Extremal Quantile Regression for Sample Selection Models, with an Application to Black-White Wage Gap

Preliminary

Xavier D’Haultfoeuille* Arnaud Maurel† Yichong Zhang‡

February 15, 2013

Abstract

We consider the estimation of a semiparametric location-scale model subject to endogenous selection, without instruments nor large support regressor. Following D’Haultfoeuille & Maurel (2013), identification relies on independence between covariates and selection, when the potential outcome tends to infinity. In this context, we propose a simple estimator based on extremal quantile regressions. We establish asymptotic normality by extending the results of Chernozhukov (2005) to allow for selection. Finally, we apply our method to the estimation of the black-white wage gap among males from the NLSY79, revisiting the results of Neal & Johnson (1996).

*CREST. E-mail address: xavier.dhaultfoeuille@ensae.fr.
†Duke University and IZA. E-mail address: apm16@duke.edu.
‡Duke University. E-mail address: yz98@duke.edu.
1 Introduction

Since the seminal work of [Heckman (1974)], endogenous selection has been recognized as one of the key methodological issues arising in the analysis of microeconomic data. The usual strategy to deal with selection is to rely on instruments that determine selection but not the outcome (see, e.g., [Ahn & Powell 1993] [Das et al. 2003] and [Newey 2009]). Because the validity of such exclusion restriction is always debatable, and instruments may be difficult to find in some applications, alternative solutions have been proposed in the literature. Arguably, identification at infinity is the most influential of these. In particular, [Chamberlain (1986)] showed that if some individuals have an arbitrary large probability of selection, and the model is linear, then one can use these individuals for identification. [Lewbel (2007)] generalized this result by proving that identification can be achieved without imposing any structure on the outcome equation, provided that a special regressor has a support which includes that of the error term, and under restrictions on the selection equation. Again, in many applications, it may be hard to find a such regressor. [D’Haultfoeuille & Maurel (2013)] have recently shown that identification at infinity was actually possible without such a covariate. The starting point is that if selection is truly endogenous, then we can expect the effect of the outcome on selection to dominate those of the covariates for large values of the outcome. Following this idea, they prove identification if, basically, the selection variable is independent of the covariates at the limit, i.e., when the outcome tends to the upper bound of its support.

The aim of this paper is to build on this paper and develop semiparametric inference for this model. For application purposes, we consider a linear location-scale model for the outcome equation, instead of the nonparametric location-scale model considered by [D’Haultfoeuille & Maurel (2013)]. The model is left unrestricted otherwise. In particular, we impose no structure on the selection equation, apart from the independence at the limit condition stated above. One of the challenges for estimation in this framework is that the identification proof of [D’Haultfoeuille & Maurel (2013)] is not fully constructive. Our first insight is to show that under independence at infinity, an extremal quantile regression ignoring selection issues allows one to recover combinations of the structural parameters. The structural parameters can then be estimated in a second step by a simple minimum distance estimator. This insight is important for at least two reasons. First, the estimator is very simple to implement. In particular, and contrary to most of the semiparametric estimators for sample selection models, it does not require a nonparametric first-step estimator. Second, the asymptotics of extremal quantile regressions, in the absence of sample selection, have been thoroughly studied in an influential paper by [Chernozhukov (2005)], and we can rely on his analysis to develop asymptotic inference in our setting.

To do so, we extend his results on intermediate order extremal quantile regressions to our setting. A challenge is that, because of selection, extreme conditional quantiles are not exactly...
linear here, but only equivalent to a linear form as the quantile index \( \tau(T) \) tends to zero. Hence, we face a trade-off that is typical in non or semiparametric analysis. Choosing a moderately small quantile index decreases the variance of the estimator but at the price of a higher bias. Conversely, choosing a vary small quantile index mitigates the bias, but increases the variance. We provide conditions under which both vanish asymptotically, and the estimators of the structural parameters are asymptotically normal. As in the case without selection examined by Chernozhukov (2005), their rate of convergence are not standard, on the other hand, and depends on the tail behavior of the error term and the rate of convergence of the quantile index \( \tau(T) \). These results require an appropriate choice of the quantile index, in a similar way nonparametric kernel regressions require appropriate bandwidth choices. But contrary to the kernel estimator case, the rate of convergence of the quantile index that ensures consistency depends in a complicated way on the data generating process. A similar issue arises in the estimation at infinity of the intercept of sample selection models (see Andrews & Schafgans 1998 and Schafgans & Zinde-Walsh 2002), as well as in the estimation of extreme value index (see Drees & Kaufmann 1998 and Danielsson et al. 2001). This is a difficult problem. In the paper, we propose a data-driven procedure that simply consists of minimizing the minimum distance criterion used for estimation with respect to the quantile index. Simulations suggest that this procedure works well in practice.

Finally, we apply our method to the estimation of the black-white wage gap among males from the National Longitudinal Survey of Youth 1979 (NLSY79), revisiting the influential results of Neal & Johnson (1996). To the extent that labor force participation differs markedly between blacks and whites, as was already noted by Butler & Heckman (1977), correcting for selection is crucial to estimate consistently the black-white wage gap. Our estimation method suggests that the black-white wage differential is actually larger than what has been found under the imputation method proposed by Neal & Johnson (1996) and Johnson et al. (2000).

The paper is organized as follows. Section 2 presents the set-up and discusses the identification results in D’Haultfoeuille & Maurel (2013). Section 3 defines the estimator and establishes the main asymptotic normality results. Section 4 discusses some simulation results. Section 5 applies our results to the estimation of the black-white wage gap among males using the NLSY79 data. Section 6 concludes. All the proofs are collected in the appendix.

2 The set-up and identification

Let \( Y^* \) denote the outcome of interest and \( X \in \mathbb{R}^d \) denotes the vector of covariates, excluding the constant. We suppose that they are related through the location-scale model

\[
Y^* = X'\beta + (1 + X'\delta)\varepsilon, \tag{2.1}
\]
where we suppose that \(1 + X'\delta > 0\). Note that because we do not standardize \(\varepsilon\), we can always suppose that there is no intercept and fix the constant in the multiplier of \(\varepsilon\) to one. Our focus is primarily to infer \(\beta\) and \(\delta\). The issue is that we face an endogenous selection here, so that only \((D,Y = D Y^*, X)\) is observed, where \(D\) denotes the selection dummy. Moreover, we do not suppose that any instrument affecting \(D\) but not directly \(Y^*\) is available. Instead, identification is achieved under the following conditions. Hereafter, we denote by \(F_U\) and \(S_U\) the cumulative distribution function (cdf) and survival function of any random variable \(U\).

**Assumption 1.** (Exogeneity) \(X \perp \varepsilon\).

**Assumption 2.** (Covariates) \(X\) has a compact support \(\text{Supp}(X)\). Let \(X = [1, X']', Q_X = E(XX')\) is full rank.

**Assumption 3.** (Tail of the residual) (i) \(\sup(\text{Supp}(\varepsilon)) = \infty\), (ii) \(S_{\exp(\varepsilon)}\) is not slowly varying at infinity\(^2\) and (iii) \(S_\varepsilon\) is in the attraction domain of extreme value (EV) distributions.

**Assumption 4.** (Independence at the limit) There exists \(h > 0\) such that for all \(x \in \text{Supp}(X)\), \(\lim_{y \to \infty} P(D = 1 | X = x, Y^* = y) = h > 0\).

We refer readers to D’Haultfoeuille & Maurel (2013) for the discussion of most of the assumptions and only comment on three conditions here. First, compact support is not required for identification but is needed when we apply extreme regression quantile techniques. In practice, econometricians can trim the dataset to make the support compact. Assumption \(\text{3(ii)}\) is satisfied if for instance \(E(\exp(\beta \varepsilon)) < \infty\) for some \(\beta > 0\). Assumption \(\text{3(i)}\) and (iii) are not necessary for identification but will be used subsequently. Part (iii) is mild and satisfied by many standard continuous cdf, including the normal one.

The next theorem is a direct application of the identification results in D’Haultfoeuille & Maurel (2013). Theorem 2.1 in their paper imply that \(x'\beta\) and \(x'\delta\) are identified. Assumption \(\text{2}\) then ensures that \(\beta\) and \(\delta\) are identified.

**Theorem 2.1.** Under Assumptions \(\text{2,3,4}\) \(\beta\) and \(\delta\) are identified.

### 3 Estimation

#### 3.1 Definition of the estimator

We suppose to have in hand a sample \((D_t, Y_t, X_t)_{t=1...T}\) of \(T\) i.i.d. random variables distributed as \((D, Y, X)\) and thus satisfying the previous restrictions. For any random variable \(U\), let us denote by \(\bar{U} = -U\) and let \(Q_U\) denote the quantile function of \(U\), \(Q_U(u) = \inf\{u : F_U(u) > \tau\}\). The

---

2 Recall that a slowly varying function \(L\) at \(+\infty\) satisfies \(L(tx)/L(x) \to 1\) as \(x \to \infty\), for all \(t > 0\).
starting point for identification is that under Assumptions 1 and 4, we have (see D’Haultfoeuille & Maurel 2013)

\[ F_{\tilde{Y}|X}(y) \sim hF_{\tilde{\epsilon}}((y + x'\beta)/(1 + x'\delta)) \]  

(3.1)

as \( y \to -\infty \). Now, the key insight is that if one also imposes Assumption 3, we can invert both sides and still maintain the equivalence. This is convenient because it implies that the conditional quantile regression is asymptotically linear.

**Theorem 3.1.** Under Assumptions 1 - 4 as \( \tau \to 0 \),

\[ Q_{\tilde{Y}|X}(\tau|x) \sim \gamma(\tau) + x'\beta(\tau) \]  

(3.2)

where \( \gamma(\tau) = Q_{\tilde{\epsilon}}(\tau/h) \) and \( \beta(\tau) = -\beta + \gamma(\tau)\delta \).

