ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Huber, Martin; Melly, Blaise

Working Paper

A test of the conditional independence assumption in sample selection models

Working Paper, No. 2012-11

Provided in Cooperation with:
Department of Economics, Brown University

Suggested Citation: Huber, Martin; Melly, Blaise (2012) : A test of the conditional independence assumption in sample selection models, Working Paper, No. 2012-11, Brown University, Department of Economics, Providence, RI

This Version is available at:
https://hdl.handle.net/10419/102613

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

A test of the conditional independence assumption in sample selection models

Martin Huber, Blaise Melly

First draft: December 2006, Last changes: September 2012

Abstract:

Identification in most sample selection models depends on the independence of the regressors and the error terms conditional on the selection probability. All quantile and mean functions are parallel in these models; this implies that quantile estimators cannot reveal any—per assumption non-existing—heterogeneity. Quantile estimators are nevertheless useful for testing the conditional independence assumption because they are consistent under the null hypothesis. We propose tests of the Kolmogorov-Smirnov type based on the conditional quantile regression process. Monte Carlo simulations show that their size is satisfactory and their power sufficient to detect deviations under realistic data generating processes. We apply our procedures to female wage data from the 2011 Current Population Survey and show that homogeneity is clearly rejected.

Keywords: sample selection, quantile regression, independence, test

JEL classification: C12, C13, C14, C21

This paper was previously circulated under the titles "Quantile regression in the presence of sample selection" and "Sample selection, heteroscedasticity, and quantile regression". We would like to thank the editor (Edward Vytlacil), four referees, Stefan Hoderlein, Frank Kleibergen, Michael Lechner, and seminar participants at Brown University, University of St. Gallen, the labor market seminar of the University of Zürich in Engelberg, the conference 'Inference and Tests in Econometrics' in Marseille, and the COST A23 conference in Paris for very useful comments that helped improve the paper.

Addresses for correspondence: Martin Huber, SEW, University of St. Gallen, Varnbüelstrasse 14, 9000 St. Gallen, Switzerland, martin.huber@unisg.ch; Blaise Melly, Brown University, Department of Economics, Providence, RI, USA, blaise_melly@brown.edu.

1 Introduction

Estimation of economic models is frequently complicated by the problem of sample selection: the variable of interest is only observed for a non-random subsample of the population. A prominent example in labor economics consists in the estimation of the determinants of female wages. Individuals are assumed to offer positive labor supply only if their potential wage exceeds their reservation wage. Starting with Gronau (1974) and Heckman (1974, 1976 and 1979), the last 40 years have seen a proliferation of work adressing this difficult problem. Although the suggested estimators have progressively weakened the distributional and parametric assumptions originally made, none of them is robust to the presence of heteroscedasticity or higher order dependence between the error terms and the covariates (after controlling for the selection probability).

However, dependence in general—and heteroscedasticity in particular—is ubiquitous in the fields where sample selection models are applied. As suggested by Mincer (1973) in his famous human capital earnings model, residual wage dispersion should increase with experience and education. In line with this finding, the large majority of the quantile regression applications in labor economics find significant heterogeneity in the returns to education and experience. Thus, the conditional independence assumption has neither theoretical nor empirical support in most economic applications. In this paper we discuss the importance of this assumption in sample selection models, its implication for the quantile functions, and we suggest a test of its validity.

Buchinsky (1998a and 2001) is the first to propose a sample selection correction for quantile regression. He extends the series estimator of Newey (2009) to the estimation of quantiles. The suggested estimator has been applied to many data—among others in both original articles—to analyze the heterogeneous effects of the explanatory variables on the distribution of the outcome. The first contribution of this paper is to show that this estimator is consistent only in two uninteresting cases: when all quantile regression slopes are equal or when selection is random. The reason is that Buchinsky (1998a) assumes that the error terms are independent from the regressors given the selection probability. This implies that all quantile slope coefficients and the mean coefficients are identical; i.e., it excludes heterogeneous effects even though their analysis has been the main motivation for using quantile regression in recent applications.

The estimator proposed by Buchinsky is nevertheless useful for two reasons, which were the initial motivations for quantile regression: it is robust and can be used as a basis for a test of independence. Koenker and Bassett (1978) also assume independence in their seminal paper and motivate quantile regression purely from a robustness and efficiency point of view in the presence of non-Gaussian errors. Similarly, the estimator of Buchinsky has a bounded influence function (in the direction of the dependent variable) so that it is more robust than mean regression.

Chronologically, the next motivation for quantile regression (in Koenker and Bassett 1982) was to provide a basis for a test for heteroscedasticity. The second contribution of this paper is to apply this

testing idea to sample selection models. In the absence of sample selection, the lack of independence between the error terms and the regressors does not affect the consistency of the estimators; it is only a minor nuisance for inference. In the presence of sample selection, the conditional independence assumption is an identifying assumption; its violation leads to the inconsistency of most sample selection estimators (including those only concerned with mean effects). To the best of our knowledge, ours is the first test of this identifying assumption.

Under the null hypothesis of conditional independence the slope coefficients are constant as a function of the quantile, and the procedure proposed by Buchinsky (1998a) consistently estimates them. Under the alternative hypothesis, the estimates do not converge to the true values but will be a nontrivial function of the quantile. These two properties justify testing the conditional independence assumption by testing whether the coefficients vary across quantiles. In order to detect any violation of the null hypothesis, our testing approach is based on the Kolmogorov-Smirnov and Cramér-von-Mises statistics, applied to the entire conditional quantile process (after trimming the tails and conditioning on the selection probability).

The presence of an unknown nuisance parameter under the null hypothesis jeopardizes the nuisance-free asymptotic distribution of these statistics; Koenker and Xiao (2002) call this situation the Durbin problem. Chernozhukov and Fernández-Val (2005) overcome this complication by estimating the critical values using resampling procedures. Since the computation of the estimates is quite demanding, Chernozhukov and Hansen (2006) resample the influence functions instead of recomputing the whole process. They apply this procedure to standard and IV quantile regression respectively; we demonstrate that their conditions for the validity of the tests are satisfied in our sample selection model. Monte Carlo simulations calibrated to fit characteristics of typical applications show that this procedure performs well in reasonably sized samples.

Finally, we apply our tests to examine whether violations of the conditional independence assumption are an empirically relevant phenomenon. We use data from the 2011 outgoing rotation groups of the Current Population Survey to estimate a traditional Mincer wage regression for women. The number of children (measured in three age brackets) is used as instrument for selection and excluded from the outcome equation. This is a typical application of sample selection models; a prominent recent example is given by Mulligan and Rubinstein (2008). We reject the conditional independence assumption at any conventional significance level. The working paper version of this article (Huber and Melly 2011) contains two other applications with similar conclusions. Therefore, we suspect that this problem is not limited to a few cases but is widespread in fields where sample selection models have been used.[1]

This paper contributes to the literatures on sample selection and quantile regression. The ability to consistently estimate econometric models in the presence of non-random sample selection is one of the

[1] The codes for the simulations and applications and the datasets used in this paper can be downloaded at http://www.econ.brown.edu/fac/Blaise_Melly/code_R_selection.html. The interested researchers can, therefore, easily verify whether our claim is true or not in their applications.

most important innovations in microeconometrics, as illustrated by the Nobel Prize received by James Heckman. Gronau (1974) and Heckman (1974, 1976 and 1979) addressed the selectivity bias and proposed fully parametric estimators, assuming that the residuals are independent and jointly normally distributed. This approach may yield inconsistent results if the distribution of the error term is misspecified. Therefore, Cosslett (1991), Gallant and Nychka (1987), Powell (1987), and Newey (2009) proposed semiparametric estimators for the sample selection model. They relaxed the distributional assumption but kept the single index structure in both the selection and the outcome equation. In addition, Ahn and Powell (1993) dropped the index structure in the selection equation. More recently, Das, Newey, and Vella (2003) considered fully nonparametric sample selection models.

Quantile regression has progressively emerged as the method of choice to analyze the effects of variables on the distribution of the outcome. In the absence of selection, Koenker and Bassett (1978) proposed a parametric (linear) estimator for conditional quantile models. Due to its ability to capture heterogeneous effects, its theoretical properties have been studied extensively and it has been used in many empirical studies; see, for example, Powell (1986), Guntenbrunner and Jurečková (1992), Buchinsky (1994), Koenker and Xiao (2002), Angrist, Chernozhukov, and Fernández-Val (2006), and Chernozhukov, Fernández-Val, and Melly (2012). Buchinsky (1998b), Koenker and Hallock (2000), and Koenker (2005) provide a comprehensive discussion of quantile regression models and recent developments.

The remainder of this paper is organized as follows. In Section 2 we describe the sample selection model of Buchinsky (1998a) and discuss the role of the independence assumption in sample selection models. Section 3 outlines the test procedure and proves its asymptotic validity. In Section 4 Monte Carlo simulations document the power and size properties of the proposed test. Section 5 offers an empirical application showing the relevance of the problem at the heart of this paper. Section 6 concludes.

2 The conditional independence assumption in sample selection models

2.1 The model

We consider the quantile sample selection model of Buchinsky (1998a), which is similar to the semiparametric models for the mean in Cosslett (1991), Powell (1987), and Newey (2009). The outcome equation and the latent selection function are assumed to be linear in the covariates like in the seminal work of Heckman (1974, 1976, and 1979). The error terms in both equations are independent of the covariates conditional on the selection probability but, in contrast to the model of Heckman, their joint distribution is unknown. At this point, we would like to emphasize that the choice of linear outcome and selection equations are made for completeness and to simplify the comparison with important existing estimators.

The importance of the independence assumption, its implication for quantile models, and its testability are all preserved in more general models. This includes models with a nonparametric selection equation as in Ahn and Powell (1993) or a nonparametric outcome equation as in Das, Newey, and Vella (2003), albeit a different testing approach may be required in more general cases.

Bearing this in mind, we maintain the following assumption, which is equation 2 in Buchinsky (1998a):

Assumption 1 (potential outcome) *The potential outcome of interest, Y^*, is determined by*

$$Y^* = X_2'\beta_0 + U$$

where X_2 denotes a vector of regressors, β_0 is the vector of slope coefficients and U is an error term.

Note that X_2 does not include a constant term because the latter is only identified through an identification-at-infinity argument. The constant is incorporated in the error term, U. We do not observe the latent variable Y^* but only Y, which is defined in Assumption 2.

Assumption 2 (observation rule and selection equation) *The observed outcome, Y, is determined by*

$$Y = Y^* \text{ if } D = 1 \text{ and not observed otherwise.}$$

$$D = 1\left(X_1'\gamma_0 + \varepsilon \geq 0\right),$$

where $1(\cdot)$ is the indicator function, X_1 is a superset of X_2, and *ε is a second error term.*

The indicator function for being selected, D, is determined by a latent single-index crossing model, which corresponds to equation (5) in Buchinsky (1998a). We do not impose any parametric restriction on the distribution of (U, ε) but we require X_1 to include at least one continuous variable which is not in X_2 and has a non-zero coefficient in the selection equation.

