi(1, U, p)] = x'\left(\beta(1, p) + p\frac{\partial\beta(1, p)}{\partial p}\right).$$

*Proof.* We prove identification of $\mathbb{E}[\varphi(D, U, V)|D = 1, V = p]$. The proof for $\mathbb{E}[\varphi(D, U, V)|D = 0, V = p]$ is similar. Recall that we have normalized $V$ to be uniformly distributed. By definition,

$$x'\mathbb{E}[\varphi(1, U, V)|p \geq V] = x'\beta(1, p).$$

From the normalization on $V$ and Assumption 1(ii) it follows that

$$x' \int_0^p \int_{-\infty}^{\infty} \varphi(1, u, v)\, \mu(du)\, dv/p = x'\beta(1, p), \tag{3.5}$$

where $\mu(du)$ is the marginal probability measure of $u$. Multiplying both sides by $p$ gives

$$x' \int_0^p \int_{-\infty}^{\infty} \varphi(1, u, v)\, \mu(du)\, dv = x'\beta(1, p)p$$

and differentiating both sides with respect to $p$ using Leibnitz' rule reveals that

$$x' \int_{-\infty}^{\infty} \varphi(1, u, p)\, \mu(du) = x'\beta(1, p) + px'\frac{\partial\beta(1, p)}{\partial p}.$$

If $p$ is a limit point of the support of $P$ both $\beta(1, p)$ and $\partial\beta(1, p)/\partial p$ are identified from observations at $P = p$. The left hand side is the object of interest. □

Finally, notice that the proof relies on the monotonicity of $D$ in $P$ implied by the selection model which allows us to formulate (3.5). See also Klein (2006) for a discussion and an analysis of the case in which monotonicity does not hold, but is wrongly assumed.

### *3.3.3 Estimation*

We have established in our discussion that from the model and the conditions of Theorem 1 it follows that

$$\mathbb{E}[Y|D = d, P = p, X = x] = x'\beta(d, p)\ ,\ \ d \in \{0, 1\},$$

where $\beta(d, p)$ is a coefficient vector with coefficient functions $\beta_k(d, p)$, $k = 1, \ldots, K$. Both depend on the observable $D$, and $P$ which is identified from observations. This is a version of the varying coefficient model which was suggested by Cleveland, Grosse, and Shyu (1991) and Hastie and Tibshirani (1993).

In a first step, we parametrically estimate $P$. For the second step we assume that the coefficient functions are bounded and have bounded second derivatives which allows us to estimate them by local linear smoothing. See, for example Fan and Zhang (1999) and Xia and Li (1999) for details as well as a proof of consistency and results on rates of convergence of the estimator. This estimation procedure is usually motivated by a Taylor expansion of the coefficient function in $\tilde{p}$ about $\tilde{p} = p$ which yields

$$\beta_k(d, \tilde{p}) = \beta_k(d, p) + \frac{\partial\beta_k(d, p)}{\partial p}(\tilde{p} - p) + \frac{1}{2}\frac{\partial^2\beta_k(d, \bar{p})}{\partial p^2}(\tilde{p} - p)^2,$$

where $\bar{p}$ is a point between $p$ and $\tilde{p}$. We select all observations with $D = d$ and index them by $i$, $i = 1, \ldots, n$. Our estimator of $\beta(d, p)$ and $\partial\beta(d, p)/\partial p$ is the solution of $a$ and $b$ to the following minimizer

$$\arg\min_{a,b} \left\{ \sum_{i=1}^{n} K\left(\frac{p_i - p}{h}\right) \cdot \left(y_i - \begin{bmatrix} x_i \\ (p_i - p) \cdot x_i \end{bmatrix}' \begin{pmatrix} a \\ b \end{pmatrix}\right)^2 \right\},$$

where $K(\cdot)$ is a kernel function with the usual properties and $h$ is the bandwidth. Since fitted values $p_i$ were parametrically estimated in a first step we do not expect them to have an impact on the distribution of the second step estimator in a first order asymptotic sense. However, we obtain confidence intervals, accounting for the first step estimation error, using a bootstrap procedure.

Estimates of the objects of interest can be obtained from these estimates of $\beta(d, p)$ and $\partial\beta(d, p)/\partial p$ using the formulas from Theorem 1.

## 3.4 College Education and Wages in the U.K.

### 3.4.1 Data

We implement the estimation procedure which was proposed in Section 3.3 for U.K. data from the NCDS. The NCDS is conducted by the Centre for Longitudinal Studies at the Institute of Education in London. It is a longitudinal data set and keeps detailed records for all those living in Great Britain who were born between March 3 and 9, 1958. The data were collected in 1958, in 1965 (when members were aged 7 years), in 1969 (age 11), in 1974 (age 16), in 1981 (age 23), in 1991 (age 33) and 1999-2000 (age 41-42). The NCDS has gathered data from respondents on child development from birth to early adolescence, child care, medical care, health, physical statistics, school readiness, home environment, educational progress, parental involvement, cognitive and social growth, family relationships, economic activity, income, training, and housing.

Recently, BDS study these data using IV estimation, a control function estimator, and matching techniques. For a more detailed data description and variable definitions the reader is referred to their paper.

Their, as well as our, outcome of interest is log hourly wages in 1991, this is at the age of 33. We select individuals who at least completed their A levels, from which 51.4% are higher education graduates. We say that an individual completes his A levels if he completed at least one A level which is generally obtained at the end of secondary school, see BDS for details. Notably, we distinguish between college graduates ($D = 1$), who have completed some kind of higher education, and those who have obtained A levels only ($D = 0$). We focus on employed males and select individuals with non-missing verbal and math ability test scores. This leaves us with 1501 observations.

The NCDS contains a host of accurately measured variables including information about the type of secondary school that was attended and a number of family background variables. In the U.K., secondary school is attended from the age of 11 to 12 on for 7 years. The individuals in our sample were born in 1958 so that they entered secondary school in the late 1960s. At that time the public school system was changing. Until then, there were two basic types of public secondary schools in the U.K., Secondary modern and Grammar schools. Secondary modern schools were intended for children who would be going into a trade and focussed on practical skills. Grammar schools were intended to prepare pupils for higher education. In the 1960s, comprehensive schools were promoted as an alternative and started to partly replace the old system providing complete and general education. But in fact, which route was pursued for the school system highly depended on the respective local authority. Nowadays, there is a mixture of types of public schools. Alongside public schools there are prestigious Private schools such as

| | no college (49%) | | college (51%) | |
|---|---|---|---|---|
| | mean | std. | mean | std. |
| LOG HOURLY WAGE AT THE AGE OF 33 | 2.04 | 0.40 | 2.32 | 0.37 |
| FATHER'S INTEREST IN THE EDUCATION OF THE CHILD AT THE AGE OF 7 | | | | |
| expects too much | 0.01 | 0.07 | 0.03 | 0.17 |
| very interested | 0.28 | 0.45 | 0.43 | 0.50 |
| some interest | 0.24 | 0.43 | 0.23 | 0.42 |
| MOTHER'S INTEREST IN THE EDUCATION OF THE CHILD AT THE AGE OF 7 | | | | |
| expects too much | 0.03 | 0.17 | 0.05 | 0.21 |
| very interested | 0.38 | 0.49 | 0.56 | 0.50 |
| some interest | 0.44 | 0.50 | 0.31 | 0.46 |
| ABILITY MEASURES | | | | |
| math ability at 7 | 55.17 | 23.97 | 66.40 | 21.53 |
| math ability at 11 | 57.39 | 16.06 | 67.25 | 14.26 |
| verbal ability at 7 | 80.15 | 20.74 | 90.45 | 13.32 |
| verbal ability at 11 | 59.52 | 20.52 | 72.05 | 16.98 |
| SECONDARY SCHOOL TYPE | | | | |
| Secondary Modern | 0.15 | 0.36 | 0.09 | 0.29 |
| Comprehensive school | 0.52 | 0.50 | 0.42 | 0.49 |
| Grammar | 0.08 | 0.28 | 0.21 | 0.41 |
| Private | 0.04 | 0.20 | 0.10 | 0.29 |
| other | 0.02 | 0.14 | 0.01 | 0.11 |
| missing school information | 0.19 | 0.39 | 0.17 | 0.37 |
| SOCIAL CLASS OF THE FATHER | | | | |
| professional | 0.03 | 0.17 | 0.10 | 0.30 |
| intermediate | 0.16 | 0.37 | 0.23 | 0.42 |
| skilled non-manual | 0.09 | 0.29 | 0.09 | 0.29 |
| skilled manual | 0.34 | 0.48 | 0.25 | 0.44 |
| semi-skilled non-manual | 0.01 | 0.09 | 0.01 | 0.07 |
| semi-skilled manual | 0.09 | 0.29 | 0.06 | 0.23 |
| unskilled | 0.19 | 0.39 | 0.18 | 0.39 |
| missing/unemployed/no father | 0.08 | 0.27 | 0.08 | 0.28 |
| FAMILY BACKGROUND VARIABLES WHEN THE CHILD WAS 16 | | | | |
| father's years of education | 7.84 | 4.32 | 8.35 | 5.11 |
| missing | 0.21 | 0.41 | 0.23 | 0.42 |
| mother's years of education | 8.04 | 4.17 | 8.20 | 4.65 |
| missing | 0.20 | 0.40 | 0.22 | 0.41 |
| father's age | 44.17 | 11.69 | 45.06 | 10.89 |
| missing | 0.05 | 0.22 | 0.04 | 0.19 |
| mother's age | 42.41 | 8.76 | 42.75 | 8.83 |
| missing | 0.02 | 0.15 | 0.03 | 0.16 |
| mother was employed | 0.58 | 0.49 | 0.55 | 0.50 |
| number of siblings | 1.70 | 1.57 | 1.47 | 1.41 |
| REGION WHEN THE CHILD WAS 16 | | | | |
| North Western | 0.10 | 0.30 | 0.10 | 0.30 |
| North | 0.07 | 0.26 | 0.08 | 0.27 |
| East and West Riding | 0.06 | 0.24 | 0.07 | 0.25 |
| North Midlands | 0.08 | 0.28 | 0.07 | 0.25 |
| Eastern | 0.07 | 0.26 | 0.09 | 0.28 |
| London and South East | 0.14 | 0.34 | 0.14 | 0.35 |
| Southern | 0.06 | 0.23 | 0.07 | 0.25 |
| South Western | 0.07 | 0.26 | 0.07 | 0.25 |
| Midlands | 0.09 | 0.28 | 0.07 | 0.26 |
| Wales | 0.06 | 0.23 | 0.06 | 0.23 |
| other | 0.20 | 0.40 | 0.19 | 0.39 |

*Tab. 3.1:* Summary Statistics.

Eton college, which are sometimes still referred to as "public schools" since they are open for the paying public as opposed to a religious school.[9]

In our analysis we proxy social class by the type of occupation of the father when the child was 16. Categories are professional, intermediate, skilled and semi-skilled non-manual as well as skilled or semi-skilled manual, and unskilled.

Table 3.1 contains summary statistics for our data. Notably, wages are higher for college graduates, and as compared to college non-graduates more college graduates (i) went to Grammar or Private school, (ii) have a father who is professional or intermediate, and (iii) have better educated parents on average.

### *3.4.2 First Stage Estimates*

The first stage of our two stage estimator consists of fitting values of $P$ by estimating a probit model. Our set of variables in the selection equation, $Z$, consists of the parent's interest in the education of the child, math and reading ability test scores at the age of 7 and 11, indicator variables for secondary school type, the father's social class when the child was 16, as well as other family background variables and, in some specifications, region. As was shown in Section 3.3. an exclusion restriction is not needed for identification of the CASF in our model—unlike for nonparametric identification as in HV.

Note that whereas the interpretation of the estimated probit coefficients as *ceteris paribus* effects heavily relies on the distributional assumptions in a probit model, the fitted values of the propensity score are less sensitive to violations of those assumptions once we interpret the usual probit model as a reduced form.[10] As suggested by the literature, and in order to undertake a sensitivity analysis, we estimated these fitted values by ordinary least squares, see Kelejian (1971) and the discussion in Angrist and Krueger (2001). However, our results did not change qualitatively.

Table 3.2 contains coefficient estimates for 5 different specifications. Throughout, the direction of the impact is as expected and in line with the literature which takes a closer look at the channels though which parents' education is transmitted to the children, see Goldberger (1989) and Haveman and Wolfe (1995) for an overview and discussions. Column (1) is the full specification in which indicator variables for region were included. Column (2) is the same specification except that secondary school type was left out because it could arguably be endogenous. This is the case whenever conditional on measured ability and all other controls in $Z$, those who know already that they will be more likely to go to college attain

[9] See, e.g., http://en.wikipedia.org/wiki/Education_in_the_United_Kingdom (February 2006).

[10] Willis and Rosen (1979) invoke a set of assumptions which allows them to estimate both a reduced form and a structural probit.

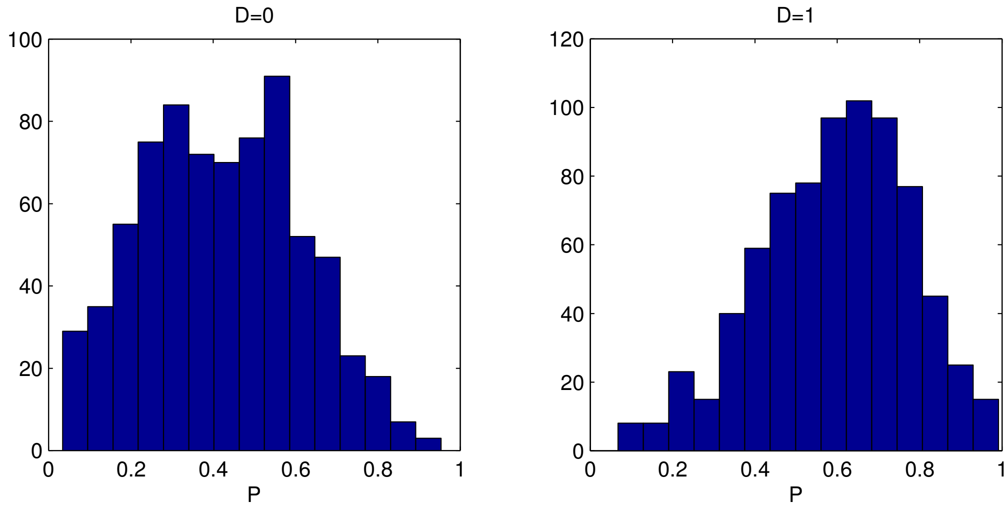


*Fig. 3.1:* Sample distribution of the propensity score conditional on $D$.

a special kind of secondary school, e.g. Grammar school. The remaining coefficients are largely unchanged. In the first two specifications, the region indicator variables were all insignificant. Column (3) and (4) contain estimates obtained from the specification in (1) and (2), respectively, except that these indicator variables were left out. Again, in comparison to the first two columns, the estimates remained largely unchanged. For our final specification in column (5) we left out the mother's interest in the education of the child since it is highly correlated with father's interest.