Theorem 3.1 suggests that extremal quantile regression can be used to estimate \((\gamma(\tau), \beta(\tau))\), for small \( \tau \). We thus define

\[ (\tilde{\gamma}(\tau), \tilde{\beta}(\tau)) = \arg \min_{\gamma,\beta} \sum_{t=1}^{T} \rho_{\tau}(\tilde{Y}_t - \gamma - X'_t\beta), \]

where \( \rho_{\tau}(u) = u(\tau - I_{u<0}) \) is the usual check function. Then, to recover \( \beta \) and \( \delta \), we can simply compute \( \beta(l\tau) \) for \( l \neq 1 \) and use the relationships

\[ \delta = \frac{\beta(l\tau) - \beta(\tau)}{\gamma(l\tau) - \gamma(\tau)} \]

\[ \beta = -\beta(\tau) + \gamma(\tau)\delta. \]  

(3.3)

We basically follow this route, except that for an efficiency matter we use \( J \) reduced form estimators \((\beta(l_1\tau), ..., \beta(l_J\tau))\) rather than just two. Let us consider

\[ g_{T}(\delta) = \begin{pmatrix} \beta(l_1\tau) - \tilde{\beta}(\tau) - (\tilde{\gamma}(l_1\tau) - \tilde{\gamma}(\tau))\delta \\ \vdots \\ \beta(l_J\tau) - \tilde{\beta}(\tau) - (\tilde{\gamma}(l_J\tau) - \tilde{\gamma}(\tau))\delta \end{pmatrix} \]  

(3.4)

Let \( W_T \) be a \( Jd \times Jd \) positive definite matrix. We estimate \( \delta \) by

\[ \hat{\delta} = \arg \min_{\delta} g_T(\delta)'W_Tg_T(\delta). \]  

(3.5)

Finally, we estimate \( \beta \) by taking the empirical counterpart of (3.3), averaged over the different \( l_j\tau \):

\[ \hat{\beta} = -\beta(\tau) + \gamma(\tau)\hat{\delta}, \]  

(3.6)
where \( \hat{\beta}(\tau) = \frac{1}{p} \sum_{j=1}^{p} \hat{\beta}(l_j \tau) \) and similarly for \( \hat{\gamma}(\tau) \). Note that we do not estimate simultaneously \( \beta \) and \( \delta \), since it turns out that they have different rates of convergence. The theory of standard minimum distance estimation then does not apply in such a case, which is why we consider the estimation of \( \delta \) and \( \beta \) sequentially rather than simultaneously.

Our estimator depends on a choice of \( \tau \), \( (l_1, \ldots, l_J) \) and \( W_T \). We derive in the following subsection the optimal weighting matrix, which can be consistently estimated. Simulations reveal that the choice of \( (l_1, \ldots, l_J) \) does not matter much in practice. On the other hand, an appropriate choice of \( \tau \) is crucial. We discuss a data-driven procedure for that purpose in Subsection 3.3.

### 3.2 Asymptotic properties

We now turn to the asymptotic properties of \( (\hat{\beta}, \hat{\delta}) \). We rely for that purpose on the asymptotic properties of extremal quantile regressions, which have been established by Chernozhukov (2005).

The following assumption ensures that we can apply his framework. Hereafter, we let \( U = \tilde{Y} + X' \beta \) and \( u = \tilde{\varepsilon} \). \( RV_\alpha(x) \) denotes the set of regularly varying functions with index \( \alpha \) at \( x \).

**Assumption 5.** (Quantile Density) (i) \( \frac{\partial Q_{U|X}(\tau|x)}{\partial \tau} \sim \frac{\partial Q_u(\tau/h)}{\partial \tau} (1 + x' \delta) \) uniformly in \( x \in X \) and (ii) \( \frac{\partial Q_u(\tau)}{\partial \tau} \in RV_{-\xi - 1}(0) \)

Theorem 3.1 ensures that \( Q_{U|X}(\tau|x) \sim Q_u(\tau/h)(1 + x' \delta) \). Part (i) of Assumption 5 refines this result by assuming that the equivalence holds for the derivative as well. The pointwise equivalence is true under nondecreasing quantile density. Uniformity is local since we assume \( X \) is compact. Part (ii) is a von Mises type conditions usually assumed in extreme value theory. See the remarks after Condition R3 in Chernozhukov (2005) for a discussion.

The next assumption is specific to our context and does not appear in Chernozhukov (2005). It ensures the consistency and asymptotic normality of our estimators of \( \beta \) and \( \delta \). In the following, we let

\[
\mu_T(\tau) = \frac{E \left[ (\tau - 1 \{ \tilde{Y}_t \leq \gamma(\tau) + X_t' \beta(\tau) \} X_t \right]}{\sqrt{T}}
\]

Roughly speaking, \( \mu_T(\tau) \) measures the accuracy of (3.2). In particular, \( \mu_T(\tau) = 0 \) if the equivalence in (3.2) is replaced by an equality.

**Assumption 6.** (Rate of convergence of the quantile index) \( \tau(T) \) satisfies, as \( T \to \infty \), (i) \( \tau(T) \to 0 \), \( T\tau(T)/Q_u(\tau) \to \infty \) and (ii) \( T\mu_T(l_j \tau) \to 0 \) for \( j = 1 \ldots J \).

Assumption 6 (i) is first a sufficient condition for \( \tau(T)T \to \infty \) which indicates we will perform an extremal quantile regression of intermediate order (namely, with \( \tau(T) \to 0 \) and \( \tau(T)T \to \infty \)) to estimate \( \beta(\tau) \). The reason why we use intermediate order instead of extreme order (where

\[\text{In other words, } F \in RV_\alpha(x) \text{ if for any fixed } t, \frac{F(tu)}{F(u)} \to t^\alpha \text{ as } u \to x.\]
\( \tau(T)T \to k \) is that applying the latter framework we would lose consistency of \( \hat{\beta} \). The stronger than intermediate order condition stated in Assumption 6(i) implies the normalizing factor of our estimator \( \hat{\beta} \) will grow to \( \infty \) which means the estimator is consistent. Intermediate order quantile theory has also the nice feature that it guarantees asymptotic normality rather than convergence towards a non-standard, DGP dependent, distribution. Part (ii) of Assumption 6 ensures that the bias stemming from selection vanishes quick enough as \( T \) grows. This bias is due to the fact that the conditional quantile is not linear but only asymptotically linear. Note that the convergence rate of the bias correction term is hard to compute in our general setting, so Part (ii) is a high level condition. Proposition 3.1 states the low level conditions on the structure of selection equation to ensure the existence of \( \tau \) who satisfies our high level Assumption 6.

**Theorem 3.2.** Under Assumptions \( \mathbb{1} - \mathbb{6} \) and if \( W_T \xrightarrow{p} W \) symmetric positive definite and nonstochastic,

\[
\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Omega_\delta) \\
\frac{\sqrt{T}}{\hat{\gamma}(\tau)}(\hat{\delta} - \delta) \xrightarrow{d} \mathcal{N}(0, \Omega_\delta)
\]

where

\[
\Omega_\delta = (G'WG)^{-1}G'W(I_J \otimes \Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_J \otimes \Delta')WG(G'WG)^{-1},
\]

\( I_J \) is the \( J \times J \) identity matrix and \( G, \Delta, \Omega_0 \) and \( \Gamma \) are defined in the appendix. One thing to note is that \( \Delta \) and \( \Omega_0 \) are functions of our structure estimator \( \delta \). By plugging in any consistent estimator of \( \delta \), we can compute the consistent estimator of our covariance matrix. The optimal weighting matrix satisfies \( W^*_\delta = ((I_J \otimes \Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_J \otimes \Delta'))^{-1} \).

**Proposition 3.1.** Suppose that \( D = 1\{\phi(X) + g(\varepsilon) - v \geq 0\} \) where \( v \perp \perp (\varepsilon, X) \), \( g(t) > at^b \) for large \( t \) and some \( a > 0, b > 0 \) and \( S_\delta \in RV_{-\alpha} \) with \( \alpha > 0 \), \( ab > 1 \). Then, if Assumption 3(ii) also holds, there exists \( \zeta > 0 \) such that \( \tau(T) = \frac{\log(T)^2}{T} \) satisfies Assumption 6.

In Appendix 7.5, we show that if the unobservable \( v \) in the selection equation is varying more rapidly than the unobservable \( u \) in the outcome equation, polynomial rate of convergence is possible. In Appendix 7.6, we show that Assumption 6 also holds under a generalized Roy model with possibly more than two sectors. In Appendix 7.7, we link the rate of convergence with conditional copula of unobservables in the selection and outcome equation and show that under general setup, Gaussian copula induces polynomial convergence rate.

### 3.3 Choice of the quantile index

The estimators of \( \beta \) and \( \delta \) are consistent and asymptotically normal provided that they are based on \( \hat{\beta}(\tau(T)) \) with \( \tau(T) \) satisfying the bias-variance trade-off of Assumption 6. An issue at this stage
is that admissible rates of convergence towards 0 for \( \tau(T) \) are unknown, since they depend on \( \mu_T(\tau) \) which is itself unknown. A similar issue arises in the estimation of extreme value index (see Drees & Kaufmann, 1998 and Danielsson et al., 2001) or the estimation at infinity of the intercept of sample selection models (see Andrews & Schafgans, 1998). We propose here to estimate \( \tau(T) \) using our minimum distance criterion. In other words, letting \( g_T(\delta, \tau) \) be defined as in (3.5) but with an explicit dependence in \( \tau \), we let
\[
(\hat{\delta}, \hat{\tau}) = \arg \min_{\delta, \tau} g_T(\delta, \tau)'Wg_T(\delta, \tau),
\] (3.7)
and \( \hat{\beta} \) is still defined by (3.6) but with \( \tau \) replaced by \( \hat{\tau} \). Because the estimation of \( \delta \) is almost immediate for a fixed \( \tau \), we solve in practice (3.7) by a grid search on \( \tau \), and standard optimization on \( \delta \) for each value of the grid.

4 Simulations

In the simulation study, we consider the following model:
\[
Y^* = \beta_1 X_1 + \beta_2 X_2 + (\delta_0 + \delta_1 X_1 + \delta_2 X_2)\varepsilon \\
D = 1 \{1.8 + Y^* + 0.3X_1 + 0.2X_2 + \eta > 0\},
\]

Where \( X_1 \) is a Bernoulli(0.5), \( X_2 \) is a normal distribution truncated on \((-1.8, 1.8)\) with variance 1. \((\varepsilon, \eta)\) are independent standard normal. The parameters are \( \beta_1 = 0.4, \beta_2 = 0.5, \delta_0 = 1, \delta_1 = -0.2, \delta_2 = -0.1 \). One key feature of this data generating process is that it leads to a selection rate of about 84%, which matches the selection rate in our application.

The sample sizes are \( T = 500, 1000, 2000 \) and 4000. For each replication, the lower bound of \( \tau' \)'s is \((0.005, 0.01, 0.015, 0.02, 0.025)\) (indexed by \( i \)) and upper bound is \((0.05, 0.075, 0.1, 0.125, 0.15)\) (indexed by \( j \)). For each pair of lower and upper bounds, \((\text{lower}(i), \text{upper}(j))\), we use \( A+1 \) different \( \tau' \)'s to form the equations used to estimate our structural parameters, where \( A = 4 \) or 8. For example, if we choose the lower bound to be 0.005 and the upper bound to be 0.05, and \( A \) to be 4, then the set of \( \tau' \)'s consists of the evenly spaced \( \tau' \)'s \((0.005, 0.02, 0.035, 0.05)\) plus the mid point 0.0275. Then, we use the mid-point to normalize all the \( \tau' \)'s, which yields the set of equations indexed by
\[
l = (1.0000, 0.1818, 0.7273, 1.2727, 1.8182)
\]
We also want to point out that the variance of our estimator is a function of the population parameter \( \delta \). In practice we use a two-stage procedure: in the first stage, we estimate \( \hat{\delta} \) and in the second stage, we reestimate \( \hat{\delta} \) by plugging in the \( \hat{\delta} \) estimated by high lower and upper bound from
the first stage when computing the optimal weighting matrix. The minimum distance estimator for each replication is then chosen to be the estimator among the 50 \((5 \times 5 \times 2)\) of them produced by the second stage which minimizes the distance. We report below the simulation results in terms of bias, Mean-Squared Error (MSE) and coverage for our minimum distance estimator.