Assumption 3 (conditional independence) *The joint density of (U, ε) is independent of X_1 conditionally on the latent selection index,*

$$f_{U,\varepsilon}(\cdot|X_1) = f_{U,\varepsilon}(\cdot|X_1'\delta_0).$$

Assumption 3 is the conditional independence that, combined with the additivity of the outcome in Assumption 1, is at the center of this paper. It corresponds to Assumption E in Buchinsky (1998a).[2]

[2] In some sample selection models concerned with identifying the mean function, only a conditional moment restriction (such as $E\left[U|X_1, D=1\right] = E\left[U|X_1'\delta_0, D=1\right]$) is imposed without assuming full independence. However, the latter often serves as a justification for the moment restriction. Departures from full independence that still satisfy the moment condition are not substantial. E.g., the moment condition allows for heteroscedastic measurement errors affecting the dependent variable but not for heteroscedastic wage functions, see the discussion in Newey and Powell (2003).

2.2 Identification and testability

We first show how Assumptions 1 to 3 identify the parameter β_0. Let $Q_V(\tau|W)$ denote the τth conditional quantile of V given W, for any random variables V and W. Then, for any $0 < \tau < 1$,

$$\begin{aligned} Q_Y(\tau|X_1, D=1) &= X_2'\beta_0 + Q_U(\tau|X_1, D=1) \\ &= X_2'\beta_0 + Q_U(\tau|X_1, X_1'\gamma_0 \geq -\varepsilon) \\ &= X_2'\beta_0 + Q_U(\tau|X_1'\gamma_0, X_1'\gamma_0 \geq -\varepsilon) \\ &= X_2'\beta_0 + Q_U(\tau|X_1'\gamma_0, D=1) \\ &\equiv X_2'\beta_0 + h_\tau(X_1'\gamma_0). \end{aligned}$$

The first line holds by Assumption 1, the second line by Assumption 2, the third line by Assumption 3, and the fourth line again by Assumption 2. In the last line we define $h_\tau(X_1'\gamma_0) \equiv Q_U(\tau|X_1'\gamma_0, D=1)$.

This result has three important consequences. First, it implies that β_0 is identified and can be estimated by the τth quantile regression of Y on X_2 and a nonparametric function of $X_1'\hat{\gamma}$ in the selected sample ($\hat{\gamma}$ being a consistent estimate of γ_0). Second, since this is true for *any* quantile τ, Assumptions 1 to 3 imply that the slopes of all quantile regressions are identical. In addition, these slopes are also equal to the slopes of the corresponding mean regression. However, in the majority of cases where quantile methods are applied the researcher is particularly interested in the heterogeneity of the coefficients across the distribution. In the sample selection model, the identifying assumptions exclude such a possibility.

Third, this result also implies that the model defined by Assumptions 1 to 3 can be tested by comparing the slope coefficients obtained at different quantiles. Heterogeneity of the slope coefficients across several quantiles points to the violation of (at least) one identifying assumption. We suggest such a test based on the quantile estimator of Buchinsky (1998a) in Section 3. Our test bears great relevance for empirical work, as the conditional independence assumption is a necessary condition for the consistency of the estimators suggested in Heckman (1979), Cosslett (1991), Gallant and Nychka (1987), Powell (1987), and Newey (2009), to name only a few. Even though the importance of this assumption in sample selection models has not remained unnoticed in the literature, see for instance Angrist (1997), we appear to be the first ones to suggest a formal test.

Although the null hypothesis of our test is the validity of Assumptions 1 to 3, it seems particularly powerful against deviations from the conditional independence in Assumption 3 (together with the additivity of the outcome equation in Assumption 1). Because of that and for brevity, we will simply call it a test of the independence assumption. It is, however, clear that in some cases this test may also be able to detect violations of the exclusion restriction or of the correct parametrization of both equations even if it is not the most direct way to test these assumptions.

2.3 Two examples

Two examples will convey the intuition about the inconsistency of traditional estimators when independence is violated and about the testability of this assumption. These two examples correspond to the data generating processes used in Section 4 to assess the finite sample properties of our tests. We first consider a simple multiplicative heteroscedastic model. If we replace Assumption 1 with

$$Y^* = X_2'\beta_0 + (\delta + X_2'\delta_X) \cdot U, \tag{1}$$

then it follows that

$$E\left[Y|X_1, D = 1\right] = X_2'\beta_0 + (\delta + X_2'\delta_X) \cdot E\left[U|X_1'\gamma_0, D = 1\right].$$

We see that controlling for the selection probability (as done by sample selection mean estimators) is no longer sufficient to identify β_0. If we fix $X_1'\gamma_0$ to a specific value, the slope estimates in the selected sample will converge to

$$\beta_0 + \delta_X \cdot E\left[U|X_1'\gamma_0, D = 1\right].$$

There is no bias only if (i) $\delta_X = 0$ (i.e. Assumptions 1 to 3 are satisfied) or (ii) $E\left[U|X_1'\gamma_0, D = 1\right] = 0$ (i.e. selection is random).

The same type of bias also arises for the quantile regression estimator of Buchinsky (1998a):

$$Q_Y\left(\tau|X_1, D = 1\right) = X_2'\beta_0 + (\delta + X_2'\delta_X) \cdot Q_U\left(\tau|X_1'\gamma_0, D = 1\right).$$

On the bright side, this demonstrates the possibility to test the validity of our model because the slope of the τth quantile regression in the selected population after fixing $X_1'\gamma_0$ is

$$\beta_0 + \delta_X \cdot Q_U\left(\tau|X_1'\gamma_0, D = 1\right).$$

Thus, the slope is constant as a function of τ if and only if $\delta_X = 0$, i.e. if and only if conditional independence holds. In other words, our test is asymptotically capable to detect any deviation from independence when the true model is given by (1).

Figure 1 illustrates this example with 500 simulated realizations of (X_2, Y^*) from model (1) with $\beta_0 = 1$. X_1 has been chosen such that the conditional selection probability is 0.5 for each observation. Selection is positive, which can be seen from the fact that we observe more realizations (symbolized by boxes around the crosses) above than below the median. The error terms are independent in the left panel and heteroscedastic in the right panel ($\delta = 0$ vs. $\delta = 0.5$).[3] In the first case, selection induces a shift in the conditional median of Y but this bias is constant across observations due to the constant participation probability, resulting in a correct slope. In the heteroscedastic case, the bias is increasing as a function of X_2 because the variance of the errors is increasing with X_2, resulting in an excessive slope. In general,

[3] More details about the data generating process can be found in Section 4.1.

the sign of the bias corresponds to the product of the sign of δ_X and the direction of the selection. Figure 1 makes also clear that the independence assumption can be tested by comparing the slopes of several quantile regressions.

Figure 1 here

After this stylistic example we consider a more realistic one. We now replace Assumption 1 with a linear quantile model for all conditional quantiles:

$$Y^* = \alpha(U) + X_2'\beta(U), \; U|X_1 \sim U(0,1),$$

where $u \mapsto \alpha(u) + x_2'\beta(u)$ is strictly increasing in u for each value of x_2 in the support of X_2. In this model,

$$Q_Y(\tau|X_1, D=1) = \alpha(Q_U(\tau|X_1'\gamma_0, D=1)) + X_2\beta(Q_U(\tau|X_1'\gamma_0, D=1)). \tag{2}$$

Thus, if we fix $X_1'\gamma_0$, the slope of the quantile function is given by $\beta(Q_U(\tau|X_1'\gamma_0, D=1))$. Again, there is no bias only if (i) $\beta(\tau) = \beta(0.5)$ for all τ (i.e. Assumptions 1 to 3 are satisfied) or (ii) $Q_U(\tau|X_1'\gamma, D=1) = \tau$ (i.e. selection is random at the quantile of interest).

We suggest to test the model defined by Assumptions 1 to 3 by testing whether $\beta(Q_U(\tau|X_1'\gamma_0, D=1))$ is constant across all quantiles τ. Our test asymptotically detects any violation of independence under a minor additional assumption: If there exist at least two quantiles $\tilde{\theta}$ and $\bar{\theta}$ with $\beta\left(\tilde{\theta}\right) \neq \beta\left(\bar{\theta}\right)$ and $\Pr\left(D=1|U=\tilde{\theta}\right) > 0$ and $\Pr\left(D=1|U=\bar{\theta}\right) > 0$, then the slopes of the $\tilde{\tau} = \Pr\left(U \leq \tilde{\theta}|D=1\right)$ and $\bar{\tau} = \Pr\left(U \leq \bar{\theta}|D=1\right)$ conditional quantile regressions in the selected sample will differ. In other words, when the true model is a linear quantile regression model then our test has power against any deviation from independence as long as the heterogeneous quantiles are not completely unobserved.

Figure 2 shows the biases that can arise in such a data generating process when the parameters take realistic values, which are defined in detail in Section 4.2. The true quantile function is the male wage function in the 2011 Current Population Survey, the selection equation is the female selection equation during that year, and the copula between U and ε is Gaussian. We plot the values to which the estimator of Buchinsky (1998a) converges when the correlation between U and ε is set to -0.99, -0.5, 0, 0.5, and 0.99. The true values correspond to the values for zero correlation. The quantile regression coefficients—in particular the way they change with the quantile of interest—are quite different for different correlations. In an application, we do not know the correlation between U and ε and this value is not identified without distributional assumptions, possibly leading to an incorrect analysis. Furthermore, the difference between the coefficients in two time periods is also biased when the correlation changes over time. The main finding in Mulligan and Rubinstein (2008) is precisely that selection into work changed from negative to positive.

Figure 2 here

3 Test Procedure

The discussion in Section 2 shows that we can test the validity of the independence assumption by testing whether the coefficients on X_2 in the quantile regression of Y on X_2 and $h_\tau\left(X_1'\gamma_0\right)$ are constant as a function of the quantile. Our testing approach is based on the Kolmogorov-Smirnov and Cramér-von-Mises statistics, which are defined on the empirical quantile regression process. We follow Buchinsky (1998a) and approximate the unknown function $h_\tau\left(X_1'\gamma_0\right)$ by a power series in $X_1'\gamma_0$. We estimate the unknown parameter γ_0 using the Klein and Spady (1993) estimator for binary choice models. In Section 3.1 and 3.2 we describe step by step the estimation and testing algorithms. In Section 3.3 we give the formal inference results.

3.1 Estimation

The first step consists of estimating the selection equation, i.e. the parameter γ_0. The formal inference results do not depend on the specific estimator of γ_0 but only require $\sqrt{n}$ consistency and asymptotic linearity of the influence function. In our simulations and application we estimate the selection equation by the semiparametric binary choice estimator suggested in Klein and Spady (1993):

$$\hat{\gamma} \equiv \arg\max_{\gamma} \sum_{i=1}^{n} \left\{ D_i \log[\hat{E}(D|X_{1i}'\gamma)] + (1-D_i)\log[1-\hat{E}(D|X_{1i}'\gamma)] \right\}, \tag{3}$$

where

$$\hat{E}(D|X_{1i}'\gamma) = \frac{\sum_{j=1}^{n} D_j \kappa((X_{1i}'\gamma - X_{1j}'\gamma)/b_n)}{\sum_{j=1}^{n} \kappa((X_{1i}'\gamma - X_{1j}'\gamma)/b_n)}. \tag{4}$$

b_n is a bandwidth that depends on the sample size n and $\kappa(\cdot)$ denotes the kernel function. This estimator satisfies our regularity conditions under the assumptions detailed in Klein and Spady (1993) and it attains the semiparametric efficiency bound for this model. Heteroscedasticity is allowed to depend on the regressors only through the linear index.

