DOI 10.1515/2156-6674.1001 — Journal of Econometric Methods 2012; 1(1): 2–14 DE GRUYTER

Research Article

Tatiana Komarova, Thomas A. Severini and Elie T. Tamer

Quantile Uncorrelation and Instrumental Regressions

Abstract: We introduce a notion of *median uncorrelation* that is a natural extension of mean (linear) uncorrelation. A scalar random variable Y is median uncorrelated with a k-dimensional random vector X if and only if the slope from an LAD regression of Y on X is zero. Using this simple definition, we characterize properties of median uncorrelated random variables, and introduce a notion of multivariate median uncorrelation. We provide measures of median uncorrelation that are similar to the linear correlation coefficient and the coefficient of determination. We also extend this median uncorrelation to other loss functions. As two stage least squares exploits mean uncorrelation between an instrument vector and the error to derive consistent estimators for parameters in linear regressions with endogenous regressors, the main result of this paper shows how a median uncorrelation assumption between an instrument vector and the error can similarly be used to derive consistent estimators in these linear models with endogenous regressors. We also show how median uncorrelation can be used in linear panel models with quantile restrictions and in linear models with measurement errors.

Keywords: quantile regression, endogeneity, instrumental variables, correlation.

Author Notes: We thank the editor and two anonymous referees for providing us with constructive comments and suggestions. We also appreciate feedback from seminar participants at the London School of Economics, University College London, the University of Toronto and Queen Mary University of London, and feedback from the participants of the Canadian Econometric Study Group and all UC Econometrics Conference.

Tatiana Komarova: London School of Economics and Political Science, E-mail: t.komarova@lse.ac.uk
Thomas A. Severini: Northwestern University, E-mail: severini@northwestern.edu
Elie T. Tamer: Northwestern University, E-mail: tamer@northwestern.edu

1 Introduction

We introduce a concept of quantile uncorrelation, or $\mathscr{L}_1$-uncorrelation, between two random variables that is a natural extension of the well-known mean uncorrelation, or $\mathscr{L}_2$-uncorrelation. We term this type of uncorrelation, "*median uncorrelation*," which is the counterpart of the familiar mean (linear) uncorrelation, or simply uncorrelation. We characterize the relationship between random variables that are uncorrelated in this manner. We provide a series of properties that imply or are implied by median uncorrelation. Naturally, for example, independence of two random variables implies median uncorrelation (or in this case $\mathscr{L}_p$-uncorrelation for any $p \geq 1$). Also, this uncorrelation is not symmetric, and is nonadditive, but it retains an important invariance property.

We extend our definition to median uncorrelation between random *vectors* which results, indirectly, in a multivariate version of a quantile restriction. We also derive an asymmetric *correlation measure*, based on this notion of quantile uncorrelation, that takes values in [–1, 1] with a value of zero for uncorrelation. In addition, we provide another correlation measure that is the analog of the coefficient of determination, or R^2, in linear regressions. We also extend this concept to cover $\mathscr{L}_p$-uncorrelation for $p \geq 1$.

As two stage least squares is based on exploiting linear uncorrelation between the error and an excluded random variable (the instrument), we also show that this uncorrelation leads naturally, and under easily interpretable conditions, to "instrumental" regressions with median uncorrelation. These are analogs of Basmann and Theil's two stage least squares, or *2SLS*, (Theil (1953) and Basmann (1960)) as derived from the usual mean uncorrelation between two random variables. As in the classical 2SLS, median uncorrelation leads to an estimator that is derived by taking a "sample analogue" of the median uncorrelation measure. This estimator, similar to one used by Chernozhukov and Hansen (2006) (or CH), is consistent provided that this uncorrelation holds (along with other standard assumptions). Other applications are natural counterparts of existing least squares methods. For example, by exploiting this uncorrelation further, we show that as instrumental variable methods can be used in mean-based models to remedy the problem of classical measurement error, variables obeying our median uncorrelation condition can be used as instruments to obtain estimates of parameters in linear models with measurement error under quantile restrictions. Furthermore, panel data quantile regression of differenced data delivers consistent estimates of parameters of interest without making assumptions on the individual effects under median uncorrelation restrictions. So, this uncorrelation gives support for running standard quantile regression of first differenced outcomes on first differenced regressors, under an absolute loss function to obtain consistent estimates of the slope parameters in linear models.

An important feature of the concept of median uncorrelatedness is the fact that it is defined in terms of the linear predictor, and hence is explicitly a "linear concept". Basically, it shares this property with best linear predictors in that, heuristically, a random variable is median uncorrelated with another if the latter is not "useful" as a *linear predictor* of the former under absolute loss. Finally, this notion of median uncorrelation is general and is loss function based.

There is a large literature in econometrics on best predictor problems. Manski (1988) delineates estimators derived from prediction problems from various loss functions. There, best linear predictors are derived and consistent estimators are provided that are based on the analogy principle. The linear model based on quantile restrictions is equally well studied starting with the work of Koenker and Bassett (1978); see also Koenker (2005). There has also been a series of papers dealing with the presence of endogenous regressors in models with quantile restrictions. Amemiya (1981) proposed a two-staged least absolute deviation estimator. See also Powell (1983). Then, based on method of moments, Honoré and Hu (2004) provide methods that can be used to do inference on parameters defined though separable moment models (that can be nonlinear). CH (see also Chernozhukov and Hansen (2005)) in a series of papers shed new light on a general class of monotonic models with conditional quantile restrictions. They provide sufficient point identification conditions for these models, and also an estimator that they show is consistent under those conditions. CH study also the asymptotic properties of their estimator and characterize its large sample distribution. The estimator based on our median uncorrelation assumption is the same as the one used in CH. Finally, Sakata (2007) and Sakata (2001) in interesting work, provide estimators based on an $\mathscr{L}_1$ loss function for instrumental regression models[1]. Both these papers use a condition that is closer to conditional median independence, but the approach in spirit is similar to ours.

In Section 2, we provide first a few elementary definitions that lead to median uncorrelation. After defining median uncorrelation, Section 3 characterizes this uncorrelation concept in terms of various properties of the joint distribution of random variables. Section 4 shows how median uncorrelation leads to natural estimators in linear models with endogenous regressors. Section 5 provides notions of median correlation among random variables. We provide in Section 6 simple applications of our median uncorrelation concept to linear quantile regression with measurement error and to panel data quantile regression. Section 7 concludes.

2 Definition and Properties

Let T be a scalar random variable and let S be a k-dimensional random vector such that $E||S||<\infty$. We are interested in the following optimization problem since it is key in defining our concept of median uncorrelation:

$$\min_{(\alpha,\beta)} E|T-\alpha-S'\beta|.$$

where we assume[2] that $E|T|<\infty$. This is done for simplicity of notation. Define $M(T,S)\subset\Re^k$ as the set of solutions to this optimization problem with respect to β:

$$M(T,S)\equiv\left\{\beta:\exists\alpha \text{ such that } (\alpha,\beta)=\arg\min_{(\tilde{\alpha},\tilde{\beta})} E|T-\tilde{\alpha}-S'\tilde{\beta}|\right\}.$$

In general, one can find distributions in which $M(T,S)$ is a set. However, under weak conditions, $M(T,S)$ is a singleton; see part 3 of Proposition 2.1 below. Notice that for a fixed β,

$$E|T-S'\beta-Med(T-S'\beta)|=\min_{\alpha} E|T-\alpha-S'\beta|,$$

where

$$Med(z)\equiv\inf\{t:P(z\le t)\ge 0.5\}.$$

Therefore,

$$M(T,S)=\arg\min_{\beta} E|T-S'\beta-Med(T-S'\beta)|.$$

1 For other approaches to estimation in quantile regression with endogeneity, see Ma and Koenker (2006), Lee (2007), and Chesher (2003).
2 Without this assumption, we can rewrite the objective function as $\min_{(\alpha,\beta)}\{E|T-\alpha-S'\beta|-E|T-\alpha_0-S'\beta_0|\}$ for some fixed (α_0,β_0).

