Labour Economics 90 (2024) 102565

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Semiparametric distribution regression with instruments and monotonicity

Dominik Wied[1]

Institute for Econometrics and Statistics, University of Cologne, Germany

ARTICLE INFO

JEL classification:
C26
J30

Keywords:
Control function
Endogeneity
Isotonic regression
Mincer-type equations

ABSTRACT

This paper proposes IV-based estimators for the semiparametric distribution regression model in the presence of an endogenous regressor, which are based on an extension of IV probit estimators and the idea of control functions. We discuss the causal interpretation of the estimators and two methods (monotone rearrangement and isotonic regression) to ensure a monotonically increasing distribution function. Asymptotic properties and simulation evidence are provided. An application to income equations with German SOEP data reveals statistically significant and heterogeneous differences to the inconsistent non-IV-based estimator.

1. Introduction

The semiparametric distribution regression (DR) model introduced by Foresi and Peracchi (1995) has become a popular model for conditional distributions if other quantities than only the conditional expectation are of interest. An important feature of this model is that no distribution assumptions on the response are made, e.g. Y is not assumed to be normally distributed, conditionally on covariates. At the same time, the model provides interpretable functional forms between the regressors and the outcome, while estimating the conditional response distribution semiparametrically. From the estimated distribution function, quantiles could be directly obtained by inversion.

A typical application is the topic of conditional wage (or, more general, income) distributions, where upper or lower quantiles are supposed to be modeled. Chernozhukov et al. (2013) and Rothe and Wied (2013) show that the DR model might be better suited than quantile regression for handling certain characteristics of income data such as genuine point masses in the distribution of incomes, nonlinearities around the minimum wage and rounding effects. In our empirical application on German SOEP data from the year 2020 below, one such genuine point is the income of 450 Euro per month, as the "450-Euro-job" was a popular type of part-time job in this year. The appealing property is that e.g. censoring points do not have to be included ex ante as in the case of censored quantile regression, but are detected by the estimation itself. Chernozhukov et al. (2013) show how the model can be used for estimating counterfactual distributions, Rothe and Wied (2020) propose a method for estimating conditional densities and quantile partial effects in this model. Troster and Wied (2021) consider DR for dynamic data, Delgado et al. (2022) in the context of duration analysis, Wang et al. (2022) develop a bivariate DR model and apply this to insurance data. See also Koenker et al. (2013) for a comparison of quantile and distribution regression.

A restriction of the literature up to now is that the regressors are assumed to be exogenous. For example, Rothe and Wied (2013) consider a version of Mincer's earnings function by explaining the logarithmic wage with the years of education and the years of experience among others, not taking into account that, for example, the years of education might be an endogenous regressor. This does not mean that the DR estimates in such approaches are not useful. They do estimate conditional distribution functions consistently, but there is no control for unobserved confounders. The years of education might be correlated with the unobserved ability or motivation of the employee. So, one would estimate the distribution only for a subset of the population, e.g. for employees with both high educational status and high ability/motivation. The novelty of the present approach is the control for confounders.

There are some recent papers on DR estimation with endogenous regressors. Briseño-Sanchez et al. (2020) consider DR estimation based on instrumental variables, but they use parametric models based on splines among others. Their approach fits in the GAMLSS framework (generalized additive models for location, scale shape) which means that their focus lies on explaining moments of the distribution. In contrast, the focus on the present paper is the distribution function itself. Chernozhukov et al. (2020) estimate structural functions in triangular models using DR techniques, but also here, the distribution function is not of immediate interest. For example, they consider average

E-mail address: dwied@uni-koeln.de.

[1] I am grateful to the Co-Editor Bernd Fitzenberger, to two referees, to Alexander Mayer and to the participants of the Statistische Woche 2023 and the economics research seminar at the University of Venice for helpful comments. Moreover, I thank Yu Ting Hsiao and Jie Lu for excellent research assistance.

https://doi.org/10.1016/j.labeco.2024.102565
Received 30 November 2023; Received in revised form 22 February 2024; Accepted 11 May 2024
Available online 20 May 2024

structural functions for the conditional mean, whereas we are interested in average structural functions for the distribution itself. Chernozhukov et al. (2022) discuss semiparametric DR models in the context of sample selection.

The present paper proposes IV-based estimators for the semiparametric DR model which use the idea of control functions and are based on a particular latent variable model. Taking into account that the DR model is fitted by pointwise estimators of simple binary outcome models, we adapt consistent estimators for binary outcome models with endogenous regressors. On the one hand, we consider maximum likelihood estimation, which is asymptotically efficient, on the other hand, we propose a computationally better tractable three-step estimator. For both estimators, consistency and convergence to Gaussian limit processes are proved. As these estimator are unconstrained, monotonicity is not guaranteed. We discuss two methods for enforcing monotonicity in a second step, monotone rearrangement and isotonic regression. Also a method for estimating the conditional density based on local linear regression is proposed.

In the following, we first present the model (Section 2), then the estimation procedures including a causal interpretation and asymptotic results (Section 3). Afterwards. we consider the monotonizing methods (Section 4) and some simulation evidence (Section 5). An application to a Mincer-type income regression (Section 6) demonstrates the importance in empirical practice to take endogeneity into account for estimating DR models and to use the new method. Section 7 makes some suggestions for future research. The Appendix gives further technical insights.

2. General semiparametric distribution regression

Consider an outcome variable Y and regressors $X_1, \dots, X_k$. In the semiparametric DR model, the conditional distribution function of Y given the set of regressors X is modeled by $F_{Y|X}(y|x) = \Lambda(x'\beta(y))$ for some link function Λ such as the distribution function of the standard normal distribution Φ and some function $\beta(y)$. Although the function Λ must be chosen in advance, in this paper, the model is called semiparametric: Usually, in the literature, no explicit restrictions on $\beta(y)$ such as continuity are imposed and there is a parameter for every y. Anyway, it is clear that $\beta(y)$ has to be chosen such that $F_{Y|X}(y|x)$ fulfills the properties of a conditional distribution function if the model is correctly specified. Given x, the function must be monotonically increasing and converge to 1 and 0 for $y \to \infty$ and $y \to -\infty$, respectively.

In the simple linear model $Y = 1 + X_2 + U$, where U is distributed with distribution function Λ and independent of X_2, it holds $F_{Y|X}(y|x) = \Lambda(y - 1 - x_2)$ (with $x = (1, x_2)$) such that $\beta(y) = (y - 1, -1)$ and the monotonicity condition is fulfilled. If, for example, Y describes wages of employees and there is minimum wage at some value y^*, the function $\beta(y)$ might contain a discontinuity point at y^*. For example, if 10% of the employees with characteristics x receive minimum wage y^*, $F_{Y|X}(y|x)$ would jump from 0 to 0.1 at y^*, which implies a jump in $\beta(y)$. This illustrates that Y may be discretely distributed even if the link function is continuous. There is no one-to-one connection between the distribution of Y and the link function.

Based on an i.i.d. sample of length n, $F_{Y|X}(y|x)$ can be consistently estimated by maximum likelihood estimation similarly as a probit model would be estimated, for example. The estimation is performed separately for each y and requires that the regressors are exogenous. To be precise, one introduces the indicator functions $I_y := 1\{Y \le y\}$ with $E(I_y|X = x) = \Lambda(x'\beta(y))$. The model can be interpreted as a latent variable model with $I_y = 1$ if $\tilde{I}_y := X'\beta(y) + U \ge 0$ and $I_y = 0$ otherwise. The random variable U is distributed with distribution function Λ and exogeneity means that X is independent of U.