In the second step, the function $h_\tau\left(X_1'\gamma_0\right)$ is approximated by a power series expansion. As suggested by Buchinsky (1998a), we use a power series expansion of the inverse Mill's ratio of the normalized estimated index but other bases of power functions also satisfy our assumptions. The order of the approximation, J, must increase with the sample size to ensure the consistency of the estimator. $\hat{\beta}(\tau)$ solve the following optimization problem:

$$\left(\hat{\beta}\left(\tau\right), \hat{\delta}\left(\tau\right)\right) = \arg\min_{(\beta,\delta)} \sum_{i:d_i=1} \hat{t}_i \rho_\tau\left(y_i - x_{2i}'\beta - \Pi_J\left(x_{1i}'\hat{\gamma}\right)\delta\right) \tag{5}$$

where $\rho_\tau(A) = A(\tau - 1\left(A \leq 0\right))$ is the check function suggested by Koenker and Bassett (1978). Furthermore, $\hat{t}_i$ is a trimming function,

$$\hat{t}_i = 1\left(x_{1i}'\hat{\gamma} \in \mathcal{X}\right),$$

where $\mathcal{X}$ is a one-dimensional compact set. Trimming is required for proving several asymptotic results (some of which we borrow from Lee 2007), even though we do not expect it to be important empirically. $\Pi_J\left(x_{1i}'\hat{\gamma}\right)$ is a polynomial vector in the inverse Mill's ratio, $\Pi_J(x_{1i}'\hat{\gamma}) = (1, \lambda(x_{1i}'\hat{\gamma}), \lambda(x_{1i}'\hat{\gamma})^2, ..., \lambda(x_{1i}'\hat{\gamma})^J)$, where $\lambda(\cdot) = \phi\left(\cdot\right)/\Phi\left(\cdot\right)$ is the ratio of the normal pdf to the normal cdf.

3.2 Tests

Our testing problem is given by

$$H_0 \quad : \quad \beta\left(\tau\right) = \beta\left(0.5\right), \quad \forall\ \tau\ \epsilon\ \mathcal{T} \tag{6}$$

$$\text{versus } H_1 \quad : \quad \beta\left(\tau\right) \neq \beta\left(0.5\right), \quad \text{for some } \tau\ \epsilon\ \mathcal{T}, \tag{7}$$

where $\mathcal{T}$ is a closed subset of $[e, 1-e]$, $0 < e < 1$, and $\beta\left(\tau\right)$ denotes the true τ quantile regression coefficient defined as

$$\beta\left(\tau\right) = \arg\min_{\beta} E\left[\rho_\tau\left(Y_i - X_{2i}'b - h_\tau\left(X_1'\gamma_0\right)\right)\right].$$

Buchinsky (1998a) gives the joint asymptotic distribution of $\hat{\beta}\left(\tau\right)$ for a finite number of quantiles. Based on his results, we can use a finite number of quantile regressions and apply a Wald test as proposed by Koenker and Bassett (1982) in the absence of sample selection. Even asymptotically, this test does not have power against all deviations from the null because the number of quantiles considered is fixed. In addition, it is tempting to choose the quantiles used by the test after having seen the results, which may distort the size of the test. Our tests use the whole quantile process and are therefore consistent against all global departures from H_0.

We measure the deviations from the null hypothesis by the Kolmogorov-Smirnov (KS) and Cramér-von-Mises (CM) statistics applied to the empirical process $\hat{\beta}\left(\tau\right) - \hat{\beta}\left(0.5\right)$:

$$T_n^{KS} = \sup_{\tau\epsilon\mathcal{T}} \sqrt{n}||\hat{\beta}\left(\tau\right) - \hat{\beta}\left(0.5\right)||_{\hat{\Lambda}_\tau} \qquad \text{and} \qquad T_n^{CM} = n\int_{\mathcal{T}} ||\hat{\beta}\left(\tau\right) - \hat{\beta}\left(0.5\right)||_{\hat{\Lambda}_\tau}^2 d\tau, \tag{8}$$

where $||a||_{\hat{\Lambda}_\tau}$ denotes $\sqrt{a'\hat{\Lambda}_\tau a}$ and $\hat{\Lambda}_\tau$ is a positive weighting matrix satisfying $\hat{\Lambda}_\tau = \Lambda_\tau + o_p(1)$, uniformly in τ. Λ_τ is positive definite, continuous and symmetric, again uniformly in τ.

Inference requires the knowledge of the asymptotic distributions of T_n^{KS}, T_n^{CM}. In Section 3.3 we derive the asymptotic distribution of the quantile regression process $\sqrt{n}\left(\hat{\beta}\left(\tau\right) - \beta\left(\tau\right)\right)$, which would allow us to derive the distribution of the test statistic if the null hypothesis was not estimated. The presence of a nuisance parameter (the median coefficients vector) jeopardizes this possibility.

As suggested by Chernozhukov and Fernández-Val (2005) we could calculate the critical values by resampling the centered inference process $\hat{\beta}\left(\tau\right) - \hat{\beta}\left(0.5\right)$. But the repeated computation of the coefficients (in our case especially of $\hat{\gamma}$) for each bootstrap sample can be quite costly, in particular when the sample size is large. For this reason, we follow Chernozhukov and Hansen (2006) and resample the score functions

instead, which are linear approximations of the empirical processes. We show in Appendix A that

$$\sqrt{n}(\hat{\beta}(\tau) - \hat{\beta}(0.5)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i(\tau) + o_p(1). \tag{9}$$

A precise definition of the score function $s_i(\tau)$ as well as its estimator $\hat{s}_i(\tau)$ is given in the next subsection. We draw B samples of size b and denote such samples as Υ_j, $1 \leq j \leq B$. The KS and CM statistics for the jth iteration are

$$T_{n,b,j}^{KS} \equiv \sup_{\tau \epsilon \mathcal{T}} \sqrt{b} ||1/b \sum_{i \epsilon \Upsilon_j} \hat{s}_i(\tau)||_{\hat{\Lambda}_\tau} \qquad \text{and} \qquad T_{n,b,j}^{CM} \equiv b \int_{\mathcal{T}} ||1/b \sum_{i \epsilon \Upsilon_j} \hat{s}_i(\tau)||_{\hat{\Lambda}_\tau}^2 d\tau.$$

The critical values of a size α test are the $1-\alpha$th quantiles of $T_{n,b,j}^{KS}$ and $T_{n,b,j}^{CM}$, i.e. $c_{n,b}^{KS}(\alpha) = \inf\left\{c : B^{-1}\sum_{j=1}^{B} 1\left(T_{n,b,j}^{KS} \leq c\right) \geq 1-\alpha\right\}$ and similarly for the CM statistic. The level α tests reject H_0 if $T_n^{KS} > c_{n,b}^{KS}(\alpha)$ and $T_n^{CV} > c_{n,b}^{CM}(\alpha)$.

3.3 Formal inference results

We prove that our tests have asymptotically the correct size and are consistent against all global alternative hypotheses by verifying the conditions of Theorem 5 in Chernozhukov and Hansen (2006). To do so, we heavily borrow from the results in Lee (2007) who considers an instrumental variable estimator that is almost identical to the Buchinsky sample selection estimator from a computational point of view. To establish these results, we make the following additional assumptions:

Assumption 4 (sampling) *The data $\{(y_i, x_{1i}, d_i) : i = 1, ..., n\}$ are i.i.d. realizations of (Y, X_1, D) and take values in a compact set.*

Assumption 5 (non-singularity) (a) *Define*

$$\begin{aligned} \Delta_{b,\tau} &= E[f_U(Q_U\left(\tau|x'_{1i\gamma 0}\right)|x'_{1i}\gamma_0))m_i m'_i], \\ m_i &= t_i d_i\left(x_{2i} - E\left[X_2|x'_{1i}\gamma_0, D=1\right]\right). \end{aligned}$$

$\Delta_{b,\tau}$ is non-singular uniformly in $\tau \in \mathcal{T}$.

(b) *Define*

$$\begin{aligned} \Phi_J(\tau) &= E\left[t_i f_U\left(Q_U\left(\tau|X'_1\gamma_0, D=1\right)\right) P_J P'_J | D=1\right] \\ P_J &= D \cdot \left[X_2, \Pi_J\left(X'_1\gamma_0\right)\right]. \end{aligned}$$

Uniformly in $\tau \in \mathcal{T}$ the smallest eigenvalue of $\Phi_J(\tau)$ is bounded away from zero for all J, and the largest eigenvalue of $\Phi_J(\tau)$ is bounded for all J.

Assumption 6 (continuity) (a) *The joint distribution of (U, ε) is absolutely continuous with respect to Lebesgue measure. The density of U in the selected sample, $f_U(u|X_1, D=1)$, is bounded above and away from zero and is twice continuously differentiable in u uniformly in u and in X_1.*

(b) *$h_\tau(X_1'\gamma_0)$ is r-times continuously differentiable with respect to $X_1'\gamma_0$ uniformly in $\tau \in \mathcal{T}$.*

(c) *As a function of $X_1'\delta_0$, $E[f_U(Q_U(\tau|X_1'\gamma_0)|X_1'\gamma_0))X_2|X_1'\gamma_0, d_i = 1]$ is continuously differentiable uniformly in $\tau \in \mathcal{T}$.*

Assumption 7 (number of approximating functions) *$J = Cn^\upsilon$ for some constant C satisfying $0 < C < \infty$ and some υ satisfying $\frac{1}{2r} < \upsilon < \frac{1}{9}$.*

Assumption 8 (first stage) *There exists an invertible matrix Δ_P and a bounded $k_i \equiv K(x_{1i}, d_i, \gamma)$ such that $E[k_i] = 0$, $E[k_i k_i']$ exists and is non-singular and*

$$\sqrt{n}(\hat{\gamma} - \gamma_0) = \Delta_P^{-1} \frac{\sum_{i=1}^n k_i}{\sqrt{n}} + o_p(1).$$

There is an estimator $\hat{\Delta}_P$ of Δ_P such that $\hat{\Delta}_P \to \Delta_P$.

