VECTOR QUANTILE REGRESSION

G. CARLIER, V. CHERNOZHUKOV, AND A. GALICHON

Abstract. We propose a notion of conditional vector quantile function and a vector quantile regression.

A conditional vector quantile function (CVQF) of a random vector Y, taking values in $\mathbb{R}^d$ given covariates $Z = z$, taking values in $\mathbb{R}^k$, is a map $u \mapsto Q_{Y|Z}(u, z)$, which is monotone, in the sense of being a gradient of a convex function, and such that given that vector U follows a reference non-atomic distribution F_U, for instance uniform distribution on a unit cube in $\mathbb{R}^d$, the random vector $Q_{Y|Z}(U, z)$ has the distribution of Y conditional on $Z = z$. Moreover, we have a strong representation, $Y = Q_{Y|Z}(U, Z)$ almost surely, for some version of U.

The vector quantile regression (VQR) is a linear model for CVQF of Y given Z. Under correct specification, the notion produces strong representation, $Y = \beta(U)^\top f(Z)$, for $f(Z)$ denoting a known set of transformations of Z, where $u \mapsto \beta(u)^\top f(Z)$ is a monotone map, the gradient of a convex function, and the quantile regression coefficients $u \mapsto \beta(u)$ have the interpretations analogous to that of the standard scalar quantile regression. As $f(Z)$ becomes a richer class of transformations of Z, the model becomes nonparametric, as in series modelling. A key property of VQR is the embedding of the classical Monge-Kantorovich's optimal transportation problem at its core as a special case.

In the classical case, where Y is scalar, VQR reduces to a version of the classical QR, and CVQF reduces to the scalar conditional quantile function. Several applications to diverse problems such as multiple Engel curve estimation, and measurement of financial risk, are considered.

Keywords: Vector quantile regression, vector conditional quantile function, Monge-Kantorovich-Brenier.

JEL Classification: C14; C21. **AMS Classification**: 49Q20, 49Q10, 90B20

Date: First draft: October 2013. This draft: June 30, 2014. This is still a preliminary version of the paper. We thank Xuming He, Roger Koenker, Steve Portnoy, and workshop participants at Oberwolfach and Université Libre de Bruxelles for helpful comments. Jinxin He, Jérôme Santoul, and Simon Weber provided excellent research assistance. Carlier is grateful to INRIA and the ANR for partial support through the Projects ISOTACE (ANR-12-MONU-0013) and OPTIFORM (ANR-12-BS01-0007). Galichon's research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement 313699, and from FiME, Laboratoire de Finance des Marchés de l'Energie (www.fime-lab.org). Chernozhukov gratefully acknowledges research support via an NSF grant.

1. Introduction

Quantile regression provides a very convenient and powerful tool for studying dependence between random variables. The main object of modelling is the conditional quantile function (CQF) $(u,z) \mapsto Q_{Y|Z}(u,z)$, which describes the u-quantile of the random scalar Y conditional on a k-dimensional vector of regressors Z taking a value z. Conditional quantile function naturally leads to a strong representation via relation:

$$Y = Q_{Y|Z}(U,Z), \quad U \mid Z \sim U(0,1),$$

where U is the latent unobservable variable, normalized to have a uniform reference distribution, and is independent of regressors Z. The mapping $u \mapsto Q_{Y|Z}(u,Z)$ is monotone, namely non-decreasing, almost surely.

Quantile regression (QR) is a means of modelling the conditional quantile function. A leading approach is linear in parameters, namely, it assumes that there exists a known $\mathbb{R}^p$-valued vector $f(Z)$, containing transformations of Z, and a ($p \times 1$ vector-valued) map of regression coefficients $u \mapsto \beta(u)$ such that

$$Q_{Y|Z}\left(u \mid z\right) = \beta\left(u\right)^{\top} f(z),$$

for all z in the support of Z and for all quantile indices u in $(0,1)$. This representation highlights the vital ability of QR to capture differentiated effects of the explanatory variable Z on various conditional quantiles of the dependent variable Y (e.g., impact of prenatal smoking on infant birthweights).[1] The model is flexible in the sense that, even if the model is not correctly specified, by using more and more suitable terms $f(Z)$ we can approximate the true CQF arbitrarily well. Moreover, coefficients $u \mapsto \beta(u)$ can be estimated via tractable linear programming method (Koenker and Basset, 1978).

The principal contribution of this paper is to extend these ideas to the cases of vector-valued Y, taking values in $\mathbb{R}^d$. Specifically, a vector conditional quantile function (CVQF) of a random vector Y, taking values in $\mathbb{R}^d$ given the covariates Z, taking values in $\mathbb{R}^k$, is a map $(u,z) \mapsto Q_{Y|Z}(u,z)$, which is monotone with respect to u, in the sense of being a

[1] Quantile regression has found many applications to econometrics (wage structure, program evaluation, demand analysis, income inequality), biometrics, finance, technometrics, and other areas of statistical analysis. Applications include inter alia the study of the wage structure (Buchinski 1994, 1997; Chamberlain 1994; Poterba and Rueben 1995; Angrist, Chernozhukov, and Fernandez-Val, 2006); endogenous treatment effect in program evaluation (Abadie, Angrist and Imbens 2001; Chernozhukov and Hansen, 2005); demand analysis (Deaton 1997); wealth inequality (Gosling, Machin and Meghir 2000); health economics (Abrevaya 2001); and finance (Engle and Manganelli 2004; White, Kim, and Manganelli 2008; Adrian and Brunnermeier 2011; Chernozhukov and Umantsev 2001), and many more, see Koenker (2005)'s monograph.