The next proposition characterizes elements of the set $M(T,S)$ and also gives conditions under which $M(T,S)$ is a singleton. We collect Proofs to results in the Appendix.

Proposition 2.1 *The following hold:*

1. *Let* $\beta^* \in \Re^k$ *Then* $\beta^* \in M(T,S)$ *if and only if for any* $\alpha \in \Re$, $\beta \in \Re^k$,

$$\left|E\left[(\alpha+S'\beta)sgn\left(T-S'\beta^*-Med\left(T-S'\beta^*\right)\right)\right]\right|$$
$$\le E\left[\left|\alpha+S'\beta\right|1\left(T-S'\beta^*-Med\left(T-S'\beta^*\right)=0\right)\right] \quad (2.1)$$

where here and in the rest of the paper we define $sgn(\cdot)$ as

$$sgn(x)=\begin{cases}1, & x>0\\ 0, & x=0\\ -1, & x<0\end{cases}$$

2. *Let* $\beta^* \in \Re^k$ be *such that* $P(T-S'\beta^*-Med(T-S'\beta^*)=0)=0$. *Then* $\beta^* \in M(T,S)$ *if and only if*

$$E[S\, sgn(T-S'\beta^*-Med(T-S'\beta^*))]=0. \quad (2.2)$$

3. *Suppose that any* $\beta \in \Re^k$ *satisfies* $P(T-S'\beta-Med(T-S'\beta)=0)=0$. *Then* $M(T,S)$ *is a singleton if and only if equation*

$$E[S\, sgn(T-S'\beta-Med(T-S'\beta))]=0$$

has a unique solution. This solution is $M(T,S)$.

We use Equation (2.2) as the basis for a measure of median correlation introduced in Section 5.

The next definition introduces the notion of median uncorrelation of a random vector with another random vector. Here, and in the remainder of the paper, we take $M(T,S)=\beta^*$ to mean that $M(T,S)$ contains the single value β^*.

Definition 2.1 (Median Uncorrelation) *Let W denote an l-dimensional random vector. We will say that W is* ***median uncorrelated with*** *S if*

$$M(c'W,S)=0 \qquad \text{for all} \quad c\in\Re^l. \quad (2.3)$$

The definition above is loss function based. So, it naturally carries over to quantiles other than the median, by simply changing the absolute loss to asymmetric loss by using the "check function." Moreover, implicit in this definition, is a formulation for multivariate quantiles. In particular, when defining this uncorrelation property meant for scalar quantiles to the multivariate case, we require that median uncorrelation holds for *any* linear combination of the elements of the multivariate vector, as in (2.3). Finally, a key property that this "loss" function maintains is the *invariance property* below.

Lemma 2.1 (Invariance) *For any constant vector* $b\in\Re^k$ *and any constant scalar* a,

$$M(T+a+S'b,S)=M(T,S)+b. \quad (2.4)$$

This property plays a key role below. Linearity of $T-\alpha-S'\beta$ in the objective function is essential for this invariance property to hold. The concept of uncorrelation we introduced is intimately tied to linear models and is similar to the relationship between uncorrelation in the least squares setup and its relationship to linear models. Median uncorrelation is median linear uncorrelation.

3 Characterizations of Median Uncorrelation

In this section, we provide key insights that explore further the meaning of median uncorrelation in Definition 2.1 above. The following characterization theorem collects a set of properties that are helpful in gaining intuition about median uncorrelation.

Theorem 3.1 (Properties of Median Uncorrelation) *The following hold:*

A. *A sufficient condition for an l-dimensional random vector W to be median uncorrelated with a random vector S is that* $Med(c'W|s)=Med(c'W)$ *for all* $c\in\Re^l$.

B. *If W is median uncorrelated with S, it does not necessarily follow that S is median uncorrelated with W.*

C. *A sufficient condition for W to be median uncorrelated with S is that the conditional characteristic function of W given S is real.*

D. *Consider a scalar random variable T and any random vector S. Assume that M(T,S) is a singleton. Then T can be written as*

$$T=\alpha_0+S'M(T,S)+\delta,$$

where $M(\delta,S)=0$, *and* α_0 *is any constant.*

E. *For a scalar random variable T and random vectors S and Z in* $\Re^k$, *assume that* $P(T-Med(T)=0)=0$ *and* $M(T,S+Z)$ *is a singleton. Then*

$$M(T,S)=M(T,Z)=0 \Rightarrow M(T,S+Z)=0.$$

F. *Suppose that for a scalar random variable T and a non-degenerate binary random variable S the median of* $T|S=1$ *and the median of* $T|S=0$ *are unique. The following hold:*

$$M(T,S)=0 \Leftarrow Med(T|S=1)=Med(T|S=0);$$

if $P(T - Med(T) = 0) = 0$, *then*

$M(T, S) = 0 \Rightarrow Med(T|S=1) = Med(T|S=0)$.

Property (A) can be directly derived from the definition and states median independence as a sufficient condition for median uncorrelation. (B) means that the definition of median uncorrelation is not symmetric. This is in direct contrast with mean uncorrelation which is a symmetric property. Property (D) is important and it states that any scalar random variable T can be decomposed into a linear combination of S's and another random variable that is median uncorrelated with S. This is a direct result of the invariance property in (2.4) above. Moreover, this is similar to the linear *mean* decomposition in best linear prediction examples. See (3.1) below. Property (E) illustrates an additivity property of median uncorrelation: If T is median uncorrelated with S and Z, then it is median uncorrelated with their sum $S+Z$. Property (F) states that under weak restrictions, T is median uncorrelated with a binary variable S if and only if T is median independent of S.

Evidently, if W is median uncorrelated with S, then S is not useful in the $\mathscr{L}_1$ prediction of linear functions of W.

3.1 Comparison to mean uncorrelation

It is helpful to compare the median uncorrelation with the well-known mean uncorrelation.

Consider the optimization problem

$$\min_{(\alpha,\beta)} E\left(T-\alpha-S'\beta\right)^2,$$

where $ET^2 < \infty$, $E||S||^2 < \infty$. Under the usual rank condition on S, this problem has a unique solution. Denote its solution with respect to β as $L(T, S)$. This is the $\mathscr{L}_2$ analogue of $M(T, S)$.

It is easy to show that, for scalar S, for example, $L(T, S) = Cov(T, S)Var(S)^{-1}$. In addition, W with the values in $\Re^l$ and S are (mean) uncorrelated if, for any $c \in \Re^l$, $L(c'W, S) = 0$ since

$$L(c'W, S) = Var(S)^{-1}Cov(S, W)c.$$

Properties in Theorem 3.1 have the following $\mathscr{L}_2$ versions.

$\mathscr{L}_2$ **Properties**. The following hold:

A. A sufficient condition for an l-dimensional random vector W to be (mean) uncorrelated with a k-dimensional random vector S is that $E(c'W|S) = E(c'W)$ for all $c \in \Re^l$. This holds, in particular, if W is mean independent of S.

B. If W is uncorrelated with S, then S is uncorrelated with W.

C. A sufficient condition for W to be uncorrelated with S is that the conditional characteristic function of W given S is real.

D. For a scalar random variable T and a k-dimensional random vector S, variable T can be represented as follows:

$$T = \alpha_0 + S'L(T, S) + \delta^*, \tag{3.1}$$

where $L(\delta^*, S) = 0$ and α_0 is any constant.