These assumptions are the sample selection equivalent of the assumptions in Lee (2007). Assumption 5(a) is satisfied if X_1 includes at least one continuous variable which is not in X_2 and has a non-zero coefficient in the selection equation (and the components of X_2 are not linearly dependent). Assumption 5(b) ensures the non-singularity of the second moment matrix of the estimator. Assumption 6(a) is standard in quantile models because quantiles are well-defined and have a standard asymptotic distribution only for continuous random variables. It corresponds to Assumption C in Buchinsky (1998a). Assumptions 6(b) and 6(c) require that the approximated functions are sufficiently smooth. Assumption 7 restricts the growth rate of J. Together with Assumption 6(b), it requires that $h_\tau(X_1'\gamma_0)$ is at least five times differentiable ($r \geq 5$). Assumption 8 imposes regularity conditions on the estimation of γ_0. The estimator suggested by Klein and Spady (1993) satisfies this condition with

$$\begin{aligned} \Delta_P &= E\left[\left.\frac{\partial P_i(\gamma)}{\partial \gamma}\right|_{\gamma=\gamma_0} \left.\frac{\partial P_i(\gamma)'}{\partial \gamma}\right|_{\gamma=\gamma_0} \frac{1}{P_i(\gamma_0)(1-P_i(\gamma_0))} \right] \\ \text{and } k_i &= \left.\frac{\partial P_i(\gamma)}{\partial \gamma}\right|_{\gamma=\gamma_0} \frac{D_i - P_i(\gamma_0)}{P_i(\gamma_0)(1-P_i(\gamma_0))}, \end{aligned}$$

where $P_i(\gamma) = \Pr(D = 1|X_1'\gamma = x_{1i}'\gamma_0)$.

Under Assumptions 1 to 8 the influence function of $\hat{\beta}(\tau)$ is given by (see Appendix A for details)

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) = \Delta_{b,\tau}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\ell_i(\tau) + \Delta_\gamma \Delta_P^{-1} k_i) + o_p(1), \tag{10}$$

where

$$\begin{aligned} \ell_i(\tau) &= (\tau - 1(y_i < x_{2i}'\beta_0 + h_\tau(x_{1i}'\gamma_0)))m_i, \\ \Delta_\gamma &= E[t_i f_\varepsilon(Q_\varepsilon(\tau|x_{1i}'\gamma_0) \frac{dh_\tau(x_{1i}'\gamma_0)}{d\gamma_0} m_i]. \end{aligned}$$

$s_i(\tau)$ is simply the difference between the influence function of $\hat{\beta}(\tau)$ and the influence function of $\hat{\beta}(0.5)$. We estimate the influence functions by their sample analogs. In particular, we estimate the densities using

the kernel method as in Powell (1986), Buchinsky (1998a), and Lee (2007):

$$\begin{aligned}
\hat{\ell}_i(\tau) &= (\tau - 1\left(y_i < x_{2i}'\hat{\beta}(\tau) + \Pi_J(x_{1i}'\hat{\gamma})\hat{\delta}(\tau)\right))\hat{m}_i, \\
\hat{\Delta}_{b,\tau} &= (nh_n)^{-1}\sum_{i=1}^{n}\hat{t}_i\kappa\left(\frac{\hat{\varepsilon}_i(\tau)}{h_n}\right)\hat{m}_i\hat{m}_i', \\
\hat{\Delta}_{\gamma} &= (nh_n)^{-1}\sum_{i=1}^{n}\hat{t}_i\kappa\left(\frac{\hat{\varepsilon}_i(\tau)}{h_n}\right)\frac{d\Pi_J(x_{1i}'\hat{\gamma})}{d\gamma}\hat{m}_i, \\
\hat{\varepsilon}(\tau) &= y_i - x_{2i}'\hat{\beta}(\tau) - \Pi_J(x_{1i}'\hat{\gamma})\hat{\delta}(\tau),
\end{aligned}$$

where $\hat{m}_i$ is the vector of residuals from the regression of X_2 on $\Pi_J(X_1'\hat{\gamma})$ evaluated at x_{1i} and multiplied by $\hat{t}_i d_i$. We impose the following restrictions on the kernel and bandwidth:

Assumption 9 (kernel and bandwidth) The kernel function κ has support $[-1,1]$, is bounded and symmetrical about 0 and satisfies $\int_{-1}^{1}\kappa(u)\,du = 1$, $\int_{-1}^{1}u\kappa(u)\,du = 0$, and $\int_{-1}^{1}u^2\kappa(u)\,du < \infty$. $h_n = C_h n^{-\mu}$ for some positive finite C_h and some μ satisfying $\upsilon/2 < \mu < (1-4\upsilon)/4$.

Under these conditions our tests are asymptotically correctly sized and consistent against any global alternative hypothesis:

Proposition 1 *Suppose Assumptions 1 to 9 hold. Then as $B \to \infty$, $b \to \infty$, $n \to \infty$: (1) Under the null hypothesis $P\left(T_n^{KS} > c_{n,b}^{KS}(\alpha)\right) \to \alpha$ and $P\left(T_n^{CM} > c_{n,b}^{CM}(\alpha)\right) \to \alpha$; (2) Under the alternative hypothesis $P\left(T_n^{KS} > c_{n,b}^{KS}(\alpha)\right) \to 1$ and $P\left(T_n^{CM} > c_{n,b}^{CM}(\alpha)\right) \to 1$.*

Chernozhukov and Hansen (2006) develop new inference tools in the context of their instrumental variable quantile regression estimator but they state the conditions in a form that is general enough to apply to other estimators. We prove Proposition 1 by verifying the conditions for inference of their Theorem 5. Similar tests based on resampling the estimates (instead of the scores) are also valid by Theorem 1 in Chernozhukov and Fernández-Val (2005) because their conditions are weaker than the conditions that we check in Appendix A. In addition to the statement of our Proposition 1, Chernozhukov and Fernández-Val (2005) also show that this test has nontrivial power against local alternatives.

3.4 Practical details

Smoothing parameters. We follow a pragmatic approach and use the generalized cross validation criterion (GCV) proposed in Craven and Wahba (1979) to select the smoothing parameters—b_n and J—required to implement our estimators. The same method was used in Buchinsky (1998a) and Newey, Powell, and Walker (1990), among others. We do not claim any optimality of this method for our estimation and inference problem but it works well in our simulations.[4]

[4] The optimal choice of smoothing parameters in semiparametric models remains an important topic for future research. Härdle, Hall, and Ichimura (1993) and Delecroix, Hristache, and Patilea (2006) suggest to estimate the bandwidth and the

Discretization. The theoretical results still hold if, instead of $\mathcal{T}$, we use a grid $\mathcal{T}_n$ with the largest cell size converging to 0 as $n \to \infty$. In the simulations and application, we use for simplicity a grid with mesh size 0.01 but the results are not sensitive to the choice of the mesh size as long as it is small enough. The differences between the results with a mesh of size 0.005, 0.01, and 0.02 are completely negligible.

Choice of Λ_τ. In order to improve the power of the test we set $\Lambda_\tau = Var\left[s_i(\tau)\right]^{-1}$. Note that by construction $Var\left[s_i(0.5)\right] = 0$, which forces us to cut out the part of $\mathcal{T}_n$ close to the median. In our simulations and application we disregard $[0.46, 0.54]$. We estimate $Var\left[s_i(\tau)\right]$ by $Var\left[\hat{s}_i(\tau)\right]$.

Choice of b *and* $\mathcal{T}$. We experiment in our simulations and application with several rules for the choice of the block size, b. Though our empirical results are not sensitive to the choice of the block size, the bootstrap has better size and power properties in our simulations while the tests are conservative when the subsample size is smaller. Concerning $\mathcal{T}$, in principle we would like to use a large range of quantiles in order to increase the asymptotic power of the tests. However, the quantile regression estimator performs poorly in the tails of the distribution if we do not make strong assumptions about the density of Y. In the simulations we experiment with several ranges ($[0.05, 0.95]$, $[0.1, 0.9]$ and $[0.2, 0.8]$).

4 Simulations: Power and size properties of the tests

In this section, we present Monte Carlo evidence about the size and power properties of the tests that we have proposed in Section 3. We first consider a simple data generating process (DGP) that was used by Koenker and Xiao (2002) and Chernozhukov and Fernández-Val (2005) with the difference that we observe only a nonrandom sample from the population. Second, we will use a more realistic multivariate example calibrated to the data used in Section 5.

4.1 A simple heteroscedastic DGP

The first DGP is the following linear multiplicative heteroscedastic model:

$$\begin{aligned} D &= 1\left(X_2 + W + \varepsilon > 0\right), \\ Y^* &= X_2 + (1 + X_2\delta) \cdot U, \\ X_2 &\sim N(0,1), \quad W \sim N(0,1) \end{aligned} \tag{11}$$

where

$$(U, \varepsilon) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}\right).$$

parameters of the binary equation simultaneously. However, they do not show the optimality of the procedure. Although they are probably not optimal, as noted by Newey (2009) our choice of J should satisfy the rate condition of Assumption 7. In addition our choice of b_n is in the range of values allows by Delecroix, Hristache, and Patilea (2006) for $\sqrt{n}$ consistency.

We consider three different values for the parameter δ, which gauges heteroscedasticity. Under the null hypothesis, $\delta = 0$, such that the regressor X_2 is independent of the error U and has a pure location shift effect. This case allows us to analyze the empirical size of our tests. We also evaluate the power of our tests in two location scale shift models with $\delta = 0.2$ and 0.5, such that errors are heteroscedastic.

In the simulations, we consider five sample sizes from $n = 400$ to $n = 6400$. We run 1000 Monte Carlo replications and draw 250 bootstrap samples within each replication. The theoretical level of significance is set at 5%. For the sake of brevity, we only report the rejection frequencies for the bootstrap, i.e., for the block size $m = n$. The results for subsampling (i.e., for some m smaller than n) are comparable and available from the authors upon request. Results are presented for three different regions $\mathcal{T}$ over which the quantile coefficients are estimated: $[0.05, 0.95]$, $[0.1, 0.9]$, and $[0.2, 0.8]$.

Table 1

Empirical rejection frequencies for 5% bootstrap tests

Normal distribution, 1000 replications, 250 bootstrap draws

	Kolmogorov-Smirnov test								
	$\gamma = 0$			$\gamma = 0.2$			$\gamma = 0.5$		
τ ϵ	[0.05,0.95]	[0.10,0.90]	[0.20,0.80]	[0.05,0.95]	[0.10,0.90]	[0.20,0.80]	[0.05,0.95]	[0.10,0.90]	[0.20,0.80]
$n = 400$	0.016	0.012	0.013	0.133	0.054	0.024	0.774	0.494	0.169
$n = 800$	0.027	0.020	0.013	0.556	0.333	0.147	0.998	0.988	0.881
$n = 1600$	0.036	0.025	0.012	0.936	0.849	0.585	1.000	1.000	1.000
$n = 3200$	0.041	0.036	0.030	1.000	0.996	0.927	1.000	1.000	1.000
$n = 6400$	0.049	0.042	0.035	1.000	1.000	1.000	1.000	1.000	1.000
	Cramér-von-Mises test								
	$\gamma = 0$			$\gamma = 0.2$			$\gamma = 0.5$		
τ ϵ	[0.05,0.95]	[0.10,0.90]	[0.20,0.80]	[0.05,0.95]	[0.10,0.90]	[0.20,0.80]	[0.05,0.95]	[0.10,0.90]	[0.20,0.80]
$n = 400$	0.008	0.007	0.007	0.149	0.105	0.048	0.912	0.790	0.465
$n = 800$	0.019	0.019	0.016	0.571	0.447	0.236	1.000	0.999	0.973
$n = 1600$	0.032	0.029	0.018	0.957	0.900	0.721	1.000	1.000	1.000
$n = 3200$	0.033	0.027	0.028	1.000	0.999	0.969	1.000	1.000	1.000
$n = 6400$	0.049	0.046	0.043	1.000	1.000	1.000	1.000	1.000	1.000

The empirical rejection frequencies reported in Table 1 suggest that the bootstrap score tests have good size and power properties. In the presence of independent errors ($\gamma = 0$), both the KS and CM tests are conservative, at least for the sample sizes considered. However, the empirical size slowly converges to the theoretical size of 5% as the sample size increases. Under heteroscedasticity, the rejection probabilities correctly converge to 100% as n becomes larger. As expected, this happens at a faster pace for $\gamma = 0.5$ than for $\gamma = 0.2$. The power properties of the KS and CM tests are rather similar, albeit the latter become relatively more powerful in larger samples and/or for a higher γ. The empirical power increases as the

range of quantiles considered increases and this holds true for both test statistics and both values of γ. Summing up, the KS and CM tests seem to perform well in finite samples with Gaussian errors. Under sample sizes of several thousand observations, they are powerful in any scenario considered.