3. Semiparametric distribution regression with instruments

3.1. Model and estimation procedure

There are different IV-based approaches for estimating binary outcome models if X and U are not independent. We focus on the two-step estimator based on the idea of control functions introduced by Rivers and Vuong (1988)[2] and present an adapted three-step estimator for the case of estimating such models separately for each y. In particular cases, as we describe below, this estimator is numerically equivalent to the original maximum likelihood estimator, which is asymptotically efficient for fixed y, but computationally less appealing. The maximum likelihood estimator was brought forward and discussed by Amemiya (1978), Newey (1987), Rivers and Vuong (1988), is explained in detail in Wooldridge (2002), Section 15.7.2 and Hansen (2022), Section 25.12.

We focus on the case of one endogenous regressor. Using the notation from the literature, X denotes a k-dimensional vector with exogenous regressors, Y_2 is scalar, endogenous and Z is a l-dimensional vector with exogenous instruments. Moreover, we consider a variable V which is a possible confounder (such as ability/motivation) and has distribution function F_V. In our situation, the goal is to estimate

$$\int E(I_y|X = x, Y_2 = y_2, V = v) dF_V(v) =: F^V_{Y|X,Y_2}(y|x, y_2), \tag{3.1}$$

i.e., we first control for possible confounders and integrate these confounders out afterwards. If Y_2 is exogenous conditioned on V, the integral in (3.1) can be interpreted as the conditional distribution function of Y given X and a hypothetically exogenized Y_2. Standard probit would be a suitable estimator for the conditional distribution function $E(I_y|X = x, Y_2 = y_2)$, but this is not the term we are interested in, see Remark 2 in Appendix A. The term in (3.1) is an example for an average structural function as considered in Blundell and Powell (2004) or Chernozhukov et al. (2020).

For estimating (3.1), we introduce a latent variable model similarly as for the standard DR model. The assumption is that $F^V_{Y|X,Y_2}(y|x, y_2) = \int P(I^*_y \ge 0|X = x, Y_2 = y_2, V = v) dF_V(v)$ with

$$\begin{aligned} I^*_y &= X'\beta_1(y) + Y_2\beta_2(y) + U(y) \\ Y_2 &= X'\gamma_1 + Z'\gamma_2 + V. \end{aligned} \tag{3.2}$$

Here, $U(y)$ and V are latent variables with mean zero, for which exists a decomposition

$$U(y) = \alpha_1 V + \alpha_2 \epsilon(y), \tag{3.3}$$

where $\epsilon(y)$ is independent from (X, Y_2, V), $E(V|X, Z) = 0$, $U(y)$ has variance 1, $\epsilon(y)$ is independent of V, has variance 1 and has some continuously differentiable distribution function $\tilde{\Lambda}$. Both the distribution functions of U and $\epsilon(y)$ are assumed to be known, i.e. they are an input for the estimation method. No a priori information is needed for the distribution of V. In this decomposition, we have endogeneity if $\alpha_1 \neq 0$.

Similarly as for the standard DR model, the parameters for the regressors X and Y_2 as well as the latent error variable $U(y)$ depend on y, which yields much flexibility despite the fact that the link function, i.e. the distribution function of U, is fixed to Λ and also the distribution of $\epsilon(y)$ has to be fixed.[3] All other parameters and random variables in the model do not depend on y in order to have a sparse parametrization. More remarks on the model can be found in Appendix A.

For estimation purposes, we use the i.i.d. sample $(I_{y,i}, Y_{2,i}, X_i, Z_i)$ with $I_{y,i} = 1\{Y_i \le y\}$. The two step estimator by Rivers and Vuong

[2] Blundell and Smith (1989) propose an estimator with a related goal, i.e. they focus on simultaneous equations in limited dependent variable models.

[3] In the simulations and the empirical application, we choose the standard normal distribution, respectively.

(1988) is explained in detail in Wooldridge (2002), Section 15.7.2. The estimator requires a non-trivial adjustment to our situation, however, because it does not directly estimate the parameters $\beta_1(y)$ and $\beta_2(y)$ consistently.

The estimator is based on the decomposition $U(y) = \alpha_1 V + \alpha_2 \epsilon(y)$ which leads to the equation

$$I_y^* = X'\beta_1(y) + Y_2\beta_2(y) + V\alpha_1 + \alpha_2\epsilon(y). \tag{3.4}$$

This means that, for fixed y, standard estimation of binary choice models with given distribution function of $\epsilon(y)$ with $Var(\epsilon(y)) = 1$ consistently estimates the parameter vector $\theta(y) := (\tilde{\beta}_1(y), \tilde{\beta}_2(y), \tilde{\alpha}_1)$ with $\tilde{\beta}_1(y) := \frac{\beta_1(y)}{\sqrt{\alpha_2}}, \tilde{\beta}_2(y) := \frac{\beta_2(y)}{\sqrt{\alpha_2}}, \tilde{\alpha}_1 := \frac{\alpha_1}{\sqrt{\alpha_2}}$ under Assumption 1.1 in Appendix A. As V is not observable, this term is replaced with the residuals of an OLS regression of Y_2 on X and Z.

The parameter α_2 is not known, so that these estimators cannot be used directly. However, they can be used to consistently estimate the conditional expectation

$$E(I_y|X = x, Y_2 = y_2, V = v) = \Lambda(x'\tilde{\beta}_1(y) + y_2\tilde{\beta}_2(y) + v\tilde{\alpha}_1).$$

Summing up the previous discussion, it turns out that the following three steps have to be performed for getting a consistent estimator for $F^V_{Y|X,Y_2}(y|x, y_2)$ from (3.1) for fixed y:

1. Run an OLS regression of Y_{2i} on X_i and Z_i and obtain residuals V_i, $i = 1, \dots, n$.
2. Run a binary choice estimation of $I_{y,i}$ on X_i, Y_{2i} and V_i and obtain the estimators $\widehat{\tilde{\beta}_1(y)}$, $\widehat{\tilde{\beta}_2(y)}$ and $\widehat{\tilde{\alpha}_1(y)}$.
3. The final estimator is then given by

$$\hat{F}^V_{Y|X,Y_2}(y|x, y_2) := \frac{1}{n}\sum_{i=1}^{n} \Lambda(x'\widehat{\tilde{\beta}_1(y)} + y_2\widehat{\tilde{\beta}_2(y)} + V_i\widehat{\tilde{\alpha}_1(y)}).$$

Due to the transformation of Y to $1\{Y \le y\}$ for the estimation, the estimator $\hat{F}^V_{Y|X,Y_2}(y|x, y_2)$ can attain at most n different values for fixed X and Y_2. The differences arise at the different outcomes Y_i, so that it is reasonable to evaluate the estimated distribution function at all Y_i, if computationally feasible.

3.2. Asymptotic result

For each y, the estimator of the transformed parameters can be equivalently calculated by maximizing the likelihood function or by minimizing some norm of the score function. Thus, the estimator falls into the framework of Z-estimators analyzed in Chernozhukov et al. (2013) and one can derive consistency and asymptotic normality, both pointwisely and uniformly in y. Then, under some additional assumptions as described in the Appendix, we obtain

Theorem 1. *Let Assumption 1 be fulfilled. Then it holds that*

$$\sqrt{n}\left(\hat{F}^V_{Y|X,Y_2}(\cdot|x, y_2) - F^V_{Y|X,Y_2}(\cdot|x, y_2)\right) \tag{3.5}$$

converges to a Gaussian process $\mathbb{G}(\cdot)$ in $l^\infty(\mathcal{U})$, the space of all bounded functions indexed by a compact subinterval $\mathcal{U}$ of $\mathbb{R}$.

The proof of this theorem shows that the limit process depends both on the limit properties of $\hat{\theta}(y)$, which is the estimator for $\theta(y)$, and the shape of the function $F^V_{Y|X,Y_2}(y|x, y_2)$.

