ENSAE ParisTech and M2 APE

Recent Developppments in Semi and Non-Parametric Econometrics

Professors : X. D'Haultfœuille, A. Simoni

Quantile regression for panel data

A discussion of Koenker (2004) Quantile Fixed Effect penalized estimator

---

Jeremy L'Hour[1], Louise Paul-Delvaux[2] & Nicolas Saleille[3]

Contents

---

[1] jeremy.l.hour@ensae-paristech.fr

[2] louise.paul.delvaux@ensae-paristech.fr

[3] nicolas.saleille@ensae-paristech.fr

I Quantile regression in panel data: position of the problem

A major feature of OLS estimation is its focus on average effects. In many cases, one might want to measure heterogeneous effects of the covariates X on the outcome variable y. This is for instance the case when regressing some expenditure (e.g. food expenditure) on income: depending on the income quantile, the effect of income on expenditure varies. In particular, consumption levels of agents within low income quantiles should react more to an increase in their income.

Quantile regressions is a way to model heterogeneity. In a standard quantile regressions, we estimate the effect of covariates on the conditional quantile of the outcome, assuming that the τ-conditional quantile of the residuals is null:

$$Y = X'\beta_\tau + \varepsilon_\tau, \quad q_\tau(\varepsilon|X) = 0 \quad \Rightarrow \quad q_\tau(Y|X) = X'\beta_\tau \tag{1}$$

This approach is flexible since the vector of parameters depends on the considered quantile τ. To estimate such regressions, we use the following property:

$$\beta_\tau \in \arg\min_\beta \mathbb{E}\left[\rho_\tau(Y - X'\beta)\right] \tag{2}$$

where $\rho_\tau(u) = (\tau - \mathbb{1}(u < 0)) \times u$ is the check function. Estimation is then conducted for a given i.i.d. sample $(Y_i, X_i)_{i\in\{1,\ldots n\}}$ through optimization methods to solve the empirical counterpart of (2):

$$\hat{\beta}_\tau \in \arg\min_\beta \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - X_i'\beta) \tag{3}$$

If quantile regressions are easily performed when observations are i.i.d., it is no longer the case when one has to deal with panel data. In this setting, we observe n individuals (each indexed by i) at T distinct points in time (each indexed by t). We consider the quantile regression model:

$$y_{it} = x_{it}'\beta_\tau + \alpha_i + \epsilon_{it,\tau} \tag{4}$$

In this specification, the individual error term is decomposed into two components:

- The random variable α_i can be interpreted as the unobserved idiosyncratic characteristics of individual i that also affect y_{it} and is time invariant.
- $\epsilon_{it,\tau}$ is the individual time-varying remaining error term.

In a wage equation, one can think of α_i as the ability, intelligence or social skills of the worker; in a price equation, α_i could be non-measurable or qualitative characteristics of the good, such as the quality of the neighborhood for a flat. Notice that there are n different α_i which means that if we have to estimate them by standard methods, we already consume n degrees of freedom. We focus on the case where the random effect assumption does not hold, *i.e.* when the unobserved heterogeneity causes a problem by violating the usual exogeneity assumption. This model is known as the fixed effect model. In this framework, panel data are essential to obtain consistent estimators of β_τ, the parameter of interest. However, the α_i are nuisance parameters that are difficult to estimate consistently unless $T \to +\infty$. Moreover, independence between observations within the same cluster cannot be assumed any longer.

The quantile regression model is of interest in panel data, for example when one is interested in evaluating the effect of a job training program on earnings while accounting for unobserved individual heterogeneity. Indeed, the job training program may have a differentiated impact at different points of the earnings distribution. But we also know that earnings are influenced by unobserved factors such as the ability of the worker that may also play a role in self-selection into the training program. Another example has been developed in Ohinata and van Ours (2013) which analyzes how the share of immigrant children in the classroom affects the educational achievment of native Dutch children. Panel quantile regression allow the authors to control for potential selectivity in the choice of parents who might prefer having their child in a school with a low share of immigrant children.

Koenker (2004) proposes a methodology to model panel data with a quantile regression. The main feature of panel data is to take into account unobserved heterogeneity coming from individuals. Applying quantile regressions to panel data is a way to allow for various forms of heterogeneity simultaneously: namely, an individual effect and a quantile effect. In this context, the model writes:

$$Q_\tau(y_{it}|x_{it}) = \alpha_i + x'_{it}\beta_\tau \tag{5}$$

where the individual effect α has a pure location shift effect on the conditional quantile. We are interested in estimating q quantiles: $\tau_1, ..., \tau_q$. To estimate this model, one has to solve:

$$\min_{\alpha,\beta} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{T} w_k \cdot \rho_{\tau_k}\left(y_{it} - \alpha_i - x'_{it}\beta_{\tau_k}\right) \tag{6}$$

Where the $w_1, ..., w_q$ are given weights that give the importance of the k-th quantile in the estimation of the fixed effects. Assume that x_{it} is a vector of length p. We have to estimate $n+q\times p$ parameters: n fixed effects, p coefficients for the effects of the covariates at the first quantile of interest β_{τ_1}, p at the second quantile β_{τ_2}, etc... which can become quickly very complicated.

In Section II, we present the estimator of Koenker (2004) and the underlying assumptions. In Section III we state and explain the main theoretical results. In Section V we apply the estimator to wage data from Vella and Verbeek (1998). In Section IV we expose the main limitations and study an extension of the estimator.

II Koenker (2004)'s solution: penalized quantile regression

2.1 The estimator

To solve program (6) in a computationally feasible manner, Koenker (2004) proposed a ℓ_1-penalized procedure that solves:

$$\min_{\alpha,\beta} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{T} w_k \cdot \rho_{\tau_k}\left(y_{it} - \alpha_i - x'_{it}\beta_{\tau_k}\right) + \lambda \sum_{i=1}^{n} |\alpha_i| \tag{7}$$

Where λ is the tuning parameter that balances the trade-off between achieving a good fit and minimizing the ℓ_1-norm of the fixed effect vector. For $\lambda \to 0$ we get the fixed effect estimator, for $\lambda \to \infty$ we obtain a model without fixed effects since we will have $\hat{\alpha}_i \to 0$ as even small value for the fixed effects are heavily penalized.