Moreover, we left out some of the insignificant indicator variables for secondary school type, social class and family background. Our estimates show throughout that parents' interest has a significant impact on the probability of attending college, so do the ability measures, whether the child went to Grammar school, and whether the father is professional.

Figure 3.1 shows the the sample distributions of the fitted values of $P$. For both $D = 0$ and $D = 1$ the support is almost equal to the full unit interval. Note that the distributions differ between $D = 0$ and $D = 1$. This shows that the variables in $Z$ have explanatory power.

### *3.4.3 Second Stage Implementation*

In the second stage, drawing on Section 3.3's results we estimate the mean coefficient functions, $\beta(d, p)$ and their derivatives with respect to $p$. For smoothing in the direction of $p$ we use an Epanechnikov kernel and estimated the coeffi-

| | (1) | | (2) | | (3) | | (4) | | (5) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | coeff. | *t*-stat. | coeff. | *t*-stat. | coeff. | *t*-stat. | coeff. | *t*-stat. | coeff. | *t*-stat. |
| FATHER'S INTEREST IN THE EDUCATION OF THE CHILD | | | | | | | | | | |
| expects too much | 1.12 | 3.09 | 1.11 | 3.11 | 1.13 | 3.15 | 1.13 | 3.18 | 1.23 | 3.72 |
| very interested | 0.13 | 1.16 | 0.12 | 1.10 | 0.12 | 1.11 | 0.11 | 1.04 | 0.27 | 3.17 |
| some interest | 0.26 | 2.73 | 0.25 | 2.61 | 0.26 | 2.68 | 0.24 | 2.57 | 0.22 | 2.47 |
| MOTHER'S INTEREST IN THE EDUCATION OF THE CHILD | | | | | | | | | | |
| expects too much | 0.16 | 0.70 | 0.15 | 0.65 | 0.17 | 0.74 | 0.15 | 0.69 | | |
| very interested | 0.20 | 1.44 | 0.21 | 1.57 | 0.22 | 1.59 | 0.23 | 1.70 | | |
| some interest | -0.02 | -0.13 | -0.01 | -0.05 | 0.00 | -0.01 | 0.01 | 0.05 | | |
| ABILITY MEASURES | | | | | | | | | | |
| math ability at 7 | 0.00 | 1.90 | 0.00 | 2.17 | 0.00 | 1.98 | 0.00 | 2.28 | 0.00 | 2.24 |
| math ability at 11 | 0.01 | 3.83 | 0.01 | 3.87 | 0.01 | 3.86 | 0.01 | 3.93 | 0.01 | 3.79 |
| verbal ability at 7 | 0.01 | 3.97 | 0.01 | 4.16 | 0.01 | 4.01 | 0.01 | 4.15 | 0.01 | 4.04 |
| verbal ability at 11 | 0.00 | 1.36 | 0.00 | 1.63 | 0.00 | 1.28 | 0.00 | 1.55 | 0.00 | 1.19 |
| SECONDARY SCHOOL TYPE RELATIVE TO COMPREHENSIVE SCHOOL | | | | | | | | | | |
| Secondary Modern | 0.01 | 0.09 | | | 0.00 | -0.02 | | | | |
| Grammar | 0.27 | 2.32 | | | 0.27 | 2.36 | | | 0.29 | 2.71 |
| Private | 0.11 | 0.71 | | | 0.11 | 0.72 | | | 0.12 | 0.78 |
| other | -0.33 | -1.17 | | | -0.32 | -1.15 | | | | |
| missing school information | -0.07 | -0.59 | | | -0.07 | -0.59 | | | | |
| SOCIAL CLASS OF THE FATHER RELATIVE TO UNSKILLED | | | | | | | | | | |
| professional | 0.57 | 2.63 | 0.62 | 2.84 | 0.58 | 2.65 | 0.62 | 2.86 | 0.45 | 2.56 |
| intermediate | 0.17 | 1.06 | 0.22 | 1.37 | 0.18 | 1.09 | 0.23 | 1.41 | 0.04 | 0.45 |
| skilled non-manual | 0.13 | 0.72 | 0.14 | 0.80 | 0.14 | 0.75 | 0.15 | 0.83 | | |
| skilled manual | 0.16 | 1.05 | 0.17 | 1.09 | 0.16 | 1.06 | 0.17 | 1.10 | | |
| semi-skilled non-manual | -0.01 | -0.03 | 0.01 | 0.02 | 0.03 | 0.07 | 0.06 | 0.13 | | |
| semi-skilled manual | 0.14 | 0.75 | 0.14 | 0.72 | 0.15 | 0.78 | 0.14 | 0.76 | | |
| missing/unemployed/no father | 0.14 | 0.63 | 0.04 | 0.23 | 0.07 | 0.37 | -0.05 | -0.31 | | |
| FAMILY BACKGROUND VARIABLES WHEN THE CHILD WAS 16 | | | | | | | | | | |
| father's years of education | 0.02 | 0.39 | 0.02 | 0.37 | 0.02 | 0.38 | 0.02 | 0.37 | 0.03 | 1.42 |
| missing | 0.79 | 2.12 | 0.83 | 2.26 | 0.78 | 2.11 | 0.83 | 2.26 | 0.36 | 1.45 |
| mother's years of education | 0.03 | 0.48 | 0.03 | 0.53 | 0.02 | 0.41 | 0.02 | 0.45 | | |
| missing | -0.47 | -1.21 | -0.43 | -1.11 | -0.48 | -1.25 | -0.44 | -1.15 | | |
| father's age | 0.00 | -0.10 | 0.00 | -0.11 | 0.00 | -0.21 | 0.00 | -0.22 | | |
| missing | -0.49 | -0.95 | -0.48 | -0.94 | -0.55 | -1.07 | -0.54 | -1.06 | | |
| mother's age | 0.02 | 1.19 | 0.01 | 1.11 | 0.02 | 1.23 | 0.01 | 1.14 | | |
| missing | 0.97 | 1.68 | 0.93 | 1.62 | 0.99 | 1.74 | 0.96 | 1.68 | | |
| mother was employed | -0.09 | -1.04 | -0.09 | -1.08 | -0.09 | -1.05 | -0.09 | -1.06 | -0.08 | -1.00 |
| number of siblings | -0.04 | -1.58 | -0.04 | -1.56 | -0.05 | -1.60 | -0.04 | -1.59 | -0.04 | -1.48 |
| interaction father's education x age | 0.00 | 0.65 | 0.00 | 0.78 | 0.00 | 0.64 | 0.00 | 0.75 | | |
| interaction mother's education x age | 0.00 | -0.98 | 0.00 | -0.99 | 0.00 | -0.94 | 0.00 | -0.93 | | |
| REGION WHEN THE CHILD WAS 16 | | | | | | | | | | |
| indicator variables | yes | | yes | | no | | no | | no | |
| CONSTANT | -3.34 | -5.69 | -3.48 | -6.03 | -3.21 | -5.62 | -3.32 | -5.94 | -2.59 | -8.40 |
| McFadden *R*-squared | 0.14 | | 0.14 | | 0.14 | | 0.14 | | 0.13 | |

*Tab. 3.2:* First stage probit coefficient estimates.

| | estimate | 95% conf. int. | |
|---|---|---|---|
| ATE population | 0.46 | 0.04 | 0.89 |
| ATE treated | 0.26 | -0.11 | 0.64 |
| ATE untreated | 0.63 | 0.03 | 1.22 |
| OLS | 0.21 | 0.17 | 0.25 |
| IV | 0.43 | 0.09 | 0.75 |
| BDS | 0.24 | 0.21 | 0.28 |
| additive | 0.40 | 0.05 | 0.74 |

*Tab. 3.3:* Comparison of the estimated average treatment effect (ATE) for different subpopulations to OLS and IV estimates as well as the BDS matching estimates, and the additive model of Carneiro and Lee (2005).

cient vectors at 101 grid points between 0 and 1. As we have seen, this is a one-dimensional nonparametric problem. The bandwidths were chosen using a standard leave-one-out cross validation procedure. It turns out that the optimal bandwidth for $D = 0$ is infinitely large, implying estimation of a fully interacted model without any smoothing. For $D = 1$ the optimal bandwidth is 1.7. The required rank conditions hold in our data, i.e. the weighted $n \times 2K$ matrix of explanatory variables and interaction terms is of rank $2K$ at all evaluation points $p$.

From these estimates, which we provide with hats in the remainder, we calculate the vector of conditional average *ceteris paribus* effects for a given $d$ and $v$, $\widehat{\mathbb{E}}[\varphi(d, U, v)]$, and the CASF, $x'\widehat{\mathbb{E}}[\varphi(d, U, v)]$ as well as other identifiable features of interest. In our bootstrap procedure for respective confidence intervals we acknowledge the fact that the propensity score is estimated in a first step by estimating it within every one of 1,000 bootstrap replications. For illustration, Figure 3.2 in the appendix contains estimates of the CASF and the marginal treatment effect for a representative individual with median characteristics. In particular, this representative individual went to comprehensive school, its father has 9 years of education and is neither professional nor intermediate, his mother is employed, and he has 1 sibling. Next, we go though the results in detail.

### 3.4.4 Average Returns to College Education

We calculate average returns using (3.3), replacing $x$ by respective population means. For the average treatment effect on the untreated and treated, we simulate the distribution of $V$ conditional on $D$ by exploiting the structure of the selection model. For example, if we observe an individual with $D = 0$ and $P = p$, we would draw values of $V$ from a uniform distribution on $(p, 1]$. Respective confidence intervals account for the simulation error. In Table 3.3, we compare our estimates to estimates obtained from an OLS regression, two stage IV estimates, as well as matching estimates obtained by BDS. The OLS estimate can be interpreted as the average difference in earnings observed in the population once we control for differences in covariates. This observed difference in earnings can be traced back to a selection effect and a causal effect of a higher education degree. Not surprisingly, the OLS estimate is very close to the matching estimate of BDS since matching is built on the assumption that conditional on observables, $D$ is independent of the error term in the outcome equation. As covariates in the wage equation we used the variables from the final specification in Table 3.2, except for the father's interest in the education of the child.

Commonly, the linear IV estimate is interpreted as estimating the average treatment effect of those who are induced to attend college by the variables that are excluded from the outcome equation, see, e.g. the discussion in BDS and Imbens and Angrist (1994) as well as Card (2001). In our specification, following BDS, we have excluded the father's interest in the education of the child from the outcome equation. The estimate obtained from the additive model is close to our estimate for the population.

Notably, we estimate the average treatment effect to be lower for those who actually attend college. This difference can partly be explained by differences in observables since the average treatment effect depends on those observables in our model.

In general, all estimates which are obtained from a two stage (ours, OLS, IV, additive) procedure are relatively imprecise. We suppose that this is due to the first stage estimation error which is carried over into the second stage.

### *3.4.5 Average* Ceteris Paribus *Effects*

Panel (1) in Table 3.4 contains estimates of average *ceteris paribus* effects and respective 95% confidence intervals. The set of covariates we included into the second step is the same as the one in the final specification for the first step, except that we leave out the father's interest in the education of the child. We also calculated estimates for alternative specifications, but the results did not change qualitatively.

The top rows contain estimates for $D = 0$ and the bottom rows for $D = 1$. Statistically significant determinants of wages are whether the father was professional, which resulted in a large wage increase both for $D = 0$ and even more so for $D = 1$ and the father's years of education, but only for $D = 0$. In general, our estimates were quite imprecise. Yet, as we have already seen above, this is also the case for the standard linear IV estimates. Therefore, we feel that this lack of precision is not a property of our estimation procedure, but a feature of our data.