4. Monotonicity

While the proposed estimators from the last section are consistent under appropriate assumptions, there is no reason to assume that the estimated conditional distribution functions are monotonically increasing in y in finite samples. This might be a drawback for interpretation purposes, e.g. if the estimators are used for calculating conditional quantiles and it turns out that the estimated 90%-quantile is smaller than the estimated 80%-quantile. Oliveira (2023) points out the problem of missing monotonicity for estimating minimum wage effects. We discuss two methods to fix this, monotone rearrangement as well as isotonic regression.[4] Note that, as in Wüthrich (2019), the monotonizations have to be performed separately for each value of (x, y_2).

4.1. Monotone rearrangement

Chernozhukov et al. (2010) propose a monotone rearrangement approach (see also Dette et al., 2006), mainly for quantile regression in order to ensure that estimated conditional quantiles do not cross. As discussed in Chernozhukov et al. (2013), this approach can also be applied to distributional regression. Using the conditional quantile function $Q^V_{Y|X,Y_2}(u|x, y_2) := \inf_y\{F^V_{Y|X,Y_2}(y|x, y_2) \ge u\}$, it is based on the identity

$$F^V_{Y|X,Y_2}(y|x, y_2) = \int_0^1 \mathbf{1}\{Q^V_{Y|X,Y_2}(u|x, y_2) \le y\}du, \tag{4.1}$$

so that in a first step the conditional quantile function needs to be estimated, before it is appropriately integrated. This leads to the estimator

$$\tilde{F}^V_{Y|X,Y_2}(y|x, y_2) = \int_0^1 \mathbf{1}\{\hat{Q}^V_{Y|X,Y_2}(u|x, y_2) \le y\}du$$

with the estimated conditional quantile function[5]

$$\hat{Q}^V_{Y|X,Y_2}(u|x, y_2) = \inf_y\{\hat{F}^V_{Y|X,Y_2}(y|x, y_2) \ge u\}.$$

The asymptotic properties of this estimator are well understood. As discussed in Chernozhukov et al. (2010), given a result like Eq. (3.5) from the last section, the convergence rate (in our case $\sqrt{n}$) carries over due to the Hadamard differentiability of the operator from (4.1) and an application of the functional delta method. Moreover, it is possible to estimate the limit process by a bootstrap approximation.

4.2. Isotonic regression

An alternative to the monotone rearrangement is the application of an isotonic regression, which can be applied directly on the functional estimator. This estimation procedure is discussed in Barlow et al. (1972) and Robertson et al. (1988), for example.[6] By construction, the estimated distribution function only changes its value at the observed $Y_1, \dots, Y_n$ and is constant between these points. The idea is to replace the points $\hat{F}_i := \Lambda(x'\hat{\beta}_1(Y_i) + y_2\hat{\beta}_2(Y_i))$ by points $\tilde{\hat{F}}_i$ that are close to $\hat{F}_i$, but fulfill the monotonicity restriction. This means that one solves the quadratic minimization problem

$$min_{\tilde{\hat{F}}_1,\dots,\tilde{\hat{F}}_n} \sum_{i=1}^{n}\left(\tilde{\hat{F}}_i - \hat{F}_i\right)^2$$

under the constraint

$$\tilde{\hat{F}}_i \le \tilde{\hat{F}}_j \text{ for } Y_i \le Y_j. \tag{4.2}$$

The problem can be solved numerically with the *pool adjacent violators algorithm*, an implementation in software packages such as R (command *isoreg* in the package *stats*) is available. The computational complexity for given n is $O(n)$ for already sorted data, see Best and Chakravarti (1990). A potential drawback is the tendency to obtain flat functions, which leads to a bias in finite samples, if the true distribution function is strictly increasing.

[4] Foresi and Peracchi (1995) discuss in their Section 2.1 some other possibilities to get monotone estimators of the distribution function, but do not elaborate on them in more detail.

[5] A researcher only interested in conditional quantiles could of course directly use this estimator.

[6] Henzi et al. (2021) also consider isotonic distributional regression, but they consider monotonicity in (x, y_2)-, and not in y-direction.

Table 1
Average squared bias, squared variance and squared MSE of the two monotonizing approaches with non-IV probit and IV probit.

Values for $x = y_2$	n	Monotone rearrangement			Isotonic regression		
		Bias^2	Var	MSE	Bias^2	Var	MSE
Non-IV							
1	100	0.0119	0.0085	0.0205	0.0102	0.0083	0.0185
	200	0.0108	0.0036	0.0145	0.0100	0.0035	0.0136
	400	0.0099	0.0017	0.0116	0.0094	0.0017	0.0116
2	100	0.0076	0.0169	0.0245	0.0043	0.0161	0.0205
	200	0.0049	0.0076	0.0125	0.0038	0.0075	0.0113
	400	0.0048	0.0039	0.0087	0.0044	0.0039	0.0083
IV							
1	100	0.0004	0.0098	0.0102	$5 \cdot 10^{-5}$	0.0094	0.0094
	200	$8 \cdot 10^{-5}$	0.0044	0.0045	$2 \cdot 10^{-5}$	0.0042	0.0043
	400	$1 \cdot 10^{-5}$	0.0022	0.0022	$< 1 \cdot 10^{-5}$	0.0021	0.0021
2	100	0.0007	0.0121	0.0128	$7 \cdot 10^{-5}$	0.0103	0.0104
	200	$4 \cdot 10^{-5}$	0.0049	0.0050	$< 1 \cdot 10^{-5}$	0.0047	0.0047
	400	$1 \cdot 10^{-5}$	0.0023	0.0024	$< 1 \cdot 10^{-5}$	0.0023	0.0023

Having obtained a monotonically increasing distribution function for the points

$Y_1, \dots, Y_n$, forecasts for other values of y might be obtained by linear interpolation, for example. Also conditional quantiles can be calculated in this way.

If (4.2) already holds for the $\hat{F}_i, i = 1, \dots, n$, $\sum_{i=1}^{n} \left(\tilde{\tilde{F}}_i - \hat{F}_i \right)^2$ is equal to 0. So, it is intuitive that the monotonized estimator is consistent if the true conditional distribution function is monotonically increasing and the estimated distribution function is uniformly consistent (over y).

5. Simulations

We simulate from the model $Y^* = \max(2, \tilde{Y})$ and

$$\tilde{Y} = 1 + X + Y_2 + U$$
$$Y_2 = 1 + X + Z + V,$$

where X and Z are i.i.d. $N(0,1)$-distributed and (U, V) is bivariate normally distributed with zero mean and covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Here, X represents the exogenous regressor, Y_2 the endogenous regressor, which is correlated with U, and Z the exogenous instrument. We consider a censored $\tilde{Y}$, which mimics the application of modeling wages with a minimum wage and highlights the appealing property of distributional regression of detecting such censoring points. In this case, $F_{Y|X,Y_2}(y|x, y_2) = \Phi(y - 1 - x - y_2)$ for $y \geq 2$ and 0 elsewhere. We fix $\rho = 0.7$ and calculate $\tilde{F}_{Y|X,Y_2}(y|x, y_2)$ (monotone rearrangement) and $\tilde{\tilde{F}}_{Y|X,Y_2}(y|x, y_2)$ (isotonic regression) for $x = y_2 = 1$ and $x = y_2 = 2$. As $E(X) = 0$ and $E(Y_2) = 1$, X and Y_2 are further away from their expectations in the latter case. The grid points for y are equidistant in the interval $[1, 5]$ with 50 grid points in total. For the rearrangement, the quantile levels are equidistant in the interval $[0.01, 0.99]$ with 99 grid points in total. To mimic the setting of the empirical application, the sample sizes are $n = 100, 200, 400$. For each case, 1000 Monte Carlo replications are performed. The results are compared with the standard probit estimates that ignore the endogeneity.

As we are concerned with uniform convergence to the true function (see Theorem 1), we consider the average squared bias, the average variance and the average MSE of $\tilde{F}_{Y|X,Y_2}(y|x, y_2)$ and $\tilde{\tilde{F}}_{Y|X,Y_2}(y|x, y_2)$ over the grid of 41 y-values. Tables 1 shows the results.