Clearly, if W is uncorrelated with S, then S is not useful in the $\mathscr{L}_2$ prediction of linear functions of W.

The main technical differences between median uncorrelation and uncorrelation are that (1) median uncorrelation is not symmetric, (2) if W_1 and W_2 are both uncorrelated with S, then the vector (W_1, W_2) is uncorrelated with S, while the same is not true for median uncorrelation, (3) a condition for W and S to be uncorrelated can be given in terms of W alone (i.e., $Cov(W, S) = 0$) without reference to linear functions and (4) the additivity of $L(W, S)$, i.e., $L(W_1 + W_2, S) = L(W_1, S) + L(W_2, S)$, which often greatly simplifies technical arguments. This latter difference basically means that if W_1 is uncorrelated with S and W_2 is uncorrelated with S, then $W_1 + W_2$ is uncorrelated with S. Two simple results in Proposition 3.1 below compare the median uncorrelation with the usual mean uncorrelation.

Proposition 3.1 *Let T be a scalar random variable and S be a random vector in $\Re^k$.*

1. *If V, a scalar random variable, is independent of S, then*

   $cov(T+V, S) = cov(T, S)$,

   but, in general,

   $M(T+V, S) \neq M(T, S)$.

2. *If V, a random vector in $\Re^k$, is independent of T, then*

   $cov(T, S+V) = cov(T, S)$,

   but, in general,

   $M(T, S+V) \neq M(T, S)$.

4 Median Uncorrelation and Instrumental Regression

This is the main section of the paper in which we exploit the median uncorrelation concept to define estimators for parameters in linear models with endogenous variables. The estimator (and the model) is defined via the uncorrelation assumption in the same way as some versions of 2SLS are defined from the mean uncorrelation.

Consider the following model:

$$Y=\alpha_0+X'\beta_0+\varepsilon, \tag{4.1}$$

where Y and ε are real-valued random variables, X is a k-dimensional random vector with a positive definite covariance matrix, α_0 is an unknown scalar parameter, and β_0 is an unknown slope vector. The parameter of interest is β_0. Assume that ε has median 0, but that

$$Med(\varepsilon|x)\neq 0,$$

where $Med(\cdot|\cdot)$ denotes the conditional median. The problem here is that this conditional median is allowed to depend on X. There are many reasons for this type of "endogeneity" in economic models. Classical work on demand and supply analysis in linear (in parameter) models motivate many early works in linear models with mean restrictions where instrumental variables assumptions were used to eliminate least squares bias that arises from this endogeneity. See Theil (1953), Basmann (1960) and Amemiya (1985) and references therein. There are a set of papers that deal with endogeneity in linear quantile based models. See for example Amemiya (1981) for a 2 stage interpretation of the 2SLS, and Chernozhukov and Hansen (2005) for an approach to inference in quantile based models, both linear and nonlinear, in the presence of endogenous regressors. Finally, also, Sakata (2007) provides a similar approach to ours for estimating models based on $\mathscr{L}_1$ loss which also involves instrumental variables.

Recall that the 2SLS strategy is based on finding an instrument vector Z such that $E[Z\varepsilon]=0$, and using this uncorrelation (moment) condition to derive a consistent estimator for β_0. In this section, we extend this intuition to median uncorrelation whereas we assume the presence of a random vector Z, which we call a vector of instruments, that obeys a median uncorrelation assumption (see Assumption A.1 below). This median uncorrelation property, similarly to its counterpart $E[Z\varepsilon]=0$, leads naturally to a simple estimator for β_0. So, the intuition for obtaining an instrument here, is similar to 2SLS in that one looks for an excluded variable that is median uncorrelated with the outcome, i.e., cannot linearly explain the outcome based on a linear median regression (here the outcome means the outcome after projection on the other regressors). Finally, our approach is closely related also to Sakata (2007) who provides a novel approach to inference in this setup. There, the IV estimator is defined through an implication of a conditional median independence assumption. Below, we state the main assumption here.

Assumption A.1 *Let there be a d-dimensional random vector Z such that:*

1. *There exists a $k\times d$ constant matrix of full rank γ, with $d\geq k$, such that*

   $$X=\gamma Z+\delta$$

   for some random vector δ.
2. *$(\delta,\varepsilon)'$ is median uncorrelated with Z.*

First, we require that the dimension of Z be at least equal to the dimension of X. This is the necessary condition for point identification. The key assumption is part 2 of A.1 where we require that not only ε be median uncorrelated with Z and δ be median uncorrelated with Z, but also that $(\delta,\varepsilon)'=(X-\gamma Z,\varepsilon)'$ be *jointly* median uncorrelated with Z (since the fact that $M(\varepsilon,Z)=0$ and $M(\delta,Z)=0$ does not imply that $(\delta,\varepsilon)'$ is median uncorrelated with Z.)

Given Assumption A.1, we are able to easily prove the following theorem, which constitutes the main result in this section.

Theorem 4.1 (Main Result) *Consider the function*

$$\psi(\beta)=M(Y-X'\beta,Z). \tag{4.2}$$

Let assumption A.1 hold. Then

$$\psi(\beta)=0\Leftrightarrow\beta=\beta_0.$$

Proof: Note that by assumption A.1, we have

$$Y=\alpha_0+Z'\gamma'\beta_0+\delta'\beta_0+\varepsilon.$$

Let

$$m\in M(Y-X'\beta,Z)=M(\alpha_0+Z'\gamma'(\beta_0-\beta)+\delta'(\beta_0-\beta)+\varepsilon,Z).$$

By the invariance property in Lemma 2.1, there exists $m_0\in M(\delta'(\beta_0-\beta)+\varepsilon,Z)$

such that

$$m=\gamma'(\beta_0-\beta)+m_0.$$

Note that $\delta'(\beta_0-\beta)=(\beta_0-\beta)'\delta$. Hence, since $(\delta,\varepsilon)'$ is median uncorrelated with Z, $m_0=0$. It follows that $m=\gamma'(\beta-\beta_0)$ and, hence, that

$$\psi(\beta)=\gamma'(\beta-\beta_0).$$

Since $d\geq k$ and γ is full column rank by assumption A.1, we have

$$\psi(\beta)=0\Leftrightarrow\beta=\beta_0,$$

which proves the theorem. ☐

log Wage	S	IQ	Experience	Tenure	Age
Least Squares	.057(7.4)	.0041(3.5)	.0138(3.11)	.0054(1.9)	.014(2.76)
2SLS	.015(.84)	.017(3.47)	.013(2.88)	.003(1.21)	.02(3.26)
Quantile Reg (.5)	.05(4.75)	.005(2.75)	.008(1.43)	.008(2)	.018(2.4)
MIR	-.000(-.08)	.024(7.18)	.014(2.39)	.0032(.77)	.019(2.32)

Table 1: Returns to Schooling when Controlling for Endogenous Ability.

The theorem can be used as the basis for an estimation method for β_0. Note that in case we use the least squares function $L(\cdot,\cdot)$ instead of $M(\cdot,\cdot)$, we get exactly Basmann's interpretation of the 2SLS estimator of β_0. Moreover, note that the estimator based on the result in Theorem 4.1 is the same as the one used by Chernozhukov and Hansen (2005). Let $\hat{Y}$ denote an $n\times 1$ vector of realizations of Y, let $\hat{X}$ denote an $n\times k$ matrix of realizations of X and let $\hat{Z}$ denote an $n\times d$ matrix of realizations of Z. Define $\hat{M}\left(\hat{Y},\hat{Z}\right)$ to be the vector $c\in\Re^d$ that minimizes

$$\sum_j |\hat{Y}_j - a - \hat{Z}_j' c|$$

when minimizing over (a, c). Then, $\hat{\beta}$ is defined as the solution in b to

$$\hat{M}\left(\hat{Y}-\hat{X}b,\hat{Z}\right)=0.$$

$\hat{\beta}$ can be obtained, as in CH, by minimizing

$$\hat{\beta}=\arg\min_{b\in\Re^k} \|\hat{M}\left(\hat{Y}-\hat{X}b,\hat{Z}\right)\|_A,$$

where $\|\cdot\|_A$ is the weighted by A Euclidian norm.