<table>
<thead>
<tr>
<th>T</th>
<th>(\delta_1)</th>
<th>(\delta_2)</th>
<th>(\beta_1)</th>
<th>(\beta_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>-0.016</td>
<td>-0.017</td>
<td>0.025</td>
<td>0.017</td>
</tr>
<tr>
<td>1000</td>
<td>0.003</td>
<td>0.013</td>
<td>-0.007</td>
<td>-0.013</td>
</tr>
<tr>
<td>2000</td>
<td>-0.011</td>
<td>-0.001</td>
<td>0.021</td>
<td>-0.003</td>
</tr>
<tr>
<td>4000</td>
<td>0.011</td>
<td>-0.003</td>
<td>-0.016</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 1: Bias

<table>
<thead>
<tr>
<th>T</th>
<th>(\delta_1)</th>
<th>(\delta_2)</th>
<th>(\beta_1)</th>
<th>(\beta_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.047</td>
<td>0.039</td>
<td>0.132</td>
<td>0.104</td>
</tr>
<tr>
<td>1000</td>
<td>0.018</td>
<td>0.022</td>
<td>0.047</td>
<td>0.063</td>
</tr>
<tr>
<td>2000</td>
<td>0.009</td>
<td>0.008</td>
<td>0.028</td>
<td>0.023</td>
</tr>
<tr>
<td>4000</td>
<td>0.004</td>
<td>0.003</td>
<td>0.011</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Table 2: MSE

<table>
<thead>
<tr>
<th>T</th>
<th>(\delta_1)</th>
<th>(\delta_2)</th>
<th>(\beta_1)</th>
<th>(\beta_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>92%</td>
<td>91%</td>
<td>92%</td>
<td>93%</td>
</tr>
<tr>
<td>1000</td>
<td>90%</td>
<td>90%</td>
<td>93%</td>
<td>93%</td>
</tr>
<tr>
<td>2000</td>
<td>93%</td>
<td>94%</td>
<td>92%</td>
<td>96%</td>
</tr>
<tr>
<td>4000</td>
<td>91%</td>
<td>97%</td>
<td>93%</td>
<td>99%</td>
</tr>
</tbody>
</table>

Table 3: 95% Coverage

It can be seen from these results that the variance dominates the bias, for all sample sizes. It follows that the sampling distribution is close to standard normal (see the QQ-plots below), and inference based on asymptotic normality provides a fairly good coverage. Although the optimal distance estimators involves an optimization procedure for each replication, and therefore the asymptotic properties are unknown at this point, this suggests that the normalized optimal distance estimator is also standard normal and provide good coverage. We use this data-driven method to select the set of \(\tau\)’s in the application.
Although the previous simulation results suggest that bias may not be a first order issue in this context, we also implement a subsampling method to correct for the bias and try and improve the coverage of our estimator. The data generating process is the same as above, with a total sample size $T = 2000$. For each replication, we randomly select 200 subsamples with size 500. For each of these 200 subsamples, we compute the minimum distance estimator of $\delta$ and $\beta$. Thus, for each replication, we obtain for each parameter a 200-element sequence. We then take the median of the sequence and construct the confidence interval for this replication as the interval centered at this median, with standard deviation computed using the full sample and standard normal critical value. The 95% coverage for this confidence intervals is reported below. Overall, the coverage is quite good and the confidence intervals seem to be conservative for all parameters.

<table>
<thead>
<tr>
<th>95% coverage</th>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>99%</td>
<td>98%</td>
<td>97%</td>
<td>98%</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: 95% coverage with bias correction
5 Application

In this section, we apply our method to estimate the black-white wage gap among males from the NLSY79, and revisit the influential work of Neal & Johnson (1996) on this question. The total sample size is 2,121, with an overall employment rate over the period of interest (1990-1991) equal to 84.4%.\(^4\) Basic descriptives for our sample are presented in the table below:

<table>
<thead>
<tr>
<th></th>
<th>Full sample</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Black</td>
<td>Hispanic</td>
</tr>
<tr>
<td>AFQT Score</td>
<td>-0.543</td>
<td>-0.250</td>
</tr>
<tr>
<td>Std.</td>
<td>0.708</td>
<td>0.854</td>
</tr>
<tr>
<td>Number of schooling years by 1991</td>
<td>12.443</td>
<td>12.212</td>
</tr>
<tr>
<td>Std.</td>
<td>1.957</td>
<td>2.182</td>
</tr>
<tr>
<td>Mother high-school graduate</td>
<td>0.442</td>
<td>0.299</td>
</tr>
<tr>
<td>Father high-school graduate</td>
<td>0.460</td>
<td>0.327</td>
</tr>
<tr>
<td>Mother college graduate</td>
<td>0.050</td>
<td>0.045</td>
</tr>
<tr>
<td>Father college graduate</td>
<td>0.061</td>
<td>0.088</td>
</tr>
</tbody>
</table>

Table 5: Data Description

Overall, the patterns are consistent with Neal & Johnson (1996).\(^5\) In particular, whites have substantially higher AFQT scores, as well as higher educational level, both for themselves and their parents. We first replicate Tables 1 and 4 in Neal & Johnson (1996) by running four median log-wage regressions. In the first two columns, we replicate Table 1 in Neal & Johnson (1996) by simply dropping the unemployed individuals from the estimation. The last two columns replicate Table 4 by imputing a zero log-wage to the unemployed individuals. As discussed in Neal & Johnson (1996) and Johnson \textit{et al.} (2000), under the assumption that, conditional on the set of observable characteristics included in the regression, the potential wage for any individual who did not work neither in 1990 nor in 1991 is below the median, this imputation method yields consistent estimates. Importantly, the identifying condition of independence at the limit used in our paper generalizes this assumption by replacing the median by some extremal quantile of the wage distribution.

\(^4\)As in Neal & Johnson (1996), an individual is considered unemployed if she did not work in 1990 nor in 1991. The log-wage is then defined as the log of the mean real wages in 1990 and 1991 for workers who worked in both years, and the log of the real wage in the year of employment for those who worked only one year.

\(^5\)It should be noted that our data does not match exactly that used in Neal & Johnson (1996), in part because the NLSY data are changed over time to correct for coding errors. Specifically, we use the same data as in Graham \textit{et al.} (2012).
As put forward by Neal & Johnson (1996), columns (1) and (2) show that the estimated black-white wage gap drops substantially, from 27.5% to 8.1%, after adding controls for ability, namely AFQT and \( AFQT^2 \). It is also worth noting that the estimated black-white wage differential changes after addressing the selection issue with the imputation method proposed in Neal & Johnson (1996). For instance, columns (2) and (4) show that the estimated black-white wage gap increases in absolute value by 2 points.

We now investigate how the results are changed when we use our estimation method. In our extremal quantile framework, we choose respectively for the lower and upper bounds of the quantiles \( \tau \)'s \([0.04, 0.06, 0.08, 0.1]\) and \([0.15, 0.2, 0.25, 0.3]\). We choose these bounds such that they are wide enough for minimum distance estimation to work well, and the lower bounds are not too small to ensure that we are in the intermediate order case (see Chernozhukov, 2005 who suggests as a rule of thumb \( \tau T \geq 20 \)). We report in Table 7 below the estimation results from our proposed method:6

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta )</td>
<td>( \beta )</td>
<td>( \beta )</td>
<td>( \beta )</td>
</tr>
<tr>
<td>Black</td>
<td>-0.275</td>
<td>-0.081</td>
<td>-0.306</td>
<td>-0.101</td>
</tr>
<tr>
<td>std</td>
<td>(0.026)</td>
<td>(0.038)</td>
<td>(0.030)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.325</td>
<td>-0.224</td>
<td>-0.157</td>
<td>-0.006</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.117</td>
<td>0.006</td>
<td>-0.092</td>
<td>0.035</td>
</tr>
<tr>
<td>std</td>
<td>(0.030)</td>
<td>(0.043)</td>
<td>(0.036)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.176</td>
<td>-0.057</td>
<td>-0.079</td>
<td>0.090</td>
</tr>
<tr>
<td>Age</td>
<td>0.080</td>
<td>0.046</td>
<td>0.050</td>
<td>0.027</td>
</tr>
<tr>
<td>std</td>
<td>(0.012)</td>
<td>(0.017)</td>
<td>(0.014)</td>
<td>0.015</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.056</td>
<td>0.104</td>
<td>0.013</td>
<td>0.079</td>
</tr>
<tr>
<td>AFQT</td>
<td>0.207</td>
<td>0.215</td>
<td></td>
<td></td>
</tr>
<tr>
<td>std</td>
<td>(0.020)</td>
<td>(0.017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.168</td>
<td>0.246</td>
<td>0.181</td>
<td>0.248</td>
</tr>
<tr>
<td>( AFQT^2 )</td>
<td>-0.016</td>
<td>-0.032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>std</td>
<td>(0.016)</td>
<td>(0.014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.047</td>
<td>0.016</td>
<td>-0.059</td>
<td>-0.004</td>
</tr>
</tbody>
</table>

Table 6: Median log-wage regressions
Table 7: Extremal quantile log-wage regressions

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>-0.125</td>
<td>0.140</td>
</tr>
<tr>
<td>std</td>
<td>(0.071)</td>
<td>(0.115)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.264</td>
<td>0.015</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.018</td>
<td>0.116</td>
</tr>
<tr>
<td>std</td>
<td>(0.074)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.126</td>
<td>0.162</td>
</tr>
<tr>
<td>Age</td>
<td>0.032</td>
<td>-0.024</td>
</tr>
<tr>
<td>std</td>
<td>(0.014)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.005</td>
<td>0.059</td>
</tr>
<tr>
<td>AFQT</td>
<td>0.242</td>
<td>-0.027</td>
</tr>
<tr>
<td>std</td>
<td>(0.029)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.185</td>
<td>0.298</td>
</tr>
<tr>
<td>AFQT$^2$</td>
<td>-0.025</td>
<td>-0.012</td>
</tr>
<tr>
<td>std</td>
<td>(0.020)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.065</td>
<td>0.014</td>
</tr>
</tbody>
</table>

The estimation results (controlling for AFQT) from our extremal quantile regression suggest that the size of the black-white wage gap (12.5%) is actually larger than the estimated gap obtained under the imputation method proposed by Neal & Johnson (1996). Comparing these results with those obtained under the imputation method in Table 6 (Column 4) shows that the estimated black-wage wage gap goes up by 2.4 points when using our approach. It is worth noting that this change is of the same order of magnitude as the one going from columns (2) to (4) in Table 6.