With the IV approach, the average MSE is dominated by the variance and is similar for both procedures with a slight advantage for the isotonic regression. Bias, variance and MSE are slightly higher if x and y_2 are further away from their expectations and half when the sample size is doubled. This suggests that, in this setup, the convergence rate of both estimators is $\sqrt{n}$.[7] The variance of the non-IV approach also halves with doubled sample size and slightly exceeds that of the IV approach for $x = y_2 = 1$, but is considerably biased as expected. So, its MSE is much higher than that of the IV approach.

6. Application to a Mincer-type income equation

We consider income data from the German SOEP from the year 2020 (Sozio-oekonomisches Panel/Socio-Economic Panel 2022) with $n = 1694$ individuals and estimate a Mincer-type regression to estimate the returns of education. We do not have precise information about the hourly wage, but we have the variable *gross income per month per employee*, measured in EUR. The logarithm of is explained by the *years of education* and the *years of working experience* (the latter both linearly and quadratically). To adjust for the working time, we additionally use the explanatory variable *part time*, which is 1, if the employee works in full time throughout the year.

The variable *educ* is assumed to be endogenous, as it might be correlated with unobserved variables such as ability or motivation. The variable *exper* is calculated from "current age minus age when education was finished" compare Krenz (2008). So there might be some correlation with unobserved characteristics as well, but we assume that this is negligible. In particular, the correlation between *educ* and *exper* is rather small with 0.157. So we consider the variable *exper* as being exogenous.

In the SOEP, the education level is categorized. For this analysis, the information was transformed to one metric variable, yielding values of 10 (lower secondary education), 13 (secondary education), 14 (post-secondary non tertiary education) and 16 (first stage of tertiary education). In the study, only individuals who provided information on all these variables and who had at least one year of working experiences, were included. (In cases, where education information for only one parent is available, the other variable was set to 10.) This way, the original sample consisting of over 30000 individuals is reduced to $n = 1694$.

A possible instrument for the years of education is the years of education of the mother. In this dataset, the first stage F-statistic is given by approximately 67 so that the instrument can be assumed to be sufficiently strong. See Wooldridge (2016) for some discussion why this model might be reasonable.

First, we estimate a simple linear model with OLS and with IV:

$$log(income)_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + \beta_4 partt_i + \varepsilon_i.$$

Table 2 shows the estimated coefficients as well as the estimated conditional expectations for the 10%, 50% and 90% quantiles of *educ* (10, 13 and 16) and *exper* (1, 4 and 11, respectively. The variable *partt* is always set to its mean (0.48). This way, the expected incomes are calculated for three groups of employees, the low-educated/low-experienced, the middle-educated/middle-experienced and the high-educated/high-experienced.

Similarly as in other studies with this type of instrument, the IV estimate for *educ* is smaller than the OLS estimate (while the standard error is larger). The intuition is that both the years of education and an unobserved variable which measures ability and/or motivation are positively correlated with the income. There is a differentiated discussion about this in the literature: While the IV estimates are often

[7] In other contexts, the convergence rate of isotonic regression is smaller than $\sqrt{n}$, for example $n^{1/3}$ in Abrevaya (2005). In these cases, the standard bootstrap (drawing with replacement) might behave erratically, see Patra et al. (2018). The Monte Carlo evidence suggests that such problems should not be expected in the present context, at least not for the setting considered in the empirical application. The intuition is that in our case, the isotonic regression is just a finite-sample correction in second step of an estimator which asymptotically fulfills the monotonicity restriction.

Table 2
Estimated regression coefficients and conditional expectations for the linear model, standard errors in parentheses.

$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{E}(log(income))$		
					$ed = 10$ $ex = 1$	$ed = 13$ $ex = 4$	$ed = 16$ $ex = 11$
OLS							
5.0704	0.1331	0.0319	−0.0007	0.8256	6.8347	7.3191	7.8664
(0.1165)	(0.0090)	(0.0068)	(0.0002)	(0.0379)	(0.0365)	(0.0182)	(0.0381)
IV							
5.9048	0.0664	0.0421	−0.0009	0.8382	7.0183	7.3295	7.7237
(0.4833)	(0.0385)	(0.0090)	(0.0002)	(0.0391)	(0.1096)	(0.0194)	(0.0890)

larger if institutional characteristics are used as instruments (Card, 2000), Lemke and Rischall (2003) and Breitung et al. (2024) give evidence that the OLS estimators might be biased upwards in larger models (number of individuals and regressors). Indeed, a robustness analysis with the years of education of the father yields results which point into the same direction. Moreover, the p-value of the Sargan test in the standard IV model using both instruments together is large (0.64), which indicates that both instruments are appropriate.

The conditional expectations increase if higher values of *educ* and *exper* are considered. Interestingly, the results for OLS and IV are similar for the 50% quantiles. For the 10% quantiles, the IV estimate is larger than the OLS estimate, for the 90% quantile, the IV estimate is smaller. There seems to be a tendency that the variability in terms of the regressor values is lower for the IV estimation. These results will be confirmed and extended by the DR analysis.

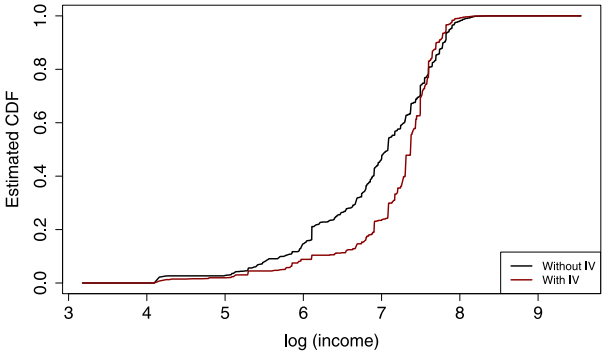
Fig. 6.1 shows the estimated conditional distribution functions for both non-IV and IV, again for the $10\%, 50\%$ and 90% quantiles of *educ* and *exper*. The estimated distribution functions are evaluated at all outcomes Y_i. In all cases, the monotonized version based on isotonic regression discussed in the last section is considered. For higher values of *educ* and *exper*, the distribution functions are shifted more and more to the right. For the 50% quantile, the two functions are rather similar. For the 10% quantile, the IV curve generally lies to the right of the non-IV curve, where the largest differences are visible for values of *log(income)* between 1 and 2. For the 90% quantile, the IV curve generally lies to the left below the 5%-quantile of *log(income)* and to the right above that. This plot particularly illustrates the usefulness of considering the whole distribution, not only the conditional mean.

The plots for the 10% and the 50% quantile of the regressors also show some jump points, in particular for *income* = 450 which implies *log(income)* = 6.1. This fits to the data, where some point mass lies on this value: 61 out of 1694 individuals report this number. This high number is not surprising because the "450 Euro-job" was a typical type of part-time job in Germany in the year 2020 (Bundeszentralamt für Steuern, 2024).

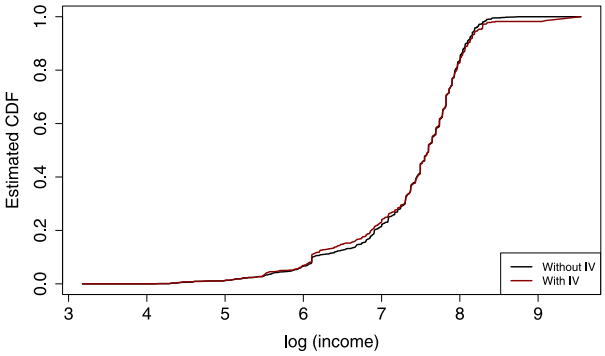
For completeness, Fig. 6.2 shows the estimated DR curve for the 90% quantiles without monotonization, illustrating why it makes sense to add the monotonizing step.