| | (1) average *ceteris paribus* effect | | | (2) test for unobserved heterogeneity | | | (3) bias in estimates when additivity is imposed | | |
|---|---|---|---|---|---|---|---|---|---|
| | estimate | 95% conf. int. | | est. slope | 95% conf. int. | | bias | 95% conf. int. | |
| | | | | NO COLLEGE DEGREE | | | | | |
| ABILITY MEASURES | | | | | | | | | |
| math ability at 7 | 0.001 | -0.001 | 0.003 | 0.000 | -0.008 | 0.009 | 0.000 | -0.002 | 0.002 |
| math ability at 11 | 0.002 | -0.002 | 0.006 | 0.000 | -0.014 | 0.016 | 0.000 | -0.006 | 0.003 |
| verbal ability at 7 | 0.003 | -0.002 | 0.007 | 0.007 | -0.007 | 0.018 | -0.002 | -0.006 | 0.002 |
| verbal ability at 11 | 0.000 | -0.004 | 0.003 | -0.001 | -0.014 | 0.012 | 0.000 | -0.005 | 0.002 |
| SECONDARY SCHOOL TYPE RELATIVE TO COMPREHENSIVE SCHOOL | | | | | | | | | |
| Grammar | 0.021 | -0.230 | 0.275 | 0.099 | -1.063 | 1.267 | -0.003 | -0.293 | 0.210 |
| Private | 0.159 | -0.099 | 0.418 | 0.040 | -1.260 | 1.295 | -0.026 | -0.362 | 0.158 |
| SOCIAL CLASS OF THE FATHER RELATIVE TO UNSKILLED | | | | | | | | | |
| professional | 0.293 | 0.007 | 0.757 | -1.662 | -3.386 | -0.245 | -0.316 | -0.663 | 0.095 |
| intermediate | -0.022 | -0.116 | 0.087 | -0.224 | -0.680 | 0.257 | 0.017 | -0.130 | 0.074 |
| FAMILY BACKGROUND VARIABLES WHEN THE CHILD WAS 16 | | | | | | | | | |
| father's years of education | 0.034 | 0.007 | 0.064 | -0.008 | -0.145 | 0.114 | -0.010 | -0.041 | 0.016 |
| missing | 0.331 | 0.053 | 0.660 | 0.062 | -1.305 | 1.358 | -0.109 | -0.441 | 0.165 |
| mother was employed | 0.023 | -0.053 | 0.119 | 0.007 | -0.328 | 0.426 | -0.003 | -0.035 | 0.137 |
| number of siblings | -0.019 | -0.053 | 0.011 | -0.022 | -0.126 | 0.084 | 0.006 | -0.016 | 0.048 |
| CONSTANT | 1.269 | 0.566 | 2.014 | -0.318 | -1.970 | 1.368 | 0.253 | -0.542 | 0.901 |
| | | | | COLLEGE DEGREE | | | | | |
| ABILITY MEASURES | | | | | | | | | |
| math ability at 7 | -0.002 | -0.010 | 0.005 | -0.003 | -0.011 | 0.006 | 0.003 | -0.004 | 0.012 |
| math ability at 11 | 0.004 | -0.010 | 0.015 | 0.002 | -0.012 | 0.015 | -0.003 | -0.014 | 0.012 |
| verbal ability at 7 | -0.011 | -0.024 | 0.004 | -0.010 | -0.022 | 0.003 | 0.010 | -0.004 | 0.024 |
| verbal ability at 11 | -0.013 | -0.024 | 0.000 | -0.014 | -0.025 | 0.000 | 0.013 | -0.003 | 0.021 |
| SECONDARY SCHOOL TYPE RELATIVE TO COMPREHENSIVE SCHOOL | | | | | | | | | |
| Grammar | 0.238 | -0.396 | 0.758 | 0.303 | -0.464 | 0.932 | -0.259 | -0.752 | 0.399 |
| Private | 0.528 | -0.079 | 1.139 | 0.569 | -0.197 | 1.331 | -0.435 | -1.048 | 0.170 |
| SOCIAL CLASS OF THE FATHER RELATIVE TO UNSKILLED | | | | | | | | | |
| professional | 0.794 | 0.054 | 1.540 | 1.035 | 0.080 | 2.090 | -0.823 | -1.611 | -0.130 |
| intermediate | 0.174 | -0.286 | 0.633 | 0.152 | -0.364 | 0.686 | -0.188 | -0.721 | 0.200 |
| FAMILY BACKGROUND VARIABLES WHEN THE CHILD WAS 16 | | | | | | | | | |
| father's years of education | -0.006 | -0.093 | 0.063 | -0.014 | -0.124 | 0.071 | 0.015 | -0.046 | 0.110 |
| missing | 0.108 | -0.873 | 0.911 | 0.006 | -1.161 | 0.943 | -0.012 | -0.692 | 1.077 |
| mother was employed | -0.001 | -0.277 | 0.298 | -0.075 | -0.374 | 0.269 | 0.033 | -0.253 | 0.323 |
| number of siblings | -0.017 | -0.123 | 0.074 | 0.003 | -0.108 | 0.102 | -0.004 | -0.071 | 0.126 |
| CONSTANT | 4.126 | 2.578 | 5.583 | 1.998 | 0.474 | 3.147 | -2.078 | -3.821 | -0.773 |

*Tab. 3.4:* Average *ceteris paribus* effects, test for unobserved heterogeneity, and estimates for the bias from imposing additivity.

Panel (2) contains the result of a test for unobserved heterogeneity. We say that unobserved heterogeneity is present whenever the impact of a component of $X$, including the constant, depends on $V$. Therefore, the null hypothesis is that the derivative of the conditional average *ceteris paribus* effect with respect to $V$ is zero at all $V = v$. This implies that the linear approximation to the slope is zero. (2) contains estimated linear approximations to the slope of $\widehat{\mathbb{E}}[\varphi(d, U, v)]$, as well as bootstrapped confidence intervals. Notice that here, we face two sources of estimation error. First, the error that stems from estimating the conditional average *ceteris paribus* effect itself and second, the error from estimating the linear approximation to its slope. The presence of essential heterogeneity is significant at the 5% level if 0 lies outside the confidence interval. Using this test, we find evidence for essential heterogeneity in the impact of the father being professional for both $D = 0$ and $D = 1$ and overall for $D = 1$, *via* the constant term.[11]

This essential heterogeneity has the interpretation of a nonseparability between the effect of $X$ and $V$ on $Y$. It is a key advantage of the techniques developed in this paper to allow us to control for this nonseparability. In panel (3), we raise the question whether imposing the absence of this nonseparability, i.e. imposing the additive model of Carneiro, Heckman, and Vytlacil (2005) and Carneiro and Lee (2005), results in biases of average *ceteris paribus* effects. We report estimates of the bias that results from imposing separability. The estimates were obtained by comparing our estimator to a simple series estimator of the additive model $Y = \mu(D, U, V) + X'_{-1}\beta(D, U)$ in which the effect of $X$ on $Y$ is not allowed to depend on $V$. A cross validation yields that only a linear term in $P$ should be included into the regression of $Y$ on $X$ conditional on $D$ in order to calculate estimates of average *ceteris paribus* effects. Clearly, this proceeding is far less elaborate than the double-residual regression procedure that is carried out in, e.g., Carneiro, Heckman, and Vytlacil (2005) and Carneiro and Lee (2005). Therefore, we prefer to interpret our estimates of the biases only as rough estimates or first approximations. However, the results in panel (2) indicate already that the additive model is misspecified for our data so that it is not surprising that we estimate the bias to be significant for the impact of the father being professional and the constant term for $D = 1$.

### *3.4.6 Conditional Average* Ceteris Paribus *Effects and Sorting*

Figure 3.3 and 3.4 in the Appendix contain estimates of conditional average *ceteris paribus* effects. They show the respective dependence of the impact of co-

[11] Carneiro, Heckman, and Vytlacil (2005) and Heckman, Urzua, and Vytlacil (2004) suggest to test for essential heterogeneity by checking whether the expected value of wages given $P$ and $X$ is linear in $P$ by fitting polynomials to the data. Using this test, we were not able to reject the null of no essential heterogeneity for our data.

| | fraction | 95% conf. int. | |
|---|---|---|---|
| level, $D = 0$ | 0.62 | 0.04 | 0.93 |
| level, $D = 1$ | 0.65 | 0.32 | 0.96 |
| marginal treatment effect | 0.54 | 0.47 | 0.60 |

*Tab. 3.5:* Fractions of observations for which the CASF (level) and the marginal treatment effect is increasing in $V$. Linear approximations to the slope were calculated.

variates on wages as a function of $D$ and $V$. Notice that according to the selection model low values of $V$ induce individuals to attend college so that we should think of *low* values of $V$ as representing *high* unobservable ability. For example, whereas the impact of the father being professional on wages is increasing in unobservable ability for $D = 0$, it is decreasing for $D = 1$.

Since $X$ varies across individuals, it is helpful to take a closer look at the dependence of the marginal treatment effect on $V$ when $X$ varies across individuals. Carneiro and Lee (2004, footnote 3) point out that individuals base their selection into educational on their *comparative advantage* with respect to monetary benefits if the marginal treatment effect is higher for those individuals who go to college, i.e. if the marginal treatment effect is falling in $V$ conditional on observables $X$.[12] Variation in covariates induces variation in the slope of the marginal treatment effect. Therefore, we estimated a linear approximation to the slope of the marginal treatment effect for every individual.

Table 3.5 contains the fractions of the population for which, respectively, the slope of the CASF and the marginal treatment effect are positive. In order to obtain those numbers, linear approximations to the slope were estimated. The numbers indicate that the way wages depend on what we labelled unobserved ability, $V$, is nontrivial.

As for the slope of the levels, the slope is positive in about 60 per cent of the cases. A positive slope implies that

$$x'\mathbb{E}[\varphi(0, U, V)|D = 1] < x'\mathbb{E}[\varphi(0, U, V)|D = 0]$$
$$x'\mathbb{E}[\varphi(1, U, V)|D = 1] < x'\mathbb{E}[\varphi(1, U, V)|D = 0].$$

Hence, the numbers indicate that in about 60 per cent of the cases those who actually graduated from college earn less compared to what those, who did not

[12] See, e.g., Roy (1951) for the impact of selection of individuals based on their comparative advantage on the income distribution, Sattinger (1978) for an empirical study of respective comparative advantages of individuals in the performance of tasks, Willis and Rosen (1979) for a parametric study of the returns to college education in the presence of such selection, as well as Carneiro and Lee (2004) for a semiparametric analysis. Heckman and Sedlacek (1985, 1990) develop models of the sectoral allocation of workers based on comparative advantage.

graduate from college, would earn, had they been forced to do so. Conversely, those who did not go to college earn more than those who did go to college would have earned, had they been prevented from doing so. This is in line with our earlier finding that treatment effects are higher for college non-graduates compared to college graduates. However, notice that this is only an analysis of monetary benefits, neglecting the costs of attending college which could have been prohibitively high for those who did not in fact attend college.

Surprisingly, only for about 46 per cent of the individuals the slope of the marginal treatment effect is negative. Hence, the comparative advantage hypothesis is only supported for these 46 per cent of the individuals. For about 54 per cent of the individuals, the slope is positive. This is in contrast to the findings in previous studies including Willis and Rosen (1979) and Carneiro and Lee (2004). One explanation could be that both of these studies do not allow the effect of $V$ on wages to depend on $X$. In fact, as we have seen in Table 3.3, such estimates would be biased for our data.

We shall end with the conjecture that the comparative advantage hypothesis, which is a central concept in Economics, could well be reconciled with these findings once nonmonetary costs and benefits are included in the analysis. Just to give an example, it could well be that a college degree is associated with nonpecuniary benefits such as the pleasure of being educated which represent an *additional* return that has not been focussed on in this study. Clearly, such nonmonetary costs and benefits might again well be correlated with unobserved ability, family background, and social class. After all, we understand our results as evidence for nontrivial sorting patterns that are not solely based on monetary considerations. Therefore, we strongly believe that more research, and other data, are of need in order to better understand the sorting patterns into educational levels.

## *3.5 Concluding Remarks*

In this paper, we have proposed and implemented a semiparametric estimator for expected wage levels and their dependence on the endogenous schooling choice.

The virtue of our approach to the problem lies in dimensionality reduction along the dimension of the usually higher dimensional vector of exogenous covariates. Moreover, we are able to circumvent the problem of limited support of the propensity score given the vector of covariates since we require only conditions on the unconditional support of $P$ to hold. At the same time, we do not impose any limiting restrictions on the joint distribution of unobservables.

The estimator we propose is a two step version of a local linear regression estimator. The usefulness of our approach was shown in turn of the empirical analysis. In particular, our results suggest that differences in wages can be attributed to differences in observables in interaction with unobserved ability. In previous studies, e.g. by Carneiro, Heckman, and Vytlacil (2005) and Carneiro and Lee (2005), this complementarity between observables and unobservables was largely neglected for reasons of tractability. In this paper, we have suggested an estimation procedure which does allow for such effects on the one hand and which is easily implementable on the other.

The results of the empirical analysis are manyfold. First, we find that measured ability, social class, secondary school type, and family background have explanatory power for the decision to attend college. Second, with an estimate of 0.46 for the population, we find the monetary return to college education to be sizable with returns for college graduates being lower than for college non-graduates. Third, our estimates do not support the hypothesis of sorting into schooling based on comparative advantage with respect to the monetary returns. Forth and last, we find nonseparabilities between the impact of observables, e.g. whether the father is professional, and unobserved ability on wages and show that biases arise once an additive structure is imposed. We feel that this shows the usefulness of our approach.

## *Appendix*

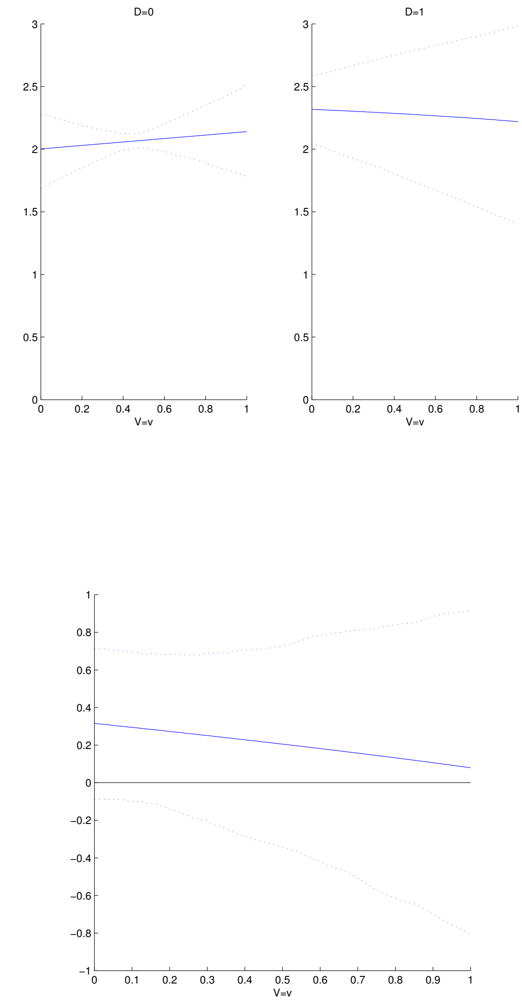


*Fig. 3.2:* Point estimates and bootstrapped 95% confidence intervals of the conditional average structural function (top) and the marginal treatment effect (bottom). Reported for a representative individual with median characteristics.