2.2 Justification of the penalized estimator

Before discussing this estimator, we state a result from regular linear panel regression that justifies the idea of a penalized estimator. Firstly, we rewrite the usual linear panel data model in its matrix form:

$$y = X\beta + Z\alpha + \epsilon \tag{8}$$

Where Z represents a matrix full of 1s and 0s so as to identify individual i's effect in the model. Now, assume ϵ to be distributed as $\mathcal{N}(0_{nT}, R)$ and α as $\mathcal{N}(0_{nT}, Q)$ independently from ϵ. Hence, the whole error term u ($u = Z\alpha + \epsilon$), is distributed as $\mathcal{N}\left(0_{nT}, R + ZQZ^T\right)$. The GLS estimator of β is given by:

$$\hat{\beta}^{GLS} = \left(X^T\left(R + ZQZ^T\right)^{-1}X\right)^{-1}X^T\left(R + ZQZ^T\right)^{-1}y \tag{9}$$

Note that such an estimator is also the solution of the following program:

$$\min_{\beta,c} \|y - X\beta - Zc\|^2_{R^{-1}} + \|c\|^2_{Q^{-1}} \tag{10}$$

Proof of this result can be found in the article. We can see that in this case, the optimal estimator of β is the result of a penalized Least-Squares program that shrinks individual effects α towards zero. We can notice that this penalty depends on the assumed distribution of α. Starting from this observation, we could as well define another penalty reflecting a different prior belief about the distribution of α and achieve a better estimate of β even though the dimension of β is not necessarily large[4].

This result opens to the door to the use of other penalty functions that possess desirable properties. The choice of the ℓ_1-penalty in the present article is interesting regarding several aspects, as advocated in the Lasso of Tibshirani (1994). The first one is that it is the smallest convex norm which makes it computationally appealing. Indeed, a norm that would penalize the number of non-zero elements in α in the same way as does the BIC in a variable selection framework would yield a NP-hard problem. Moreover, for a level of penalty which is large enough, the result will be exactly sparse in the sense that a number of elements in α will be exactly set to zero since the ℓ_1-norm has a kink at zero. The ℓ_1-penalty reflect a belief of approximate sparsity regarding α. Economically speaking, this assumption amounts to say that the unobserved heterogeneity is rare and that the vast majority of people have an α_i equal to zero.

III Theorical results

In this section, we decompose the two main theorems of the paper, provide intuition for their assumptions and detail the proof of one of them. As both of them use Knight (1998) identity, we provide a proof of this result in Appendix A.

[4]For the link between prior distribution in the Bayesian framework and penalized regressions, see for example (Hastie *et al.*, 2009, p. 61). The penalty we have just proposed corresponds indeed to a Normal prior distribution. The ℓ_1 penalty corresponds to a Laplace prior distribution.

3.1 Theorem 1

This first proof is a simple case where only a single quantile is considered and does not consider the penalized estimator. It helps however exposing the main arguments and proof techniques. The idea is to find the asymptotic distribution of the following quantity:

$$G\left(\hat{\alpha}, \hat{\beta}\right) = \sum_{i=1}^{n}\sum_{t=1}^{T} \rho_\tau\left(y_{it} - \hat{\alpha}_i - x_{it}'\hat{\beta}_\tau\right) \tag{11}$$

Notations:

x_{it}	:	independant variables for individual i at time t. The associated $nT \times p$ matrix is X.
y_{it}	:	outcome for individual i at time t. Its distribution function conditional to x_{it} is denoted F_{it} and the associated density is f_{it}.
$(\hat{\beta}_\tau, \hat{\alpha})$	:	minimizers of the function $G(.,.)$
$\xi_{it,\tau}$	:	true predicted part of the model, defined as $\xi_{it,\tau} = \alpha_i + x_{it}'\beta_\tau$.
$\hat{\delta}_0$	:	vector of length n defined as $\hat{\delta}_0 = \sqrt{T}(\hat{\alpha} - \alpha)$
$\hat{\delta}_1$	:	vector of length p defined as $\hat{\delta}_1 = \sqrt{nT}(\hat{\beta}_\tau - \beta_\tau)$
$\hat{\delta}$	:	$= (\hat{\delta}_0, \hat{\delta}_1)$
z_{it}	:	vector of length n with $z_{it}[i] = 1$ and 0 otherwise so as to identify the fixed effect associated with individual i. We can notice that this vector is not time varying.
Z	:	$nT \times n$ matrix equals to $I_n \otimes e_T$ where e_T is a vector of length T full of 1.
$D_0^{n,T}$	:	$= \frac{\omega}{T}\left(Z : X/\sqrt{n}\right)'\left(Z : X/\sqrt{n}\right)$. If exists, the limit of this matrix will be denoted D_0.
$D_1^{n,T}$	:	$= \frac{1}{T}\left(Z : X/\sqrt{n}\right)' diag(f_{ij}(\epsilon_{ij}(\tau)))\left(Z : X/\sqrt{n}\right)$. If it exists, the limit of this matrix will be denoted D_1.
v_{it}	:	$= z_{it}'\delta_0 + x_{it}'\delta_1/\sqrt{n}$ (for convenience)

Now we will consider α and β_τ the true values of the parameters and work on the quantity $G\left(\hat{\alpha}, \hat{\beta}\right) - G(\alpha, \beta)$. With the above notations, it can be re-written:

$$G\left(\hat{\alpha}, \hat{\beta}\right) - G(\alpha, \beta) = \sum_{i=1}^{n}\sum_{t=1}^{T} \rho_\tau\left(y_{it} - \xi_{it,\tau} - v_{it}/\sqrt{T}\right) - \rho_\tau\left(y_{it} - \xi_{it,\tau}\right) \tag{12}$$

Theorem 1 relies on three assumptions that can be found in the article. Instead of re-stating them, we explain the intuition.

Assumption: *A1*

We deal with an independent sample across both the individuals and time periods, which is a regular assumptions in quantile regressions in the cross-section setting. This assumption however appears to be very strong in the panel data case as one can hardly assume that observations are independent within individual clusters. Conditional densities of the outcome are assumed to have bounded derivatives at $\xi_{it,\tau}$. The author states that it is usual in quantile regression. However, we find it very strong and notice that to assess the asymptotic distribution of similar estimators, Knight (1998) uses weaker assumptions. Also, the regular quantile regression only assumes the existence of a conditional density at $\xi_{it,\tau}$, but nothing regarding its derivative.