To give more evidence about the difference between non-IV and IV estimation, pointwise confidence bounds for the differences of the conditional distribution functions are calculated and plotted in Fig. 6.3. This is done by bootstrap, i.e. by drawing with replacement $B = 100$ times from the individuals. For each y, the confidence interval to the level of significance 90% is calculated. This yields a Hausman-type statistical test for the relevance of the IV approach: If 0 is not contained in the interval, one can conclude that the two estimators of the distribution functions are statistically significantly different. Assuming that the instrument is exogenous and correlated with the endogenous regressor, the IV-based estimator is then the only valid one.

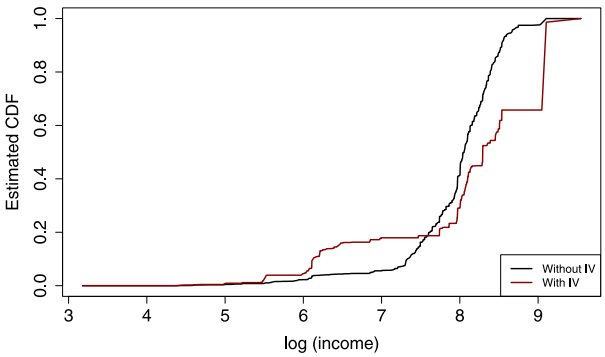
The confidence bounds essentially confirm the analysis from Fig. 6.1. For the 10% as well as for the 90% quantile, the bounds do not contain 0 for some subsets of the ranges of $log(income)$ described above. For the 50% quantile, 0 is mostly contained.



(a) 0.1-quantile of the regressors



(b) 0.5-quantile of the regressors



(c) 0.9-quantile of the regressors

Fig. 6.1. Estimated conditional distribution functions.

Summed up, the message is that IV-estimation of DR models does make a difference compared to non-IV-estimation. If the regressor *educ* is hypothetically exogenized, for large values of the regressors, the upper conditional quantiles of the income tend to be larger, for small values, the lower conditional quantiles tend to be larger. This means that we have less inequality across the regressor values for the lower conditional quantiles and more for the upper conditional quantiles.

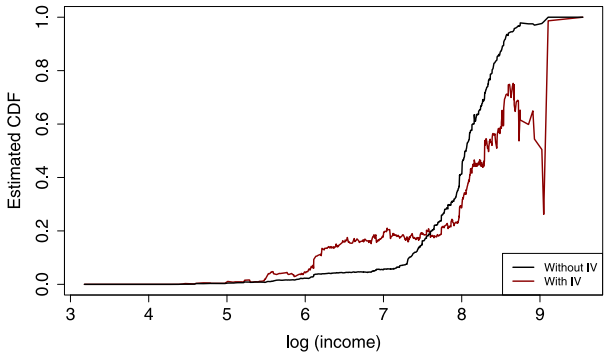


Fig. 6.2. Estimated conditional distribution function without monotonicity constraint for the 0.9-quantile of the regressors.

7. Conclusion and outlook

The paper proposes a new consistent estimator for the semiparametric DR model which allows for endogenous regressors and where monotonicity is enforced. The method is easy to implement and should be appealing to practitioners. Apparently, the proposed procedure only works well if the instruments are sufficiently strong. To circumvent the problem of choosing appropriate instruments, it might be an idea for future research to adapt the procedure proposed by Breitung et al. (2024) for linear regression models to DR models. Here, rank-based transformations of non-normal regressors are used as additional regressors and no external instruments are necessary to obtain consistent parameter estimators. Other tasks for future research would be a framework for handling non-i.i.d. data, e.g. individuals in different clusters with different variances. Also relaxing the linearity assumption for the first stage in the spirit of Blundell and Powell (2004) or Imbens and Newey (2009) might be interesting.

CRediT authorship contribution statement

Dominik Wied: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

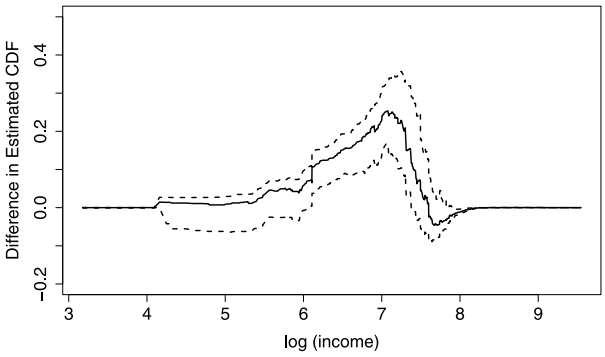
Data availability

The data was downloaded from the German SOEP, see the beginning of Section 6.

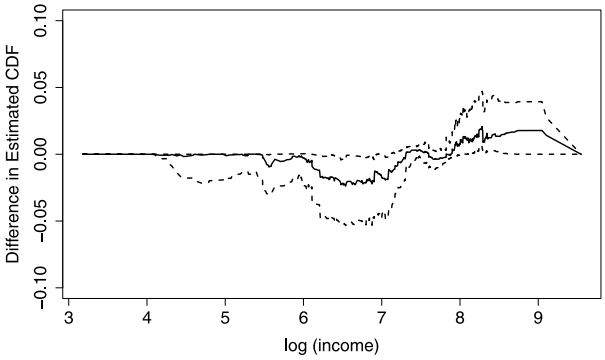
Appendix A. Additional remarks on the model

This section contains some remarks about the model assumptions.

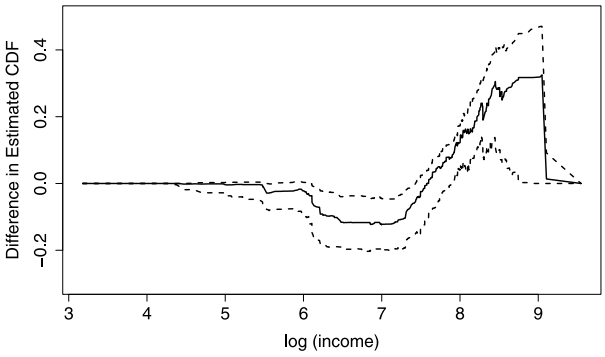
1. A sufficient condition for the existence of the decomposition (3.3) is that the random variables $U(y)$ and V are jointly normally distributed conditionally on X and Z,
$$\begin{pmatrix} U(y) \\ V \end{pmatrix} |(X,Z) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right). \tag{A.1}$$
In this case, with $\rho := \frac{\sigma_{12}}{\sigma_2}$, a possible parametrization is $\alpha_1 = \frac{\rho}{\sigma_2}$ and $\alpha_2 = \sqrt{1-\rho^2}$. The requirement of continuous V then implies that Y_2 is also continuously distributed.
2. In the case of (A.1), we have $I_y^* = X'\beta_1(y) + Y_2\beta_2(y) + \alpha_1 V + \alpha_2 \epsilon(y)$. As
$$\int \Phi(a+bx)\phi(x)dx = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right),$$



(a) 0.1-quantile of the regressors



(b) 0.5-quantile of the regressors



(c) 0.9-quantile of the regressors

Fig. 6.3. Difference between estimated conditional distribution functions and confidence bounds.

see Ellison (1964), we then have

$$F^V_{Y|X,Y_2}(y|x,y_2) = P(U(y) \le x'\beta_1(y) + y_2\beta_2(y)) = \Phi(x'\beta_1(y) + y_2\beta_2(y)), \tag{A.2}$$

which would the same as $E(Y \le y|X = x, Y_2 = y_2)$ if $U(y)$ were independent of X and Y_2, see Remark 1.3 below. So, (A.2) can be interpreted as the expression for the conditional distribution function in the hypothetical case that Y_2 is exogenized. From (A.2) we observe that consistent estimation of the parameters in (3.2) leads to consistent estimation of $F^V_{Y|X,Y_2}(y|x,y_2)$ by the continuous mapping theorem.