We investigate below the sensitivity of our results to the assumptions made to recover the intercept, by assuming that the error term of the outcome equation is normally distributed. From the extremal quantile regression, we can estimate the quantile of the error term if the quantile is sufficiently large. Then, using the normality assumption, we can recover the location parameter by running a regression of estimated quantiles on normal quantiles. Finally, we take the intercept as given and estimate the structural parameters using our extremal quantile approach. Table 8 below shows the estimation results for this specification.

---

7The quantiles we use to recover the location and scale parameters are $0.01 \times (1 : 10)$. 

13
<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black std</td>
<td>-0.145</td>
<td>0.087</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.318</td>
<td>0.027</td>
</tr>
<tr>
<td>Hispanic std</td>
<td>0.001</td>
<td>0.072</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.178</td>
<td>0.179</td>
</tr>
<tr>
<td>Age std</td>
<td>0.036</td>
<td>-0.015</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.002</td>
<td>0.069</td>
</tr>
<tr>
<td>AFQT std</td>
<td>0.246</td>
<td>-0.017</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.176</td>
<td>0.315</td>
</tr>
<tr>
<td>AFQT^2 std</td>
<td>-0.023</td>
<td>-0.008</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.072</td>
<td>0.026</td>
</tr>
</tbody>
</table>

Table 8: Extremal quantile log-wage regressions (intercept estimated under normality)

Although the estimation results display a similar pattern as in Table 7, the point estimates for Black and Hispanic dummies are quite different. With this new specification, we find in particular that the estimated black-white wage gap is equal to 14.5%, which corresponds approximately to a 4.5 point increase relative to estimated wage gap under the imputation approach. Finally, we report in Table 9 below the estimation results with bias-corrected confidence intervals, using the subsampling method discussed in Section 4. Note that, as suggested by the Monte Carlo simulation results, these confidence intervals are likely to be conservative.
<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>-0.145</td>
<td>0.087</td>
</tr>
<tr>
<td>std</td>
<td>(0.088)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.225</td>
<td>0.120</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.001</td>
<td>0.072</td>
</tr>
<tr>
<td>std</td>
<td>(0.091)</td>
<td>(0.074)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.076</td>
<td>0.282</td>
</tr>
<tr>
<td>Age</td>
<td>0.036</td>
<td>-0.015</td>
</tr>
<tr>
<td>std</td>
<td>(0.017)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.032</td>
<td>0.099</td>
</tr>
<tr>
<td>AFQT</td>
<td>0.246</td>
<td>-0.017</td>
</tr>
<tr>
<td>std</td>
<td>(0.036)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.148</td>
<td>0.287</td>
</tr>
<tr>
<td>AFQT$^2$</td>
<td>-0.023</td>
<td>-0.008</td>
</tr>
<tr>
<td>std</td>
<td>(0.025)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>-0.076</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Table 9: Extremal quantile log-wage regressions with bias-corrected confidence intervals

6 Conclusion

This paper proposes a new estimation method for a semiparametric location-scale model in the presence of sample selection, without instruments nor large support regressor. Following a recent paper by D’Haultfoeuille & Maurel (2013), identification relies on independence between covariates and selection, when the potential outcome tends to infinity. This condition is likely to hold in many situations. Interestingly, this identifying assumption generalizes the condition used by Neal & Johnson (1996) and Johnson et al. (2000) to correct for differential labor market participation rates when estimating the black-white wage gap. Building up on this identification strategy, we propose a simple estimator based on extremal quantile regressions. We establish asymptotic normality by extending the results of Chernozhukov (2005) to allow for selection. Finally, we apply our method to the estimation of the black-white wage gap among males from the NLSY79. Our estimation results point to a larger black-white wage gap that what has been found under the imputation method proposed by Neal & Johnson (1996).
References


7 Appendix

Throughout the appendix, let $L(t)$ be a generic slowly varying function and $K$ be a generic constant.

7.1 Preliminary lemmas

Lemma 7.1. Assumption 3 (ii)-(iii) imply that $S_\varepsilon$ is rapidly varying at $-\infty$, i.e. its EV index $\xi$ is 0.

Proof. Suppose $S_\varepsilon$ is not rapidly varying. By Assumption 3 (iii) and the assumption that $\sup(\text{Supp}(\varepsilon)) = \infty$, $S_\varepsilon$ is in the attraction domain of type 2 EV distribution, i.e. $S_\varepsilon \in RV_{-\xi-1}(+\infty)$ with EV index $\xi > 0$. We also have

$$\frac{S_{\exp(\varepsilon)}(xt)}{S_{\exp(\varepsilon)}(t)} = \frac{S_\varepsilon(x_t \log(t))}{S_\varepsilon(\log(t))}$$

(7.1)

where $x_t = \frac{\log(t) + \log(x)}{\log(x)} \to 1$ as $t \to +\infty$. Thus the RHS of Equation 7.1 converges to 1 since $S_\varepsilon \in RV_{-\xi-1}(+\infty)$, which means $S_{\exp(\varepsilon)}(t)$ is slowly varying. Contradiction. □

Let $h_2(t) = F'_\varepsilon(t)/(1 - F_\varepsilon(t))$ and $H_2(t) = \int_0^t h_2(s)ds$.

Lemma 7.2. Under Assumption 3 (ii), (iii) and assume $h_2 \in RV_{\rho_2}(\infty)$, then $\rho_2 \geq 0$ and $H_2 \in RV_{\rho_2+1}(\infty)$.

Proof. Suppose not, $\rho_2 < 0$, $h_2(t) = \frac{L(t)}{t^{-\rho_2}}$ with $L(t)$ being slowly varying (when $\rho_2 = -1$, $L(t) \to \infty$ as $t \to \infty$). So $\exists t_0$ such that for $t \geq t_0$, $L(t) \leq t^{-\rho_2/2}$. Thus for $x > 1$,

$$1 \geq \frac{1 - F_{\exp(\varepsilon)}(xt)}{1 - F_{\exp(\varepsilon)}(t)} = \frac{1 - F_\varepsilon(\log(x) + \log(t))}{1 - F_\varepsilon(\log(t))}$$

$$= \exp(-\int_{\log(t)}^{\log(xt)} \frac{L(s)}{s^{-\rho_2}} ds)$$

$$\geq \exp(-(\rho_2/2 + 1)((\log(xt))^{\rho_2/2+1} - (\log(t))^{\rho_2/2+1}))$$

$$\to 1 \text{ as } t \to \infty.$$ 

This means that $1 - F_{\exp(\varepsilon)}(t)$ is slowly varying, a contradiction. □

7.2 Proof of Theorem 3.1

Let $U_x(y) \equiv 1/P(Y > y|X = x)$, $V_x(y) \equiv 1/hS_\varepsilon((y - x\beta)/(1 + x'\delta))$. Then from Equation 3.1, $U_x(y) \sim V_x(y)$. We aim to show $U_x^\leftarrow(\tau) \sim V_x^\leftarrow(\tau)$, where $G^\leftarrow(\tau) = \inf\{x : G(x) > \tau\}$. For that purpose, we suppose that there exists $\varepsilon_0 > 0$ such that

$$V_x^\leftarrow(y)/U_x^\leftarrow(y) \geq 1 + \varepsilon_0,$$

(7.2)
and shows that this leads to a contradiction. The reasoning is similar for the other inequality.

First, by lemma 7.1, $S_0$ is in the attraction domain of Type 1 EV distribution. This implies that $V \equiv 1/S_0$ is $\Gamma$-varying (see Resnick 1987 Proposition 0.10), i.e. $\lim_{t \to \infty} \frac{V(t + x f(t))}{V(t)} = e^x$ for some auxiliary function $f$. Define $f_x(y) = f((y - x'\beta)/(1 + x'\delta))$. Then

$$\frac{V_x(z + tf_x(z))}{V_x(z)} = \frac{V\left[\frac{z-x'\beta}{1+x'\delta} + tf\left(\frac{z-x'\beta}{1+x'\delta}\right)\right]}{V\left[\frac{z-x'\beta}{1+x'\delta}\right]} \to e^t$$

as $z \to \infty$. Thus $V_x(z)$ is $\Gamma$-varying with auxiliary function $f_x$. Furthermore, $U_x(y) \sim V_x(y)$, so that

$$\frac{U_x(z + tf_x(z))}{U_x(z)} = \frac{U_x(z + tf_x(z))}{V_x(z + tf_x(z))} \frac{V_x(z + tf_x(z))}{V_x(z)} \to e^t.$$

So $U_x$ is also $\Gamma$-varying with the same auxiliary function. $f_x$ also satisfies (see Resnick 1987 Ex. 0.4.3.10)

$$\lim_{z \to \infty} \frac{f_x(z)}{z} \to 0. \quad (7.3)$$

Combining (7.2) and (7.3), we obtain

$$\frac{V_x^+(y)}{U_x^+(y)} \geq 1 + \varepsilon_0 \frac{f_x(U_x^-(y))}{U_x^-(y)},$$

Now, because $y \sim V_x(V_x^+(y))$ and $y \sim U_x(U_x^+(y))$ (see Resnick 1987 page 28), for any $\varepsilon_1/0$, there exists $y$ large enough such that

$$y(1 + \varepsilon_1) \geq V_x(V_x^+(y)) \geq V_x(U_x^+(y)) \geq V_x(U_x^+(y) + \varepsilon_0 f(U_x^-(y))) \geq (1 - \varepsilon_1)U_x(U_x^+(y)) + \varepsilon_0 f(U_x^-(y)) = (1 - \varepsilon_1)^2 e^{\varepsilon_0} U_x(U_x^+(y)) \geq (1 - \varepsilon_1)^3 e^{\varepsilon_0} y.$$

Therefore, $1 \geq \frac{(1-\varepsilon_1)^3}{1+\varepsilon_1} e^{\varepsilon_0}$. Letting $\varepsilon_1$ tend to zero lead to a contradiction.