Assumption: *A2*

This assumption makes sure that the asymptotic distribution of the quantile FE estimator will exist by stating that its variance will be positive definite:

$$\lim_{T\to\infty, n\to\infty} D_0^{n,T} = D_0 \in \mathcal{S}^+$$

$$\lim_{T\to\infty, n\to\infty} D_1^{n,T} = D_1 \in \mathcal{S}^+$$

Assumption: *A3*

This statement assumes that the norm of the regressors x_{it} are bounded from above - say for the above proofs by M - which makes the proof of the theorem easier.

Theorem: *Asymptotic Normality of the non-penalized quantile FE estimator*

Under conditions A1-A3, if it exists $a > 0$ such that $n^a/T \to 0$, the $\hat{\delta}_1 = \sqrt{nT}(\hat{\beta}_\tau - \beta_\tau)$ minimizer of (12) converges in distribution to a Gaussian random vector with mean zero and covariance matrix given by the lower p by p block of the matrix $D_1^{-1} D_0 D_1^{-1}$.

Proof: : We introduce $G\left(\hat{\alpha}, \hat{\beta}\right) - G\left(\alpha, \beta\right) = Z_{nT}(\hat{\delta})$

First step: Rewrite $Z_{nT}(\hat{\delta})$ using Knight identity

We use the result of Knight (1998) proven in Appendix A with $u = y_{it} - \xi_{it,\tau}$ and $v = v_{it}/\sqrt{T}$. This allows us to decompose:

$$Z_{nT}(\delta) = Z_{nT}^{(1)}(\delta) + Z_{nT}^{(2)}(\delta)$$

with:

$$\begin{aligned} Z_{nT}^{(1)}(\delta) &= -\tfrac{1}{\sqrt{T}} \textstyle\sum_{i=1}^n \sum_{t=1}^T \psi_\tau(y_{it} - \xi_{it,\tau}) v_{ij} \\ Z_{nT}^{(2)}(\delta) &= \textstyle\sum_{i=1}^n \sum_{t=1}^T \int_0^{v_{it}/\sqrt{T}} \mathbb{1}\left(y_{it} \le \xi_{it,\tau} + s\right) - \mathbb{1}\left(y_{it} \le \xi_{it,\tau}\right) \mathrm{d}\, s \\ &= \tfrac{1}{\sqrt{T}} \textstyle\sum_{i=1}^n \sum_{t=1}^T \int_0^{v_{it}} \mathbb{1}\left(y_{it} \le \xi_{it,\tau} + u/\sqrt{T}\right) - \mathbb{1}\left(y_{it} \le \xi_{it,\tau}\right) \mathrm{d}\, u \end{aligned} \tag{13}$$

where $\psi_\tau(x) = \tau - \mathbb{1}\left(x < 0\right)$ and the last transformation is obtained using the change of variable $u = s\sqrt{T}$.

Second step: Apply Central Limit Theorem to $Z_{nT}^{(1)}(\delta)$

The first term is asymptotically Gaussian as a consequence of Central Limit Theorem. This might not be easy to see as we are dealing with a sum of nT terms and in v_{it}, $z_{it}'\hat{\delta}_0$ is not divided by $\sqrt{n}$. However, one needs to remember that z_{it} as one and only one non-null element which means that the i^{th} element of δ_0 appears only T times in the sum. Moreover, conditions A2 and A3 imply that we can apply the CLT.

$$Z_{nT}^{(1)}(\delta) \xrightarrow{n\to\infty, T\to\infty} -\delta' B$$

Where B is a a zero mean Gaussian vector with convariance matix equals to D_0.

Third step: Find the limiting distribution of $Z_{nT}^{(2)}(\delta)$
(a) Evaluate the asymptotic expected value of $Z_{nT}^{(2)}(\delta)$
Conditional to x_{it}, we have:

$$E\big(Z_{nT}^{(2)}(\delta)\big) = \tfrac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\sqrt{T}\int_0^{v_{it}}\Big(F_{it}(\xi_{it,\tau}+u/\sqrt{T})-F_{it}(\xi_{it,\tau})\Big)\,\mathrm{d}\,u \qquad (14)$$

Using a Taylor series expansion with explicit formulae for the remainder, we can rewrite:

$$F_{it}\left(\xi_{it,\tau}+u/\sqrt{T}\right)-F_{it}\left(\xi_{it,\tau}\right)=\frac{u}{\sqrt{T}}f_{it}(\xi_{it,\tau})+\int_0^{u/\sqrt{T}}(u/\sqrt{T}-v)f'_{it}(\xi_{it,\tau}+v)\,\mathrm{d}\,v$$

Introducing this equality in (14), we can study the two terms separately:

- By integration on u, the first term becomes:

$$\tfrac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\sqrt{T}\int_0^{v_{it}}\tfrac{u}{\sqrt{T}}f_{it}(\xi_{it,\tau})\,\mathrm{d}\,u = \tfrac{1}{2T}\sum_{i=1}^{n}\sum_{t=1}^{T}f_{it}(\xi_{it,\tau})v_{it}^2$$

  Considering that $v_{it}=z'_{it}\delta_0+x'_{it}\delta_1/\sqrt{n}$ this double sum is a quadratic form in δ. The associated matrix is D_1^n and we have:

$$\begin{aligned}\tfrac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\sqrt{T}\int_0^{v_{it}}\tfrac{u}{\sqrt{T}}f_{it}(\xi_{it,\tau})\,\mathrm{d}\,u &= \tfrac{1}{2}\delta' D_1^{n,T}\delta\\ &\xrightarrow{n\to\infty,T\to\infty}\tfrac{1}{2}\delta' D_1\delta\end{aligned}$$

  Thanks assumption (A2) and by continuity of the quadratic operator.