3. Assuming joint normality of $(U(y), V, X, Z)$, $U(y) = \psi Y_2 + C(y)$ for $\psi = \frac{\sigma_{12}}{\sigma^2_{Y_2}} = \frac{\rho\sigma_2}{\sigma^2_{Y_2}}$, where $C(y)$ is independent of Y_2 and $\sigma^2_{Y_2} = Var(Y_2)$. Then, similarly as in Li et al. (2022),
$$P(Y \leq y|X = x, Y_2 = y_2) = \Phi\left(\frac{x'\beta_1(y) + \left(\beta_2(y) + \psi\right) y_2}{\sqrt{1-\tau^2}}\right)$$
with $\tau = \frac{\sigma_{12}}{\sigma_{Y_2}} = \frac{\rho\sigma_2}{\sigma_{Y_2}}$. This is the quantity that standard probit would estimate, but this is not the quantity we are interested in. The present paper aims at estimating $\int P(Y \leq y|X = x, Y_2 = y_2, V = v) dF_V(v)$, which is a different object.
4. For fixed y and under condition (A.1), the estimator obtained in Step 2 in the end of Section 3.1 is the two step estimator by Rivers and Vuong (1988). In the case of just identified models (one instrument for the endogenous regressor), this estimator is then numerically equal to the maximum likelihood estimator for $\theta(y)$.[8]

Appendix B. Assumptions and Proof of Theorem 1

The assumptions required for Theorem 1 are formulated in terms of the score function which is based on the binary choice model (3.4) $\Psi : \Theta \times \mathcal{I} \to \Theta$ with Θ being a compact subset of $\mathbb{R}^{k+2}$ and an open interval $\mathcal{I}$ that covers a compact interval $\mathcal{U}$. Define
$$\Psi_i^*(\theta(y), y, v) = \left(I_{y,i}\frac{f_i(y)}{F_i(y)} + (1 - I_{y_i})\frac{-f_i(y)}{1-F_i(y)}\right)(X_i, Y_{2i}, v)'$$
with $F_i(y) = \tilde{\Lambda}(X_i'\tilde{\beta}_1(y) + Y_{2i}\tilde{\beta}_2(y) + v\tilde{\alpha}_1)$, $f_i(y) = dF_i(y)/d\theta(y)$ and $\theta(y) = (\tilde{\beta}_1(y), \tilde{\beta}_2(y), \tilde{\alpha}_1)$.

Then it holds (see Greene, 2017, eq. 17–17) that $\Psi(\theta(y), y) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \Psi_i^*(\theta(y), y, \hat{V}_i)$. In the remaining part of this appendix, we denote the true parameter by $\theta_0(y)$. We impose similarly to Lemma E.1 and Lemma E.3 in Chernozhukov et al. (2013).

Assumption 1.

1. (a) $\Psi : \Theta \times \mathcal{I} \mapsto \Theta$ is continuous, and $\theta \mapsto \Psi(\theta, y)$ is the gradient of a convex function in θ for each $y \in \mathcal{U}$, (b) for each $u \in \mathcal{U}$, $\Psi\left(\theta_0(y), y\right) = 0$, (c) $\frac{\partial}{\partial(\theta', y)}\Psi(\theta, y)$ exists at $\left(\theta_0(y), y\right)$ and is continuous at $\left(\theta_0(y), y\right)$ for each $y \in \mathcal{U}$, and $\dot{\Psi}_{\theta_0(y),y} := \frac{\partial}{\partial\theta'}\Psi(\theta, y)\Big|_{\theta_0(y)}$ obeys $\inf_{y\in\mathcal{U}}\inf_{\|h\|=1}\left\|\dot{\Psi}_{\theta_0(y),y}h\right\| > c_0 > 0$.
2. The function class $\mathcal{D} := \{\Psi_i^*(\theta, y, V_i) : \theta \in \Theta, y \in \mathcal{I}\}$ is Donsker with square integrable envelope.
3. (a) For fixed y and $\theta(y)$, $\Psi_i^*(\theta, y, v) =: f_y(v)$ as a function of v is two times continuously differentiable with $\frac{1}{n}\sum_{i=1}^n \frac{d^2 f_y(\tilde{V}_i)}{d^2 v} = O_p(1)$, where $\tilde{V}_i$ lies between V_i and $\hat{V}_i$. The analogous property is true for $E(f_y(v))$. (b) The regressor matrix for the first stage $\mathbf{A}_n := (\mathbf{X}_n\mathbf{Z}_n)$ has full column rank and is independent from $\mathbf{V}_n := (V_1, \ldots, V_n)$, where $\mathbf{X}_n = (X_1, \ldots, X_n)'$, $\mathbf{Z}_n = (Z_1, \ldots, Z_n)'$. (c) $E(df_y(V_i)/dV_i \cdot \mathbf{A}_i)$ is finite.

Ass.1.1 is required for the pointwise convergence of $\hat{\theta}(y)$ to $\theta_0(y)$. A crucial point here is the positive definiteness of the derivative matrix of the score vector in Ass. 1.1.(a), from which the existence of a unique Z-estimator follows. Results from Newey (1987) (Assumption A.3.(v)) or Amemiya (1978), Section 6, yield that this holds in the just identified case if (A.1) holds, if $E_{XZ} := E((X_i, Z_i)(X_i, Z_i)')$ is invertible, if the true parameters lie inside of the parameter space and if $\gamma_2 \neq 0$, see Rivers and Vuong (1988). The invertibility condition implies that weak instruments might be problematic for the estimation procedure.

Ass.1.2 concerns the uniform convergence of $\hat{\theta}(y)$ to $\theta_0(y)$. Due to the boundedness of Θ, the (then assumed to be finite) norm of the matrix E_{XZ} can be chosen as the envelope in Ass. 1.2., see Step 3 in the proof of Theorem 5.2 in Chernozhukov et al. (2013) and Example 19.7 in van der Vaart (1998). Then the Donsker property holds with the observation that the function class $\mathcal{D}$ is a Lipschitz transformation of VC classes.

Note that Ass. 1.2 is formulated in terms of the score function for the case that $\hat{V}_i$ is replaced by V_i. To ensure that the resulting estimation error behaves in an unproblematic way, the high-level assumption Ass. 1.3 is imposed. This is fulfilled e.g. for a probit link function and if standard assumptions concerning the asymptotic normality of OLS estimators are fulfilled.

Proof of Theorem 1. Denote $l^\infty(\mathcal{A})^p$ the space of p-dimensional bounded functions with index sex $\mathcal{A}$. We first show that, with Assumption 1.2, Assumption 1.3 and a Taylor expansion $\sqrt{n}(\hat{\Psi}(\cdot,\cdot) - \Psi(\cdot,\cdot)) \Rightarrow_d A$ in $l^\infty(\Theta\times\mathcal{U})^{k+2}$, where A is a Gaussian process. It holds with a Taylor expansion
$$\begin{aligned}
\sqrt{n}(\hat{\Psi}(\theta(y), y) - \Psi(\theta(y), y)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Psi_i^*(\theta(y), y, \hat{V}_i) - E(\Psi_i^*(\theta(y), y, \hat{V}_i)) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Psi_i^*(\theta(y), y, V_i) - E(\Psi_i^*(\theta(y), y, V_i)) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{df_y(V_i)}{dV_i}(\hat{V}_i - V_i) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{dE(f_y(V_i))}{dV_i}(\hat{V}_i - V_i) + D_n \\
&=: A_n(y) + B_n(y) + C_n(y) + D_n(y).
\end{aligned}$$
With an application of the Cauchy–Schwarz inequality, $D_n(y) = o_p(1)$. $A_n(y)$ converges in distribution to a stochastic process due to the Donsker property (Ass. 1.2). It holds
$$B_n(y) = \frac{1}{n}\left(\frac{df_y(V_1)}{dV_1} \ldots \frac{df_y(V_n)}{dV_n}\right)\mathbf{A}_n\left(\frac{1}{n}\mathbf{A}_n'\mathbf{A}_n\right)^{-1}\frac{1}{\sqrt{n}}\mathbf{A}_n\mathbf{V}_n,$$
which converges to a normally distributed random variable. The analogous argument is valid for $C_n(y)$.