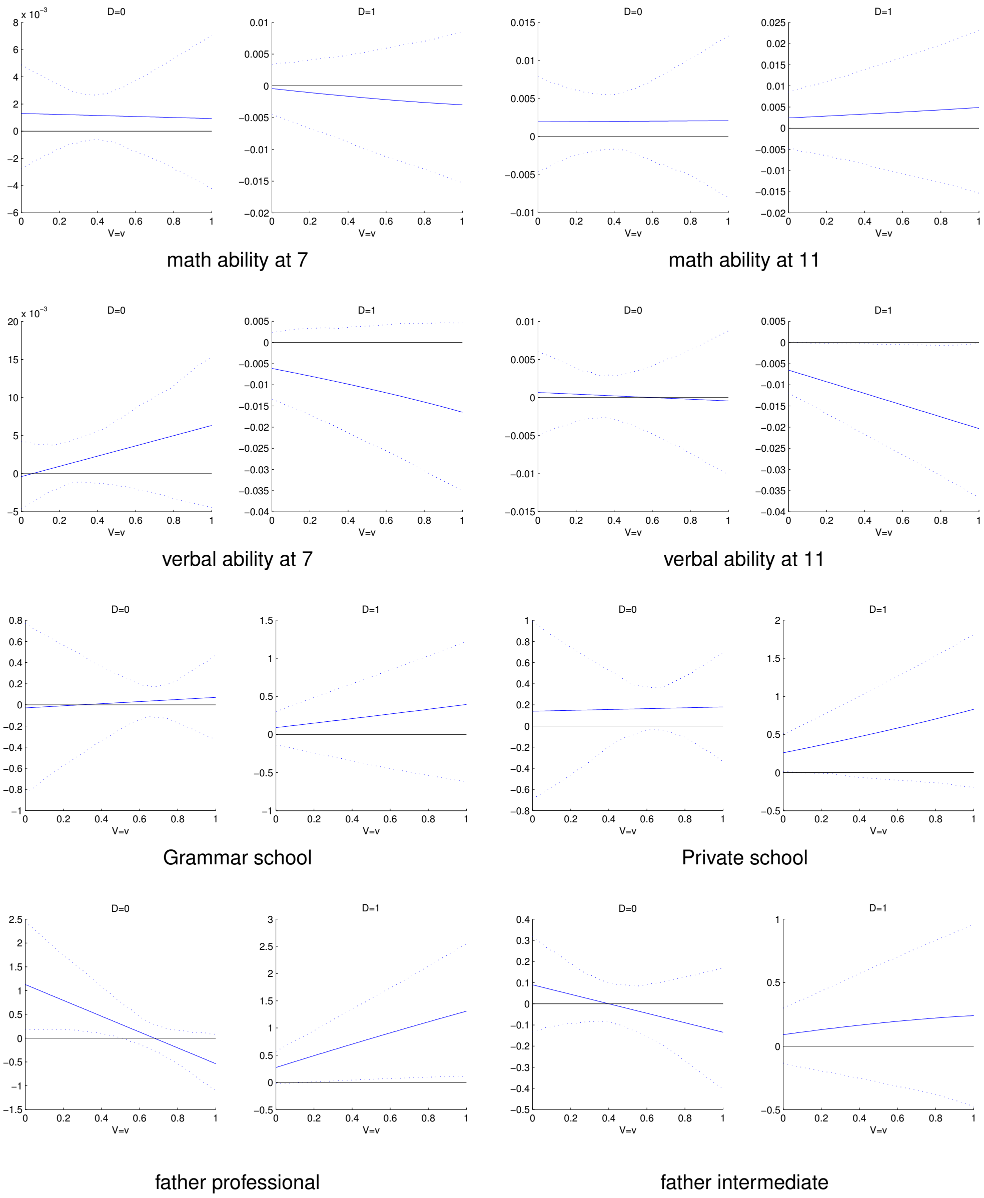


*Fig. 3.3:* Conditional average *ceteris paribus* effects 1/2. Point estimates and bootstrapped 95% confidence intervals.

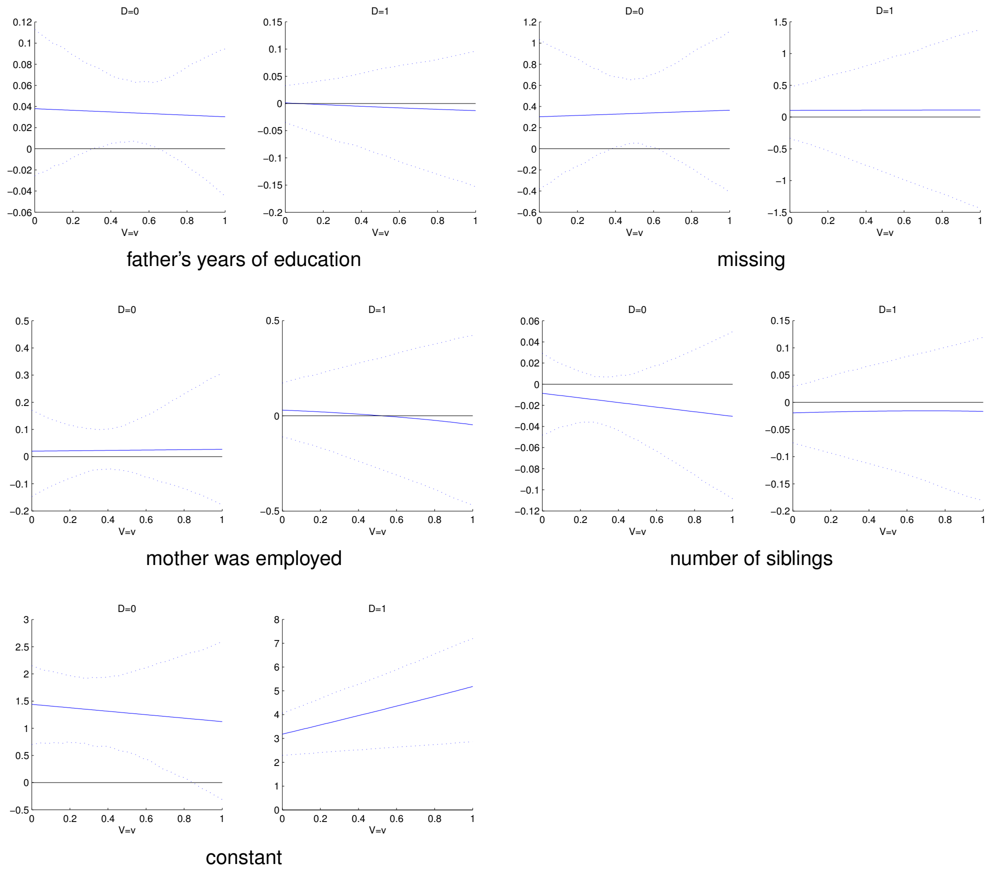


*Fig. 3.4:* Conditional average *ceteris paribus* effects 2/2. Point estimates and bootstrapped 95% confidence intervals.

# BIBLIOGRAPHY

AAKVIK, A., J. J. HECKMAN, AND E. J. VYTLACIL (2005): "Treatment Effects for Discrete Outcomes When Responses to Treatment Vary Among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs," *Journal of Econometrics*, 125(1-2), 15–51.

ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70(1), 91–117.

ANGRIST, J. D., K. GRADDY, AND G. W. IMBENS (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67(3), 499–527.

ANGRIST, J. D., AND A. B. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics*, 106(4), 979–1014.

——— (2001): "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 15(4), 69–85.

ASHENFELTER, O., AND J. D. MOONEY (1968): "Graduate Education, Ability, and Earnings," *Review of Economics and Statistics*, 50(1), 78–86.

BECKER, G. S. (1967): "Human Capital and the Personal Distribution of Earnings," Woytinski Lecture No. 1, Ann Arbor: University of Michigan Press.

BJÖRKLUND, A., AND R. MOFFITT (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, 69(1), 42–49.

BLUNDELL, R., L. DEARDEN, AND B. SIANESI (2005): "Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS," *Journal of the Royal Statistical Society, Series A*, 168(3), 473–512.

BLUNDELL, R., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics*, ed. by L. Hansen, Amsterdam. North Holland.

CARD, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1160.

CARNEIRO, P., J. J. HECKMAN, AND E. VYTLACIL (2005): "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education," Mimeograph.

CARNEIRO, P., AND S. LEE (2004): "Comparative Advantage and Schooling," Mimeograph.

——— (2005): "Ability, Sorting and Wage Inequality," c*e*mmap Working Paper CWP 16/05, University College London.

CHAMBERLAIN, G. (1977): "Education, Income, and Ability Revisited," *Journal of Econometrics*, 5, 241–259.

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

CHERNOZHUKOV, V., G. W. IMBENS, AND W. K. NEWEY (2004): "Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions," MIT Working Paper.

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71(5), 1405–1441.

——— (2005): "Nonparametric Identification Under Discrete Variation," *Econometrica*, 73(5), 1525–1550.

CLEVELAND, W. S., E. GROSSE, AND W. M. SHYU (1991): "Local Regression Models," in *Statistical Models in S*, ed. by J. M. Chambers, and T. J. Hastie, pp. 309–376. Wadsworth/Brooks-Cole, Pacific Grove, CA.

DAROLLES, S., J.-P. FLORENS, AND E. RENAULT (2003): "Nonparametric Instrumental Regression," IDEI Working Paper 228.

DAS, M. (2005): "Instrumental Variables Estimation of Nonparametric Models with Discrete Endogenous Regressors," *Journal of Econometrics*, 124(2), 335–361.

FAN, J., AND W. ZHANG (1999): "Statistical Estimation in Varying Coefficient Models," *Annals of Statistics*, 27(5), 1491–1518.

FISHER, R. A. (1935): *The Design of Experiments*. Oliver & Boyd, Edinburgh.

GAREN, J. (1984): "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable," *Econometrica*, 52(5), 1199–1218.

GOLDBERGER, A. S. (1972): "Structural Equation Methods in the Social Sciences," *Econometrica*, 40(6), 979–1001.

——— (1989): "Economic and Mechanical Models of Intergenerational Transmission," *American Economic Review*, 79(3), 504–513.

GRILICHES, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45(1), 1–22.

GRILICHES, Z., AND W. M. MASON (1972): "Education, Income, and Ability," *Journal of Political Economy*, 80(3, Part 2: Investment in Education: The Equity-Efficiency Quandary), S74–S103.

HANSEN, W. L., B. A. WEISBROD, AND W. J. SCANLON (1970): "Schooling and Earnings of Low Achievers," *American Economic Review*, 60(3), 409–418.

HASTIE, T., AND R. TIBSHIRANI (1993): "Varying-Coefficient Models," *Journal of the Royal Statistical Society Series B (Methodological)*, 55(4), 757–796.

HAVEMAN, R., AND B. WOLFE (1995): "The Determinants of Children's Attainments: A Review of Methods and Findings," *Journal of Economic Literature*, 33(4), 1829–1878.

HECKMAN, J. J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 49, 931–959.

HECKMAN, J. J., AND R. ROBB (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman, and B. Singer, chap. 4, pp. 159–244. Cambridge Univeristy Press, Cambridge, MA.

——— (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in *Drawing Inference from Self-Selected Samples*, ed. by H. Wainer, pp. 61–113. Springer, Heidelberg, New York, with discussion.

HECKMAN, J. J., AND G. SEDLACEK (1985): "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93(6), 1077–1125.

——— (1990): "Self-Selection and the Distribution of Hourly Wages," *Journal of Labor Economics*, 8(1, Part 2: Essays in Honor of Albert Rees), S329–S363.

HECKMAN, J. J., S. URZUA, AND E. VYTLACIL (2004): "Understanding Instrumental Variables in Models with Essential Heterogeneity," Memeograph.

HECKMAN, J. J., AND E. J. VYTLACIL (1998): "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling," *Journal of Human Resources*, 33(4), 974–987.

——— (1999): "Local Instrumental Variables and Latent Variables Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730–4734.

——— (2000a): "Local Instrumental Variables," in *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, ed. by C. Hsiao, K. Morimune, and J. Powell, pp. 1–46. Cambridge University Press, Cambridge.

——— (2000b): "Local Instrumental Variables," NBER Working Paper No. T0252.

——— (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

IMBENS, G. W., AND W. K. NEWEY (2003): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," MIT Working Paper, Presented at the 2003 EC2 conference held in London.

KELEJIAN, H. H. (1971): "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables," *Journal of the American Statistical Association*, 66(334), 373–374.

KLEIN, T. J. (2006): "Heterogeneous Treatment Effects: Instrumental Variables without Monotonicity?," Mimeograph.

Leibowitz, A. (1974): "Home Investments in Children," *Journal of Political Economy*, 82(2, Part 2: Marriage, Family Human Capital, and Fertility), S111–S131.

Newey, W., and J. L. Powell (2003): "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

Newey, W., J. L. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67(3), 565–603.

Newey, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimates," *Journal of Econometrics*, 79, 147–168.

Pinske, J. (2000): "Nonparametric Regression Estimation Using Weak Separability," University of British Columbia.

Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56(4), 931–954.

Rosenbaum, P. R., and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70(1), 41–55.

Roy, A. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

Sattinger, M. (1978): "Comparative Advantage in Individuals," *Review of Economics and Statistics*, 60(2), 259–267.

Solmon, L. C., and P. J. Taubman (1973): *Does College Matter? Some Evidence of the Impacts of Higher Education*. Academic Press, New York.

Taubman, P. J. (1973): "Schooling, Ability, Non Pecuniary Rewards, Socioeconomic Background and the Lifetime Distribution of Earnings," in *The Distribution of Economic Well-Being*, ed. by T. F. Juster, pp. 419–510. Ballinger Publishing Company, Cambridge.

Taubman, P. J., and T. J. Wales (1972): "Mental Ability and Higher Educational Attainment in the Twentieth Century," NBER and Carnegie Commission on Higher Education.

Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70(1), 331–341.

Weisbrod, B. A., and P. Karpoff (1968): "Monetary Returns to College Education, Student Ability, and College Quality," *Review of Economics and Statistics*, 50(4), 491–497.

WILLIS, R. J., AND S. ROSEN (1979): "Education and Self-Selection," *Journal of Political Economy*, 87(5, Part 2: Education and Income Distribution), S7–S36.

XIA, Y., AND W. K. LI (1999): "On the Estimation and Testing of Functional-Coefficient Linear Models," *Statistica Sinica*, 9, 735–757.

YATES, F. (1937): "The Design and Analysis of Factorial Experiments," Discussion paper, Technical Communication no. 35 of the Commonwealth Bureau of Soils.