- Under assumption (A1) considering $M'=\sup_{1\le i\le n,1\le t\le T}||f'_{it}||$, we have:

$$\forall i,\forall t,\int_0^{u/\sqrt{T}}(u/\sqrt{T}-v)f'_{it}(\xi_{it,\tau}+v/\sqrt{T})\,\mathrm{d}\,v\le\int_0^{u/\sqrt{T}}(u/\sqrt{T}-v)M'\,\mathrm{d}\,v=\frac{1}{2}M'\frac{u^2}{T}$$

  Such as:

$$\begin{aligned}\tfrac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\sqrt{T}\int_0^{v_{it}} &\int_0^{u/\sqrt{T}}(u/\sqrt{T}-v)f'_{it}(\xi_{it,\tau}+v)\,\mathrm{d}\,v\,\mathrm{d}\,u\\ &\le\tfrac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\tfrac{M'}{2\sqrt{T}}\int_0^{v_{it}}u^2\,\mathrm{d}\,u\\ &\le\tfrac{1}{\sqrt{T}}\times\tfrac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}v_{it}^3\times\tfrac{M'}{6}\end{aligned}$$

Here we can notice that:

$$\begin{aligned}\tfrac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}v_{it}^3 &\le \max_{1\le i\le n,1\le t\le T}|v_{it}|\sum_{i=1}^{n}\sum_{t=1}^{T}v_{it}^2\\ &\le \max_{1\le i\le n,1\le t\le T}|v_{it}|\tfrac{1}{\omega}\delta' D_0^{n,T}\delta\end{aligned}$$

With $\max_{1\le i\le n,1\le t\le T}|v_{it}|\le||\delta_0||+p||\delta_1||M$ thanks assumption (A3). Assumption (A2) tells us that $D_0^{n,T}\xrightarrow{n\to\infty,T\to\infty}D_0$ such as $\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}v_{it}^3$ tends to a constant. As a product of a $\frac{1}{\sqrt{T}}$ and a quantity which tends to a constant, we have this second term which tends to 0 when n and T tend to ∞. Thus, we have:

$$E\big(Z_{nT}^{(2)}(\delta)\big)\xrightarrow{n\to\infty,T\to\infty}\frac{1}{2}\delta' D_1\delta$$

(b) Show that the variance of $Z_{nT}^{(2)}(\delta)$ tends to 0

$$Var(Z_{nT}^{(2)}(\delta))\le\frac{2}{\sqrt{T}}\max_{1\le i\le n,1\le t\le T}|v_{it}|E(Z_{nT}^{(2)}(\delta))$$

This quantity tends to 0 as the expectation of $Z_{nT}^{(2)}(\delta)$ is finite and $max(v_{it})$ is bounded by assumption A3.

Fourth step: Conclusion of the heuristic proof
The previous steps proved that the limiting form of the objective function is:

$$Z_\infty(\delta)=-\delta' B+\frac{1}{2}\delta' D_1\delta$$

with B a zero-mean Gaussian vector with covariance matrix D_0. This asymptotic function has a simple minimizer:

$$\frac{\partial Z_\infty}{\partial \delta} = \frac{1}{2}(D_1' + D_1)\delta - B$$

$$\frac{\partial Z_\infty}{\partial \delta} = 0 \quad \Rightarrow \quad \delta^\infty = D_1^{-1}B$$

If we focus on the δ_1 component of δ - we can not discuss δ_0 as the number of elements goes towards infinity as $n \to \infty$ - we have $\hat{\delta}_1^{n,T} \to \delta_1^\infty$. This is established thanks the convexity of Z_{nT} and the uniqueness of Z_∞ minimizer. The limits of this heuristic proof are reached here: the infinite dimension nature of δ_0 (since the number of fixed effects to be estimated grows as n tends to ∞) casts doubts on the validity of this approach. ■

3.2 Theorem 2

Theorem 2 deals with the asymptotic distribution of the Panel FE Quantile (7), by using the same arguments used in the toy example of Theorem 1.

The assumptions are similar to those which supported Theorem 1, regarding the conditional distribution of y_{it}, the asymptotic variance-covariance matrices of the regressors, and the boundedness of the regressors.

Theorem: *Asymptotic Normality of the penalized quantile FE estimator*

> Under conditions similar to A1-A3, if it exists $a > 0$ such that $n^a/T \to 0$, and provided that $\lambda_T/\sqrt{T} \to \lambda_0$, then the $\sqrt{nT}(\hat{\beta}_{\tau_1} - \beta_{\tau_1})$, ..., $\sqrt{nT}(\hat{\beta}_{\tau_q} - \beta_{\tau_q})$ where $\hat{\beta}_{\tau_1}$, ..., $\hat{\beta}_{\tau_q}$ are the minimizers of (7), converge in distribution to a Gaussian random vector, as both n and T tend to ∞.

It is to be noted that no information is provided regarding what the quantity λ_0 might look like.

IV Limitations and extensions

4.1 Discussion of the main statistical limitations

When one deals with longitudinal data in econometrics, the general rule is that n, the number of individuals is large (a few hundreds or thousands) while the number of time periods T is small (less than ten). This implies that for an good panel estimator to be consistent, we would like to have the asymptotics in n at T fixed. However, for the present estimator to be consistent, T needs to tend to infinity which is a major drawback.

Moreover, the model (4) is a regular Fixed Effect model that assumes that fixed effects are constant across quantiles for a given individual. In other words, it means that the unobserved heterogeneity plays the same role regardless of the rank of the individual in the distribution of the outcome. It is definitely a restrictive assumption, but cannot be avoided unless the dataset contains a fair number of time periods.

Finally, proofs of the two theorems are relatively hard to follow and only heuristic for the second theorem. Since Theorem 2 also relies in asymptotics in T, the usefulness of the penalty terms is not fully acknowledged. This leaves the reader to think that the use of the penalized quantile FE estimator is not fully justified because when T is large one can directly estimate the fixed effects using Theorem 1. Hence the use of the penalty term appears to be a tool to alleviate the computational burden, but its advantage in terms of variance reduction is not fully highlighted in the theoretical part. Nevertheless, the Monte Carlo experiment shows that it performs well in terms of Root Mean Squared Error (RMSE) while introducing a small bias.

4.2 Choice of tuning parameters

The proposed estimator depends on two types of tuning parameters: (i) the weights $w_1, ..., w_q$ that control the influence of quantiles $\tau_1, ..., \tau_q$ in the estimation of the parameters α_i, (ii) the overall penalty level λ. However, the author is rather silent about the optimal choice for them.

The weights - The fixed effects are assumed to be invariant across quantiles, making the resolution of program (6) non-separable in each quantile. Indeed, if the fixed effects were to depend on the quantile, we could estimate one quantile at the time and obtain the same result. However, in the present setting, the quantile has a relative influence on the estimated α_i and it is controlled by some weights. Koenker proposes to consider the α_i as weighted L-statistics and cites Mosteller (1946) as a reference for the choice of the weights. However, during the proof of the two theorems, no reference to an optimal choice of the weights is made, while they do appear in the asymptotic variance of the Quantile FE estimator.