The results in Koenker and Xiao (2002) and Chernozhukov and Fernández-Val (2005) provide useful benchmarks. In order to compare our results with theirs we must take into account that we observe Y^* only for half of the sample. Therefore, it seems fair to compare our results for 1600 observations with theirs for 800 observations. In this case, we see for instance that with $\gamma = 0.2$ our empirical power is 0.936 and 0.957, the power in Koenker and Xiao (2002) is 0.982 (in the best case) and the power in Chernozhukov and Fernández-Val (2005) is 1. In general, the price to pay for allowing for nonrandom selection is moderate in this DGP.

4.2 A realistic multivariate DGP

In this subsection we analyze the properties of our tests in a more realistic setup. We calibrate our DGP to fit several characteristics of the CPS data sets used in the application in Section 5. We draw the Monte Carlo samples of X_1 from the female sample distribution in 2011 (containing 45,296 observations). We consider a model with four regressors (experience, experience squared, a high school dummy and a college dummy) and three exclusion restrictions (number of children in three age brackets). These covariates are only a subset of those used in our application because we also consider samples of moderate sizes, which would lead to frequent multicollinearity problems with the 19 regressors of the application.

The conditional distribution of the potential wage is generated by a linear quantile model:

$$Y^* = \alpha(U) + X_2'\beta(U), \; U|X_1 \sim U(0,1).$$

We consider two sets of functions for $\alpha(\cdot)$ and $\beta(\cdot)$. The first one satisfies the null hypothesis with a constant $\beta(\tau)$ set to the value of the median regression coefficient estimates in the male sample.[5] In the second case, we set the coefficients to the unrestricted quantile regression estimates in the male sample.[6] This ensures that the violation of the null hypothesis is realistic and allows us to see the power of our tests in a typical application. The selection equation is

$$D = 1(X_1\gamma + \varepsilon > 0),$$

where $\varepsilon \sim N(0,1)$ and γ is set to the probit estimates of the female participation equation in 2011.

[5] $\alpha(\cdot)$ and $\beta(\cdot)$ are calibrated using the male sample to avoid the sample selection problem, as the employment probability is high for men. If we took the female sample then we could find violations of the independence assumption due to the selection bias. Alternatively, we could use the female sample and correct for the selection bias but this would require making distributional assumptions which seem quite arbitrary.

[6] We approximate the $U(0,1)$ distribution of U with a discrete uniform distribution with 1000 points. This means that 1000 quantile regression are used to approximate the conditional distribution.

Finally, the copula between ε and U is Gaussian with a correlation coefficient of 0.233, which is the estimate obtained by Mulligan and Rubinstein (2008) in the period 1990-1995.

Table 2

Empirical rejection frequencies for 5% bootstrap tests

Empirical data, 1000 replications, 250 bootstrap draws

	$\tau \in [0.05,0.95]$					$\tau \in [0.10,0.90]$					$\tau \in [0.20,0.80]$				
	all	exp	$\exp^2$	hs	col	all	exp	$\exp^2$	hs	col	all	exp	$\exp^2$	hs	col
Kolmogorov-Smirnov test: size															
$n = 400$	0.033	0.006	0.007	0.017	0.005	0.026	0.004	0.003	0.021	0.000	0.028	0.002	0.004	0.015	0.002
$n = 800$	0.037	0.006	0.007	0.032	0.005	0.032	0.004	0.006	0.025	0.004	0.028	0.004	0.005	0.022	0.005
$n = 1600$	0.021	0.020	0.023	0.024	0.016	0.009	0.016	0.013	0.018	0.010	0.009	0.009	0.009	0.015	0.011
$n = 3200$	0.030	0.023	0.030	0.029	0.034	0.017	0.020	0.022	0.017	0.031	0.010	0.019	0.019	0.015	0.028
$n = 6400$	0.035	0.041	0.043	0.022	0.041	0.025	0.031	0.030	0.024	0.038	0.022	0.027	0.026	0.017	0.035
Kolmogorov-Smirnov test: power															
	$\tau \in [0.05,0.95]$					$\tau \in [0.10,0.90]$					$\tau \in [0.20,0.80]$				
	all	exp	$\exp^2$	hs	col	all	exp	$\exp^2$	hs	col	all	exp	$\exp^2$	hs	col
$n = 400$	0.026	0.006	0.004	0.019	0.009	0.025	0.004	0.002	0.020	0.003	0.023	0.004	0.002	0.019	0.002
$n = 800$	0.042	0.015	0.013	0.020	0.089	0.028	0.011	0.010	0.015	0.058	0.022	0.009	0.009	0.014	0.042
$n = 1600$	0.229	0.112	0.069	0.043	0.346	0.080	0.047	0.038	0.018	0.136	0.020	0.020	0.022	0.009	0.043
$n = 3200$	0.694	0.261	0.156	0.109	0.755	0.340	0.150	0.105	0.053	0.421	0.090	0.093	0.072	0.022	0.146
$n = 6400$	0.986	0.518	0.343	0.178	0.972	0.808	0.329	0.228	0.091	0.815	0.329	0.200	0.155	0.041	0.352
Cramér-von-Mises test: size															
	$\tau \in [0.05,0.95]$					$\tau \in [0.10,0.90]$					$\tau \in [0.20,0.80]$				
	all	exp	$\exp^2$	hs	col	all	exp	$\exp^2$	hs	col	all	exp	$\exp^2$	hs	col
$n = 400$	0.002	0.005	0.003	0.004	0.005	0.001	0.003	0.002	0.002	0.003	0.000	0.001	0.001	0.003	0.004
$n = 800$	0.004	0.012	0.009	0.007	0.012	0.003	0.011	0.009	0.006	0.010	0.002	0.009	0.007	0.005	0.009
$n = 1600$	0.011	0.020	0.023	0.018	0.022	0.003	0.022	0.020	0.011	0.020	0.003	0.014	0.016	0.010	0.015
$n = 3200$	0.026	0.039	0.040	0.029	0.042	0.020	0.038	0.040	0.020	0.041	0.013	0.032	0.035	0.015	0.041
$n = 6400$	0.034	0.043	0.035	0.038	0.043	0.032	0.034	0.034	0.032	0.039	0.022	0.029	0.029	0.024	0.038
Cramér-von-Mises test: power															
	$\tau \in [0.05,0.95]$					$\tau \in [0.10,0.90]$					$\tau \in [0.20,0.80]$				
	all	exp	$\exp^2$	hs	col	all	exp	$\exp^2$	hs	col	all	exp	$\exp^2$	hs	col
$n = 400$	0.002	0.008	0.005	0.003	0.013	0.001	0.005	0.003	0.002	0.010	0.001	0.003	0.001	0.003	0.004
$n = 800$	0.014	0.017	0.014	0.009	0.080	0.009	0.015	0.012	0.008	0.061	0.007	0.013	0.010	0.007	0.048
$n = 1600$	0.134	0.095	0.071	0.033	0.260	0.068	0.076	0.057	0.019	0.157	0.014	0.041	0.035	0.009	0.071
$n = 3200$	0.491	0.212	0.157	0.069	0.570	0.298	0.166	0.126	0.042	0.380	0.101	0.104	0.098	0.022	0.176
$n = 6400$	0.904	0.455	0.338	0.106	0.901	0.700	0.361	0.284	0.069	0.717	0.361	0.246	0.206	0.051	0.364

When there are several regressors, the identifying assumptions 1 to 3 imply that the whole vector $\beta(\tau)$ is constant as a function of τ. Thus, the tests suggested in Section 3 simultaneously use all coefficients

and weight them by the inverse of the covariance matrix (of the difference between $\hat{\beta}(\tau)$ and $\hat{\beta}(0.5)$). In addition, we can also separately test whether each element of the vector is constant as a function of the quantile. This will inform the researcher about the sources of the rejection when the global test rejects the null hypothesis.

The results in Table 2 give the empirical rejection probabilities of our tests when the coefficients are homogeneous (empirical size) and heterogeneous (empirical power). Our tests are somewhat conservative with none of the empirical sizes being above the theoretical size of 5%. However, under the null hypothesis the rejection probability does converge to the intended one even if it does not yet attain it with 6400 observations. The power of the tests is very low when $n = 400$ but attains 98.6% when $n = 6400$ for the KS statistic, which tends to be more powerful than the CM statistic. The dataset used in our application contains $45,296$ observations such that we can expect our tests to be extremely powerful. The widest range of quantiles that we considered, $[0.05, 0.95]$, delivers the best performance.

The tests of the constancy of a single coefficient are also slightly conservative. When n is large the empirical power of all tests are above 5%, which shows that all variables contribute to the rejection of the null hypothesis. The return to a college degree is the most heterogeneous coefficient with rejection rates close to those obtained when jointly testing all variables.

5 Female wage distribution in the USA

The estimation of the female wage function is the most frequent application of sample selection estimators, see e.g. Buchinsky (1998a) and Mulligan and Rubinstein (2008). We use the 2011 Merged Outgoing Rotation Groups of the Current Population Survey, which offer a good measure of hourly wages. Similarly to many examples in the wage regression literature, we restrict our sample to white non-Hispanic adults who are between 25 and 54 years old and exclude individuals who are self-employed, have allocated earnings, or work in the military, agricultural, and public sectors.

We classify females as either working or not according to whether they worked at least 35 hours during the week preceding the surveys. The dependent variable, Y, is the log hourly wage. Our vector of regressors, X_2, contains five educational dummies (at least high school, some college, associate degree, bachelor degree, and advanced degree), potential experience and potential experience squared as well as their interaction with educational attainment in years, three regional dummy variables, and an indicator for being married. The vector X_1 consists of the same elements plus the numbers of children in three age brackets (0-2, 3-5 and 6-13 years) and their interaction with the marital indicator.[7] Descriptive statistics can be found in Appendix B. We use the CPS earning weights for all our calculations.

[7] Due to data coding, in about 5% of cases we are only able to determine lower and upper bounds on the number of children in the three age brackets. We impute the average between the lower and upper bound in these cases. In contrast, the total number of children is always known.