Then, with Assumption 1.1, Condition Z in Chernozhukov et al. (2013) holds and $u \mapsto \theta_0(u)$ is continuously differentiable. Then, with Lemma E.3 in Chernozhukov et al. (2013),
$$\begin{aligned}
\sqrt{n}\left(\hat{\theta}(\cdot) - \theta_0(\cdot)\right) &= -\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot}\sqrt{n}(\hat{\Psi} - \Psi)\left(\theta_0(\cdot), \cdot\right) \\
&+ o_{\mathbb{P}}(1) \rightsquigarrow -\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot}\left[A\left(\theta_0(\cdot), \cdot\right)\right] =: G(\cdot) \qquad \text{(B.1)}
\end{aligned}$$
in $l^\infty(\mathcal{U})^{k+2}$.

We have the integral representation
$$\hat{F}^V_{Y|X,Y_2}(y|x, y_2) = \int \Lambda\left(x'\widehat{\beta_1(y)} + y_2\widehat{\beta_2(y)} + v\hat{\tilde{\rho}}\right) d\hat{F}_n(v),$$
where $\hat{F}_n(v)$ is the empirical distribution function of the OLS residuals $\hat{v}_i$. As V has distribution function F_V, the population analogon is
$$F^V_{Y|X,Y_2}(y|x, y_2) = \int \Lambda\left(x'\tilde{\beta}_1(y) + y_2\tilde{\beta}_2(y) + v\tilde{\rho}\right) dF_V(v).$$
The empirical process $\sqrt{n}\left(\hat{F}_n(\cdot) - \Lambda(\cdot)\right)$ converges to a Gaussian process (see e.g. Chen and Lockhart, 2001) in $l^\infty(\mathbb{R})$. Then the result follows from Theorem 1 and applying the functional delta method on the vector
$$\sqrt{n}\begin{pmatrix}\hat{\theta}(\cdot) - \theta_0(\cdot) \\ \hat{F}_n(\cdot) - F_V(\cdot)\end{pmatrix}. \quad \square$$

Appendix C. Special case of normally distributed latent variables

Under the normality assumption (A.1) (which has no immediate implications for the distribution of Y, as mentioned earlier) and in the case of a just identified model, the estimators for the transformed parameters $\theta(y)$ are equivalent to the estimators which are obtained

[8] This is also true for the AGLS estimator from Amemiya (1978), which is implemented in the *R*-package *ivprobit*.

from maximum likelihood estimation. In this case, the parameter vector is given by $\vartheta(y) := (\beta_1(y), \beta_2(y), \gamma_1, \gamma_2, \rho, \sigma_2^2)$ (note that σ_{12} and σ_ϵ^2 can be calculated from the other parameters). Similarly as in Hansen (2022), the likelihood is derived by factorizing the joint density of I_y and Y_2. The log-likelihood is then essentially the sum of the standard regression and the standard probit log-likelihood. It is given as $L_y(\vartheta(y)) = \sum_{i=1}^n L_{y,i}(\tilde{\vartheta}(y))$ with

$$L_{y,i}(\tilde{\vartheta}(y)) = I_{y,i} \log \Phi\left(\frac{\mu_{y,i}(\vartheta(y))}{\sigma_\epsilon}\right) + (1 - I_{y,i}) \log \Phi\left(1 - \frac{\mu_{y,i}(\vartheta(y))}{\sigma_\epsilon}\right) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma_2^2 - \frac{1}{2\sigma_2^2}\left(Y_{2,i} - X_i'\gamma_1 - Z_i'\gamma_2\right)^2.$$

It holds $\mu_{y,i}(\vartheta(y)) = X_i'\beta_1(y) + Y_{2,i}\beta_2(y) + \rho(Y_{2,i} - X_i'\gamma_1 - Z_i'\gamma_2)$ and $\sigma_\epsilon = \sqrt{1 - \rho^2\sigma_2^2}$.

For each y, the parameter estimator can be equivalently calculated by maximizing the likelihood function or by minimizing some norm of the score function. The score function appears as $\Psi_i^*(\vartheta(y), y) = \frac{\partial}{\partial\vartheta(y)} L_{y,i}(\vartheta(y)) =$

$$\begin{pmatrix} \frac{X_i}{\sigma_\epsilon} A_{y,i} \\ \frac{Y_{2,i}}{\sigma_\epsilon} A_{y,i} \\ \frac{X_i}{\sigma_\epsilon}\left(-\rho A_{y,i} + \frac{1}{\sigma_2^2}\left(Y_{2,i} - X_i'\gamma_1 - Z_i'\gamma_2\right)\right) \\ \frac{Z_i}{\sigma_\epsilon}\left(-\rho A_{y,i} + \frac{1}{\sigma_2^2}\left(Y_{2,i} - X_i'\gamma_1(y) - Z_i'\gamma_2\right)\right) \\ \frac{\partial}{\partial\rho}\frac{\mu_{y,i}(\vartheta(y))}{\sigma_\epsilon} A_{y,i} \\ \frac{\partial}{\partial\sigma_2^2}\frac{\mu_{y,i}(\vartheta(y))}{\sigma_\epsilon} A_{y,i} - \frac{1}{\sigma_2^2} + \frac{1}{2\sigma_2^4}\left(Y_{2,i} - X_i'\gamma_1 - Z_i'\gamma_2\right)^2 \end{pmatrix}$$

and

$$A_{y,i} = I_{y,i}\left(\Phi\left(\frac{\mu_{y,i}(\vartheta(y))}{\sigma_\epsilon}\right)\right)^{-1} \varphi\left(\frac{\mu_{y,i}(\vartheta(y))}{\sigma_\epsilon}\right) - (1 - I_{y,i})\left(\Phi\left(1 - \frac{\mu_{y,i}(\vartheta(y))}{\sigma_\epsilon}\right)\right)^{-1} \varphi\left(1 - \frac{\mu_{y,i}(\vartheta(y))}{\sigma_\epsilon}\right)$$

$$\frac{\partial}{\partial\rho}\frac{\mu_{y,i}(\vartheta(y))}{\sigma_\epsilon} = \frac{\sigma_\epsilon\left(Y_{2,i} - X_i'\gamma_1 - Z_i'\gamma_2\right) + \left(X_i'\beta_1(y) + Y_{2i}\beta_2(y) + \rho\left(Y_{2,i} - X_i'\gamma_1 - Z_i'\gamma_2\right)\right)\frac{\rho\sigma_2^2}{\sigma_\epsilon}}{\sigma_\epsilon^2}$$

$$\frac{\partial}{\partial\sigma_2^2}\frac{\mu_{y,i}(\vartheta(y))}{\sigma_\epsilon} = \frac{\left(X_i'\beta_1(y) + Y_{2i}\beta_2(y) + \rho\left(Y_{2,i} - X_i'\gamma_1 - Z_i'\gamma_2\right)\right)\frac{\rho\sigma_2^2}{\sigma_\epsilon}}{\sigma_\epsilon^2}$$