The overall penalty level - As in all penalized regression problems, the choice of the overall penalty level is in question and rarely receives a definitive answer. One reference to the regularization parameter is made during the proof of Theorem 2 which is $\lambda_T/\sqrt{T} \to \lambda_0$. It means that the penalty level must increase as the number of time periods increases which appears to be rather contradictory: when the penalty level increases, it shrinks more the fixed effects towards zero. On the other hand, when a growing number of periods is available, we estimate the fixed effects with better accuracy because we observe the behaviour of individual i for a longer period of time, so the need to shrink them is less stringent. Also, the paper provides absolutely no comment on what the λ_0 might be. We think that it is particularly problematic because the applied econometrician cannot necessarily use intuition to choose a good penalty level in this context. Indeed, when running a Lasso, the regularization pathSee the `lars` package in the software `R` or (Hastie *et al.*, 2009, p. 65) for a graphical example. for β provides a good intuition of the relative relevance of each regressor and they do have an economic sense. Hence, the expertise of the econometrician can help him decide the right amount of shrinkage. Here, we are penalizing the fixed effects which represents nothing but individual i's unobserved heterogeneity. As it is unobserved, the econometrician cannot have a good prior knowledge regarding how many individuals are to be shifted from the common intercept. Consequently, the choice of the overall penalty level is a predicament as the author acknowledges in the conclusion.

In a more recent contribution, Lamarche (2010) looks at the same estimator and proposes setting λ in order to minimize the asymptotic variance of the estimator. Monte Carlo simulations offer evidence that it reduces significantly the variability of the Quantile FE estimator, while not introducing too much bias.

4.3 Approximate sparsity in fixed effects

The estimator proposed by Koenker (2004) is subject to criticism when it comes to its implications in terms of the underlying behaviour of the agent it assumes. Indeed, the implicit assumption required for the use of the ℓ_1-penalty in the estimator is that the fixed effect vector α is *approximately sparse*.

The assumption about the sparsity of α appears to be slightly odd specified in this way, even though we understand that it has been introduced to alleviate the computional burden and give more estimation accuracy. Indeed, the economic thinking that underpins this assumption is that a vast majority of individuals belong to the main category for which $\alpha_i = 0$ and only a few outliers, for which we cannot tell why they are outliers, exist. This assumption does not necessarily seem unrealistic, but seems at least not general enough. In the context of high-dimensional linear panel models, Belloni *et al.* (2014) refuse to make this assumption and consider the Within transform of the model where the unobserved individual-specific heterogeneity vanishes. We understand that such a simple transformation is not necessarily possible in the quantile regression setting without making stronger assumptions. However, a related but more general approach could be to assume another type of approximate sparisty which would be grouped unobserved heterogeneity. It means that α_i could take a small finite number K of values defined by the econometrician in order to reflect appartenance to a group of individuals with similar unobserved characteristics. It is a setting which would be more suitable for analysis and interpretability, and would also have the advantage of encompassing the Koenker (2004) model. Such a setting has been investigated by Bonhomme and Manresa (2012) in a small-dimensional linear case. Their "grouped fixed effects" estimator has the appealing property that individual group membership is data-driven and allows for nice interpretation of the division into groups.

4.4 The extension of Canay (2011)

To deal with quantile regression in the panel data setting, Canay (2011) proposes an estimator that does not suffer from some of the drawbacks we have highlighted here. Namely, its estimator (i) does not assume approximate sparsity in the fixed effect vector, (ii) is not a penalized estimator, so does not require to choose a λ. However, assumptions regarding the overall form of the model are the same since the fixed effects are assumed to be location-shifters only and is consistent only as both n and T tends to infinity which limits the scope of application for this estimator in panel data in the same way that Koenker (2004) is limited.

The idea is to purge the equation from the fixed effects with a two-step approach. Firstly, estimate the fixed effects as in a usual linear fixed effects regression model using a Within transform. The estimator of the fixed effects is $\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it} - x_{it}'\hat{\beta}$. In a second step, run a usual quantile regression on the outcome purged of the fixed effects: $y_{it} - \hat{\alpha}_i$. The intuition behind this result is that as $T \to +\infty$, the fixed effects are consistently estimated. When T is large, it is almost as if we had the true α_i, so we can easily get rid of them to perform a usual quantile regression.

V An application to wage data

5.1 Presentation of the dataset from Vella and Verbeek (1998)

Data come from the National Longitudinal Survey (Youth Sample) and comprise a sample of 545 full-time working males ($n = 545$) who have completed their schooling by 1980. They are followed over the period 1980 to 1987 such as $T = 8$. In order to obtain a balanced design, the authors have chosen to exclude those who drop out. Vella and Verbeek (1998) are particularly interested in estimating the *union wage premium*, i.e. the additional wage that a given individual gains from participating in a union. Robinson (1989) reports studies that evidence that union workers average 10-30% higher pay than non-union in the United States after controlling for individual, job and labor market characteristics. However, measuring the union wage premium is not straightforward. Indeed, there are some evidence that the union status is endogenous in the wage equation since there exist unobserved factors that influence both union membership status and wages such as unobserved work skills and social skills.

Union membership is based on whether or not an individual had his wage set in a collective bargaining agreement. In the sample about 55% of the individuals did not change union status: 48.6% never had a wage set by collective bargaining and 6.2% always had. Overall description of other covariates can be found in Table 1 of Vella and Verbeek (1998).

5.2 Estimation results

We run the estimation strategy based on equation (7) described in Koenker (2004) on this data-set. The dependant variable is the log hourly-wage and we give particular attention to the union wage premium, *i.e.* the additional wage that a given individual gain from participating in a union. We set arbitrarily the penalty parameter to one ($\lambda = 1$) and we consider the quartiles $\tau = (0.25, 0.5, 0.75)$. As suggested in the reference article we set weights that favor estimation of the median $w = (0.25, 0.5, 0.25)$. Estimation results are reported in Table 1. Several coefficients are statistically significant. The number of years of schooling have a positive impact and is relatively stable between quartiles: one more year of schooling provides an increase of 10% of the log-hourly wage. The level of experience also has a positive impact, but its effect decrease from 15% to 9% depending on the quartile we consider. This suggests that experience is better rewarded for lower wages. The impact of experience squared is slightly negative, i.e. experience is related to wages through an inverse u-shape curve in all three quartiles. Being black has a large negative impact on the wage while being hispanic has a positive impact that ranges from 7% in the lower quartile to 4.5% in the highest one. Union wage premium is estimated at 6.5% at the first quartile, 6.3% at the second one, and 5.9% at the third one. These values are under the 15% labor economists usually agree on. Figure 5 displays the distribution of estimated fixed effects. The l_1 penalty parametered by $\lambda = 1$ constraints them to 0 for approximately 80 individuals.