# 4. LAST MINUTE FEEDBACK

JOINT WITH CHRISTIAN LAMBERTZ, GIANCARLO SPAGNOLO, AND
KONRAD O. STAHL

> The secret... is to wait until the 90 day feedback period is nearly up and then zap em w[ith a] negative feedback when they only have a few hours remaining to respond... That way they can't retaliate... This only wor[ks] if you are able to hold a grudge for 90 days...—*An eBay user.*[1]

## 4.1 Introduction

Feedback mechanisms in electronic markets allow partners to rate each other after a transaction. These schemes, also referred to as "reputation mechanisms," are claimed to be crucial for the success of anonymous trading platforms such as eBay.[2] On these platforms the room for opportunistic behavior on both sides of the market is particularly wide: anonymity and distance allow sellers to cheat on the quality of the good. Likewise, buyers can be dishonest concerning their payment behavior.[3]

In spite of the incentive to free ride—providing feedback appears a purely altruistic act *prima facie*—feedback is given in the better part of the transactions on eBay.[4] Therefore, it could be argued that this device plays an important role in diminishing informational asymmetries by enhancing the discipline of transacting parties. However, there is a lively discussion about the economic effects of reputation mechanisms in electronic markets.[5]

Rather than focusing on the *effects* of reputation, for example on prices or the

[1] Quote taken from a newsgroup discussion on http://www.the-gas-station.com/messages.cfm? type=normal&thread_id=49933&lastdays=2000& (February 2006).

[2] For example, in the founder's letter posted on February 26, 1996, Pierre Omidyar claims that "some people are dishonest. Or deceptive... But here, those people can't hide. We'll drive them away." See http://pages.ebay.com/services/forum/feedback-foundersnote.html (February 2006).

[3] According to the Internet Crime Complaint Center (IC3) 2004 Internet Fraud Crime Report "internet auction fraud was by far the most reported offense, comprising 71.2% of 207,449 referred complaints." See http://www1.ifccfbi.gov/strategy/statistics.asp (July 2005). Likewise, the FTC reports that "internet auction fraud is on the rise, with an increasing number of consumers complaining about sellers who deliver their advertised goods late or not at all, or deliver something far less valuable than promised." See the FTC's "Top Ten Dot Cons" on http://www.ftc.gov/bcp/conline/edcams/dotcon/auction.htm (February 2006).

[4] Resnick and Zeckhauser (2001) were among the first to investigate feedback behavior on eBay. They find that in about 52 per cent of the transactions feedback is left.

[5] See Dellarocas (2005) for a useful survey of recent research on reputation mechanisms. The effects of seller reputation on prices and the probability of selling the object are usually found to be negligible or positive. See, for example, Melnik and Alm (2002), Bajari and Hortaçsu (2003), Cabral and Hortaçsu (2005), Livingston and Evans (2004), Lucking-Reiley, Bryan, Prasad, and Reeves (2005), Houser and Wooders (forthcoming). See also Bajari and Hortaçsu (2004) as well as Resnick, Zeckhauser, Swanson, and Lockwood (2004) for an overview.

probability of selling, we focus on the *occurrence and timing* of feedbacks. In particular, we document an asymmetry in the feedback behavior previously unnoted in the literature, propose an explanation based on the micro structure of the eBay feedback mechanism, and support this explanation by findings from a large data set. Our analysis implies that the informational content of feedback records is possibly rather low. This is often neglected by eBay users.[6]

On eBay, both the seller and the buyer of an object are allowed to rate each other after a transaction. Mostly, feedbacks are positive. Moreover, it is well known that the correlation between first and second feedbacks is very high. We argue that this is at least partly driven by expectations on *feedback reciprocity*, i.e. giving a positive feedback while expecting the trading partner to reciprocate.[7] In contrast, and surprisingly, incidences of *feedback retaliation*, i.e. reacting to a negative feedback with a negative, are relatively rare in the data. In most cases, the second feedback is missing. In particular, in our data, we find that about 71 per cent of the positive feedbacks are reciprocated, whereas only about 37 per cent of the negative feedbacks are retaliated.

Feedback behavior can be influenced by several forces including the outcome of the transaction and strategic considerations. Public statements by eBay emphasize the ability of the feedback mechanism to discipline transacting parties by informing potential future trading partners about their current conduct.[8] Truthful reporting, however, may be in conflict with *strategic feedback behavior* which is present whenever agents anticipate the opponents' reactions when giving feedback. The following newsgroup discussion contains interesting insights of some eBay users. Its title is "Fix some eBay problems" and the contributions show that users are well aware of feedback retaliation.[9] One buyer reports

> Just last week, I had my first unpleasant experience in five years of eBay'ing. I received an item from a seller who had not left feedback for me (I mailed my money order the day after the auction ended). I was not happy with the item - flaws were not disclosed in the listing - and I notified the seller. After three e-mails and three phone calls went unanswered, I left negative feedback for her. She turned around and

[6] Jin and Kato (2002) find in a field experiment that "at least some buyers" overestimate the informational content of feedback score and "drastically underestimate the risk of trading online." Likewise, Resnick, Zeckhauser, Swanson, and Lockwood (2004) question whether price premia, which they find, reflect a reputation equilibrium, and should in fact not be observed in the data.

[7] Dellarocas, Fan, and Wood (2004) relate the motivation for leaving positive feedback to the user's expectation of reciprocal behavior from their trading partners.

[8] eBay states that the feedback "comments and ratings are valuable indicators of your reputation as a buyer or seller on eBay," see http://pages.ebay.com/help/feedback/questions/feedback.html (February 2006).

[9] Quotes are taken from http://ideas.4brad.com/archives/000018.html (February 2006).

> posted retaliatory negative feedback for me ruining my 100% rating. Indeed, the system needs to be improved.

Another user writes

> In the past I've not left any neg[ative] feedback as I'm afraid of revenge feedback that'll paint me as a bad trading partner... the dodgy seller ends up with getting away with it just to rip someone else off.

Yet another user notes

> As a buyer I have had problems with false item descriptions, even if you get a refund ... you end up paying postage for the item to you and back. Up till now I have not left any feedback for these bastards because of revenge.

and one concludes that

> I have been basing my purchase decisions [on eBay] on sellers' feedback scores. I had no idea these scores are so unreliable ... They are holding this feedback system out as the reason we should trust sellers, but the system has little to no basis in truth ... I suspect there are many, many people out there who have had actual monetary losses from this behavior.

This shows that at least some users are aware of possible feedback retaliation, or "revenge." Therefore, a reputation of being an imitator, who always reacts strategically to a positive feedback with a positive reply, and to a negative feedback with a negative one, could be valuable because it encourages future partners to give positive feedbacks, and discourages them from giving negative ones. eBay even sells a service to sellers allowing them to automatically reciprocate positive feedbacks.[10] Such behavior is in principle observable to other users on eBay.[11]

Our explanation for the asymmetry between the likelihood of a positive feedback being reciprocated and a negative one being retaliated is based on the institutional rules for giving feedback, combined with agents' expectations about retaliation and the assumption of attention decreasing in time. We argue that if an agent strategically anticipates the likely reaction of her counterpart and considers giving a first positive feedback, she should give it as early as possible in order to maximize the probability of a favorable reciprocation. Conversely, if an agent

[10] The price for an online seller tool which includes this service is currently $15.99 a month, see http://pages.ebay.com/sell/automation.html (February 2006) for a description.

[11] In particular, the feedbacks a user gets and the replies she leaves can be inferred from her feedback record.

considers giving negative feedback first and is concerned with possible retaliation, she should give the feedback as late as possible, i.e. she should opt for a *last minute feedback* to minimize the time that is left for the counterpart to retaliate.

This explanation implies that agents dissatisfied with their trading partners anticipate the risk of revenge, and may therefore be induced to refrain from leaving negative feedbacks all together, reducing and biasing the informational content of the reputation index towards positive outcomes. If agents who give the first feedback expect the opponent to reciprocate positive feedbacks, or retaliate negative feedbacks, then one would expect negative first feedbacks to be rare and positive ones to be common, a pattern that is usually found.[12] This point is also confirmed by a very recent study of Dellarocas and Wood (2006) showing that only about 86 per cent of eBay users are actually happy with the underlying transaction.

In order to support this explanation, we collected a large data set on eBay auctions and the ensuing feedbacks. We find that feedback is given substantially earlier if positive rather than neutral or negative. Moreover, our nonparametric analysis reveals that the probability of the trading partner reacting to the feedback is decreasing in the time the first feedback was given. It is interesting to see that this is the case no matter whether the feedback was positive or negative. As predicted, the probability that a feedback is negative increases substantially towards the last minute of the feedback period.

Towards refining our insights into the observed pattern, we look separately at buyers and sellers, and the trading partners' experience as well as at interaction effects between role and experience. In particular, we find that experience promotes strategic behavior. In the conclusion we propose simple changes in the eBay feedback mechanism that could greatly reduce this form of strategic behavior and improve its informational content.

In this work, we argue that giving negative feedback in the "last minute" and positive feedbacks early is motivated by strategic considerations. "Last minute bidding" in English auctions with fixed ending time (Roth and Ockenfels 2002) is a similar phenomenon. In both cases last minute action is exploited in order to prevent the opponents' reaction to the revelation of private information. However, if one were to consider mechanisms without fixed ending times, agents in an auction would still prefer placing a bid to abstaining.[13] On the contrary, giving a negative feedback becomes less attractive because of the fear of retaliation. Therefore, from a welfare point of view, the presence of a last minute is desirable in the context of feedbacks, whereas in the context of bids, it is not necessarily so.

[12] See, for example, Resnick, Kuwabara, Zeckhauser, and Friedman (2000), Resnick and Zeckhauser (2001), Bajari and Hortaçsu (2004), Cabral and Hortaçsu (2005), and Chwelos and Dhar (2005).

[13] For example, Amazon type auctions end only when no more bids are placed.

## 4.2 The eBay Feedback Mechanism

eBay is by far the biggest internet trading platform that brings together both private and professional buyers and sellers. In 2005, the number of listings exceeded 1.9 billion and eBay's gross merchandise volume accounted to more than 44 billion U.S. dollars.[14] Amongst other services, eBay provides a second price auction mechanism in which the seller describes the object and specifies a reservation price as well as the length of the auction period. Then, potential buyers can enter their bids.

As a matter of principle, eBay is only involved in the post auction transaction process if problems arise. In general, information on the conduct of the two parties is neither observable to us nor to future trading partners. However, eBay encourages its users to leave a feedback for each other within 90 days after the termination of the auction.[15] If a feedback is given, it consists of a positive or negative or neutral mark, and is accompanied by unformatted comments. For every user, eBay keeps a feedback record which contains all feedbacks received and left from transactions in which she was involved.[16] A recorded feedback cannot be removed unless both parties agree to. All marks are summarized in a feedback score and several summary statistics including the percentage of positive feedbacks.[17] While the feedback score can easily be observed by any partner in the bidding process, the observation of the detailed remarks is more involved.

In order to investigate the possibility of strategic feedback behavior, let us now discuss the decision to give feedback. At first, we make the simplifying assumptions that the timing of this decision is restricted to the 90 day period after the end of the auction and that there is no possibility to withdraw feedbacks. In Section 5 below, we will comment in detail on these assumptions and provide a schematic sensitivity analysis. A decision tree for the "feedback game" that starts after the end of an auction is provided in Appendix A.

For ease of the exposition, we follow the literature and occasionally group neutral and negative marks together, see e.g. Resnick and Zeckhauser (2001) and Cabral and Hortaçsu (2005). We will refer to them as negatives.[18] Moreover, in the sequel we refer to the first and second feedback as *feedback* and *reply*,

[14] See http://investor.ebay.com/news/Q405/EBAY0118-123321.pdf (February 2006).

[15] More precisely, eBay encourages, and guarantees recording feedback only within this period. An informal survey and the empirical evidence suggest that feedback thereafter is extremely rare. We will further comment on this in Section 5 below.

[16] eBay also offers internet shop services. Thus, feedbacks may also be based on experiences in trading via this channel, rather than auction trading.

[17] The feedback score is calculated as the number of users who left at least one positive feedback minus the number of users who left at least one negative feedback.

[18] However, we should emphasize that separating neutral from negative marks would not qualitatively change our results.

respectively.

We now juxtapose two modes of feedback behavior: truth telling, or non strategic feedback, as opposed to strategic, or opportunistic feedback. The former truthfully reveals information on the outcome of the transaction and thus leads to establishing credible feedback records, whereas the latter yields potentially biased reports, as they are influenced by the anticipation of the possible reaction of the trading partner.

Consider two parties in an eBay transaction, a buyer and a seller. Suppose that both are planning to interact with other partners in the future, and therefore attach positive value to their reputation. That is, both agents derive positive utility (expected pay off) from a positive, zero utility from no, and negative utility from a negative feedback received. This will be the case as long as there is some potential future trading partner believing that the feedback score is informative about the likely behavior of its holder.[19]

If the parties truthfully report their evaluation of the transaction without taking into account the reaction to their feedback in form of a reply, they should be indifferent about the timing of the feedback. However, the fact that delayed delivery or payment is seen as bad performance should imply that some negatives are given relatively late. In the empirical analysis, we disentangle this delay from a strategic delay.

On eBay, feedbacks are immediately observable to the counterpart. If the parties are influenced by strategic considerations, the timing at which feedbacks are left may thus become relevant.[20]

Suppose the two agents believe that their partners may have a tendency to reciprocate positive and retaliate negative feedbacks. This tendency to reciprocate may be due to behavioral components in agents' decision making processes, similar to the ones found by Fehr and Schmidt (1999), due to the attempt to build up a reputation as a "reciprocator" or "impersonator" in order to discourage future negative feedbacks and encourage positive ones—"the high courtesy equilibrium" of Resnick and Zeckhauser (2001)—, or due to a combination of both motives. The quotes in the Introduction indicate that such beliefs are realistic.

Suppose, in addition, that agents believe that the likelihood to receive a reply

[19] As was already pointed out in footnote 5 price effects of reputation are usually found to be nonnegative. Therefore, a "good" reputation is valuable to sellers. In principle, potential buyers in an auction could distinguish feedbacks the seller has received as a seller from feedbacks she has received as a buyer. However, it is a complex task to infer separate summary statistics from the records. See also Cabral and Hortaçsu (2005) who find that at least some sellers were able to build up their reputation as buyers. Even pure buyers can benefit from a "good" reputation record since sellers are allowed to exclude buyers from their auctions. This is possible on the basis of their subjective judgement of a bidder's reputation record.

[20] This is made explicit in Appendix A using the decision tree.

| | | |
|---|---|---|
| | positive | 0.982 |
| *feedback* | neutral | 0.008 |
| | negative | 0.01 |

2,471,459 observations.