### 7.3 Proof of Theorem 3.2

First let us introduce additional notations. Let $\theta(\tau) = (\gamma(\tau), \beta(\tau))'$, $\hat{Z}_T(l) = \alpha_T(l)(\hat{\theta}(\tau) - \theta(\tau))$, where $\alpha_T(l) = \frac{\sqrt{l_T}}{\gamma(l_T) - \gamma(l_T)} = \frac{\sqrt{l_T}}{Q_{n(l_T)} - Q_{n(l_T/\beta)}}$ for some arbitrary fixed $m > 1$ and $\alpha_T \equiv \alpha_T(1)$. Let also

$$\hat{Z}_T(l_1, \cdots, l_J) = \left(\hat{Z}_T(l_1), \hat{Z}_T(l_1), \cdots, \hat{Z}_T(l_J)\right)'$$

19
Finally, let $\Omega_0 = \log(m)^{-2}Q_H^{-1}Q_XQ_H^{-1}$ where $Q_H = E_XX'_t/(1 + X'_t\delta)$ and $L$ be the matrix with $(i,j)$ term $L_{i,j} = \frac{l_{i-1}^j\lambda_{i-1}}{\sqrt{\lambda_{i-1}l_{i-1}}}$, $i,j = 1 \cdots J + 1$ with $l_0 = 1$.

The main part of the proof (steps 1 to 3) consists in proving that the standardized first step estimator $\tilde{Z}_T(l_1, \cdots, l_J)$ is asymptotically normal. The proof of this part is close to the one of Chernozhukov (2005), and we refer to it repeatedly hereafter. We however have to take into account that (3.2) is an equivalence, not an equality as in his framework. Step 4 of the proof finally proves asymptotic normality of $\delta$ and $\beta$, and consistency of $\beta$.

1. Let $u = -\varepsilon$ and $U = \tilde{Y}_t + X'_t\beta$. As $\tau \to 0$,

$$\frac{Q_u(l\tau) - Q_u(\tau)}{Q_u(m\tau) - Q_u(\tau)} \to \frac{\log(l)}{\log(m)}, \quad \frac{Q_{U|X}(\tau|x) - Q_u(\tau/h)(1 + x'\delta)}{Q_u(m\tau) - Q_u(\tau)} \to 0.$$

For the first limit, $Q_u$ is II-varying with auxiliary function $\tau \mapsto Q_u(e\tau) - Q_u(\tau)$ (see Resnick (1987)). Then divide both the numerator and denominator by this auxiliary function and apply the II varying property.

For the second limit, let $z = Q_{U|X}(\tau|x)$. Then

\[
\tau = P(U \leq z|X = x) = P(Y \geq x'\beta + z|X = x) = P(Y^* \geq x'\beta + z, D = 1|X = x) = \int_{x'\beta + z}^{\infty} P(D = 1|Y^* = y, X = x)dP_{Y^*|X = x}(y)
\]

(7.4)

For $y$ large, $(h - \Delta)P(u(1 + x'\delta) \geq z) \leq \text{RHS of Equation (7.4)} \leq (h + \Delta)P(u(1 + x'\delta) \geq z)$. Invert both side, $(1 + x'\delta)Q_u(\tau/(h + \Delta)) \leq Q_{U|X}(\tau|x) \leq (1 + x'\delta)Q_u(\tau/(h - \Delta))$.

\[
(1 + x'\delta)\left(\frac{\log(h/(h + \Delta))}{\log(m)}\right)
\leq \lim \inf \frac{Q_{U|X}(\tau|x) - Q_u(\tau/h)(1 + x'\delta)}{Q_u(m\tau) - Q_u(\tau)}
\leq \lim \sup \frac{Q_{U|X}(\tau|x) - Q_u(\tau/h)(1 + x'\delta)}{Q_u(m\tau) - Q_u(\tau)}
\leq (1 + x'\delta)(\lim \sup \frac{Q_u(\tau/(h - \Delta)|x) - Q_u(\tau/h)}{Q_u(m\tau) - Q_u(\tau)})
\leq (1 + x'\delta)(\frac{\log(h/(h - \Delta))}{\log(m)})
\]

Let $\Delta \to 0$, both the upper and lower bounds $\to 0$. Since $X$ has bounded support, the upper and lower bound is uniform over $X$. So the convergence holds uniformly over $X$. 20
2. For any \( z = (z_1, z_2)' \) with \( z_1 \) scalar and \( z_2 \in \mathbb{R}^d, G_T(z, \tau) \xrightarrow{P} \frac{1}{2} \log(m)z'Q_X z \), where
\[
G_T(z, \tau) = \frac{\alpha_T}{\tau T} \sum_{t=1}^{T} \int_{0}^{(z_1 + X_t' z_2)/\alpha_T} (\mathbb{1}\{\tilde{Y}_t - \gamma(\tau) - X_t' \beta(\tau) \leq s\} - \mathbb{1}\{\tilde{Y}_t - \gamma(\tau) - X_t' \beta(\tau) \leq 0\})ds.
\]

By Chernozhukov (2005), \( V(G_T^2(z, \tau)) \to 0 \). Thus it suffices to prove that \( E[G_T(z, \tau)] \to \frac{1}{2} \log(m)z'Q_H z \). First,
\[
EG_T(z, \tau)
= \frac{\alpha_T}{\sqrt{\tau T}} \int_{0}^{(z_1 + X_t' z_2)/\alpha_T} (\mathbb{1}\{\tilde{Y}_t - \gamma(\tau) - X_t' \beta(\tau) \leq s / \alpha_T\} - \mathbb{1}\{\tilde{Y}_t - \gamma(\tau) - X_t' \beta(\tau) \leq 0\})ds
\]

\[
= \frac{T}{\sqrt{\tau T}} E \int_{0}^{(z_1 + X_t' z_2)} (\mathbb{1}\{\tilde{Y}_t + X_t' \beta - \gamma(\tau) - X_t' \delta Q_u(\tau/h) \leq s / \alpha_T\} - \mathbb{1}\{\tilde{Y}_t + X_t' \beta - \gamma(\tau) - X_t' \delta Q_u(\tau/h) \leq 0\})ds
\]

\[
= TE \int_{0}^{(z_1 + X_t' z_2)} P_t(\tilde{Y}_t + X_t' \beta - \gamma(\tau) - X_t' \delta Q_u(\tau/h) \leq s / \alpha_T) - P_t(\tilde{Y}_t + X_t' \beta - \gamma(\tau) - X_t' \delta Q_u(\tau/h) \leq 0) ds \tag{7.5}
\]

Notice \( P_t(\tilde{Y}_t + X_t' \beta - \gamma(\tau) - X_t' \delta Q_u(\tau/h) \leq s / \alpha_T) = P_t(U \leq (1 + X_t' \delta)Q_u(\tau/h) + s / \alpha_T) \equiv F_t(1 + X_t' \delta)Q_u(\tau/h) + s / \alpha_T) \). Just for the simplicity of notation, we denote \( f_t(\cdot) = f_{U|X_t}(\cdot) \) and \( F_t(\cdot) = F_{U|X_t}(\cdot) \). Notice that if Equation (3.1) holds with equality as in Chernozhukov (2005), we will have \( Q_t(\tau) = (1 + X_t' \delta)Q_u(\tau/h) \). But in our case, it is not strictly equal, so we need to consider the difference \( Q_t(\tau) - (1 + X_t' \delta)Q_u(\tau/h) \) and bound it using the first step.

RHS of Equation (7.3) =
\[
TE \int_{0}^{(z_1 + X_t' z_2)} \frac{f_t((1 + X_t' \delta)Q_u(\tau/h) + s / \alpha_T) - f_t((1 + X_t' \delta)Q_u(\tau/h))}{\sqrt{\tau T}} ds \tag{7.6}
\]

\[
= TE \int_{0}^{(z_1 + X_t' z_2)} \frac{s f_t((1 + X_t' \delta)Q_u(\tau/h) + O(s / \alpha_T))}{\sqrt{\tau T}} ds \tag{7.7}
\]

Since \( O(s / \alpha_T) = o(Q_u(m\tau) - Q_u(\tau)) \) and \( Q_t(\tau x) - Q_u(\tau) \) uniformly over \( x \), \( f_t((1 + X_t' \delta)Q_u(\tau/h) + O(s / \alpha_T)) = f_t(Q_t(\tau) + V_\tau) \) where \( V_\tau = o(Q_u(m\tau) - Q_u(\tau)) \). The RHS of Equation (7.7) then becomes
\[
TE \int_{0}^{(z_1 + X_t' z_2)} \frac{s f_t(Q_t(\tau) + V_\tau)}{\sqrt{\tau T}} ds \tag{7.8}
\]

Follow the same argument in Chernozhukov (2005), \( f_t(Q_t(\tau) + V_\tau) \sim f_t(Q_t(\tau)) \).

Because \( \xi = 0 \),
\[
\frac{(Q_u(m\tau) - Q_u(\tau)) f_t(Q_t(\tau))}{\tau} \sim \frac{(Q_u(m\tau) - Q_u(\tau)) f_u(Q_u(\tau))}{\tau(1 + X_t' \delta)h \xi} \sim \int_{1}^{m} \frac{f_u(Q_u(\tau))}{f_u(Q_u(s\tau))} ds \frac{1}{(1 + X_t' \delta)h \xi} \to
\]
\[ \log(m) \frac{1}{1 + X_t \delta}. \] Thus by DCT, 

Equation (7.8) = \( E(\frac{1}{2}((z_1 + z_2 X_t)(X_t' z_2 + z_1)(Q_u(m\tau) - Q_u(\tau))f_t(\tau)) \rightarrow \frac{1}{2} \log(m) z'Q_H z, \)

with \( Q_H = E(X_t X_t')/(1 + X_t \delta). \)

3. \( \hat{Z}_T(l_1, \ldots, l_J) \xrightarrow{d} \mathcal{N}(0, L \otimes \Omega_0). \)

We prove the result for \( \hat{Z}_T(1) \) only, the multivariate generalization being straightforward but notationally cumbersome. First, let

\[
W_T(\tau) = \frac{-1}{\sqrt{T}} \sum_{t=1}^{T} (\tau - 1 \{(\hat{Y}_t - \gamma(\tau) - X_t' \beta(\tau)) \leq 0\}) X_t
\]

and let \( M_{T,t}(\tau) = \frac{-1}{\sqrt{T}} (\tau - 1 \{(\hat{Y}_t - \gamma(\tau) - X_t' \beta(\tau)) \leq 0\}) X_t - \mu_T. \) Then

\[
W_T(\tau) = \sum_{t=1}^{T} M_{T,t}(\tau) + T \mu_T(\tau). \]

From Chernozhukov (2005),

\[
\sum_{t} M_{T,t}(\tau) \xrightarrow{d} \mathcal{N}(0, Q_X). \tag{7.10}
\]

By Assumption 6, \( T \mu_T(\tau) = o(1). \) Therefore, \( W_T(\tau) \xrightarrow{d} \mathcal{N}(0, Q_X). \)