The maximum likelihood estimator is asymptotically efficient for fixed y and a similar result as Theorem 1 holds also in this case. As the estimators for $\beta_1(y)$ and $\beta_2(y)$ are different for both cases, the resulting limit process is different, however. In the case of the maximum likelihood estimator, note that $F^V_{Y|X,Y_2}(y|x, y_2) := \Phi(x'\beta_{1,0}(y) + y_2\beta_{2,0}(y))$ is a differentiable transformation of the first $k+1$ components of the vector $\tilde{\theta}_0(y)$. So, we can apply Theorem A.1 in Wied et al. (2012) to express the limit process of (3.5) by means of the functional delta method and the gradient of this function. To be precise, we have $f(a, b) = \Phi(x'a + y_2 b)$ and $Df(a, b) = \begin{pmatrix} x\varphi(x'a + y_2 b) \\ Y_2\varphi(x'a + y_2 b) \end{pmatrix}$. This leads to

$$\sqrt{n}\left(\hat{F}^{ML}_{V,Y|X,Y_2}(\cdot|x, y_2) - F^V_{Y|X,Y_2}(\cdot|x, y_2)\right) \Rightarrow_d Df(\beta_{1,0}(\cdot), \beta_{2,0}(\cdot))' \begin{pmatrix} G_{1,\dots,k}(\cdot) \\ G_{k+1}(\cdot) \end{pmatrix} \text{ in } l^\infty(\mathcal{U}), \tag{C.1}$$

where $G_{1,\dots,k}(\cdot), G_{k+1}(\cdot)$ are the components of the limiting process of $\sqrt{n}(\widehat{\tilde{\theta}(y)} - \tilde{\theta}_0(y))$, compare also (B.1).

References

Abrevaya, J., 2005. Isotonic quantile regression: Asymptotics and bootstrap. Sankhya: Indian J. Stat. 67 (2), 187–199.

Amemiya, T., 1978. The estimation of a simultaneous equation generalized probit model. Econometrica 46 (5), 1193–1205.

Barlow, R., Bartholomew, D., Bremner, J., Brunk, H., 1972. Statistical Inference Under Order Restrictions. Wiley, London.

Best, M., Chakravarti, N., 1990. Active set algorithms for isotonic regression; A unifying framework. Math. Program. 47 (1-3), 425–439.

Blundell, R., Powell, J., 2004. Endogeneity in semiparametric binary response models. Rev. Econ. Stud. 71, 655–679.

Blundell, R., Smith, R., 1989. Estimation in a class of simultaneous equation limited dependent variable models. Rev. Econ. Stud. 56, 37–57.

Breitung, J., Mayer, A., Wied, D., 2024. Asymptotic properties of endogeneity corrections using nonlinear transformations. Economet. J. (forthcoming).

Briseño-Sanchez, G., Hohberg, M., Groll, A., Kneib, T., 2020. Flexible instrumental variable distributional regression. J. Roy. Statist. Soc. Ser. A 183 (4), 1553–1574.

Bundeszentralamt für Steuern, 2024. Minijob (fachaufsicht). https://www.bzst.de/de/Privatpersonen/Minijob/minijob_node.html. (21 Feb 2024).

Card, D., 2000. The causal effect of earnings on education. In: Ashenfelter, O., Card, D. (Eds.), In: Handbook of Labor Economics, vol. 3, Elsevier, Amsterdam, pp. 1802–1863.

Chen, G., Lockhart, R., 2001. Weak convergence of the empirical process of residuals in linear models with many parameters. Ann. Statist. 29 (3), 748–762.

Chernozhukov, V., Fernández-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. Econometrica 78 (3), 1093–1125.

Chernozhukov, V., Fernández-Val, I., Luo, S., 2022. Distribution regression with sample selection, with an application to wage decompositions in the UK. arXiv:1811.11603. [econ].

Chernozhukov, V., Fernández-Val, I., Melly, B., 2013. Inference on counterfactual distributions. Econometrica 81 (6), 2205–2268.

Chernozhukov, V., Fernández-Val, I., Newey, W., Stouli, S., Vella, F., 2020. Semiparametric estimation of structural functions in nonseparable triangular models. Quant. Econ. 11 (2), 503–533.

Delgado, M., García-Suaza, A., Sant'Anna, P., 2022. Distribution regression in duration analysis: An application to unemployment spells. Econom. J. 25 (3), 675–698.

Dette, H., Neumeyer, N., Pilz, K.F., 2006. A simple nonparametric estimator of a strictly monotone regression function. Bernoulli 12 (3), 469–490.

Ellison, B., 1964. Two theorems for inferences about the normal distribution with applications in acceptance sampling. J. Amer. Statist. Assoc. 59 (305), 89–95.

Foresi, S., Peracchi, F., 1995. The conditional distribution of excess returns: An empirical analysis. J. Amer. Statist. Assoc. 90, 451–466.

Greene, W., 2017. Econometric Analysis, 8th ed. Pearson.

Hansen, B., 2022. Econometrics. Princeton University Press.

Henzi, A., Ziegel, J., Gneiting, T., 2021. Isotonic distributional regression. J. R. Stat. Soc. Ser. B 83 (5), 963–993.

Imbens, G., Newey, W., 2009. Identification and estimation of triangular simultaneous equations models without additivity. Econometrica 77, 1481–1512.

Koenker, R., Leorato, S., Peracchi, F., 2013. Distributional vs. quantile regression. Einaudi Institute for Economics and Finance (EIEF) Working Paper Series 1329.

Krenz, A., 2008. Theorie und Empirie über den Wirkungszusammenhang zwischen sozialer Herkunft, kulturellem und sozialem Kapital, Bildung und Einkommen in der Bundesrepublik Deutschland. Panel Data Research 128, DIW Berlin, The German Socio-Economic Panel (SOEP), URL https://ideas.repec.org/p/diw/diwsop/diw_sp128.html.

Lemke, R., Rischall, I., 2003. Skill, parental income, and IV estimation of the returns to schooling. Appl. Econ. Lett. 10, 281–286.

Li, C., Poskitt, D., Windmeijer, F., Zhao, X., 2022. Binary outcomes, OLS, 2SLS and IV probit. Econometric Rev. 41 (8), 859–876.

Newey, W., 1987. Efficient estimation of limited dependent variable models with endogenous explanatory variables. J. Econometrics 36, 231–250.

Oliveira, C., 2023. The minimum wage and the wage distribution in Portugal. Lab. Econ. 85, 1–16.

Patra, R., Seijo, E., Sen, B., 2018. A consistent bootstrap procedure for the maximum score estimator. J. Econometrics 205 (2), 488–507.

Rivers, D., Vuong, Q., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. J. Econometrics 39 (3), 347–366.

Robertson, T., Wright, F., Dykstra, R., 1988. Order Restricted Statistical Inference. Wiley, New York.

Rothe, C., Wied, D., 2013. Misspecification testing in a class of conditional distributional models. J. Amer. Statist. Assoc. 108, 314–324.

Rothe, C., Wied, D., 2020. Estimating derivatives of function-valued parameters in a class of moment condition models. J. Econometrics 217 (1), 1–19.

Sozio-oekonomisches Panel (SOEP), 2022. Version 37, Daten der Jahre 1984–2020 (SOEP-core v37, EU-edition). http://dx.doi.org/10.5684/soep.core.v37eu.

Troster, V., Wied, D., 2021. A specification test of dynamic conditional distributions. Econometric Rev. 40 (2), 109–127.

van der Vaart, A., 1998. Asymptotic Statistics. Cambridge University Press.

Wang, Y., Oka, T., Zhu, D., 2022. Bivariate distribution regression with application to insurance data. arXiv:2203.12228v1.

Wied, D., Krämer, W., Dehling, H., 2012. Testing for a change in correlation at an unknown point in time using an extended functional delta method. Econometric Theory 28 (3), 570–589.

Wooldridge, J., 2002. Econometric Analysis of Cross-Sectional and Panel Data. MIT Press.

Wooldridge, J., 2016. Introductory Econometrics, A Modern Approach, 6th ed. Cengage Learning.

Wüthrich, K., 2019. A closed-form estimator for quantile treatment effects with endogeneity. J. Econometrics 210 (2), 219–235.