We implement our test procedure as described in Section 3. The estimates of the selection equation by the Klein and Spady estimator, reported in Table 3, show that the coefficients on all six variables excluded from the outcome equation are strongly significant. The signs of the other coefficients are also as expected with a higher employment probability for non-married and high-educated women. The bandwidth, determined by GCV, was set to 0.25 times the standard deviation of the fitted index $X_1'\hat{\gamma}$.

Table 3

Female wage distribution in the USA

	Selection equation	Outcome equation - Buchinsky (1998a)		
	Klein and Spady	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
Experience	0.013 (0.012)	0.021 (0.003)	0.031 (0.003)	0.034 (0.003)
Experience squared	-0.001 (3e-4)	-3e-4 (7e-5)	-5e-4 (7e-5)	-5e-4 (8e-5)
High school	1.848 (0.255)	0.147 (0.032)	0.260 (0.033)	0.273 (0.035)
Some college	0.216 (0.071)	0.117 (0.017)	0.132 (0.015)	0.146 (0.018)
Associate	0.630 (0.100)	0.088 (0.020)	0.128 (0.020)	0.203 (0.022)
Bachelor	0.234 (0.078)	0.229 (0.020)	0.255 (0.021)	0.230 (0.024)
Advanced	0.437 (0.097)	0.232 (0.023)	0.236 (0.023)	0.241 (0.024)
Exp.×education	-0.005 (0.002)	2e-3 (5e-4)	3e-3 (5e-4)	4e-3 (7e-4)
Exp.^{2}×education	5e-5 (6e-5)	-4e-5 (1e-5)	-6e-5 (1e-5)	-8e-5 (2e-5)
Married	-0.392 (0.065)	0.035 (0.013)	0.033 (0.012)	0.020 (0.014)
# children $[0, 2]$	-1.000			
# children $[3, 5]$	-0.530 (0.130)			
# children $[6, 13]$	-0.249 (0.067)			
Married×# children $[0, 2]$	-0.262 (0.152)			
Married×# children $[3, 5]$	-0.339 (0.113)			
Married×# children $[6, 13]$	-0.336 (0.069)			

Note: Column 1 contains the coefficients of the selection equation obtained by the Klein and Spady (1993) estimator. Columns 2-4 show the quantile regression coefficients obtained by the Buchinsky (1998a) estimator. Standard errors are reported in parentheses. Three regional dummy variables are also part of the regressors but are not shown for brevity. These coefficients are consistent only if the independence assumption is satisfied or if selection is random. Our tests in Table 4 reject the independence assumption.

Table 3 also contains the 0.25, 0.5, and 0.75 quantile regression coefficients estimated using the Buchinsky (1998a) estimator. Figure 3 shows all quantile coefficients between the 2nd and the 98th percentile of four of the most important covariates. The order of the power series in the inverse Mill's ratio, also determined by GCV, is 5, which is large compared to the literature. Note that the coefficients are consistent only if the independence assumption is satisfied or if selection is random. They are provided here because they serve—among other quantile coefficients—as a basis for our tests.

The linear term in experience is more or less monotonically increasing as we move up the distribution.

At the same time, the quadratic term is decreasing as a function of the quantile. In other words, the whole shape of the experience-wage function is changing with the quantile. This is a clear hint against the independence assumption. In addition, the returns to education are also increasing in the quantile. The returns are becoming flatter as a function of the quantile only for the highest levels of education, maybe due to a more homogeneous population.

Figure 3 here

Table 4 reports the p-values of our tests of the conditional independence assumption. The null hypothesis tested in the first row is the constancy of the whole quantile regression vector, i.e. the validity of the model defined by Assumptions 1 to 9. The remaining rows give the p-values for each coefficient separately. The results are reported for two choices of block size ($b = n$ and $b = 20 + n^{1/2.01}$) and two ranges of quantiles ($\mathcal{T} = [0.02, 0.98]$ and $\mathcal{T} = [0.2, 0.8]$).[8] The number of bootstrap replications B was set to $1,000$.

Table 4

Empirical application - test results

Statistic	Kolmogorov-Smirnov				Cramér-von-Mises			
$\tau \in$	$[0.02, 0.98]$		$[0.20, 0.80]$		$[0.02, 0.98]$		$[0.20, 0.80]$	
b (block size)	45,296	227	45,296	227	45,296	227	45,296	227
All variables	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Experience	0.000	0.000	0.002	0.011	0.000	0.000	0.013	0.019
Experience^2	0.006	0.018	0.130	0.159	0.015	0.024	0.237	0.244
High school	0.003	0.048	0.010	0.046	0.000	0.008	0.016	0.029
Some college	0.002	0.014	0.422	0.381	0.028	0.028	0.781	0.722
Associate	0.007	0.027	0.004	0.014	0.000	0.000	0.000	0.000
Bachelor	0.113	0.179	0.387	0.351	0.195	0.218	0.275	0.300
Advanced	0.188	0.219	0.303	0.270	0.386	0.349	0.628	0.553
Exp.×education	0.338	0.317	0.242	0.190	0.131	0.169	0.076	0.113
Exp.^{2}×education	0.438	0.343	0.291	0.223	0.233	0.221	0.172	0.171
Married	0.140	0.161	0.354	0.311	0.151	0.149	0.323	0.304

Note: 45,296 observations. Critical values obtained using 1,000 bootstrap repetitions.

The hypothesis of the validity of the sample selection model defined in Section 2 is clearly rejected in any set up as the p-values are always zero. A one by one investigation of the covariates shows that this rejection is mainly due to heterogeneous returns to experience and to lower level education degrees (in particular high school and associate degrees). These results and those for two other data sets reported in the working paper version of this article (Huber and Melly 2011) cast serious doubts on the validity of

[8] Not reported are the results for two other block sizes and two other ranges of quantiles, which are extremely similar.

traditional sample selections models in female wage regressions. The conclusion suggests a few tracks to explore without assuming homogeneous coefficients.

6 Conclusion

Assuming additivity and conditional independence of the error term in the outcome equation is rather restrictive. It implies that all individuals with the same selection probability react identically to changes in the observable characteristics. In economic models, the unobservable random terms are not simply measurement errors but they often have important economic interpretations. The recent econometric literature has relaxed many restrictions on the interaction between observed and unobserved variables. Advances have been reported in selection on observables, instrumental variables, and panel data models, among many others (for a discussion, see e.g. Matzkin 2007).

Somewhat surprisingly, the sample selection model has been excluded from this trend. The consistency of almost all sample selection estimators hinges on the additivity and conditional independence of the error term. This is also the case in the quantile regression model of Buchinsky (1998a). However, in the quantile regression framework the conditional independence assumption implies that the quantile slope coefficients are equal to the mean slope coefficients. In other words, the heterogeneity that we want to analyze is excluded by assumption. Applications of the sample selection correction for quantile regression that have found significant differences across the coefficients estimated at distinct quantiles have merely proven the violation of the underlying assumptions and the inconsistency of the estimator.

A general lesson to draw from this example is the dangerousness of importing one-to-one mean recipes to quantile models. Another example is given by the fitted value approach in endogenous models, which is justified only by a similar independence assumption. The fitted value IV estimators suggested by Amemiya (1982), Powell (1983), and Chen and Portnoy (1996) are, therefore, not useful for analyzing heterogeneity (which was *not* the intention of their authors). Similarly, the control function estimators in Lee (2007), Blundell and Powell (2007), and Carneiro and Lee (2009) are consistent only if the coefficients do not vary across quantiles.

Given the importance of conditional independence for the identification of (mean and quantile) sample selection models, we propose the first formal test for this assumption. Our test is based on the conditional quantile regression process estimated by the procedure of Buchinsky (1998a), which is consistent under the null hypothesis, and compares the coefficients obtained at different quantiles. Monte Carlo simulations provide evidence on the satisfactory power and size properties of our tests. We clearly reject the conditional independence assumption in an application to recent US wage data.

What can be done if the independence assumption is rejected? Unfortunately, the parameters of interest are then no longer point identified. In our companion paper (Melly and Huber 2012), we derive the sharp bounds on the quantile regression parameters when this assumption is no longer imposed.

It appears that there are only two ways to recover point identification.[9] The first one is to impose a parametric assumption on (at least) the copula function between the error terms in the selection and outcome equations. Donald (1995) makes one step in this direction and proposes a two-step estimator that allows for conditional heteroscedasticity but requires the error terms to be bivariate normally distributed. Arellano and Bonhomme (2011) obtain point identification by parametrizing only the copula while keeping the marginal distributions of the error terms nonparametric. The second way to recover point identification is to use an identification-at-infinity argument. The latter requires not only that some subpopulation is observed with probability (close to) one but also that the slope parameters are identified from this subpopulation.

Another strategy consists in changing the estimand by considering a different population. Newey (2007) analyzes a nonseparable model but shows identification only in the selected population instead of the entire population. In the absence of an exclusion restriction, Lee (2009), Lechner and Melly (2010), and Huber and Mellace (2011) provide sharp bounds for several subpopulations. This is of interest in some applications but clearly not in all. For instance, it does not allow comparing wages of different population groups like male and female or white and nonwhite individuals. These wage comparisons were precisely the reason for which Gronau (1974) developed corrections for the sample selection bias.

Appendix A: Proof of Proposition 1

We verify that the conditions for inference I.1-I.4 in Chernozhukov and Hansen (2006) are satisfied; Proposition 1 will then follow from Theorem 5 in Chernozhukov and Hansen (2006). Note that Chernozhukov and Hansen (2006) develop these new inference tools in their IV quantile regression framework but state the assumptions in a way that is general enough to cover other estimators such as standard quantile regression and, as we will show, the Buchinsky (1998a) sample selection estimator.

Assumption I.1. in Chernozhukov and Hansen (2006) is satisfied with $R(\tau) = [1, 1, ..., 1]$ being a vector of 1 of length $d_X \equiv \dim(x_{2i})$, $\theta(\tau)$ being the vector of slope coefficients $\beta(\tau)$, and $r(\tau)$ being the vector of median coefficients $\beta(0.5)$. Our continuity assumptions imply that $\beta(\tau)$ is a continuous function. Under the null hypothesis, $g(\tau) = 0$ for all τ and under the global alternative hypothesis $g(\tau) \neq 0$ for some τ.

Our estimator defined in (5) corresponds almost exactly to the estimator defined in equation (7) of Lee (2007). The first difference is that in our case the estimation is performed only in the selected population , while Lee (2007) uses the whole population. Our problem fits, however, into his framework by redefining the trimming function to be zero for observations with $D = 0$. The second difference is that the first step fitted values are estimated by the Klein and Spady (1993) estimator in our case, while Lee (2007) uses

[9] Naturally, it is also possible to add structure to the model or use auxiliary data. For instance, Chen and Khan (2003) impose de facto a new type of exclusion restriction—the availability of a regressor that affects the variance but not the location of the dependent variable—to identify a multiplicative heteroscedastic model.

quantile regression. This is not an issue because Lee (2007) states the assumptions in a way that allows for other $\sqrt{n}$ consistent and asymptotically linear estimators.