It is interesting to note that the sufficient condition for identification in CH adapted to the linear model is (in our notation) that for all Z the following has a unique solution at the true parameter β_0:

$$P\left(Y<\alpha_0+X'\beta|Z\right)=E\left[1\left[Y<\alpha_0+X'\beta\right]|Z\right]=\frac{1}{2},$$

while our median uncorrelation condition requires that the moment condition

$$E[Z\,\mathrm{sgn}(Y-X'\beta-Med(Y-X'\beta))]=0 \qquad (4.3)$$

has a unique solution at β_0.

CH's condition above can be written as

$$E[\mathrm{sgn}(Y-X'\beta_0-Med(Y-X'\beta_0))|Z]=0,$$

which obviously implies (4.3) when it is calculated at β_0. Clearly, it is a conditional statement, as opposed to an unconditional statement. But, our approach requires an (unconditional) uncorrelation assumption on the joint distribution of (δ, ε, Z).

We next state the asymptotic distribution without any conditions and refer the reader to Chernozhukov and Hansen (2005) who derived these results for details, and for ways to compute the estimator and its standard errors. Under the conditions in CH, as $n\to\infty$, we have

$$\sqrt{n}\left(\hat{\beta}-\beta\right)\xrightarrow{d}\mathcal{N}\left(0,C^{-1}D\left[C^{-1}\right]'\right),$$

where $C=E[f_\varepsilon(0|X, Z)XZ']$ and $D=\frac{1}{4}E[ZZ']$ and $\varepsilon=y-\alpha_0-X'\beta_0$.

4.1 Relationship to the 2SLS Assumptions

In the usual model with endogeneity we have

$$Y=\alpha_0+X'\beta_0+\varepsilon,$$

$$Cov(\varepsilon, X)\neq 0.$$

Here, a random vector Z is an instrument if $Cov(X, Z)$ and $Cov(Z, Z)$ have full rank and $Cov(Z, \varepsilon)=0$, or $E[Z\varepsilon]=0$ with a mean zero assumption on ε.

Let $\gamma=Cov(X, Z)Cov(Z, Z)^{-1}$ and define $\delta=X-\gamma Z$. Then,

$$X=\gamma Z+\delta.$$

Here $(\delta, \varepsilon)'$ is uncorrelated with Z because δ is uncorrelated with Z by construction and ε is uncorrelated with Z by definition. This is not true in the median case, where we need to impose the joint median uncorrelation condition in part 1 of A.1. This is the key difference between the mean and the median formulations.

4.2 Empirical illustration

We illustrate our approach above by estimating a wage regression similar to Griliches (1976) using an extract from the 1980 NLSY which contains data on wages, schooling and many other variables[3]. We are interested in the

3 For information about this sample, see Blackburn and Neumark (1992).

relationship between schooling and wages allowing for Ability proxied here with IQ to be endogenous. We use the following regression as the benchmark:

$$Ln(Wage)=\alpha+\beta_1 S+\beta_2 IQ+\beta_2 Experience+\beta_4 Tenure +\beta_5 \text{ Age}+\varepsilon,$$

where S is completed years of schooling, IQ is the IQ score and here stands for Ability, Experience is years of experience and Tenure is years of tenure. In this regression, the variable IQ is endogenous, and so, we use KWW, or "knowledge of the world" test, as an instrument for it. Above, Table 1 provides estimates for the parameter vector β_0 using a set of estimators, each imposes various assumptions on the underlying distribution of ε conditional on the regressors and the instruments.

We report least squares and two stage least square results, a median quantile regression results and MIR, which is median instrumental regression results. The Table also presents the t-stat in parentheses. In least squares, all the coefficient are significant and are useful in predicting wages. This story changes somehow when we consider 2SLS: now, it appears that schooling becomes much less important (and the result holds if we use efficient GMM). The median regression results are similar qualitatively to the least squares results. So, the interesting result to note is that the returns to schooling when we control for ability is statistically and economically significant when we do not correct for endogeneity of IQ (either least squares or median regression) and is roughly around 5%. When we control for endogeneity of IQ using KWW as an instrument, schooling becomes both economically and statistically insignificant even when we use MIR, or median uncorrelated regressions. So, the MIR results in particular show that schooling is not useful in linearly predicting wage under absolute loss when we include IQ (and other regressors) and when we allow for endogeneity as defined through the MIR model assumptions.

5 Some Measures of Median Correlation

In the case when two random variables are not median uncorrelated, we would like to be able to measure the degree of their median correlation. Two such measures are presented below. The first generalizes the usual (mean) correlation; the second generalizes the idea of the coefficient of determination.

First, we review the $\mathscr{L}_2$ case. For scalar random variables T and S, introduce the normalized random variables

$$T^* = \frac{T-E(T)}{\sigma_T},$$

$$S^* = \frac{S-E(S)}{\sigma_S}.$$

Correlation between T and S is measured by the correlation coefficient *corr(T, S)*:

$$corr(T, S)=E[|T^*||S^*|\text{sgn}(T^*)\text{sgn}(S^*)].$$

This definition requires T and S to have finite variances.

A second way to measure the linear relationship between two scalar random variables is to consider the extent to which a linear function of one random variable is useful in the prediction of the other; when applied to data, this measure is the coefficient of determination, often denoted by R^2. Thus, let

$$R^2 \equiv rsq(T,S)=1-\frac{\min_{(\alpha,\beta)} E(T-\alpha-\beta S)^2}{E(T-E(T))^2}.$$

It is well-known that $rsq(T, S)=corr(T, S)^2$.

Now, consider the $\mathscr{L}_1$ case; we begin by considering the analogue of *corr*. Suppose that $E|T|<\infty$ and[4] $E|S|<\infty$. Define $\tilde{T}$ and $\tilde{S}$ as

$$\tilde{T}=\frac{T-Med(T)}{E|T-Med(T)|},$$

$$\tilde{S}=\frac{S-Med(S)}{E|S-Med(S)|}.$$

Let *medcorr(T, S)* denote a measure of median correlation between T and S defined as

$$medcorr(T,S)\equiv E\left[|\tilde{S}|\,\text{sgn}(\tilde{T})\text{sgn}(\tilde{S})\right].$$

Note that, in general, *medcorr(T, S)* is different from *M(T, S)*.

The theorem below establishes some important properties of the *medcorr* measure.

Theorem 5.1 *Consider the random variables T and S such that $E|S|<\infty$. The following hold:*

1. $medcorr(T, S)\in[-1, 1]$.

4 We can avoid assuming $E|T|<\infty$ if *medcorr* (T,S) is defined in the following way:

$$medcorr(T, S) = E[|\tilde{S}|\text{sgn}(T-\text{Med}(T))\text{sgn}(\tilde{S})].$$

When $E|T|<\infty$, these two definitions give the same numerical value.

2. *Suppose that $M(T, S)$ is a singleton and $P(T-M(T, S)S-Med(T-M(T, S)S)=0)=0$ and $P(T-Med(T)=0)=0$. Then* $sgn(medcorr(T, S))=sgn(M(T, S))$.