*Tab. 4.1:* Sample probabilities for the type of feedback.

to their feedback is decreasing with the time at which feedback is left after the end of the auction. This is a reasonable expectation given that attention is costly and is likely to be fading in time. For obvious reasons this probability tends to zero when the first feedback is given very close to the expiration of the 90 day period after which feedbacks cannot be left any more—in the last minute.

Then, agents willing to post a negative feedback for a non performing partner will find it convenient to wait and do it in the last minute in order to minimize the likelihood that their counterpart notes the negative feedback and has time to retaliate. Conversely, agents willing to post a first positive feedback will find it convenient to do it early in order to maximize the likelihood that their counterpart notes and reciprocates it.

Therefore, *ceteris paribus*, we expect for feedbacks that are left strategically that the first feedback will be given early when it is positive, in order to *encourage* positive reciprocation. Conversely, we expect it to be given late, or not at all, when it is negative—in order to *reduce* the likelihood of receiving a retaliatory negative feedback. Because of this, we also expect positive feedbacks to be reciprocated more often than negative feedbacks to be retaliated.

## 4.3 Feedback Patterns

The data for the empirical analysis were collected in the second quarter of 2005 from the eBay platform. Starting from randomly drawn auctions we created a data set consisting of 2,471,459 auction records including respective feedbacks and their timing. By construction, the data include auctions for which at least one feedback was left. It is a random sample with respect to the category of the auctioned good which we think is appropriate for the purpose of this empirical analysis since we want to study feedback behavior *in general*. The data collection procedure is described in more detail in Appendix B.

Table 4.1 contains sample probabilities for the feedback being positive, neu-

| | | *reply* | | | |
|---|---|---|---|---|---|
| | | positive | neutral | negative | missing |
| unconditional | | 0.697 | 0.003 | 0.006 | 0.295 |
| | positive | 0.709 | 0.002 | 0.002 | 0.288 |
| *feedback* | neutral | 0.044 | 0.096 | 0.042 | 0.818 |
| | negative | 0.025 | 0.010 | 0.367 | 0.598 |

2,471,459 observations.

*Tab. 4.2:* Unconditional and conditional sample probabilities for the reply.

tral, or negative.[21] Observe that 98 per cent of all feedbacks given are positive. Resnick and Zeckhauser (2001) report a similar table and find that at least one feedback is left in 52 per cent of the transactions. If reporting was truthful and non strategic, the other 48 per cent of the feedbacks could reasonably be assumed to be missing at random. Otherwise, it could well be that neutrals or negatives are hiding behind these missing feedbacks.

Table 4.2 contains unconditional and conditional sample probabilities for the reply being positive, neutral, negative, or missing. In 70 per cent of the cases a reply is left. In about 71 per cent of the cases we observe that a positive feedback is reciprocated whereas only in about 37 per cent of the cases a negative feedback is retaliated.

We have argued in Section 2 that the relationship between the timing and type of the feedback is key in trying to understand these empirical phenomena.

In Figure 4.1, we have plotted the dependence between the reply and both the time of the feedback and its type. This was done by nonparametric local linear regressions of indicator variables for the type of the reply on the time of the feedback.[22] All graphs show that the later the feedback is given the less likely it is that a reply is given *at all*. More precisely, the probability that a reply is missing is increasing in time. This observation is independent of the type of the feedback.

In Figure 4.2, we document the timing of feedbacks. It shows empirical distribution functions of the time the feedback is given conditioned on the type of feedback. In particular, we find that in a first order stochastic dominance sense feedback is given earlier if it is positive rather than neutral, and in turn neutral

[21] Recall that this refers to the first feedback, as opposed to the second feedback which we call "reply."

[22] We used a Gaussian kernel. It turned out that the choice of the bandwidth did not have a substantial impact on these estimates. Here, we chose them *ad hoc*. Notice that the bootstrapped confidence intervals are extremely narrow due to the size of the data set.

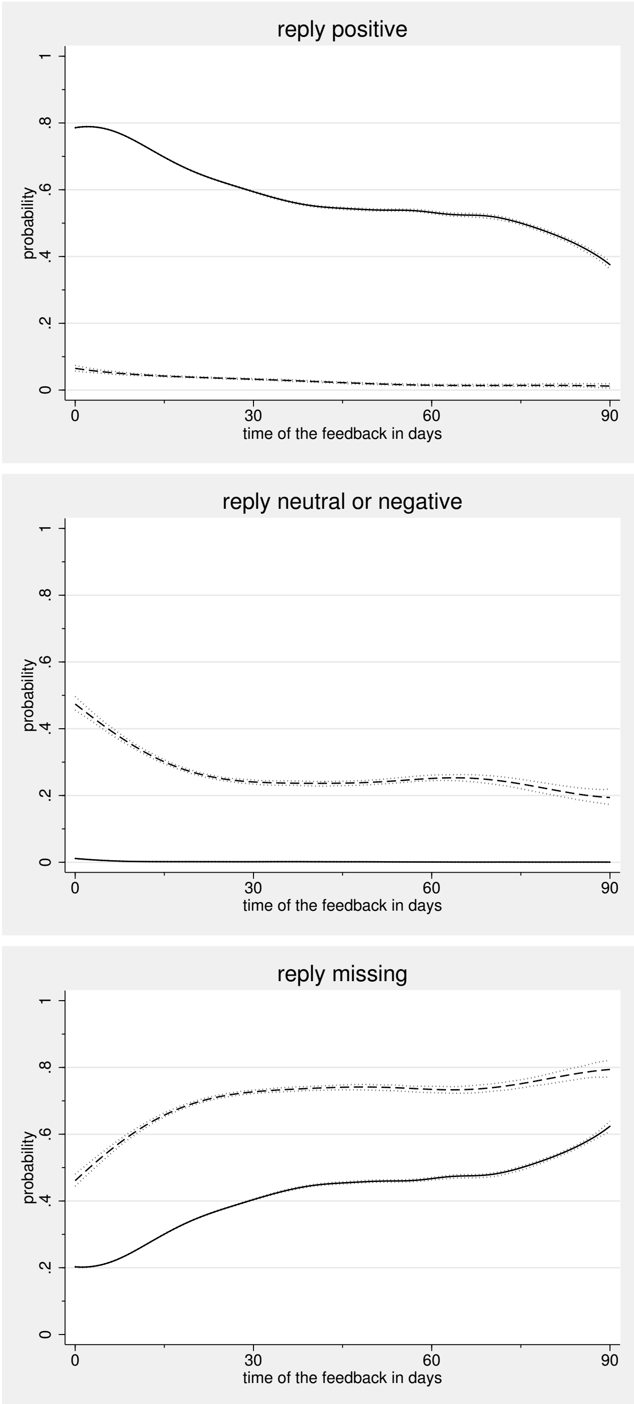


*Fig. 4.1:* The probability of a positive (top), negative or neutral (middle), and missing (bottom) reply given a positive (solid line) and negative or neutral (dashed line) feedback against time of the feedback. Local linear regressions and bootstrapped 95 per cent confidence intervals (100 replications).

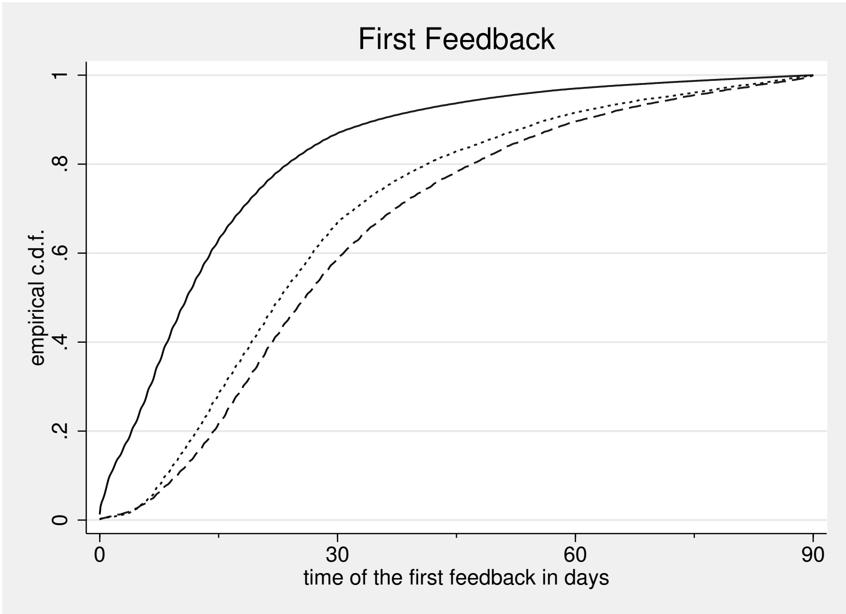


*Fig. 4.2:* Empirical cumulative distribution functions for the timing of the feedback given that it is positive, neutral, or negative (from left to right). Note that negative and neutral feedbacks are given later in a first order stochastic dominance sense.

rather than negative. With respective *p*-values of 0 this is confirmed by one-sided Kolmogorov-Smirnov tests.

These estimates are complemented with estimated conditional probabilities of the feedback being positive as well as the probability of a feedback being neutral or negative, conditional on the time of the feedback, respectively. Recall that most feedback is positive and is left relatively early within the 90 day period. However, Figure 4.3 shows that the later the feedback is left, the more likely it is to be negative or neutral—even culminating into a spike right at the end of the 90 day period. Hence, there is last minute feedback in the sense that feedback left in the "last minute" is much more likely to be negative.

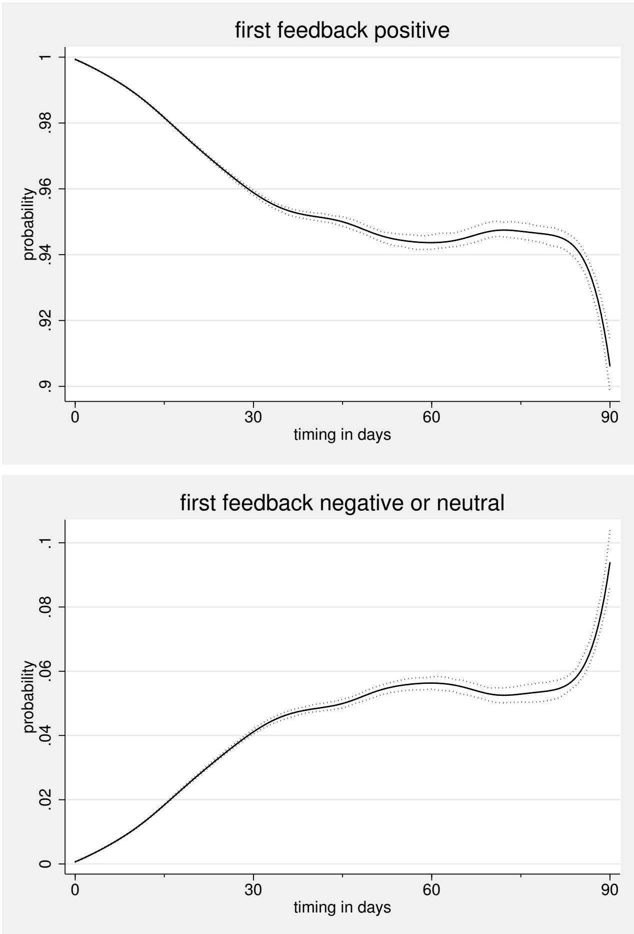


*Fig. 4.3:* The probability of a positive (top) and neutral/negative (bottom) feedback against time. Local linear regressions and bootstrapped 95 per cent confidence intervals (100 replications).

## 4.4 Strategic Feedback Behavior

We have documented the existence of a robust "Last Minute Feedback" effect for negative feedbacks, and of a symmetric early feedback effect for positive ones. In Section 2, we have argued that feedbacks are possibly given late when they are negative because agents strategically postpone the time at which the negative feedback is left to minimize the likelihood of retaliation. However, even if feedback behavior was non strategic, negative or neutral marks could be given later simply because the transaction was delayed and *therefore*, a negative or neutral feedback is left. These effects could entirely stem from transactions characterized by late delivery in which a truthful negative report is posted late. Conversely, those transactions characterized by timely delivery on both sides are likely to produce truthful positive feedbacks that are posted early. Resnick and Zeckhauser (2001, Table 2) have analyzed the feedback comments belonging to a sample of negative or neutral marks. On one hand, they find that 11 per cent of the complaints were about slow shipment. Additionally, in 23 per cent of the cases buyers claimed not to have received the item after they had paid for it. Hence, there is at least some scope for delays. On the other hand, however, in 24 per cent of the cases the good was shipped in time but was in poor condition, thus giving room for truthful negative and timely feedback. While these observations contribute to the explanation of the observed pattern, they quantitatively work in the same direction as the incentive to act strategically, and thus to postpone negative or neutral marks. In this section we present evidence strongly suggesting that delayed performance is not the only driving force of late negative feedbacks, i.e. that agents do indeed leave feedbacks strategically.

### 4.4.1 Last Minute Feedback

Figure 4.3 shows that the probability that a negative or neutral feedback is left increases in the first 30 to 40 days after the end of an auction. This increase could be explained by information revelation over time in problematic, possibly delayed, transactions which result in a negative or neutral feedback. Thereafter, until the last day of the 90 day feedback period, the probability of a negative or neutral mark seems not to depend on the timing of the feedback. However, *it increases dramatically on the 90th day after the end of an auction*. As for statistical inference, we have regressed an indicator variable for a negative or neutral feedback on a spline function in the time of the feedback, controlling for experience of the trading partners, and on whether the feedback was left by the buyer or the seller. It reveals that the probability that a given feedback is negative or neutral on the last half a day of the 90 day period after the end of the auction increases by about 6 per cent on average. This increase is highly significant at any level. This can

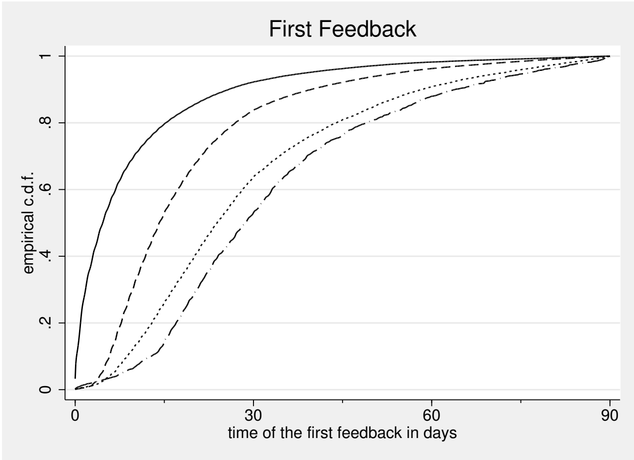


*Fig. 4.4:* Empirical cumulative distribution functions for the timing of the feedback given that it is given by the seller and that it is positive positive, buyer and positive, buyer and negative or neutral, seller and negative or neutral (from left to right).

hardly be reconciled with non strategic behavior since that would require that all of a sudden more negative or neutral than positive information on the trading partner would be revealed on the second half of the 90th day, compared to the 50 day period preceding this day.