Now, according to Chernozhukov (2005), Equation (9.43), \( \hat{Z}_T(1) \) minimizes \( \Psi_T(z, \tau) = W_T(\tau)' z + G_T(z, \tau). \) \( G_T(z, \tau) \) is convex in \( z \) and by Step 2, \( G_T(z, \tau) \rightarrow \frac{1}{2} \log(m) z'Q_H z. \) Moreover, \( W_T(\tau) \Rightarrow \mathcal{N}(0, Q_X). \) Let \( \theta_T(\tau) = -\log(m)^{-1} Q_H^{-1} W_T(\tau). \) \( \theta_T(\tau) \) is stochastically bounded by (7.10). By applying the convexity lemma and the argument in Pollard (1991), \( \hat{Z}_T(1) - \theta_T(\tau) = o_p(1). \) Thus,

\[
\hat{Z}_T(1) = -\log(m)^{-1} Q_H^{-1} W_T(\tau) + o_p(1) \xrightarrow{d} \mathcal{N}(0, \Omega_0).
\]

4. Asymptotic normality of \( \hat{\sigma} \) and \( \hat{\beta} \), and consistency of \( \hat{\beta}. \)

Next, we consider

\[
\frac{\hat{\gamma}(m\tau) - \hat{\gamma}(\tau)}{(Q_u(m\tau) - Q_u(\tau))} = \frac{\hat{\gamma}(m\tau) - \hat{\gamma}(\tau)}{\gamma(m\tau) - \gamma(\tau)} = \frac{\hat{\gamma}(m\tau) - \gamma(m\tau)}{\gamma(m\tau) - \gamma(\tau)} + 1 + \frac{\gamma(\tau) - \hat{\gamma}(\tau)}{\gamma(m\tau) - \gamma(\tau)}
\]

\[
\hat{\gamma}(m\tau) - \gamma(m\tau) = \frac{e^T \alpha_T(\hat{\beta}(m\tau) - \theta(m\tau))}{\sqrt{T}} = o_p(1)
\]

Similarly, the third term is also \( o_p(1). \)
\[\alpha T Q_u(\tau) = \sqrt{\tau T} \frac{Q_u(\tau)}{Q_u(m \tau)} \to \infty \text{ because } Q_u(\tau) \text{ is slowly varying.}\]

\[\frac{\hat{\gamma}(\tau)}{Q_u(\tau)} = 1 + \frac{\alpha T e'_1(\hat{\theta}(\tau) - \theta(\tau))}{\alpha T Q_u(\tau)} = 1 + o_p(1) \quad (7.12)\]

With \(g_T(\delta)\) defined in Equation (3.4), we have

\[G_T = \frac{\partial g_T(\delta)}{\partial \delta} = \begin{pmatrix}
\hat{\gamma}(l_1 \tau) - \hat{\gamma}(\tau) \\
\vdots \\
\hat{\gamma}(l_J \tau) - \hat{\gamma}(\tau)
\end{pmatrix} \otimes I_{d-1}\]

and

\[g_T(\delta_0) = \begin{pmatrix}
(\hat{\beta}(l_1 \tau) - \beta(l_1 \tau)) - (\hat{\beta}(\tau) - \beta(\tau)) \\
(\hat{\beta}(l_1 \tau) - \beta(l_1 \tau)) - (\hat{\beta}(\tau) - \beta(\tau)) \\
\vdots \\
(\hat{\beta}(l_J \tau) - \beta(l_J \tau)) - (\hat{\beta}(\tau) - \beta(\tau)) \\
(\hat{\beta}(l_J \tau) - \beta(l_J \tau)) - (\hat{\beta}(\tau) - \beta(\tau))
\end{pmatrix} \delta_0\]

Notice that

\[G_T \otimes \begin{pmatrix}
\log(l_1) \\
\log(m) \\
\vdots \\
\log(l_J) \\
\log(m)
\end{pmatrix} \to G = \begin{pmatrix}
-I_d & I_d & 0 & \cdots & 0 \\
-I_d & 0 & I_d & 0 & 0 \\
\vdots & 0 & 0 & \ddots & 0 \\
-I_d & 0 & 0 & 0 & I_d
\end{pmatrix} \quad (7.12)\]

Let \(\Delta = (-\delta_0; I_{d-1})\). \(\hat{\Gamma} = \begin{pmatrix}
-I_d & I_d & 0 & \cdots & 0 \\
-I_d & 0 & I_d & 0 & 0 \\
\vdots & 0 & 0 & \ddots & 0 \\
-I_d & 0 & 0 & 0 & I_d
\end{pmatrix}\) and \(\Gamma = \begin{pmatrix}
-I_d & \frac{1}{\sqrt{l_1}} I_d & 0 & \cdots & 0 \\
-I_d & 0 & \frac{1}{\sqrt{l_2}} I_d & 0 & 0 \\
\vdots & 0 & 0 & \ddots & 0 \\
-I_d & 0 & 0 & 0 & \frac{1}{\sqrt{l_J}} I_d
\end{pmatrix}\)

be two \(dJ \times d(J + 1)\) matrices.

\[\alpha T g_T(\delta_0) = (I_J \otimes \Delta) \hat{\Gamma} \hat{Z}_T(l_1, \cdots, l_J)\]
Linear representation:
\[
\sqrt{T}(\hat{\delta} - \delta_0) = \left( \frac{G_T}{Q_u(m_T) - Q_u(\tau)} \right)^T W_T \left( \frac{G_T}{Q_u(m_T) - Q_u(\tau)} \right)^{-1} \left( \frac{G_T}{Q_u(m_T) - Q_u(\tau)} \right)^T W_T G_T \left( \hat{\delta} - \delta_0 \right)
\]

Then
\[
\sqrt{T}(\hat{\delta} - \delta) = (G'WG)^{-1}G'W(I_J \otimes \Delta)\Gamma \hat{Z}_T(l_1, \ldots, l_J) + o_p(1) \Rightarrow N(0, \Omega_\delta)
\]

By what precedes and the fact that \(Q_u(\tau)\) is a slowly varying function (Lemma 7.1), \(\gamma(\tau) \rightarrow 1\) in probability. Therefore, we can replace the normalizing factor in the previous display by \(\sqrt{T}/\gamma(\tau)\).

\[
\frac{\sqrt{T}}{Q_u(\tau)}(\hat{\beta} - \beta) = \sqrt{T}(\hat{\delta} - \delta) + o_p(1) \Rightarrow N(0, \Omega_\delta)
\]

\[
\tau - EP(Y \geq -X'\beta(\tau)|X = x) = \tau - P(u \leq Q_u(\tau/h), D = 1|X = x)
\]
\[
= h \int_{-\infty}^{Q_u(\tau/h)} (1 - P(D^* = 1|X = x, u = e))dF_u(e|x)
\]
\[
= h \int_{-\infty}^{Q_u(\tau/h)} (1 - F_u(g(-e) + \phi(x)))dF_u(e)
\]
\[
\leq h \int_{-\infty}^{Q_u(\tau/h)} (1 - F_u(a(-e)^b + \phi(x)))dF_u(e)
\]

Let \(h_1(t) = F'_v(t)/(1 - F_v(t))\),
\[
(1 - F_v(t)) = (1 - F_v(t_0)) \exp(- \int_{t_0}^{t} h_1(u)du) \tag{7.14}
\]

Similarly, \(h_2(t) = F'_\varepsilon(t)/(1 - F_\varepsilon(t))\),
\[
(1 - F_\varepsilon(t)) = (1 - F_\varepsilon(t_0)) \exp(- \int_{t_0}^{t} h_2(u)du) \tag{7.15}
\]
Let us assume $h_2(t) \in RV_{\rho_2}$. By Lemma 7.2, $\rho_2 \geq -1$. By Karamata theory (see Resnick (1987)), $H_2(t) = \int_0^t h_2(u) du \in RV_{\rho_2+1}$.

Because $S_u \in RV_{-\alpha}$, by Karamata representation (Resnick (1987), Page 17), $h_1(u) = \frac{\alpha(u)}{u}$, where $\alpha(u) \to \alpha$ as $u \to \infty$.

$$P(v \geq a t^b + \phi(X))(-\log(P(\varepsilon \geq t)))^m \sim P(v \geq a t^b)(-\log(P(\varepsilon \geq t)))^m \sim L(t)^{m(\rho_2+1)-ab}$$ (7.16)

If $m(\rho_2+1) < \alpha b$, then $P(v \geq a t^b + \phi(x))(-\log(P(\varepsilon \geq t)))^m \to 0$. Thus $\exists K > 0$,

$$P(v \geq a t^b + \phi(x)) \leq \frac{K}{(-\log(P(\varepsilon \geq t)))^m}$$ (7.17)

Plug Equation (7.17) into Equation (7.13), we get

$$\text{RHS of Equation (7.13)} \leq K \int_{-\infty}^{Q_u(\tau/h)} \frac{1}{(-\log(F_u(e)))^m} dF_u(e)$$

$$= K \int_0^{\tau/h} \frac{1}{(-\log(p))^m} dp$$

$$= O\left(\frac{\sigma}{(-\log(\tau))^m}\right)$$ (7.18)

$$T|\mu_T| = O\left(\frac{\sqrt{T}}{(-\log(\tau))^m}\right)$$ (7.19)

We also need to consider $\frac{\sqrt{T}}{Q_u(\tau)}$ which is the normalizing factor of $\beta$. $F_u$ is rapidly varying, which implies $Q_u(\tau)$ is slowly varying. From equation (7.15), we get

$$Q_u(\tau) \sim H_2^+(-\log(\frac{\tau}{1-F_\varepsilon(t_0)})) \sim H_2^+(-\log(\tau))$$ (7.20)

Under Assumption 3(ii) and Lemma 7.2,

$$\text{RHS of Equation (7.20)} = L(-\log(\tau))(-\log(\tau))^{1/(\rho_2+1)}$$ (7.21)

Let $\tau = \frac{\log(T)^{2\zeta}}{T}$. Following Equation (7.19), $T|\mu_T| = O((\log(T))^{\zeta-m})$. From Equation (7.21), $\sqrt{T}/Q_u(\tau) = L(T)(\log(T))^{\zeta-1/(\rho_2+1)}$. We need $1/(\rho_2+1) < \zeta < m$ for arbitrary $m < \alpha b/(\rho_2+1)$. Then suitable choice of $\zeta$ and $m$ exist if $\alpha b > 1$. One sufficient condition for this is $E|v|^{1/b} < \infty$.

Remark 1. If $1-F_u(t)$ is rapidly varying, $\forall \zeta > 1/(\rho_2+1)$ and $\tau = \frac{\log(T)^{2\zeta}}{T}$ will make $T|\mu_T| = O(1)$ and $\sqrt{T}/Q_u(\tau) \to \infty$. So the convergence rate for $\beta$ is faster than $(\log(T))^m$ for any $m$. 