Our assumptions are the uniform in $\tau \in \mathcal{T}$ and conditional on $D = 1$ versions of the assumptions of Lee (2007). To simplify the notation we replace trimming with respect to X_2 with a compactness assumption on the distribution of X_2. In addition, by Assumption 3, $f_U(u|x_{1i}'\delta_0, x_{1i}) = f_U(u|x_{1i}'\delta_0)$, such that we can simplify the definition of m_i. We apply Theorem 3.1 in Lee (2007) and obtain the following influence function (equations 39 and 40 in Lee 2007):

$$\sqrt{n}(\hat{\beta}(\tau) - \beta_0) = A\Phi_J(\tau)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\ell_{2i}(\tau) + \Delta_{2\gamma}\Delta_P^{-1} k_i) + o_p(1), \tag{12}$$

where A denotes the $[d_X \times (d_X + J)]$ matrix such that $A = (I_{d_X}, 0_{d_X+J})$, where I_{d_X} is the d_X-dimensional identity matrix, 0_{d_X+J} is the $[d_X \times J]$ matrix of zeros, d_X is the dimension of x_{2i}, and

$$\begin{aligned} \ell_{2i}(\tau) &= t_i(\tau - 1(y_i < x_{2i}'\beta_0 + h_\tau(x_{1i}'\delta_0)))p_{Ji}, \\ \Delta_{2\gamma} &= E\left[t_i f_\varepsilon(Q_\varepsilon(\tau|x_{1i}'\delta_0) \frac{dh_\tau(X_1'\delta_0)}{d\delta_0} p_{Ji}\right]. \end{aligned}$$

The matrix A selects the relevant components of the $d_X + J$ score functions. The influence function can equivalently be written as in equation (10). This satisfies Assumption I.3 of Chernozhukov and Hansen (2006).[10]

We estimate the elements in the score function (10) by $\hat{\Delta}_{b,\tau}$, $\hat{\ell}_i(\tau)$, $\hat{\Delta}_\gamma$, $\hat{\Delta}_P$, and $K(x_{1i}, d_i, \hat{\gamma})$. Under assumptions 1 to 9, arguments identical to those used in Lemma A.9. in Lee (2007) imply that $\hat{\Delta}_{b,\tau}$ is uniformly consistent for the nonsingular matrix $\Delta_{b,\tau}$. Similarly, arguments identical to those used in Lemma A.11. in Lee (2007) imply that $\hat{\Delta}_\gamma$ is consistent for Δ_γ. Since the trimming function is 0 when $x_{1i}'\hat{\gamma}$ is outside of a compact set and x_{1i}' takes value in a compact set, exploiting the boundedness and monotonicity of the indicator function, it follows that the class of functions

$$T(\tau - 1(Y < X_2'\beta + h_\tau(X_1'\gamma)))M, \ \beta \in R^{d_X} \text{ and } \tau \in \mathcal{T}$$

is Donsker. In addition, Δ_γ, Δ_P^{-1}, and k_i are bounded by assumption. Since sums and products of Donsker functions are Donsker, this verifies Assumption I.4 (a) of Chernozhukov and Hansen (2006).[11] Assumption I.4 (b) of Chernozhukov and Hansen (2006) holds by our Assumptions 1 to 3 that imply that $x_{2i}'\beta_0 + h_\tau(x_{1i}'\gamma_0)$ is the τ conditional quantile of y_i given x_{1i} and $d_i = 1$. Assumption I.4. (c) of Chernozhukov and Hansen (2006) holds by the bounded density assumption.

By the central limit theorem for empirical processes indexed by Donsker classes of functions (Van der Vaart and Wellner 1996), $\Delta_{b,\cdot}\sqrt{n}(\hat{\beta}(\cdot) - \beta_0)$ converges in distribution to a zero mean Gaussian process $z(\cdot)$ with covariance function $\Sigma(\tau, \tau')$:

$$\Sigma(\tau, \tau') = [\min(\tau, \tau') - \tau\tau'] \cdot E[m_i m_i'] + \Delta_\gamma \Delta_P^{-1} var(k_i).$$

[10] Buchinsky (1998a) obtains the same score function in a more heuristic way.

[11] The verification of this condition follows almost directly from Lemma B.2.II. in Chernozhukov and Hansen (2006).

This satisfies Assumption I.2 of Chernozhukov and Hansen (2006). The result of Proposition 1 follows from Theorem 5 in Chernozhukov and Hansen (2006), parts (a) and (b), where the continuity of the distribution of the test statistic follows from part (c).

Appendix B: Descriptive statistics

Table A1

Means of the relevant variables

	All observations	Employed full-time	Not employed full-time
High school	0.904	0.951	0.865
Some college	0.626	0.694	0.570
Associate	0.447	0.521	0.387
Bachelor	0.318	0.374	0.273
Advanced	0.089	0.113	0.070
Experience	21.76	21.41	22.044
Married	0.718	0.668	0.759
# children $[0, 2]$	0.162	0.103	0.211
# children $[3, 5]$	0.204	0.138	0.259
# children $[6, 13]$	0.496	0.392	0.582
Number of obs.	45,296	20,789	24,507

References

Ahn, H., and J. Powell (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3–29.

Amemiya, T. (1982): "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–711.

Angrist, J. (1997): "Conditional Independence in Sample Selection Models," *Economics Letters*, 54, 103–112.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563.

Arellano, M., and S. Bonhomme (2011): "Quantile Selection Models," Walras-Bowley Lecture, North American Summer Meeting of the Econometric Society, St. Louis, 10 June 2011.

Blundell, R. W., and J. L. Powell (2007): "Censored regression quantiles with endogenous regressors," *Journal of Econometrics*, 141(1), 65–83.

Buchinsky, M. (1994): "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica*, 62, 405–458.

——— (1998a): "The dynamics of changes in the female wage distribution in the USA: A quantile regression approach," *Journal of Applied Econometrics*, 13, 1–30.

——— (1998b): "Recent advances in quantile regression models: A practical guidline for empirical research," *The Journal of Human Resources*, 33, 88–126.

——— (2001): "Quantile regression with sample selection: Estimating women's return to education in the U.S.," *Empirical Economics*, 26, 87–113.

Carneiro, P., and S. Lee (2009): "Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality," *Journal of Econometrics*, 149(2), 191–208.

Chen, L., and S. Portnoy (1996): "Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equations models," *Comm. Statist. Theory Methods*, 25, 1005–1032.

Chen, S., and S. Khan (2003): "Semiparametric estimation of a heteroskedastic sample selection model," *Econometric Theory*, 19, 1040–1064.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2012): "Inference on Counterfactual Distributions," cemmap working paper 05/12.

Chernozhukov, V., and I. Fernández-Val (2005): "Subsampling inference on quantile regression processes," *Sankhya: The Indian Journal of Statistics*, 67, 253–276, Part 2.

Chernozhukov, V., and C. Hansen (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132, 491–525.

Cosslett, S. (1991): "Distribution-Free Estimator of a Regression Model with Sample Selectivity," in *Nonparametric and semiparametric methods in econometrics and statistics*, ed. by W. Barnett, J. Powell, and G. Tauchen, pp. 175–198. Cambridge University Press, Camdridge, UK.

Craven, P., and G. Wahba (1979): "Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation," *Numerische Mathematik*, 31, 377–403.

Das, M., W. K. Newey, and F. Vella (2003): "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70, 33–58.

Delecroix, M., M. Hristache, and V. Patilea (2006): "On semiparametric M-estimation in single-index regression," *Journal of Statistical Planning and Inference*, 136, 730–769.

Donald, S. G. (1995): "Two-step estimation of heteroskedastic sample selection models," *Journal of Econometrics*, 65, 347–380.

Gallant, A., and D. Nychka (1987): "Semi-nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363–390.

Gronau, R. (1974): "Wage comparisons-a selectivity bias," *Journal of Political Economy*, 82, 1119–1143.

Guntenbrunner, C., and J. Jurečková (1992): "Regression Quantile and Regression Rank Score Process in the Linear Model and Derived Statistics," *Annals of Statistics*, 20, 305–330.

Härdle, W., P. Hall, and H. Ichimura (1993): "Optimal smoothing in single-index models," *Annals of Statistics*, 21, 157–178.

Heckman, J. J. (1974): "Shadow Prices, Market Wages and Labor Supply," *Econometrica*, 42, 679–694.

——— (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables, and a Simple Estimator for such Models," *Annals of Economic and Social Measurement*, 5, 475–492.

——— (1979): "Sample selection bias as a specification error," *Econometrica*, 47, 153–161.

Huber, M., and G. Mellace (2011): "Sharp bounds on causal effects under sample selection," *University of St. Gallen Discussion Paper no. 2011-34*.

Huber, M., and B. Melly (2011): "Quantile regression in the presence of sample selection," *University of St. Gallen, Economics Working Paper No 1109*.

Klein, R. W., and R. H. Spady (1993): "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 61, 387–421.

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

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

——— (1982): "Robust tests for heteroskedasticity based on regression quantiles," *Econometrica*, 50, 43–62.

Koenker, R., and K. F. Hallock (2000): "Quantile regression An introduction," *Discussion paper for the Journal of Economic Perspectives "Symposium on Econometric Tools"*.

Koenker, R., and Z. Xiao (2002): "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612.

Lechner, M., and B. Melly (2010): "Partial Identification of Wage Effects of Training Programs," *Brown University working paper*.

Lee, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76, 1071–1102.

Lee, S. (2007): "Endogeneity in quantile regression models: A control function approach," *Journal of Econometrics*, 141, 1131–1158.

Matzkin, R. (2007): "Nonparametric Identification," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, pp. 5307–5368. Elsevier, Amsterdam.

Melly, B., and M. Huber (2012): "Nonseparable sample selection models," *unpublished manuscript*.

Mincer, J. (1973): *Schooling, Experience, and Earnings*. NBER, New York.

Mulligan, C. B., and Y. Rubinstein (2008): "Selection, Investment, and Women's Relative Wages Over Time," *Quarterly Journal of Economics*, 123, 1061–1110.

Newey, W. K. (2007): "Nonparametric continuous/discrete choice models," *International Economic Review*, 48, 1429–1439.

Newey, W. K. (2009): "Two-step series estimation of sample selection models," *Econometrics Journal*, 12, S217–S229.