In addition, we can show[5] that *medcorr(T, S)* is increasing in $|M(T,S)|$. So, for example, if $M(T,S)>0$, we know that *medcorr(T, S)* is also positive, and a higher *M(T, S)* results in a higher *medcorr(T, S)*. In the extreme case where $M(T, S)=+\infty$, it is easy to see that $medcorr(T, S)=1$.

The $\mathscr{L}_1$ analogue of *rsq* is

$$medrsq(T,S)\equiv 1-\frac{\min_\beta E|T-\beta S-Med(T-\beta S)|}{E|T-Med(T)|}.$$

Note that

$$medrsq(T,S)=1-\frac{E|T-\beta_0 S-Med(T-\beta_0 S)|}{E|T-Med(T)|},$$

where β_0 is an arbitrary element of *M(T, S)*. This method was used in Koenker and Machado (1999) to measure the goodness of fit for quantile regressions. Koenker and Machado (1999) explain why *medrsq* is bounded between 0 and 1. They also show that this correlation measure takes the value of 1 where the random variable *T* and the random vector *S* are linearly perfectly correlated.

We collect some results about *medrsq* and about the relationship between *medcorr* and *medrsq* in the following theorem.

Theorem 5.2 *Consider random variables T and S such that $E|S|<\infty$ and $E|T|<\infty$. The following hold:*

1. *If $M(T, S)=0$ then $medrsq(T, S)=0$; if $medrsq(T, S)=0$ then $0\in M(T, S)$.*
2. *Suppose that $P(T-Med(T)=0)=0$. Then $medrsq(T, S)=0$ if and only if $medcorr(T, S)=0$.*

Part (1) shows that *medrsq* takes the value of zero when *T* is median uncorrelated with *S*. This is similar to the usual R^2 in linear models. Part (2) says that this median R^2 is equal to zero when the median correlation is zero.

Also, Blomqvist (1950) introduced the following measure of median correlation between random variables *T* and *S*:

$$k(T, S)=E[\text{sgn}(T-Med(T))\text{sgn}(S-Med(S))],$$

or, in terms of normalized variables,

$$k(T,S)=E\left[\text{sgn}(\tilde{T})\text{sgn}(\tilde{S})\right]$$

if $E|T|<\infty$, $E|S|<\infty$. As we can see, this measure is different from ours. In particular, *k(T, S)* is symmetric and does not satisfy the invariance property. The value of *medcorr(T, S)* measures the degree of linear relationship between *T* and *S* while *k(T, S)* represents an analog of Kendall's rank correlation because

$$k(T, S)=Pr((T-Med(T))(S-Med(S))>0)-Pr((T-Med(T))(S-Med(S))<0).$$

Next, we generalize the concept of $\mathscr{L}_1$-correlation to other loss functions. This will be a natural extension to the above results.

5.1 $\mathscr{L}_p$-correlation for any $p\geq 1$

The notion of $\mathscr{L}_1$-correlation can be generalized to the case of $\mathscr{L}_p$-correlation for any $p\geq 1$.

Definition 5.1 *For a random variable Y and for any p, $1\leq p<\infty$, define $Med_p(Y)$ as follows:*

$$Med_p(Y)\equiv\inf\{d: E[|Y-d|^{p-1}\,sgn(Y-d)]\leq 0\}.$$

Note that $Med_1(Y)=Med(Y)$ and $Med_2(Y)=E(Y)$.

Let *T* be a random variable and *S* be a random vector with values in $\Re^k$ such that $E|T|^p<\infty$ and $E\|S\|^p<\infty$. Consider the optimization problem

$$\min_{(\alpha,\beta)} E|T-\alpha-S'\beta|^p.$$

We are interested in the solutions to this problem with respect to β. Denote the set of these solutions as $M_p(T, S)$:

$$M_p(T,S)\equiv\left\{\beta:\exists\alpha \text{ such that } (\alpha,\beta)=\arg\min_{(\tilde{\alpha},\tilde{\beta})} E|T-\tilde{\alpha}-S'\tilde{\beta}|^p\right\}.$$

Notice that for a fixed β,

$$E|T-S'\beta-Med_p(T-S'\beta)|^p=\min_\alpha E|T-\alpha-S'\beta|^p.$$

Therefore,

5 A sketch of a proof for this is as follows. Since, $medcorr(\tilde{T},\tilde{S})\equiv E\left[|\tilde{S}|\text{sgn}(\tilde{T})\text{sgn}(\tilde{S})\right]$, replace $\tilde{T}$ with $\tilde{T}=\alpha'+\tilde{S}M(\tilde{T},\tilde{S})+\delta$ to get $E\left[|\tilde{S}|\text{sgn}\left(\alpha'+\tilde{S}M(\tilde{T},\tilde{S})+\delta\right)\text{sgn}(\tilde{S})\right]$ which is in turn equal to $E\left[\tilde{S}\,\text{sgn}\left(\alpha'+\tilde{S}M(\tilde{T},\tilde{S})+\delta\right)\right]$. The derivative of the latter with respect to $M(\tilde{T},\tilde{S})$ is equal to $2\int\tilde{S}^2 f_{\delta|\tilde{S}}\left(-\alpha'-\tilde{S}M(\tilde{T},\tilde{S})\right)dF_{\tilde{S}}$, which is positive.

$$M_p(T,S)=\arg\min_\beta E|T-S'\beta-Med_p(T-S'\beta)|^p.$$

The next definition introduces the notion of $\mathscr{L}_p$-uncorrelation of a random vector with another random vector.

Definition 5.2 ($\mathscr{L}_p$-uncorrelation) *Let W denote an l-dimensional random vector. We say that W is $\mathscr{L}_p$-uncorrelated with S if*

$$M_p(c'W, S) = 0 \quad \text{for all} \quad c\in\Re^l.$$

To measure $\mathscr{L}_p$-correlation of a scalar random variable T with a scalar random variable S, let us normalize these variable and define $\tilde{T}$ and $\tilde{S}$ in the following way:

$$\tilde{T}=\frac{T-Med_p(T)}{\left(E|T-Med_p(T)|^p\right)^{\frac{1}{p}}},$$

$$\tilde{S}=\frac{S-Med_p(S)}{\left(E|S-Med_p(S)|^p\right)^{\frac{1}{p}}}.$$

Define a measure of $\mathscr{L}_p$-correlation of T with S as follows:

$$medcorr_p(T,S)=E\left[|\tilde{S}||\tilde{T}|^{p-1}\operatorname{sgn}(\tilde{T})\operatorname{sgn}(\tilde{S})\right].$$

The value of $medcorr_p(T, S)$ lies in the interval [–1, 1], and it can be shown that under weak restrictions, similar to the ones in Theorem 5.1,

$$\operatorname{sgn}(medcorr_p(T, S))=\operatorname{sgn}(M_p(T, S)).$$

Note that if for some c_2, $T=c_1+c_2S$ with probability 1, then $medcorr_p(T, S)=sgn(c_2)$. It is easy to see that $medcorr_2(T, S)$ coincides with the familiar correlation coefficient $corr(T, S)$.

The $\mathscr{L}_p$ analogue of *medrsq* is defined as follows:

$$medrsq_p(T,S)\equiv 1-\frac{\min_\beta E|T-\beta S-Med_p(T-\beta S)|^p}{E|T-Med_p(T)|^p},$$

and obviously,

$$medrsq(T,S)=1-\frac{E|T-\beta_0 S-Med_p(T-\beta_0 S)|^p}{E|T-Med_p(T)|^p},$$

where β_0 is an arbitrary element of $M_p(T, S)$.