### *4.4.2 The Role of the 'Role'*

In general, we suppose sellers to be more likely to be sellers in future transactions so that they are more interested in getting a positive feedback and avoiding a negative one. In consequence, the effects we have documented so far should be more pronounced for sellers once agents act strategically, since sellers' interest in their reputation is higher. Figure 4.4 shows that *feedbacks are in fact given substantially earlier if they are positive and given by the seller, as compared to positives given by the buyer. Along these lines, we find that negative or neutral marks are given later by sellers.*[23] We interpret this as further evidence for strategic behavior.

### *4.4.3 The Effect of Experience*

In several decades of experimental economics evidence has been accumulated on the effect of players' experience in strategic interactions. An important aspect

[23] With respective $p$-values of 0 one-sided Kolmogorov-Smirnov tests indicate that positive feedback is given earlier and negative feedback is given later by sellers.

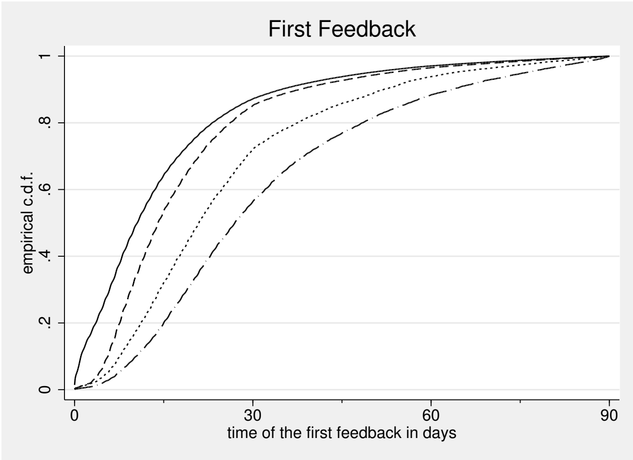


*Fig. 4.5:* Empirical cumulative distribution functions for the timing of the feedback given high experience and the feedback being positive, low experience and positive, low experience and negative or neutral, high experience and negative or neutral (from left to right).

therein is a deepened understanding of the opponent's strategic reaction to one's own action once a strategic situation is experienced repeatedly. On eBay, a proxy for experience that is easily observable is an agent's feedback score. Once feedback behavior is strategic, we should therefore again expect the *observed patterns to be more pronounced for experienced agents*. Figure 4.5 shows that this is the case in our data.[24] High experience is defined by a feedback score of at least 20. We have also run regressions in which we include the role of the agent giving feedback and its experience as explanatory variables. The results confirm this finding since the effect is statistically significant at any level. Such an analysis is sensible because experience and role are positively correlated.

### *4.4.4 Tools*

In further support of these findings, it has been suggested in various newsgroups to set up a service that automates strategic feedback timing. In a typical conversation, a user suggests[25]

[24] With respective $p$-values of 0 one-sided Kolmogorov-Smirnov tests indicate that positive feedback is given earlier and negative feedback is given later by experienced agents.

[25] See, e.g.,
http://community.auctionsniper.com/groupee/forums/a/tpc/f/785608021/m/308108399/r/3721016131 (February 2006). The quotes that follow are taken from this site.

> will someone out there please invent FEEDBACK SNIPER SOFTWARE that allows one to leave feedback (good or bad) at the last second? that way, you can leave legit[imate] bad feedback w[ith] no fear of retaliatory bad feedback left for you- thus purifying the ebay world, making ebay stock go up, and just making ebay a better community as a whole. i do not leave deserved bad feedback for fear of retaliatory bad feedback left on me!!!

And indeed, Auctionhawk, a company specialized on offering services around eBay, developed and advertised a service, for payment, to give feedback in the last minute.[26] The discussion from above, however, continues with the remark that

> It has already been suggested on this forum a handful of times. The problem is that it's not an exact 90 days. It can be several days longer.

and the reply

> random time and not 90 days, eh? that would definitely throw the idea for a loop. if we could isolate the time generator at ebay and get a handle on how these times are generated we could do it an ebay would be a purer place as crooks would think twice about fraud.

In fact, Auctionhawk has stopped advertising this feature in the meantime. Strategic implications of the randomness of the last minute will be at the center of our discussion in Subsection 4.5.1.

## 4.5 Mutual Feedback Withdrawal and Extended Feedback Periods

So far, we have abstracted from two features that are peculiar to the eBay feedback mechanism. First, feedbacks and replies can be left more than 90 days after the end of the auction and second, feedbacks and replies can be withdrawn if both parties agree to. We discuss the implications of these institutional details in the following two subsections.

### 4.5.1 Extended Feedback Periods

So far, we assumed that feedbacks can only be left within 90 days from the end of the auction. While eBay *guarantees* that feedback comments are recorded if left within 90 days after the end of the auction, it is a little known fact that this does

[26] See http://auctionbytes.com/cab/abn/y04/m08/i10/s01 (February 2006). A free reminder service for "last minute feedback" is offered by U.K. Auction Watch at http://www.ukauctionhelp.co.uk/remindme.php (February 2006).

not exclude the possibility of leaving feedback after this 90 day period. In eBay's own words: "eBay only commits to items being available for 90 days, so if it is greater than 90 days you may not be able to leave feedback."[27]

After 90 days, eBay removes the link on a member's personal "My eBay" page that encourages one to leave feedback. However, since the item number identifying a particular transaction is known in principle, one might still be able to leave feedback for a transaction by doing so manually. This finding is connected to what we experienced during our data collection: in feedback profiles auction details are linked for 90 days after the feedback was left or received. If details are not linked, this does not necessarily imply that they are not available any more.

In our random sample of auctions we found only very few cases in which feedback was left more than 90 days after the end of an auction. In particular, only 0.03 per cent of the feedbacks and 0.06 per cent of the replies were given after 90 days. If the feedback was negative, the probability that is was given after 90 days was only 0.4 per cent. In this case, the probability of a retaliative reply is 20 per cent.

From these findings we conclude that the possibility of leaving feedback later than 90 days after the end of an auction seems to be a little known secret and we doubt that many users consider the chance that even their "last minute" feedback could be followed by retaliation.

Even if it was publicly known that leaving feedback for longer than 90 days is possible, our analysis and conclusions would change only slightly. The empirical evidence in favor of the explanation given in this paper, namely the difference in the timing of positive and negative feedbacks we found, remains unchanged.

As for agents' strategic situation, what matters is the effect of such a "random last minute" on the decision to leave a feedback. For any given time at which a feedback is left, the possibility for the other agent to reply after the 90 day deadline simply increases the likelihood of receiving such a reply. Specifically, for positive feedbacks, this marginally increases the incentive to leave a positive feedback. Conversely, the extended period tends to further discourage negative feedbacks. It allows the agent to postpone a first negative feedback even more (after the 90th day). This is at the risk that the negative feedback is not registered. The benefit is a lower probability that the feedback is retaliated if the attention of the trading partner is fading over time.

All in all, we conclude that the uncertain length of the extended period in which a feedback can be given works against finding evidence for "last minute feedback" by making the "last minute" a probabilistic deadline. We still found this evidence in the data and believe our empirical results would be stronger if the period for leaving feedbacks would be of fixed length.

[27] See http://pages.ebay.com/help/feedback/questions/leaving-feedback.html (February 2006).

### 4.5.2 Mutual Feedback Withdrawal

A less well known feature of the eBay feedback mechanism is that, after feedbacks are given, they can be withdrawn upon the mutual approval by both parties. All feedbacks given up to the time of the withdrawal are removed and no more feedback can be left. If feedbacks are removed, they do not enter the calculation of the feedback score but remain in the parties' feedback histories as "withdrawn" feedbacks. Additionally, the number of withdrawn feedbacks is shown next to the members' feedback summary. Surprisingly, in our data, only 0.1 per cent of all feedbacks were withdrawn, 25 per cent of these within two days after the last feedback was left.[28]

Obviously, the possibility to withdraw feedbacks increases agents' ability to renegotiate in order to reach a final agreement. It allows agents to trade feedbacks, in particular a withdrawn negative for a withdrawn negative. This ability to renegotiate may have negative effects on the informational content of the feedback mechanism by promoting strategic, i.e. non-truthful, feedbacks.[29]

While it is observable whether feedbacks were withdrawn, we cannot observe whether the withdrawn feedbacks were positives or negatives. However, under the assumption that a good reputation record is valuable to at least one party, two positives should never be withdrawn. Therefore, we restrict our attention to situations in which at least one negative feedback was left.

#### Withdrawal of Two Negative Feedbacks

*Rational agents and backward induction.* The simplest among these situations, the one in which we expect most feedbacks to be withdrawn, is the case in which two negative feedbacks were left. Suppose both parties in the transaction value not having negative feedbacks more than the procedural cost of withdrawing feedbacks, are rational and not norm guided, and have a negligible chance of trading again in the future. Suppose further that observing and interpreting an agent's past feedback behavior—how has an agent left feedbacks and reacted to other agents'

[28] Note that in principle, we could have missed some late feedbacks and late feedback withdrawals if they occurred very late after the end of the auction, say 125 days. This is due to the design of the data collection procedure and it would be beyond the scope of this paper to improve on this. However, in view of the observation that already between the 90th and 100st day after the end of the auction incidences of feedback activity decline to a negligible number we are only missing very few feedbacks.

[29] See Dini and Spagnolo (2005a) for a general result. This is why Dini and Spagnolo (2005b) and Dellarocas, Dini, and Spagnolo (forthcoming) suggest to keep feedback mechanisms unilateral whenever it is possible: then, trading a positive feedback against a positive feedback is not an option. Provided that other forms of (e.g. monetary) trade are costly the informational content of such feedbacks increases. The empirical results in Chwelos and Dhar (2005) nicely support this policy prescription.

feedbacks in the past?—is too costly relative to the value of the transaction. Then, simple backward induction applies and suggests that two negative feedbacks will always be renegotiated and withdrawn, independent of the history that led to them.

We found the following advice in a guide entitled "How to Sell on eBay":[30]

> However much you're not supposed to do it, you really shouldn't let a buyer leave you negative feedback without leaving them a negative in return ... This might not be the 'nicest' way to do business on eBay, but it's the only realistic way to protect your flawless reputation. Don't be worried: retaliatory feedback is not against eBay's rules, however much it should be. Anyway, you're not just doing this for revenge—it's essential for the next step. Try for a Mutual Withdrawal. Since the buyer probably won't want a negative response or feedback comment on their record, you can do a simple "I'll take away my negative if you take away yours" deal. This is called mutual feedback withdrawal, and the process can be started at this page:...

*Reputation as a punisher.* For Sections 3 and 4's empirical analysis we have coded withdrawn feedbacks as negatives. Surprisingly, even under this assumption, only 22 per cent of all retaliated feedbacks were mutually withdrawn afterwards.[31] Not withdrawing a negative feedback may serve as a signalling device for an agent because she can thereby build a reputation as a "tough" punisher to deter future cheating, or receiving negative feedbacks.[32]

*Experience.* We consider it more likely that many agents involved in badly performing transactions were unexperienced and did not know about the possibility of withdrawal. To test this hypothesis we calculated withdrawal probabilities by the experience of the trading partners. We say that an agent is experienced if her feedback score is at least 10. If both trading partners are experienced, the probability of a feedback to be withdrawn was 23.5 per cent. If only one of them was inexperienced, this number declined to about 18.5 per cent. Unfortunately, even in our large data set there is no case in which both trading partners were inexperienced and a feedback was withdrawn. This shows once more that strategic behavior is promoted by experience.

*Withdrawal as a threat.* Alternatively, assume that agents believe there exist potential trading partners who do not fear retaliation because they do not value

[30] See http://www.askaboutthis.com/ebay/articles/How-to-Dispute-Unfair-Ratings-on-eBay.html (February 2006), line breaks were removed.

[31] Note, however, that our findings are robust to feedback withdrawal: dropping all observations in which feedback was withdrawn did not change our results.

[32] This requires that observing and analyzing an agent's past feedback behavior is an option for at least some potential future trading partners. We are sceptical about this possibility, given how complex it is to track and interpret an agent's past feedback behavior by manually going through her feedback history.

their own reputation. These could, e.g., be occasional buyers paying in advance. Such agents could make use of the possibility to withdraw feedbacks to signal toughness by giving a negative feedback early, before the good is delivered. The signal to the trading partner would be that their valuation for reputation is low so that they are always prepared to give a negative or retaliate a negative without withdrawing it thereafter if the delivered good does not conform to promises. This can be used to exact performance of the other trading partner.

*Withdrawal of One Negative Feedback*

In case only one negative feedback was given, naturally, only one party is interested in the withdrawal of this feedback. Suppose reports were truthful, so that one party (say the seller) was satisfied by the trading partner's performance (say prompt payment), but the other party was not (the buyer found the good of lower than expected quality). Then, the unsatisfied party that left a negative could agree to withdraw the feedbacks in exchange for a discount in price, or the seller's willingness to take back the good. However, in general, these "renegotiated" transactions are hard to enforce. Since feedbacks are already given, and can only be jointly withdrawn, there is no guarantee for the accomplishment of such a new agreement. In our example, after the seller accepts to take back the good, if she returns the money first the buyer could be inclined not to return the good, or not to agree to withdraw the feedback given that hers is positive. Likewise, if the good is sent back first, the seller may not want to return the money. These enforcement problems of *ex post* renegotiation in asymmetric situations, where no feedback mechanism is present to limit opportunism, let us believe that most withdrawn feedbacks are couples of negatives.

## 4.6 Policy Implications and Concluding Remarks

In this paper, we have highlighted empirical phenomena such as "last minute (negative) feedbacks" and have reconciled them with an idea of agents' strategic feedback behavior on eBay. In particular, we have shown evidence indicating that agents tend to anticipate reactions of the trading partner in the "feedback game" when they leave feedback.