Table 2 reports estimation results for simpler specifications. The first column corresponds to the standard within estimator (thus ignoring the possibility for heterogeneous effects within the population), while the second one fits a standard median-quantile specification (thus ignoring the panel structure of the data-set). The coefficient associated to union is positive and highly significant in all cases but the estimated value varies depending on the model (under 7% in panel-quantile regression, 7.5% in the within regression, 14% in the quantile regression). This suggests

Table 1: Koenker Panel Data / Quantile Regression estimates - $\lambda = 1$

	$(\tau_1, w_1) = (0.25, 0.25)$	$(\tau_2, w_2) = (0.5, 0.5)$	$(\tau_3, w_3) = (0.75, 0.25)$
(Intercept)	0.0052	0.3192^{***}	0.5746^{***}
	(0.0808)	(0.0745)	(0.0822)
Black	-0.1225^{***}	-0.1205^{***}	-0.1230^{***}
	(0.0262)	(0.0223)	(0.0253)
Experience	0.1511^{***}	0.1069^{***}	0.0897^{***}
	(0.0150)	(0.0099)	(0.0099)
Experience squared	-0.0066^{***}	-0.0039^{***}	-0.0030^{***}
	(0.0010)	(0.0007)	(0.0006)
Union	0.0650^{***}	0.0627^{***}	0.0589^{**}
	(0.0172)	(0.0170)	(0.0195)
Hispanic	0.0728^{***}	0.0687^{***}	0.0452^{*}
	(0.0210)	(0.0188)	(0.0204)
Health disabilities	0.0214	0.0354	0.0128
	(0.0697)	(0.0248)	(0.0431)
Married	0.0539^{***}	$0.0291^{\cdot}$	$0.0285^{\cdot}$
	(0.0162)	(0.0150)	(0.0152)
Northern Central	-0.0267	-0.0225	-0.0250
	(0.0339)	(0.0321)	(0.0333)
Rural area	-0.0312	-0.0026	-0.0069
	(0.0248)	(0.0222)	(0.0223)
Years of schooling	0.1047^{***}	0.1035^{***}	0.0999^{***}
	(0.0050)	(0.0046)	(0.0045)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Data: National Longitudinal Survey (Youth Sample) for the period 1980-1987.

ignoring individual heterogeneity conducts to an over-estimation of the union wage premium. A limitation of our result comes from the fact that we only estimated individual effects without taking into account the possibility that they are time varying. Furthermore some bias might subsist from simultaneity effects, e.g. if some workers engage in a union to get a higher wage.

5.3 Sensitivity to model parameters

Koenker (2004) provides a Monte-Carlo experiment of the penalized quantile regression and illustrates the respective shrinkage effects of l_1 and l_2 penalties. The l_2 penalty tends to smooth fixed effects while the l_1 criterion is more tolerant of large discrepancies. For sufficiently large values of λ, the l_1 penalty restricts the individual effects to be exactly at 0. Comparison with simple least squares and quantile estimators is conducted in finite samples for different data generating processes where the authors set a reasonable amount of "within" variance in the covariates. Compared to least squares, within estimators, and quantile regressions, the penalized quantile estimator is competitive in a pure location-shift model, and significantly improves the quality of estimation in models with location and scale shifts.

In order to further investigate the sensitivity of Koenker's procedure to parameters we es-

Table 2: Standard Within and Quantile Estimators

	Within	Quantile ($\tau = 0.5$)
(Intercept)	0.0000^{***}	0.6521^{***}
	(0.0000)	(0.0656)
Experience	0.1304^{***}	0.0669^{***}
	(0.0087)	(0.0094)
Health disabilities	0.0010	-0.0044
	(0.0465)	(0.0650)
Married	0.0402^{*}	0.1047^{***}
	(0.0180)	(0.0143)
Northern Central	-0.1049^{*}	0.0020
	(0.0473)	(0.0176)
Rural area	0.0571^{*}	-0.1153^{***}
	(0.0286)	(0.0185)
Union	0.0759^{***}	0.1422^{***}
	(0.0192)	(0.0158)
Experience Squared	-0.0050^{***}	-0.0017^{*}
	(0.0006)	(0.0007)
Black		-0.1236^{***}
		(0.0238)
Hispanic		0.0214
		(0.0175)
Years of schooling		0.0834^{***}
		(0.0044)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

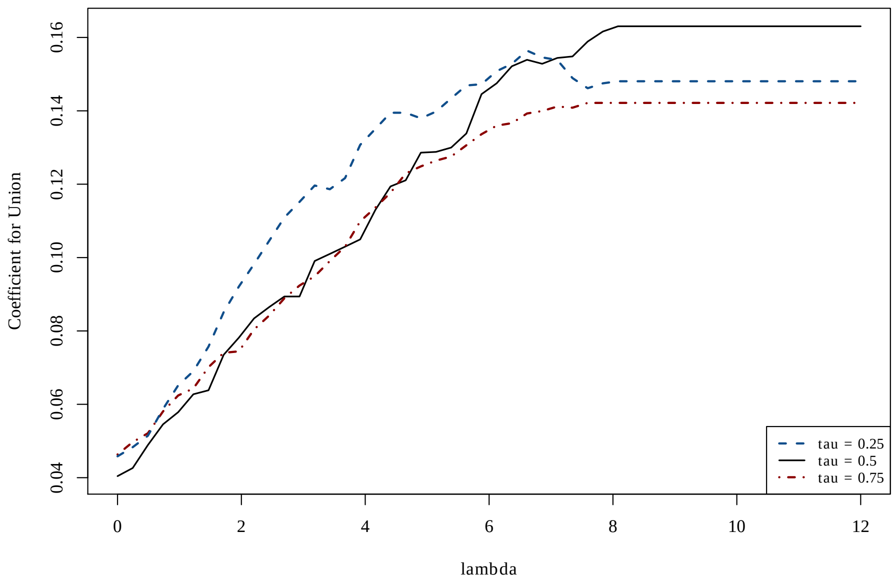
Data: National Longitudinal Survey (Youth Sample) for the period 1980-1987.

timated model (7) for several values of the penalty λ and for several values of the weights $w = (w_1, w_2, w_3)$. Our results show that model parameters have a sizable impact on estimates.