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

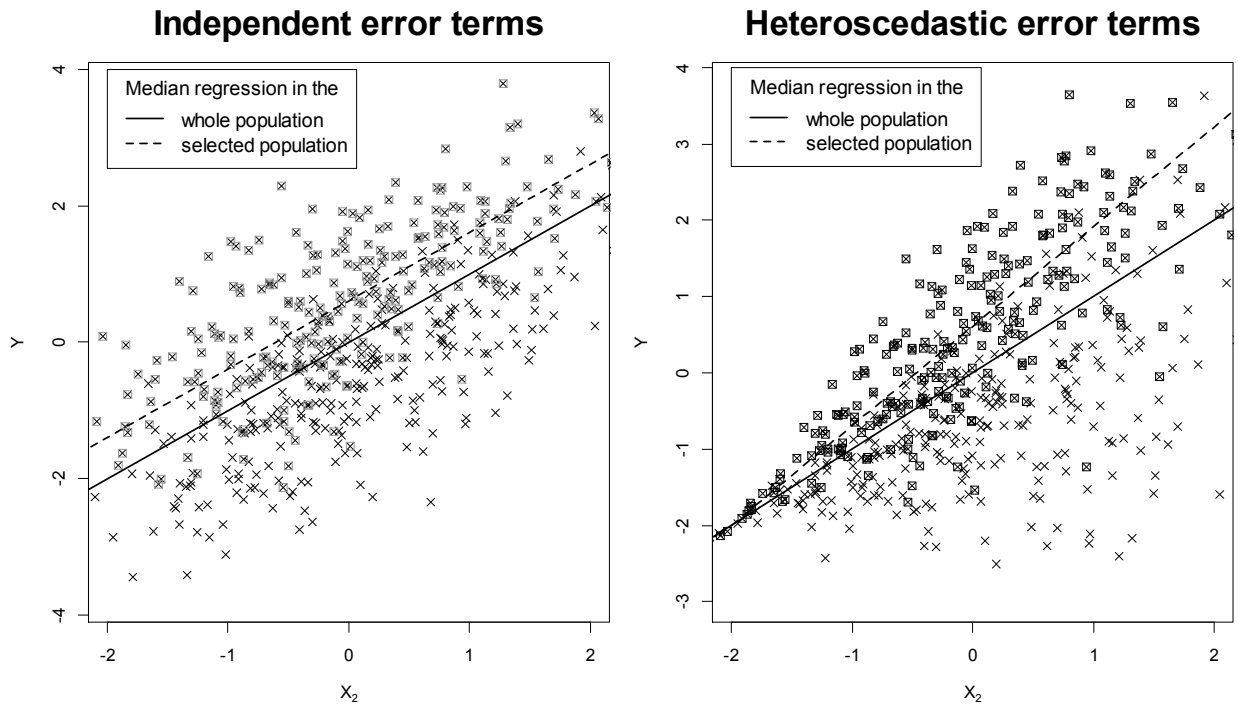
Newey, W. K., J. L. Powell, and J. Walker (1990): "Semiparametric estimation of selection models: Some empirical results," *American Economic Review*, 80, 324–328.

Powell, J. (1983): "The asymptotic normality of two-stage least absolute deviations estimators," *Econometrica*, 51, 1569–1575.

Powell, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.

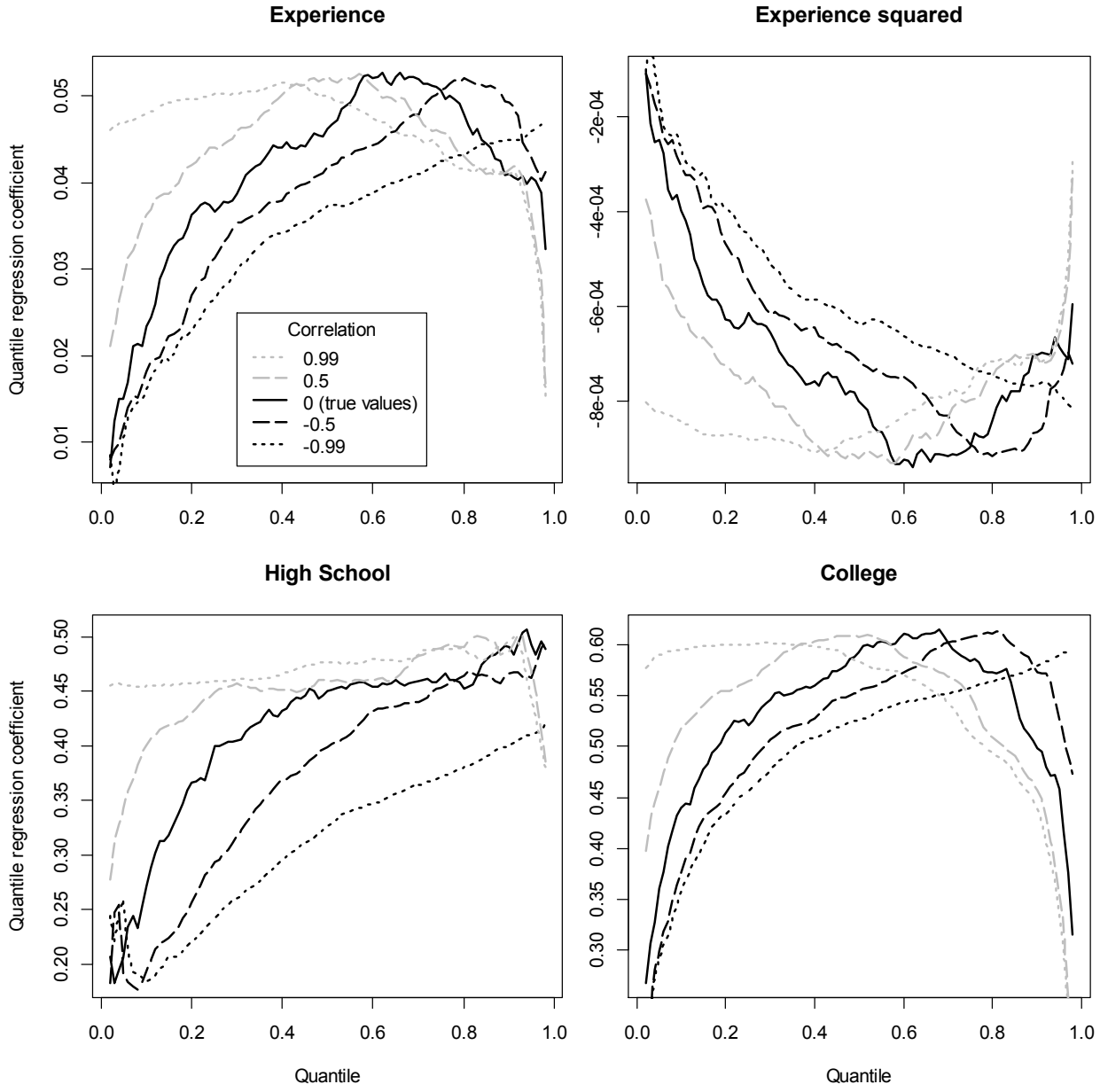
——— (1987): "Semiparametric Estimation of Bivariate Latent Variable Models," *unpublished manuscript*, University of Wisconsin-Madison.

Figure 1. Sample selection under independence and heteroscedasticity.



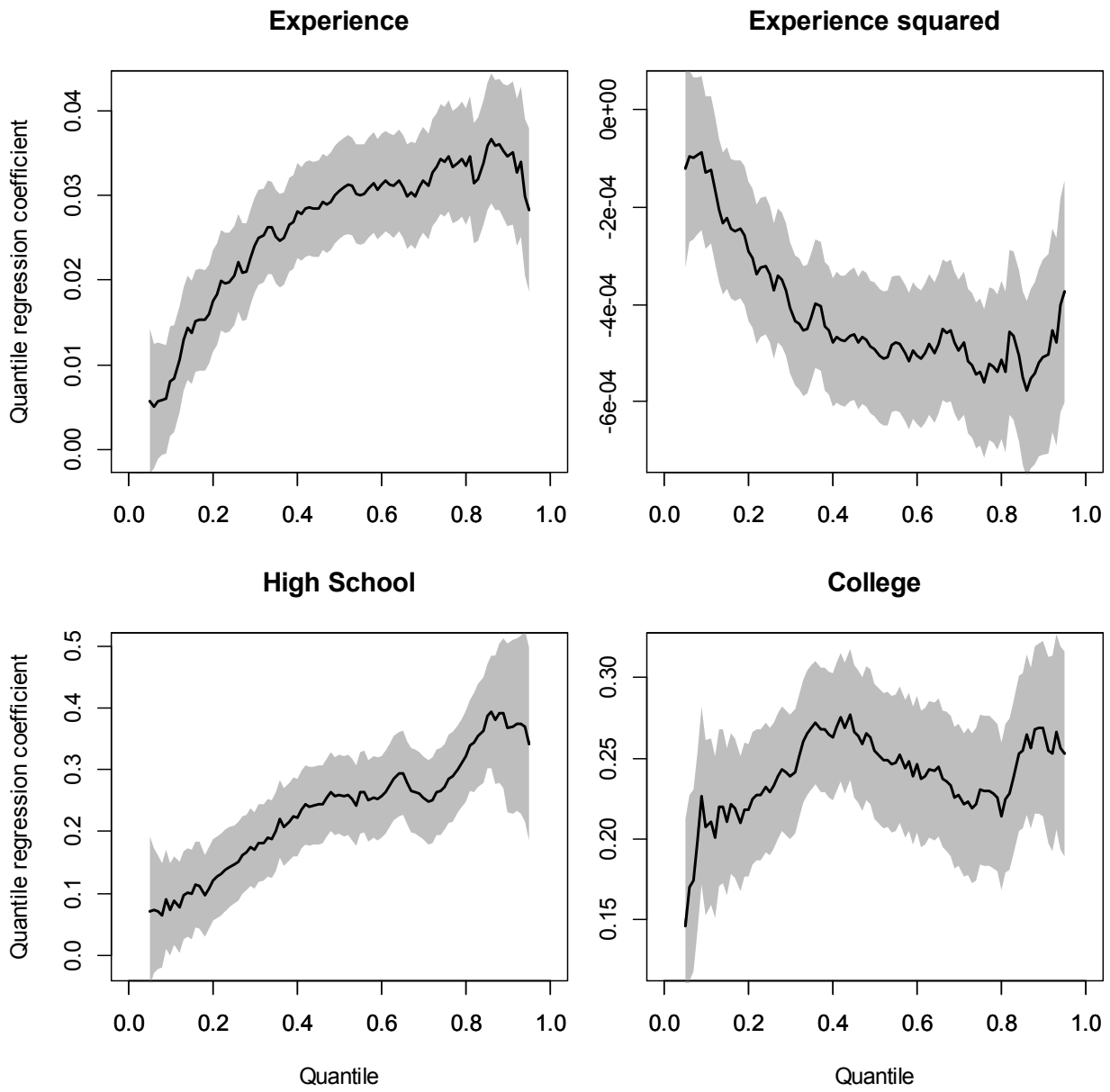
Note: Random samples of 500 observations. Only the observations surrounded by a box are observed. All observations have X_1 set such that $Pr(D=1|X_1)=0.5$.

Figure 2. Asymptotic estimates in realistic DGPs.



Note: Details about the data generating process can be found in Section 4.2. The true coefficients correspond to the male coefficients in log wage regressions in the CPS 2011. The plotted lines give the values to which the estimates of Buchinsky (1998a) converge for five different correlations between the error terms in the selection and outcome equation, $\mathcal{E}$ and U.

Figure 3. Quantile regression coefficients corrected for sample selection



Note: Quantile regression coefficients estimated using the estimator of Buchinsky (1998a). Pointwise 95% confidence intervals are also plotted. In addition to the four variables the coefficients of which are plotted, the vector of regressors also contains: three regional dummy variables, indicators for some college, associate degree, advanced degree, and being married, and interactions between experience and experience squared with education in years. These coefficients are consistent only if the independence assumption is satisfied or if selection is random. Our tests in Table 4 reject the independence assumption.