6 Other Applications of Median Uncorrelation

We provide two other applications of this median uncorrelation by mimicking implications of mean uncorrelation when dealing with measurement error in linear models under quantile restrictions, and in panel data models with quantile restrictions.

6.1 Quantile regression with measurement error

We apply the idea of median uncorrelation to linear quantile regressions with classical measurement error in the regressors. In particular, consider the model

$$Y=\alpha_0+X^{*\prime}\beta_0+\varepsilon, \quad Med(\varepsilon)=0, \tag{6.1}$$

where we assume that $M(\varepsilon, X^*)=0$ or that ε is median uncorrelated with a k-dimensional random vector X^*. We do not observe X^* directly, but we observe an error-ridden version of it, X, such that

$$X=X^*+\nu, \tag{6.2}$$

where we assume that $M(\nu,X^*)=0$. We also observe Y. To remedy the identication problem that results from the measurement error, we follow the treatment of the linear model under the mean uncorrelation and use instruments. Let there exist a d-dimensional random vector Z and a $k\times d$ constant matrix γ, with $d\geq k$, such that

$$X^*=\gamma Z+\psi \tag{6.3}$$

for some random vector ψ, and $M(\psi, Z) = 0$. Then

$$X=\gamma Z+\psi+\nu.$$

Given the results of the previous section, we can show the following result.

Theorem 6.1 *For model (6.1) suppose that we observe (Y,X) such that (6.2) holds with $M(\nu, X^*)=0$. Moreover, assume that (ε, ν, ψ) is median uncorrelated with Z and that γ in (6.3) has full rank. Then,*

$$M(Y-X'\beta, Z)=0 \Leftrightarrow \beta=\beta_0.$$

Note that the requirements of the above model are that the vector (ε, ν, ψ) is *jointly* median uncorrelated with Z. The real assumption here is that the vector of unobservables is required to be median uncorrelated with Z. In contrast, in the mean uncorrelation model, Z is mean uncorrelated with ψ by construction. So, again, as in the

2SLS generalization, it is the joint median uncorrelation that is needed.

6.2 Quantile regression with panel data

We are interested in inference on β_0 in the following model:

$$y_{it}=x_{it}'\beta_0+\alpha_i+\varepsilon_{it}, \qquad t=1, 2, \tag{6.4}$$

where α_i is the individual effect that is arbitrarily correlated with $\mathbf{x}_i=(x_{i1}', x_{i2}')'$. Denote $\Delta y_i=y_{i1}-y_{i2}$, $\Delta x_i=x_{i1}-x_{i2}$ and $\Delta\varepsilon_i=\varepsilon_{i1}-\varepsilon_{i2}$. Suppose that we have a data set of iid observations $(\mathbf{y}_i, \mathbf{x}_i)$ for $i=1,\ldots,n$, where $\mathbf{y}_i=(y_{i1}, y_{i2})'$. If we maintain the assumption that $\boldsymbol{\varepsilon}_i=(\varepsilon_{i1}, \varepsilon_{i2})'$ is median uncorrelated with $\mathbf{x}_i$, then

$$\beta_0=M(\Delta y_i, \Delta x_i).$$

Indeed, this follows from

$$E|\Delta y_i-a-\Delta x_i'\beta|=E|\Delta\varepsilon_i-a-\Delta x_i'(\beta-\beta_0)|$$

and the definition of the median uncorrelation of the vector $\boldsymbol{\varepsilon}_i$ with $\mathbf{x}_i$. We want to emphasize that we require not only $\boldsymbol{\varepsilon}_{it}$ be contemporaneously median uncorrelated with x_{it}, $t=1,2$, but also that the vector ε_i be *jointly* median uncorrelated with the vector $\mathbf{x}_i$ of explanatory variables in both periods. On the other hand, it is possible to relax this joint median uncorrelation condition in the panel setup to requiring that the random variable $\Delta\varepsilon_i$ be median uncorrelated with Δx_i.

7 Conclusion

The paper considers an analogue of the 2SLS estimator which is commonly used in econometrics for estimating regressions with endogenous variables. The 2SLS estimator is based on the assumption that even though a regressor is correlated with the error, there exists an excluded *exogenous* regressor that is (linearly) uncorrelated with the error. This regressor is called an instrument. And so, 2SLS exploits implications of this (linear) uncorrelation between the instrument and the error in the main regression to obtain a consistent estimator for the slope. This paper tries to follow the same model, but uses median uncorrelation instead. This median uncorrelation is new to our knowledge and is exactly similar to mean uncorrelation, except that it uses the absolute loss function, as opposed to the squared loss function used with the mean. We characterize properties of two vectors that are linearly median uncorrelated and then provide a measure of median uncorrelation which is bounded between -1 and 1. This is meant to mirror the typical correlation coefficient in linear models. We also provide counterparts to R^2 the coefficient of determination. Most importantly, we show that in a linear regression model where the regressors are correlated with the errors, a median uncorrelation assumption between a set of instruments and the error provides the basis for inference on the linear slope parameter β that is akin to what the 2SLS approach does under mean uncorrelation. We apply this uncorrelation concept to other examples like linear models with measurement error and quantile restrictions, and panel data quantile models.

8 Appendix

Proof of Proposition 2.1

1. First, suppose that $\beta^*\in\Re^k$ satisfies inequality (2.1) for any $\alpha\in\Re$, $\beta\in\Re^k$. Denote $m^*(S)=Med(T-S'\beta^*)+S'\beta^*$. Choose any $a\in\Re$, $b\in\Re^k$ and denote $m(S)=a+S'b$. Then

$$\begin{aligned}
&E|T-m^*(S)|-E|T-m(S)|=E[(T-m^*(S))\operatorname{sgn}(T-m^*(S))]\\
&-E|T-m(S)|\\
&\quad=E[(T-m(S))\operatorname{sgn}(T-m^*(S))]+E[(m(S)-m^*(S))\operatorname{sgn}\\
&\qquad(T-m^*(S))]\\
&\quad-E|T-m(S)|\\
&\quad\le E[(T-m(S))\operatorname{sgn}(T-m^*(S))\cdot 1(T-m^*(S)\ne 0)]\\
&\quad+E[|m(S)-m^*(S)|\cdot 1(T-m^*(S)=0)]-E|T-m(S)| \qquad (8.1)\\
&\quad\le E[|T-m(S)|\cdot 1(T-m^*(S)\ne 0)]\\
&\quad+E[|m(S)-T|\cdot 1(T-m^*(S)=0)]-E|T-m(S)|\\
&\quad\le E|T-m(S)|-E|T-m(S)|=0,
\end{aligned}$$

where the first term in (8.1) is obtained using inequality (2.1). Thus, $\beta^*\in M(T,S)$.