Sensitivity to the penalty - Figure 1 displays the evolution of the coefficient associated to covariate "Union" with respect to the penalty level λ. A larger λ means less individual heterogeneity since many individual effects are set to 0 by the l_1 criteria. Less heterogeneity conducts to larger coefficients for Union. The amplitude of the variation (union wage-premium jumps from 4% to 16% depending on λ) signals that controlling for individual characteristics matters a lot. Furthermore, the penalty level has great influence on quartile comparisons. For $\lambda < 6$, the union wage-premium is estimated to be larger in the lower quartile, while it is roughly the same for the median and upper quartiles. Surprisingly, for $\lambda > 6$, the premium is estimated to be significantly larger for the median quartile than for the upper and lower ones.

Sensitivity to the weight distribution - Figure 2 displays the evolution of the union wage-premium for each quartile with respect to its own weight. A higher weight corresponds to larger coefficients for the lower and upper quartiles, but to a lower coefficient in the case of the median. Figure 3 shows the co-evolution of "Union" coefficients together with the weights $w = (w_1, w_2, w_3)$. We have set $w_1 = w_3 = \frac{1}{2}(1 - w_2)$ so as to keep weights summing to one, and we computed estimates for varying values of the median weight w_2. A higher weight on

Figure 1: Sensibility of the coefficient for Union to the penalty parameter λ



the median leads to smaller union wage-premiums for all quartiles. These results suggest that the median contributes significantly to the estimation of individual heterogeneity. Once again depending on the choice we make for w, the relative effect of "Union" will not be the same across quartiles. However variations of the estimates due to the weights are rather limited compared to the effect of the penalty (coefficients varies between 5% and 7 %).

Figure 2: Sensibility of Union coefficients to varying weights - $\lambda = 1$

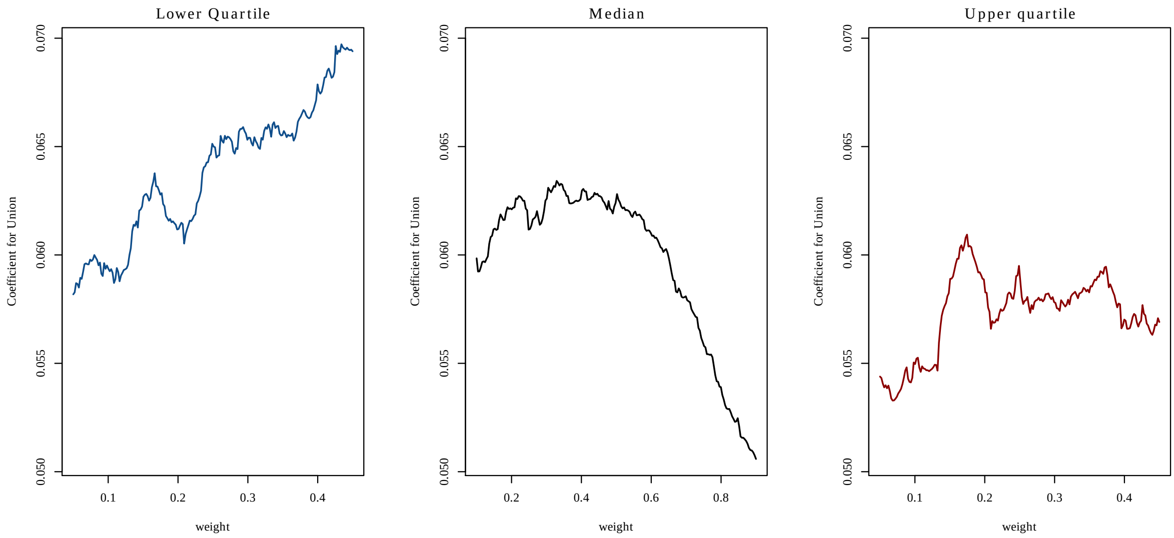
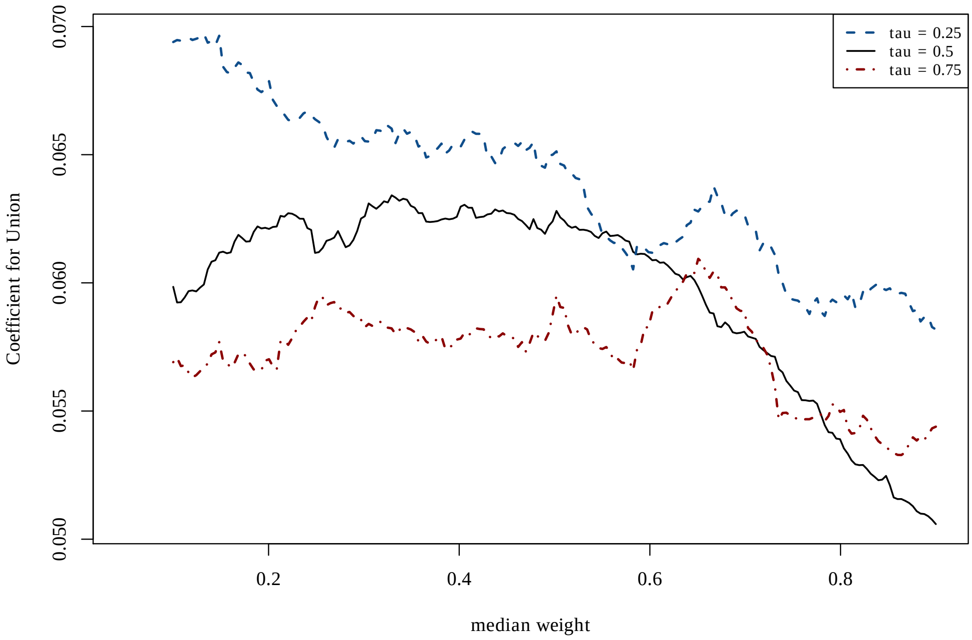


Figure 3: Sensibility of Union coefficients to varying weights - $\lambda = 1$



VI Conclusion

We have presented the quantile fixed effect estimator of Koenker (2004) that uses a penalized approach to reduce the dimensionality of the problem. We have seen that the penalized estimator appears natural in a panel data context and that it achieves a good performance in terms of variance reduction without introducing too much bias. However, through both the theoretical proof and the application to wage data, we have seen that several problems remain. Since we are dealing with panel data, the most prominent one is that the asymptotics is when both n and T tend to which almost never happens in economics. Moreover, the choice of the tuning parameters is not covered and appears to be difficult in practice. These two issues are limiting the scope of application for this estimator, leaving the applied econometrician unsatisfied when estimating quantile-dependent parameters in a panel data context.