25
7.5 Further sufficient conditions for polynomial convergence rate

Next we maintain Assumption \ref{ass:3} (ii), the conditions in Proposition \ref{prop:3.1} and consider a special case of \(1 - F_v(t)\) being rapidly varying. We will show polynomial rate of convergence is possible.

**Assumption 7.** \(1 - F_v(t)\) is rapidly varying, \(h_1(t) \to \infty\). \(h_1(t) \in RV_{\rho_1}\), \(H_1(t) = \int_{0}^{t} h_1(u) du \in RV_{\rho_1+1}\) and \(\rho_1 + 1 > (\rho_2 + 1)b\).

For any \(m > 0\),

\[
P(v \geq at^b + \phi(x)) = K \exp(-H_1(t) + mH_2(at^b + \phi(x))) \to 0 \text{ as } t \to \infty \quad (7.22)
\]

Thus

\[
\text{RHS of Equation (7.13)} \leq K \int_{-\infty}^{Q_u(\tau/\mu_t)} (F_u(e))^m dF_u(e) = O(\tau^{m+1}) \quad (7.23)
\]

\[
\frac{T|\mu_T|}{\sqrt{T}} = O(1) \quad (7.24)
\]

for any \(m > 0\).

Computation of \(\sqrt{T}/Q_u(\tau)\) remains unchanged. Let \(\tau = T^{-\zeta}\) for \(0 < \zeta < 1\), \(\exists m \geq (1 - \zeta)/2\zeta\) so that \(\sqrt{T}/\tau^{m} = O(1)\) and \(\sqrt{T}/Q_u(\tau) \sim L(T)T^{(1-\zeta)/2} \to \infty\). Thus the convergence rate for \(\beta\) is faster than \(T^{-(1-\zeta)/2}\) for any \(0 < \zeta < 1\).

**Theorem 7.1.** Under Assumption \ref{ass:3} the conditions in Proposition \ref{prop:3.1} and \ref{prop:7}, the convergence rate for \(\beta\) is faster than \(T^{-(1-\zeta)/2}\) for any \(0 < \zeta < 1\).

7.6 Sufficient conditions for Assumption \ref{ass:6} in generalized Roy model

We consider here the generalized Roy model. In particular, let us assume we have total S sector. The potential outcome and selection equation for the sth-sector are defined as

**Assumption 8.**

\[
Y^*_s = X'\beta_s + (1 + X'\delta_s)\varepsilon_s \quad (7.25)
\]

\[
\{D = s\} = \{H = s\}\{D^* = s\}, \{D^* = s\} = \{Y^*_s + G_s(X) + \eta_s \geq \max_{i \neq s} (Y^*_i + G_i(X) + \eta_i)\} \quad (7.26)
\]

We further assume a decomposition of \(\varepsilon_s\) and \(\eta_s\)

**Assumption 9.** (1) \(\varepsilon_s = \lambda_{s,1}^s \pi + \nu_{s,1}\) and \(\eta_s = \lambda_{s,2}^s \pi + \nu_{s,2}\), where \(\pi\) is a vector of common factors. (2) \(\pi\) has compact support. (3) \(\pi \perp (\{\nu_{s,1,1}^s, \nu_{s,2,1}^s\}_{s=1,\ldots,S}), \nu_{s,1} \perp (\{\nu_{i,1}^s\}_{i \neq s}, \{\nu_{i,2}^s\}_{i=1,\ldots,S}), (\varepsilon, \eta) \perp X\). (4) \(H \perp (\varepsilon, \eta, X)\) and \(P(H = s) = h_s\).
Let $D_s = 1\{D = s\}$ and $Y = \sum_{s=1}^{J} D_s Y^*_s$. Consider dataset $(D_s, D_s Y, X)$ and the identification strategy: $\lim_{y \to \infty} P(D_s = 1 | X = x, Y^*_s = y) = h_s$. Denote $\overline{U}$ and $\underline{U}$ as the upper and lower bound of random variable $U$. Following the same spirit of Equation (7.13) with $U \equiv -\varepsilon_j, g(t) = (1 + x^T \delta)t, \phi(x) = -\max_{i \neq j} (x^T (\beta_i - \beta_j) + G_i(x) - G_j(x)) + (1 + x^T \delta)X_{i,1} + X_{i,2}$ and $v = \max_{i \neq s} (x_i \delta \nu_{i,1} + \nu_{i,2} - \nu_{s,2}), v \perp \perp \varepsilon_s$

$$\tau - EP(D_s Y \geq -X^T \beta_j(x)|X = x) = \tau - P(u \leq Q_u(\tau/h_s), D_s = 1 | X = x) = h_s \int_{-\infty}^{Q_u(\tau/h_s)} (P(D_s = 0 | X = x, u = e))dF_u(e|x)$$

$$= h_s \int_{-\infty}^{Q_u(\tau/h_s)} P(Y^*_s + G_s(X) + \eta_s \leq \max_{i \neq s} (Y^*_i + G_i(X) + \eta_i)|X = x, U = e)dF_u(e) \leq h_s \int_{-\infty}^{Q_u(\tau/h_s)} (1 - F_v(g(-e) + \phi(x)))dF_u(e)$$

(7.27)

Following the same argument used in the standard Roy model example, one sufficient condition for the high level assumptions is $E|v| < \infty$, which is true if $\sum_{s} (E|\nu_{s,1}| + E|\nu_{s,2}|) < \infty$.

**Theorem 7.2.** Under Assumption 8 and 9, there exists $\tau(T)$ such that $bias \tau(\mu_T) \to 0$ and $\sqrt{T} \overline{Q_u(\tau)} \to \infty$ if $\sum_{s} (E|\nu_{s,1}| + E|\nu_{s,2}|) < \infty$.

### 7.7 Conditional copula and convergence rate

Assumption 9 assumes we have a common factor decomposition and further more, the common factors have compact support. This assumption is intuitive because our identification strategy and thus estimation method emphasize on extracting information about structure parameter from extreme quantile objects (i.e. person with very high wage). And it makes more economic sense to have this extreme high wage generated from idiosyncratic unobservables for each individual rather than common factors among them. Technically, the assumption allows us to disentangle the dependent part of the unobservables from the independent part. One disadvantage of Assumption 9 is that it exclude the multivariate normal case which is used as a benchmark in empirical exercise. In what follows, we provide another way to show the existence of $\tau$ which includes the multivariate normal case. The main idea is, instead of decomposing the unobservables into dependent and independent part, we directly model the dependence by using copulas. Then we check several widely used copulas and show convergence rate for each of them.

**Assumption 10.** (1) $D = HD^*, D^* = 1_{\phi(X) \geq g(\varepsilon, X, \nu)}$, where $\nu$ is the unobservable for the selection equation. (2) $\text{Prob}(H) = h, \text{ and } H \perp (X, \varepsilon, \nu)$. (3) $(\varepsilon, \nu) \perp X$. (5) Let $Q = F_u(u|X = x) = F_u(u)$. $P = F_{g(\varepsilon, X, \nu)}(g(\varepsilon, X, \nu)|X = x)$ be two conditionally uniformly distributed random variables
Our aim then is to show that the existence of
\[ C \] and \[ P \] makes
\[ \sqrt{Q(x)} \]
Let the conditional copula that characterize the dependence between \((Q, P)\) be \(C(q, p|X)\).

The existence of conditional copula if two marginals are continuous is established by [Patton (2006)]. We will further impose parametric assumptions on the conditional copula thus the parameters of the copula are functions of \(X\). We use this conditional copula concept mainly because we allow for \(g\) to have \(x\) as its argument. By doing so, our assumption is general enough to nest multiplicative heteroskedastic unobservables. In addition, no decomposition of \(\varepsilon\) and \(\nu\) are needed. In fact, from the following two equations, we find the conditional marginal distributions of \(u\) and \(g(\varepsilon, X, \nu)\) do not play any role in our calculation. The assumption that \(P(x)\) is bounded away from 0 and 1 is not restrictive because \(\text{Supp}(X)\) is assumed to be compact. Following Equation (7.13) with a change of variable argument and denoting \(\tau' = h\tau\), we have

\[
\frac{\tau - EP(Y \geq -X'/\beta(\tau)|X = x)}{\tau} = \frac{\tau - P(u \leq Q_u(\tau/h), D = 1|X = x)}{\tau} = \frac{\tau}{\tau} = 1 - \frac{C(\tau/h, P(x); \theta(x))}{\tau} = 1 - \frac{C(\tau', P(x); \theta(x))}{\tau'} \leq 1 - \frac{C(\tau', P; \theta(x))}{\tau'}
\]

RHS of Equation (7.28) = \(1 \int_0^{\tau'} (1 - C_1(q, P; \theta(x))) dq\)

where \(C_1\) is the partial derivative of \(C\) w.r.t. its first argument.

Our aim then is to show that the existence of \(\tau\), under different parametric specification, which makes

\[
\sqrt{T} \frac{\tau}{\tau'} E(1 - \frac{C(\tau', P; \theta(X))}{\tau'}) \rightarrow 0, \sqrt{\tau' T / Q(\tau')} \rightarrow -\infty
\]

Usually, uniform bound on \(\theta(x)\) is needed to ensure one choice of \(\tau\) which does not depend on \(x\). The uniform boundedness of \(\theta(x)\) is not a strong assumption because \(\text{Supp}(X)\) is assumed to be compact. Depending on different situations, we will use Equation (7.28) or Equation (7.29) to proceed the calculation. But before that, we first notice that our Assumption 3 implies \(C(q, P(x); \theta(x))/q \rightarrow 1\) as \(q \rightarrow 0\) for \(\forall x \in X\). If \(C(q, P(x), \theta(x))/q \rightarrow 0\), then we could use a rotation argument and define our rotated copula to be \(C'(q, P(x); \theta(x)) = q - C(q, 1 - P(x); \theta(x))\) and \(C'(q, P(x), \theta(x))/q \rightarrow 1\). If \(C(q, P(x); \theta(x))/q \rightarrow h(x)\) where \(h(x)\) is a nontrivial function of \(x\), then Assumption 3 is violated. Frank and Joe copulas are in this category. However, fortunately enough, Gaussian, Clayton,
Gumbel, Gumbel-Barnett and many other one-parameter Archimedean copula families mentioned in Nelsen (2006) Chapter 4 satisfy Assumption 4 and the choice of $\tau$ for some of them are calculated. For each of the cases below, we first list the definition of unconditional copulas with corresponding parameter restrictions. In the actual calculation, we will deal with the conditional version of the copula so that the parameter of the copula will be a function of $x$. Then further restrictions on the uniform bound for the parameter across $x$ is imposed.