Moreover, agents seem to be aware of the risk that a negative feedback is retaliated. This implies that positive feedbacks are usually given early, with the aim of stimulating reciprocation, and negative feedbacks are given late or not given at all, in fear of retaliation. Therefore, positive feedbacks are likely to be given too often and negatives are likely to be given too seldom. Hence, negative feedbacks, *if* they are given, typically contain more information than positive ones. In general, the informational content of feedback histories in such eBay type bilateral feedback mechanisms appears to be low.

Let us finally develop some ideas towards improving on the design of the feedback mechanism.[33] Our analysis suggests that the "feedback game" should be made less transparent to both parties. In particular, favorable "anonymity" should be pursued, so that both feedbacks are revealed to the trading partners and the public only if no more feedbacks can be left. This could be done after a fixed period, or after both have already given their feedback. Note that this device requires that feedback withdrawal is not possible. Otherwise, under general conditions, it is a dominant strategy for the players to always leave a negative feedback in order to be able to renegotiate after feedbacks have been revealed.

In general, the performance of buyers, if asked to pay first, is subject to little uncertainty. It is also easier to discipline them: either the full payment arrives in time, and bank transfer details can demonstrate this, or it does not. Sellers can instead "cheat" in non evident ways on a variety of aspects of their performance, and this opaqueness creates room for opportunistic behavior. Therefore, it may be worthwhile to limit feedbacks to buyers rating sellers as in Amazon auctions.[34]

We shall end with the appeal that[35]

> Sooner or later we all face this dilemma on e-Bay. Do we slag an obvious jerk with a negative feedback, only to get a retaliatory negative feedback from him. You have to decide if it's worth it. Always check out his feedback first. See if he posts retaliatory feedbacks. Avoid him like the plague if he does. In your case, seeing as how you aren't out any cash, I would just let this one slide. Let this moron fester in his own little crooked world. There are a lot of goofs out there in e-Bayland, just steer clear of them if possible. IMHO [in my humble opinion], save your negative feedbacks for the really bad experiences that cost you serious money. Cheers!

[33] Roth (2002) makes a strong case for economists helping to *design* markets and institutions.

[34] This is also suitable for e-procurement platforms. See Dini and Spagnolo (2005a, 2005b) for further details.

[35] Taken from http://antiqueradios.com/forums/Forum14/HTML/000994.html (February 2006).

## *Appendix A: Decision Tree*

Figures 4.6 and 4.7 describe the decision tree for the “feedback game” that starts right after the end of the auction. Here, we assume for simplicity that feedback can be left for 90 days. Together, Figures 4.6 and 4.7 depict the 90 days long decision tree as a continuum of decision nodes. Starting from the first instant after the auction agents 1 and 2, which we treat symmetrically, simultaneously choose whether or not to give a feedback, as long as neither of them has placed their feedback yet. The first decision node of Figure 4.6 depicts this simultaneous game for the first instant of time. Each agent can simultaneously choose among abstaining from giving feedback (0), giving a positive feedback (+), and giving a negative or neutral one (−). For instance, a pair $(+, -)$ denotes that agent 1 received a positive mark from agent 2 whereas agent 2 received a negative mark from agent 1. As usual, the information set at 2’s decision node implies that 1 and 2’s decisions are simultaneous, i.e. that neither of the two players observes the trading partner’s decision for that instant before choosing. If both 1 and 2 choose to give a feedback in an instant of time, the “feedback game” ends and the two feedbacks are recorded and become observable on the platform. If one of the two agents gives her feedback and the other does not, the other agent will be able to observe the feedback received an instant of time later, and will remain the only one with a choice to make. As long as both 1 and 2 choose not to give feedback and the last instant is not reached, the simultaneous game starts again in the following instant of time. Once the “last minute” is reached, i.e. the last instant of the 90 days in which a feedback can be given, the simultaneous decision node changes form into that described in Figure 4.7.

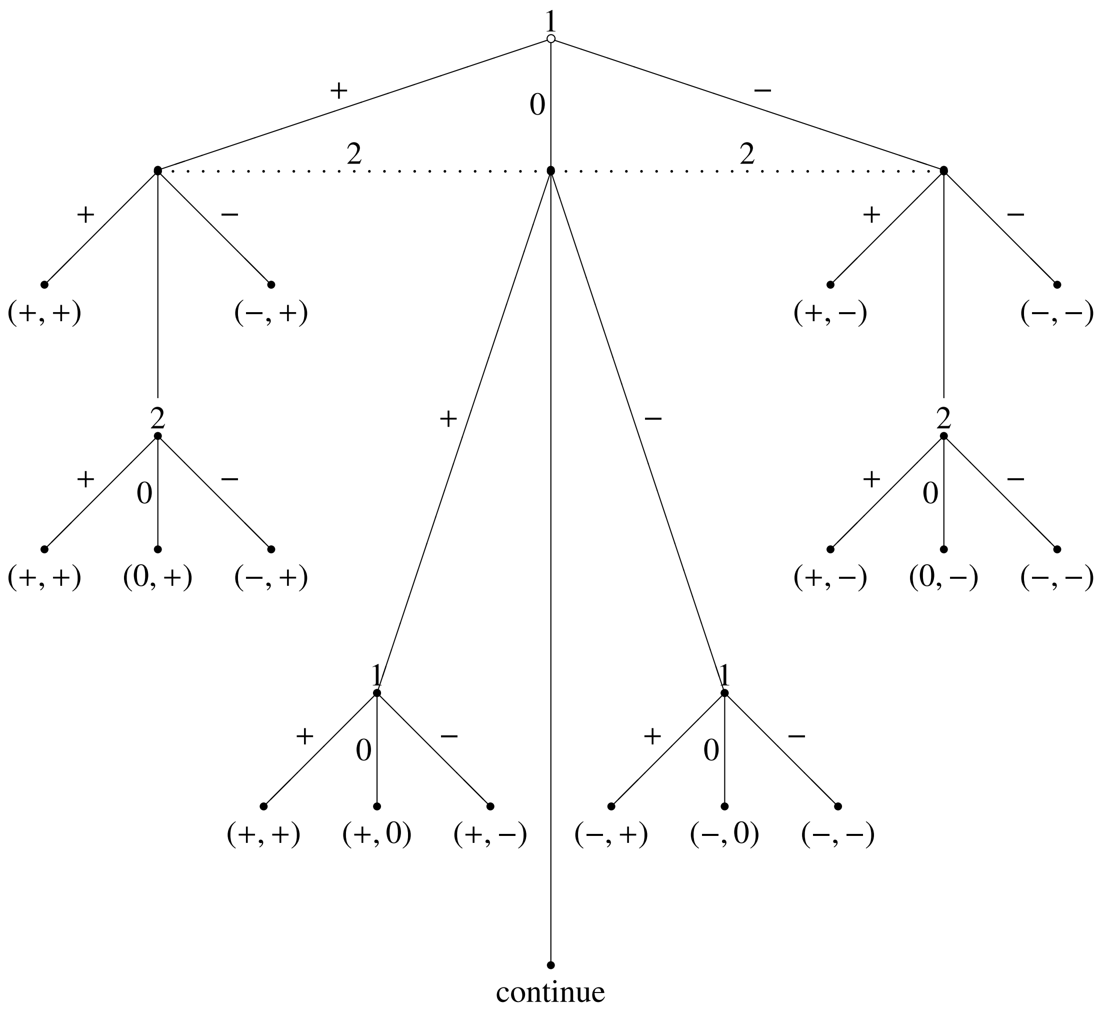


*Fig. 4.6:* Decision tree before the last minute.

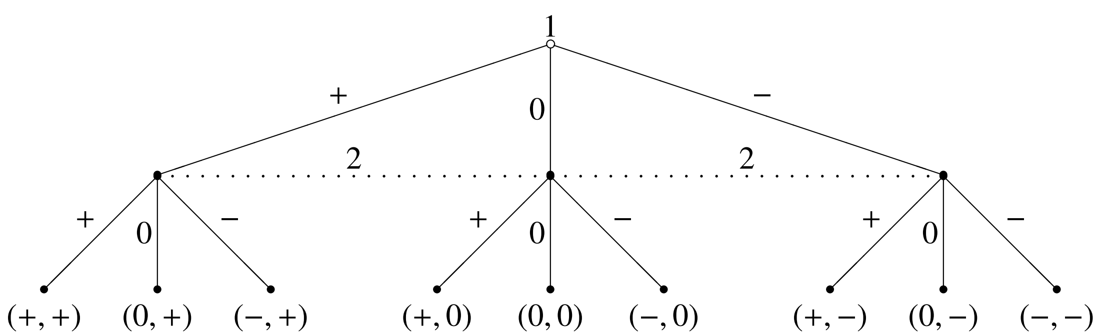


*Fig. 4.7:* Decision tree in the last minute.

## *Appendix B: Data Collection*

We first randomly drew auction numbers and downloaded the respective auction details. From these auction details we obtained the respective seller member ID and randomly selected $10,000$ sellers from the United States.

In a next step, for each seller, we used the information in her feedback profile to obtain auction details including the corresponding feedback which was received and left, and the respective timing information. By construction, since we start from a member's feedback profile, our sample consists of auction records for which at least one feedback was left by either the seller or the buyer. In order to minimize the loss of information, we included only those auctions into our data set which ended at least 100 days before the date of our data collection. Moreover, we required the auctions to have ended at most 125 days before the date of our data collection. This value is suggested by the data because after 125 days auction details might not be available any more.

We restricted our attention to standard eBay auctions. That is, we dropped auctions that belong to "eBay Motors," are "Live Auctions," serve as an "Advertisement Only," and are "Quantity Items." Moreover, we did not consider auctions that ended early.

Mutually withdrawn feedbacks were coded as negatives.

# BIBLIOGRAPHY


Bajari, P., and A. Hortaçsu (2003): “The Winner’s Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions,” *RAND Journal of Economics*, 34(2), 329–355.

——— (2004): “Economic Insights from Internet Auctions,” *Journal of Economic Literature*, 42(2), 457–489.

Cabral, L. M. B., and A. Hortaçsu (2005): “The Dynamics of Seller Reputation: Theory and Evidence from eBay,” Mimeograph.

Chwelos, P., and T. Dhar (2005): “Caveat Emptor: Differences in Online Reputation Mechanisms,” Mimeograph.

Dellarocas, C. (2005): “Reputation Mechanisms,” Working Paper, University of Maryland.

Dellarocas, C., F. Dini, and G. Spagnolo (forthcoming): “Designing Reputation Mechanisms,” in *Handbook of Procurement*, ed. by N. Dimitri, G. Piga, and G. Spagnolo, chap. 18. Cambridge University Press, in press.

Dellarocas, C., M. Fan, and C. A. Wood (2004): “Self-Interest, Reciprocity, and Participation in Online Reputation Systems,” Working Paper.

Dellarocas, C., and C. A. Wood (2006): “The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias,” Mimeograph.

Dini, F., and G. Spagnolo (2005a): “Buying Reputation on eBay,” Manuscript, Consip Research Unit and Stockholm School of Economics.

——— (2005b): “Reputation Mechanisms and Electronic Markets: Economic Issues and Proposals for Public Procurement,” in *Challenges in Public Procurement: an International Perspective*, ed. by K. V. Thai, A. Araujo, R. Y. Carter, G. Callender, D. Drabkin, R. Grimm, K. R. Ejlskov Jensen, R. E. Lloyd, C. P. McCue, and J. Telgen. Academic Press.

Fehr, E., and K. Schmidt (1999): "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, 114(3), 817–868.

Houser, D., and J. Wooders (forthcoming): "Reputation in Auctions: Theory, and Evidence from eBay," *Journal of Economics and Management Strategy*.

Jin, G. Z., and A. Kato (2002): "Blind Trust Online: Experimental Evidence from Baseball Cards," Working Paper, University of Maryland.

Livingston, J. A., and W. N. Evans (2004): "Do Bidders in Internet Auctions Trust Sellers? A Structural Model of Bidder Behavior on eBay," Working Paper, Bentley College.

Lucking-Reiley, D., D. Bryan, N. Prasad, and D. Reeves (2005): "Pennies from eBay: the Determinants of Price in Online Auctions," Working Paper, University of Arizona.

Melnik, M. I., and J. Alm (2002): "Does a Seller's Reputation Matter? Evidence from eBay Auctions," *Journal of Industrial Economics*, 50(3), 337–349.

Resnick, P., K. Kuwabara, R. Zeckhauser, and E. Friedman (2000): "Reputation systems," *Communications of the ACM*, 43(12), 45–48.

Resnick, P., and R. Zeckhauser (2001): "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. The Economics of the Internet and E-Commerce.," in *Advances in Applied Microeconomics*, ed. by M. R. Baye, vol. 11, Amsterdam. Elsevier Science.

Resnick, P., R. Zeckhauser, J. Swanson, and K. Lockwood (2004): "The Value of Reputation on eBay: A Controlled Experiment," Working Paper, Harvard Kennedy School of Business.

Roth, A. E. (2002): "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics," *Econometrica*, 70(4), 1341–1378.

Roth, A. E., and A. Ockenfels (2002): "Last-Minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from eBay and Amazon on the Internet," *American Economic Review*, 92(4), 1093–1103.

## *Ehrenwörtliche Erklärung*

Hiermit erkläre ich ehrenwörtlich, daß ich diese Dissertationsschrift selbständig angefertigt habe und mich anderer als der in ihr angegebenen Hilfsmittel nicht bedient habe. Entlehnungen aus anderen Schriften sind ausdrücklich als solche gekennzeichnet und mit Quellenangaben versehen.

Mannheim, 22.6.2006

Tobias J. Klein

# Curriculum Vitae

| | |
|---|---|
| 9/2003 - 8/2004 | University College London,<br>Ph.D. studies in Economics (visiting) |
| 11/2002 - 7/2006 | University of Mannheim,<br>Ph.D. studies in Economics |
| 8/2001 - 6/2002 | University of California, Berkeley,<br>Ph.D. courses in Economics (visiting) |
| 10/1998 - 11/2002 | University of Mannheim,<br>undergraduate studies in Economics |
| 6/1998 | Graduation from secondary school |
| 5/1979 | Born in Stuttgart, German nationality |