Now suppose that $\beta^*\in M(T,S)$. Then for any $r\in\Re$,

$$E|T-m^*(S)+rm(S)|-E|T-m^*(S)|\ge 0,$$

and therefore,

$$\liminf_{r\downarrow 0}\frac{E|T-m^*(S)+rm(S)|-E|T-m^*(S)|}{r}\ge 0.$$

Note that $\dfrac{E|T-m^*(S)+rm(S)|-E|T-m^*(S)|}{r}$

$$=E\left[|m(S)|\cdot 1(T-m^*(S)=0)\right]+\frac{1}{r}E[(|T-m^*(S)+rm(S)|-E|T-m^*(S)|)\cdot 1(T-m^*(S)\ne 0)]$$

When $T-m^*(S)\ne 0$,

$$\frac{|T-m^*(S)+rm(S)|-|T-m^*(S)|}{r}$$
$$=\frac{2(T-m^*(S))m(S)+rm^2(S)}{|T-m^*(S)+rm(S)|+|T-m^*(S)|},$$

and

$$\lim_{r\downarrow 0}\frac{|T-m^*(S)+rm(S)|-|T-m^*(S)|}{r}$$
$$=\frac{2(T-m^*(S))m(S)}{2|T-m^*(S)|}=sgn(T-m^*(S))m(S)\cdot$$

Taking into account that

$$\left|\frac{|T-m^*(S)+rm(S)|-T-m^*(S)|}{r}\right|\le|m(S)|,$$

and applying Lebesgue's dominated convergence theorem, we obtain

$$\lim_{r\downarrow 0}\inf\frac{E|T-m^*(S)+rm(S)|-E|T-m^*(S)|}{r}$$
$$=\lim_{r\downarrow 0}\frac{E|T-m^*(S)+rm(S)|-E|T-m^*(S)|}{r}$$
$$=E[|m(S)|\cdot 1(T-m^*(S)=0)]$$
$$+E|\,\mathrm{sgn}(T-m^*(S))m(S)\cdot 1(T-m^*(S)\neq 0)]\cdot$$

Then

$$-E[\mathrm{sgn}(T-m^*(S))m(S)\cdot 1(T-m^*(S)\neq 0)]\le E[|m(S)|\cdot 1(T-m^*(S)=0)].$$

If the same technique is applied to $E|T-m^*(S)-rm(S)|-E|T-m^*(S)|$, then

$$E[\mathrm{sgn}(T-m^*(S))m(S)\cdot 1(T-m^*(S)\neq 0)]|\le E[|m(S)|\cdot 1(T-m^*(S)=0)].$$

Therefore,

$$|E[\mathrm{sgn}(T-m^*(S))m(S)\cdot 1(T-m^*(S)\neq 0)]|\le E[|m(S)|\cdot 1(T-m^*(S)=0)],$$

which concludes the proof of part 1.

2. Use the result of part 1 of this proposition. Under given conditions, for any $\alpha\in\Re$,

$$E[\alpha\ \mathrm{sgn}(T-S'\beta^*-Med(T-S'\beta^*))]=0,$$

and the right-hand side in (2.1) is 0. This gives

$$E[S'\beta\ \mathrm{sgn}(T-S'\beta^*-Med(T-S'\beta^*))]=0$$

for any $\beta\in\Re^k$. Choosing $\beta=(1,0,\ldots,0)$, we obtain that

$$E[S_1\ \mathrm{sgn}(T-S'\beta^*-Med(T-S'\beta^*))]=0.$$

In a similar way we can show that for any $i=1,\ldots,k$,

$$E[S_i\ \mathrm{sgn}(T-S'\beta^*-Med(T-S'\beta^*))]=0,$$

which means that

$$E[S\ \mathrm{sgn}(T-S'\beta^*-Med(T-S'\beta^*))]=0.$$

3. This result is obvious from part 2 of this proposition.

Proof of Lemma 2.1

We prove this lemma in two steps. In the first step we show that $M(T,S)+b\subset M(T+a+S'b,S)$. In the second step, we establish that $M(T+a+S'b,S)\subset M(T,S)+b$.

First of all, note that for a given b and any a,

$$M(T+a+S'b,S)=\arg\min_{q\in\Re^k}E|T+S'(b-q)-Med(T+S'(b-q))|.$$

Let $m_1\in M(T,S)$. This implies that for any $q\in\Re^k$

$$E|T+S'(b-q)-Med(T+S'(b-q))|\ge E|T-S'm_1-Med(T-S'm_1)|.$$

Obviously, the inequality becomes the equality if $q=m_1+b$. Therefore, $m_1+b\in M(T+a+S'b,S)$.

Now let $m_2\in M(T+a+S'b,S)$. This implies that for any $\beta\in\Re^k$

$$E|T-S'\beta-Med(T-S'\beta)|\ge E|T+S'(b-m_2)-Med(T+S'(b-m_2))|.$$

The inequality becomes the equality if $\beta=m_2-b$. Therefore, $m_2-b\in M(T,S)$ and, hence, $m_2\in M(T,S)+b$.

Proof of Theorem 3.1

(A): Suppose $Med(c'W|s)\equiv c^*$. Then, we know that c^* minimizes the following problem over all (measurable) functions $g(S)$:

$$E|c'W-c^*|\le E|c'W-g(S)|.$$

In particular, this holds for any linear function of S, $\alpha+S'\beta$ with $\beta\neq 0$.

(B): Consider independent random variables S and Z such that $P(S=1)=\frac{7}{16}$, $P(S=-1)=\frac{9}{16}$, and $P(Z=1)=\frac{1}{6}$, $P(Z=0)=\frac{1}{2}$, $P(Z=-1)=\frac{1}{3}$. Define random variable W as $W=SZ$. Since

$$Med(W|S=1)=Med(W|S=-1)=0,$$

then from part (A) we conclude that W is median uncorrelated with S.

Let us now analyze whether S is median uncorrelated with W. Consider the optimization problem

$$\min_{\beta} E|S-W\beta-Med(S-W\beta)|.$$

Since $Med(S)=-1$, then the value of the objective function when $\beta=0$ is $E|S+1|=\frac{7}{8}$. Let us find the value of this objective function when $\beta=-1$. Since $Med(S+W)=Med(S+SZ)=0$, then $E|S+W-Med(S+W)|=E|S+SZ|=E|1+Z|=\frac{5}{6}$, which is smaller than $\frac{7}{8}$. Thus, $\beta=0$ cannot be a solution to the optimization problem. This implies that S is not median uncorrelated with W.

(C): This means that the conditional characteristic function of $c'W$ given S is real, which in part means that the conditional distribution of $c'W$ given S is symmetric around 0. Hence, $Med(c'W|s)=0=Med(c'W)$ for all s.

(D): Let $\delta = T-\alpha_0-S'M(T, S)$, where α_0 is any constant. Showing that $M(\delta, S)$ is equal to 0 is a direct result of the invariance property in (2.4).

(E): Since by assumption $P(T-Med(T)=0)=0$, Proposition 2.1 and conditions $M(T, S)=0$ and $M(T, Z)=0$ imply that

$$E[S\,\text{sgn}(T-Med(T))]=0,\ E[Z\,\text{sgn}(T-Med(T))]=0.$$

Then

$$E[(S+Z)\ \text{sgn}(T-Med(T))]=E[S\ \text{sgn}(T-Med(T))]+E[Z\ \text{sgn}(T-Med(T))]=0,$$

that is, $0\in M(T, S+Z)$. Since $M(T, S+Z)$ is assumed to be a singleton, $M(T, S+Z)=0$.

(F): The first part of the statement follows from (A). For the second part of the statement, note that Proposition 2.1 implies

$$E[S\,\text{sgn}(T-Med(T))]=0.$$

Given that the conditional median of $T|S=1$ is unique, we have:

$$E[S\,\text{sgn}(T-Med(T))]=0 \Rightarrow E[\text{sgn}(T-Med(T))|S=1]=0 \Rightarrow Med(T)=Med(T|S=1).$$

Because $E[\text{sgn}(T-Med(T))]=0$,

$$E[\text{sgn}(T-Med(T))|S=1]=0 \Rightarrow E[\text{sgn}(T-Med(T))|S=0]=0.$$

Taking into account that that the conditional median of $T|S=0$ is unique, we obtain that $Med(T)=Med(T|S=0)$.