Case 1: Gaussian copula

$$C_{\rho}(q,p) = \int_{-\infty}^{\Phi^{-1}(q)} \int_{-\infty}^{\Phi^{-1}(p)} \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp\left(\frac{-(t^2 - 2\rho ts + s^2)}{2(1 - \rho^2)}\right) ds dt$$

with $\rho \in [-1, 1]$. For the conditional counterpart, we further assume $\rho = \inf_{x \in \text{Supp}(x)} \rho(x) > 0$. If $\sup_{\rho} < 0$, we can switch the direction of our identification condition (i.e. instead of looking at upper quantile of log wage, we look at the lower end of it). We refer to the remarks of Assumption 3 in D’Haultfoeuille & Maurel (2013) for a detail discussion. Let $\Phi$ and $\phi$ be the cdf and pdf of standard normal distribution respectively.

$$1 - C_{1_{\rho(x)}}(q,p) = \frac{\partial}{\partial q} \int_{-\infty}^{\Phi^{-1}(p)} \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp\left(\frac{-(\Phi^{-1}(q))^2 - 2\rho \Phi^{-1}(q)s + s^2}{2(1 - \rho^2)}\right) ds$$

$$= \left. \frac{\partial \Phi^{-1}(q)}{\partial q} \phi(\Phi^{-1}(q)) \right|_{p} - \Phi\left(\frac{\Phi^{-1}(p) - \rho(x) \Phi^{-1}(q)}{\sqrt{1 - \rho^2}}\right)$$

$$= \Phi\left(\frac{\rho(x) \Phi^{-1}(q) - \Phi^{-1}(p)}{\sqrt{1 - \rho^2}}\right)$$

(7.31)

Plug Equation (7.31) into Equation (7.29) and notice that $\Phi\left(\frac{\rho(x) \Phi^{-1}(q) - \Phi^{-1}(p)}{\sqrt{1 - \rho^2}}\right)$ is increasing in $q$ if $\inf_{x} \rho(x) > 0$ and $\Phi^{-1}(\tau') < 0$ we have

$$\text{RHS of Equation (7.28)} \leq \Phi\left(\frac{\rho \Phi^{-1}(\tau') - \Phi^{-1}(p)}{\sqrt{1 - \rho^2}}\right)$$

$$\sim \frac{\sqrt{1 - \rho^2}}{\rho \Phi^{-1}(\tau')} \phi\left(\frac{\rho \Phi^{-1}(\tau') - \Phi^{-1}(p)}{\sqrt{1 - \rho^2}}\right)$$

(7.32)

$$\leq \frac{\sqrt{1 - \rho^2}}{\rho \Phi^{-1}(\tau')} \phi\left(\frac{1 - \Delta}{\sqrt{1 - \rho^2}} \sqrt{-2 \log(\tau')}\right)$$

$$= O\left(\frac{\tau^{2(1 - \Delta)^2}}{L(\tau')}\right)$$

29
where $\Delta$ is an arbitrarily positive number. We let $\tau' = T^{-\alpha}, 1 > \alpha > 1/(1 + \frac{2\rho}{\sqrt{1 - \rho^2}})$. It is easy
to verify that for this choice of $\tau'$, any $\tau = \tau' l$ with some $l > 0$ will ensure $\sqrt{\tau T}/Q_u(\tau) \to -\infty$
and $T\mu_T(\tau) \to 0$. So the existence of $\tau$ have been proven and the convergence rate is shown to be
polynomial in $T$.

Case 2:

$$C_\theta(q, p) = \max\left(\left[ q^{-\theta} + p^{-\theta} - 1 \right]^{-1/\theta}, 0 \right)$$

where $\theta > 0$. We follow the notion used in Nelsen (2006) and call it the Clayton copula family.
For the conditional counterpart, we assume $\infty > \bar{\theta} = \sup_x \theta(x) \geq \theta(x) \geq \inf_x \theta(x) = \underline{\theta} > 0$,
$\left[q^{-\theta(x)} + p^{-\theta(x)} - 1\right]^{-1/\theta(x)} > 0$ eventually as $q \to 0$. So we could ignore the max part.

RHS of Equation (7.28) \leq 1 - \left(1 + \tau^{-\theta(x)}(p^{-\theta(x)} - 1)\right)^{-1/\theta(x)} \leq K\tau^{-\theta(x)} = O(\tau^{-\theta}) \quad (7.33)$

Therefore, let $\tau' = T^{-\alpha}, 1 > \alpha > 1/(2\theta + 1)$ guarantee $\sqrt{\tau T}/Q_u(\tau) \to -\infty$ and $T\mu_T(\tau) \to 0$
and the convergence rate is shown to be polynomial in $T$.

Case 3:

$$C_\theta(q, p) = \exp(-[-(\log(q))]^\theta + (-\log(p))^\theta)^{1/\theta})$$

where $\theta > 1$. When $\theta = 1$, the copula degenerates to independent copula ($C(q, p) = qp$) which
violates our Assumption 4. For the conditional counterpart, we assume $\infty > \bar{\theta} = \sup_x \theta(x) \geq \theta(x) \geq \inf_x \theta(x) = \underline{\theta} > 1$.

RHS of Equation (7.28) \leq 1 - \exp(-\log(\tau') + \log(\tau')\left[1 + (\frac{\log(p)}{\log(\tau')})^{\theta(x)}\right])^{1/\theta(x)}

\leq -K(-\log(\tau'))(1 - (1 + (\frac{\log(p)}{\log(q)})^{\theta(x)})) \quad (7.34)$

Applying the same argument in Appendix 7.4, if $\theta - 1 > \frac{1}{\rho_2+1}$ where $\rho_2$ is defined in Lemma 7.2,
then $\tau$ exists and we have log convergence rate.

Case 4:

$$C_\theta(q, p) = q - q(1 - p) \exp(-\theta \log(q) \log(1 - p))$$

where $1 \geq \theta > 0$. This is the rotated Gumbel-Barnett copula family. For the conditional counterpart, we assume $1 \geq \bar{\theta} \geq \theta(x) \geq \underline{\theta} > 0$

RHS of Equation (7.28) \leq (1 - p) \exp(-\theta(x) \log(\tau') \log(1 - p)) \leq O(\tau'^{-\theta \log(1 - p)}) \quad (7.35)$
Thus \( \tau' = T^{-\alpha} \), \( \alpha > 1/(1 - 2\theta \log(1 - p)) \) guarantee \( \sqrt{\tau' T}/Q_u(\tau) \to -\infty \) and \( T\mu_T(\tau) \to 0 \) and the convergence rate is shown to be polynomial in \( T \).

Case 5:
\[
C_\theta(q, p) = (1 + [(q^{-1} - 1)^\theta + (p^{-1} - 1)^\theta]^{1/\theta})^{-1}
\]
where \( \theta \geq 1 \). For the conditional counterpart, we assume \( \theta(x)(\theta) \geq 1 \).

The RHS of Equation (7.28) = 1 - \( \exp((-K(1 + \log(T))^{1-\theta}) \log(T)^m) \to 0 \) for \( \forall m > 0 \), it can be easily verified that for any \( m > 1/(\rho_2 + 1) \), \( \tau' = \log(T)^m / T \) guarantee \( \sqrt{T}T/Q_u(\tau) \to -\infty \) and \( T\mu_T(\tau) \to 0 \).

Case 6:
\[
C_\theta(q, p) = \exp(1 - [(1 - \log(q))^\theta + (1 - \log(p))^\theta - 1]^{1/\theta})
\]
\( \theta \geq 0 \). For the conditional counterpart, let \( \theta(x) \geq \theta > 0 \). When \( \theta = 1 \), the copula degenerates to independent copula which violates Assumption 4. If \( \theta(x) \leq \bar{\theta} < 1 \),

Thus \( \tau' = T^{-\alpha} \), \( \alpha > 1/3 \) guarantee \( \sqrt{T}T/Q_u(\tau) \to -\infty \) and \( T\mu_T(\tau) \to 0 \) and the convergence rate is shown to be polynomial in \( T \).

Case 7:
\[
C_\theta(q, p) = (1 + [(q^{-1/\theta} - 1)^\theta + (p^{-1/\theta} - 1)^\theta]^{1/\theta})^{-\theta}
\]
where $\theta \geq 1$. For the conditional counterpart, let $\infty > \theta \geq \theta(x) \geq \theta \geq 1$.

RHS of Equation (7.28) = $1 - (\tau^{1/\theta(x)} + [(1 - \tau^{1/\theta(x)})^{\theta(x)} + \tau'(p^{-1/\theta(x)} - 1)^{1/\theta(x)}) - ^{\theta(x)}$ 

$\leq K \tau^{1/\theta}$ \hspace{1cm} (7.39)

$\alpha > 1/(1 + 2/\theta)$ guarantee $\sqrt{T}/Q_u(\tau) \to -\infty$ and $T\mu_T(\tau) \to 0$ and the convergence rate is shown to be polynomial in $T$.

Case 8:

$C_\theta(q, p) = \theta / \log(\exp(\theta/q) + \exp(\theta/p) - \exp(\theta))$

where $\theta > 0$. For the conditional counterpart, we assume $\infty > \theta \geq \theta(x) \geq \theta > 0$.

RHS of Equation (7.28) = $1 - 1/(1 + \tau'/\theta(x) \log(1 + (\exp(\theta(x)/p) - \exp(\theta(x)))) \exp(-\theta(x)/\tau')) \leq K \tau' \exp(-\theta'/\tau')$ \hspace{1cm} (7.40)

Therefore $\forall \alpha < 1$, $\tau' = T^{-\alpha}$ will guarantee $\sqrt{T}/Q_u(\tau) \to -\infty$ and $T\mu_T(\tau) \to 0$ and the convergence rate is shown to be polynomial in $T$.

Case 9:

$C_\theta(q, p) = [\log(\exp(q^{-\theta}) + \exp(p^{-\theta}) - e)]^{-1/\theta}$

where $\theta > 0$. For the conditional counterpart, let $\infty > \theta \geq \theta(x) \geq \theta > 0$.

RHS of Equation (7.28) = $1 - [1 + \tau^{\theta(x)}(x) \log(1 + \frac{\exp(p^{-\theta(x)}) - e}{\exp(\tau' - \theta(x)))}]^{-1/\theta(x)}$

$\leq K \tau^{\theta(x)}(x) \log(1 + [\exp(p^{-\theta(x)}) - e] \exp(-\tau' - \theta(x)) \hspace{1cm} (7.41)$

Therefore $\forall \alpha < 1$, $\tau' = T^{-\alpha}$ will guarantee $\sqrt{T}/Q_u(\tau) \to -\infty$ and $T\mu_T(\tau) \to 0$ and the convergence rate is shown to be polynomial in $T$. 