VII Bibliography

Belloni, A., Chernozhukov, V., Hansen, C., and Kozbur, D. (2014): "Inference in High Dimensional Panel Models with an Application to Gun Control". *ArXiv e-prints*.

Bonhomme, S. and Manresa, E. (2012): "Grouped Patterns Of Heterogeneity In Panel Data". Working Papers wp2012 1208, CEMFI.

Canay, I. A. (2011): "A simple approach to quantile regression for panel data". *The Econometrics Journal*, 14(3):368–386.

Hastie, T., Tibshirani, R., and Friedman, J. (2009): *The Elements of Statistical Learning: Data mining, Inference and Prediction*. Springer, 2 edition.

Knight, K. (1998): "Limiting distributions for L_1 regression estimators under general conditions". *Ann. Statist.*, 26(2):755–770.

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

Lamarche, C. (2010): "Robust penalized quantile regression estimation for panel data". *Journal of Econometrics*, 157(2):396–408.

Mosteller, F. (1946): "On Some Useful Inefficient Statistics". *Ann. Math. Statist.*, 17(4):377–408.

Ohinata, A. and van Ours, J. (2013): "Spillover effects of studying with immigrant students; a quantile regression approach". Technical report.

Robinson, C. (1989): "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models". *Journal of Political Economy*, 97(3):639–67.

Tibshirani, R. (1994): "Regression Shrinkage and Selection Via the Lasso". *Journal of the Royal Statistical Society, Series B*, 58:267–288.

Vella, F. and Verbeek, M. (1998): "Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men". *Journal of Applied Econometrics*, 13(2):163–183.

A Proof of Knight identity

Define the check function $\rho_\tau(x) = (\tau - \mathbb{1}(x < 0))x$ and the following function $\psi_\tau(x) = \tau - \mathbb{1}(x < 0)$.

Lemma: *Knight (1998) identity*

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v (\mathbb{1}(u \le s) - \mathbb{1}(u \le 0))\, \mathrm{d}\, s \tag{15}$$

Proof: To prove this lemma, we introduce Φ_r which is equal to the right term of the equality:

$$\Phi_r(u, v) = -v(\tau - \mathbb{1}(x < 0)) + \int_0^v (\mathbb{1}(u \le s) - \mathbb{1}(u \le 0))\, \mathrm{d}\, s$$

Φ_l is equal to the left term. We then consider four cases:

				$\mathbb{1}(u \le v)$	$\mathbb{1}(u \le 0)$
[1]	$u \le v$	&	$u \le 0$	1	1
[2]	$u \le v$	&	$u > 0$	1	0
[3]	$u > v$	&	$u \le 0$	0	1
[4]	$u > v$	&	$u > 0$	0	0

The left term can be rewritten:

$$\begin{array}{llcll}
[1] & \Phi_l(u, v) & = & [(u - v) - u](\tau - 1) & = -v(\tau - 1) \\
[2] & \Phi_l(u, v) & = & (u - v)(\tau - 1) - u\tau & = -u - v(\tau - 1) \\
[3] & \Phi_l(u, v) & = & (u - v)\tau - u(\tau - 1) & = u - v\tau \\
[4] & \Phi_l(u, v) & = & [(u - v) - u]\tau & = -v\tau
\end{array}$$

The right term can be rewritten:

$$\begin{array}{llcll}
[1] & \Phi_r(u, v) & = & -v(\tau - 1) + \int_0^v (1 - 1)\, \mathrm{d}\, s & = -v(\tau - 1) \\
[2] & \Phi_r(u, v) & = & -v\tau + \int_u^v (1 - 0)\, \mathrm{d}\, s & = -u - (\tau - 1)v \\
[3] & \Phi_r(u, v) & = & -v(\tau - 1) + \int_u^v (0 - 1)\, \mathrm{d}\, s & = u - \tau v \\
[4] & \Phi_r(u, v) & = & -v\tau + \int_0^v (0 - 0)\, \mathrm{d}\, s & = -v\tau
\end{array}$$

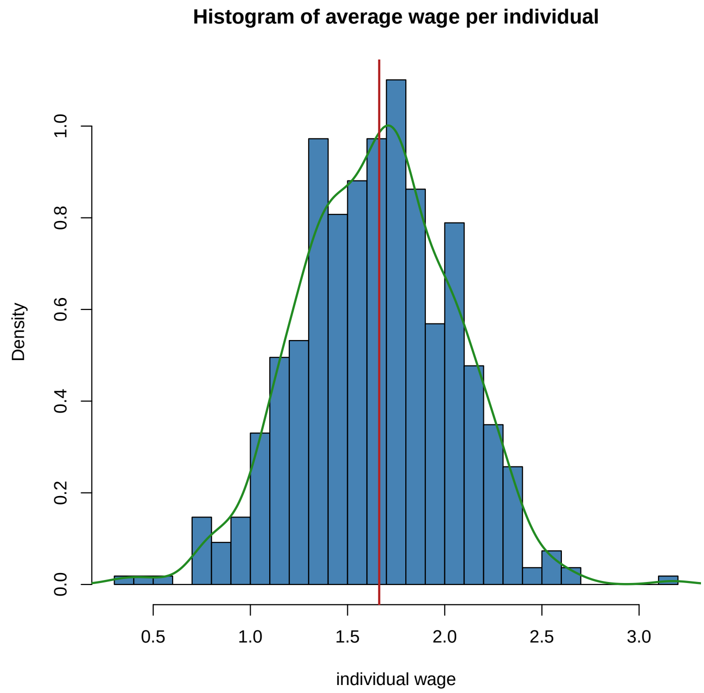
It allows us to conclude that:

$$\forall (u, v) \quad \in \quad \mathbb{R}^2, \Phi_l(u, v) = \Phi_r(u, v)$$

■

B Data description

Figure 4: Log hourly wage evolution by sector for non-union members



Data: National Longitudinal Survey (Youth Sample) for the period 1980-1987.
Note: This graph shows the distribution of log hourly wages on the period 1980-1987.

C Distribution of estimated individual effects

Figure 5: Distribution of estimated individual effects - $\lambda = 1$