Proof of Theorem 5.1

(1): This follows from

$$|medcorr(T,S)|=|E[\tilde{S}\ \text{sgn}(T-Med(T))\text{sgn}(\tilde{S})]|\le E|\tilde{S}|=1.$$

(2): First, let us prove that $M(T, S)=0 \Leftrightarrow medcorr(T, S)=0$. Taking into account the conditions of this theorem and applying Proposition 2.1, obtain that

$$\begin{aligned}M(T, S)=0 &\Leftrightarrow E\,[S\,\text{sgn}(T-Med(T))]=0\\ &\Leftrightarrow E\left[\frac{S-Med(S)}{E|S-Med(S)|}\text{sgn}(T-Med(T))\right]=0\\ &\Leftrightarrow E\left[\tilde{S}\,\text{sgn}(T-Med(T))\right]=0\\ &\Leftrightarrow medcorr(T, S)=0.\end{aligned}$$

Note that $medcorr(T, S)=medcorr(T-Med(T), \tilde{S})$ and

$$M(T,S)=\frac{M(T-Med(T),\tilde{S})}{E|S-Med(S)|},$$

and, hence, $\text{sgn}(M(T, S))=\text{sgn}(M(T-Med(T), \tilde{S}))$. Thus, it is enough to show that

$$\text{sgn}(medcorr(T-Med(T), \tilde{S}))=\text{sgn}(M(T-Med(T), \tilde{S})).$$

Denote $b^*=M(T-Med(T), \tilde{S})$. For $b^*=0$ the result is already proven.

Suppose $b^*\neq 0$. Notice that

$$\text{sgn}(medcorr(T-Med(T), \tilde{S}))=\text{sgn}(b^*)\text{sgn}(E[b^*\tilde{S}\,\text{sgn}(T-Med(T))]),$$

and therefore, the result will be proven if we establish that $E[b^*\tilde{S}\,\text{sgn}(T-Med(T))]>0$.

Denote

$$a^*=Med(T-Med(T)-b^*\tilde{S})=Med(T-b^*\tilde{S})-Med(T).$$

According to Proposition 2.1, b^* satisfies

$$E\left[\tilde{S}\ \text{sgn}(T-Med(T)-b^*\tilde{S}-a^*)\right]=0$$

Then

$$\begin{aligned}E|b^*\tilde{S}\ sgn(T-Med(T))]&=E[(b^*\tilde{S}+a^*)\text{sgn}(T-Med(T))]\\ &=E[(b^*\tilde{S}+a^*)(\text{sgn}(T-Med(T))-\text{sgn}(T-Med(T)-b^*\tilde{S}-a^*))]\\ &=2E[(b^*\tilde{S}+a^*)1(T-Med(T)>0)1(T-Med(T)-b^*\tilde{S}-a^*<0)]\\ &\quad-2E[(b^*\tilde{S}+a^*)1(T-Med(T)<0)1(T-Med(T)-b^*\tilde{S}-a^*>0)].\end{aligned}$$

Notice that both terms in the last sum are non-negative. Moreover, at least one of them is strictly positive because

$$Pr(\text{sgn}(T-Med(T))\text{sgn}(T-Med(T)-b^*\tilde{S}-a^*)=-1)>0,$$

or equivalently,

$$Pr(\text{sgn}(T-Med(T))\text{sgn}(T-b^*\tilde{S}-Med(T-b^*\tilde{S}))=-1)>0.$$

This follows from the assumptions of the theorem and part 2 of Proposition 2.1, according to which $b=\frac{b^*}{E|S-Med(S)|}$ solves the equation

$E[S\ \mathrm{sgn}(T-bS-Med(T-bS))]=0,$

while $b=0$ does not.

Thus, $E[b^*\tilde{S}\ \mathrm{sgn}(T-Med(T))]>0$.

Proof of Theorem 5.2

(1): If $M(T, S)=0$, then

$$medrsq(T,S)=1-\frac{E|T-Med(T)|}{E|T-Med(T)|}=0.$$

If $medrsq(T, S)=0$, then

$$\min_{\beta} E|T-\beta S-Med(T-\beta S)|=E|T-Med(T)|,$$

so that, clearly, $0 \in M(T,S)$.

(2): If $medrsq(T, S)=0$, then, from part (1), $0\in M(T, S)$. From Proposition 2.1 it follows that $E[S\ \mathrm{sgn}(T-Med(T))]=0$ and, hence, $medcorr(T, S)=0$. If $medcorr(T, S)=0$, then $E[S\ \mathrm{sgn}(T-Med(T))]=0$ so that by Proposition 2.1, $0\in M(T, S)$. It follows that

$$\min_{\beta} E|T-\beta S-Med(T-\beta S)|=E|T-Med(T)|.$$

Proof of Theorem 6.1

The proof of this theorem is analogous to the proof of Theorem 4.1.

Let

$m\in M(Y-X'\beta, Z)=M(\alpha_0+Z'\gamma'(\beta_0-\beta)+\psi'(\beta_0-\beta)+\varepsilon-\nu'\beta, Z).$

By the invariance property in Lemma 2.1, there exists $m_0\in M(\psi'(\beta_0-\beta)+\varepsilon-\nu'\beta, Z)$ such that

$m=\gamma'(\beta_0-\beta)+m_0.$

Note that $\psi'(\beta_0-\beta)=(\beta_0-\beta)'\psi$ and $\nu'\beta=\beta'\nu$. Hence, since $(\varepsilon, \nu, \psi)'$ is median uncorrelated with Z, $m_0=0$. It follows that $m=\gamma'(\beta_0-\beta)$, and hence, that

$M(Y-X'\beta, Z)=\gamma'(\beta_0-\beta).$

Since $d \geq k$ and γ is full column rank by assumption, then

$M(Y-X'\beta, Z)=0\Leftrightarrow\beta=\beta_0.$

References

Amemiya, T. (1981): "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–711.

Amemiya, T. (1985): *Advanced Econometrics*. Harvard University Press.

Basmann, R. L. (1960): "On the Asymptotic Distribution of Generalized Linear Estimators," *Econometrica*, 28(1), pp. 97–107.

Blackburn, M., and D. Neumark (1992): "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," *The Quarterly Journal of Economics*, 107(4), pp. 1421–1436.

Blomqvist, N. (1950): "On a Measure of Dependence Between Two Random Variables," *Ann. Math. Statistics*, 21, 593–600.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

Chernozhukov, V., and C. Hansen (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, 132(2), 491–525.

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

Griliches, Z. (1976): "Wages of Very Young Men," *The Journal of Political Economy*, 84(4), pp. S69–S86.

Honoré, B., and L. Hu (2004): "On the Performance of Some Robust Instrumental Variables Estimators," *Journal of Business and Economic Statistics*, 22(1), 30–39.

Koenker, R. (2005): *Quantile Regression*, vol. 38 *of Econometric Society Monographs*. Cambridge University Press, Cambridge.

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

Koenker, R., and J. A. F. Machado (1999): "Goodness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, 94(448), 1296–1310.

Lee, S. (2007): "Endogeneity in Quantile Regression Models: A Control Function Approach," *Journal of Econometrics*, 141(2), 1131–1158.

Ma, L., and R. Koenker (2006): "Quantile Regression Methods for Recursive Structural Equation Models," *Journal of Econometrics*, 134(2), 471–506.

Manski, C. (1988): *Analog Estimation Methods in Econometrics*. Chapman and Hall.

Powell, J. (1983): "The Asymptotic Normality of Two Stage Least Absolute Deviations Estimators," *Econometrica*, 51, 1569–1575.

Sakata, S. (2001): "Instrumental Variable Estimation Based on Mean Absolute Deviation Estimator," University of Michigan Working Paper.

Sakata, S. (2007): "Instrumental Variable Estimation Based on Conditional Median Restriction,"*Journal of Econometrics*, 141(2), 350–382.

Theil, H. (1953): "Estimation and Simultaneous Correlation in Complete Equation Systems," The Hague: Centraal Planbureau.