gradient of a convex function, which implies that

$$(Q_{Y|Z}(u,z) - Q_{Y|Z}(\overline{u},z))^\top (u - \overline{u}) \geq 0 \quad \text{for all } u, \overline{u} \in (0,1)^d, z \in \mathcal{Z}, \tag{1.1}$$

and such that the following strong representation holds with probability 1:

$$Y = Q_{Y|Z}(U,Z), \quad U \mid Z \sim U(0,1)^d, \tag{1.2}$$

where U is latent random vector uniformly distributed on $(0,1)^d$. We can also use other non-atomic reference distributions F_U on $\mathbb{R}^d$, for example, the standard normal distribution instead of uniform distribution (as we can in the canonical, scalar quantile regression case). We show that this map exists and is unique under mild conditions, as a consequence of Brenier's polar factorization theorem. This notion relies on a particular, yet very important, notion of monotonicity for maps $\mathbb{R}^d \to \mathbb{R}^d$, which we adopt here.

We define vector quantile regression (VQR) as a model of CVQF, particularly a linear model. Specifically, under correct specification, our linear model takes the form:

$$Q_{Y|X}(u \mid z) = \beta(u)^\top f(z),$$

where $u \mapsto \beta(u)^\top f(z)$ is a monotone map, in the sense of being a gradient of convex function; and $u \mapsto \beta(u)$ is a map of regression coefficients from $(0,1)^d$ to the set of $p \times d$ matrices with real entries. This model is a natural analog of the classical QR for the scalar case. In particular, under correct specification, we have the strong representation

$$Y = \beta(U)^\top f(Z), \quad U \mid Z \sim U(0,1)^d, \tag{1.3}$$

where U is uniformly distributed on $(0,1)^d$ conditional on Z. (Other reference distributions could also be easily permitted.)

We provide a linear program for computing $u \mapsto \beta(u)$ in population and finite samples. We shall stress that this formulation offers a number of useful properties. In particular, the linear programming problem admits a general formulation that embeds the optimal transportation problem of Monge-Kantorovich-Brenier, establishing a useful intellectual link to an important area of optimization and functional analysis (see, e.g. Villani, 2005).

Our paper also connects to a number of interesting proposals for performing multivariate quantile regressions, which focus on inheriting certain (though not all) features of univariate quantile regression– for example, minimizing an asymmetric loss, ordering ideas, monotonicity, equivariance or other related properties, see, for example, some key proposals (including some for the non-regression case) in Chaudhuri (1996), Koltchinskii (1997), and Serfling (2004), Hallin et al (2010), Kong and Mizera (2010), Belloni and Winkler (2011), and the references therein. Our proposal is quite different from all of these excellent proposals in that it targets to simultaneously reproduce two fundamentally different properties of quantile regression in higher dimensions – namely the deterministic coupling property (1.3) and the monotonicity property (1.1).[2] This is the reason we deliberately don't use adjective "multivariate" in naming our method. By using a different name we emphasize the major differences of our method's goals from those of the other proposals. This also makes it clear that our work is complementary to other works in this direction. We discuss other connections, including to some of our own work, as we present our main results.

We organize the rest of the paper as follows. In Section 2, we introduce and develop the properties of CVQF. In Section 3, we introduce and develop the properties of VQR as well its linear programming implementation. In Section 4, we provide computational details of the discretized form of the linear programming formulation, which is useful for practice and computation of VQR with finite samples. In Section 5, we implement VQR in an empirical example, providing the testing ground for these new concepts. We provide proofs of all formal results of the paper in the Appendix.

2. Conditional Vector Quantile Function

2.1. **Conditional Vector Quantiles as Gradients of Convex Functions.** We consider a random vector (Y, Z) defined on a complete probability space $(\Omega_1, \mathcal{A}_1, \mathrm{P}_1)$. The random vector Y takes values in $\mathbb{R}^d$. The random vector Z is a vector covariate, taking values in $\mathbb{R}^k$. Denote by F_{YZ} the joint distribution function of (Y, Z), by $F_{Y|Z}$ the (regular) conditional distribution function of Y given Z, and by F_Z the distribution function Z. We also consider random vectors V defined on a complete probability space $(\Omega_0, \mathcal{A}_0, \mathrm{P}_0)$, which are required to have a fixed reference distribution function F_U. Let $(\Omega, \mathcal{A}, \mathrm{P})$ be the a *suitably enriched* complete probability space that can carry all vectors (Y, Z) and V with distributions F_{YZ} and F_U, respectively, as well as the independent (from all other variables) standard uniform random variable on the unit interval.[3] The symbols $\mathcal{Y}$, $\mathcal{Z}$, $\mathcal{U}$, $\mathcal{YZ}$, $\mathcal{UZ}$ denote the support of F_Y, F_Z, F_U, F_{YZ}, F_{UZ}, and $\mathcal{Y}_z$ denotes the support of $F_{Y|Z}(\cdot|z)$.

We assume that the following condition holds:

[2] Note that it is not possible to reproduce all "desirable properties" of scalar quantile regression in higher dimensions, so various proposals focus on achieving different sets of properties. Thus we remark that there is no holy grail in dimension greater than 1, in reference to the following remark in Koenker (2005): "the search for a satisfactory notion of multivariate quantiles has become something of a quest for the statistical holy grail in recent years.".

[3] Formally, this product space takes the form $(\Omega, \mathcal{A}, \mathrm{P}) = (\Omega_0, \mathcal{A}_0, \mathrm{P}_0) \times (S_1, \mathcal{A}_1, \mathrm{P}_1) \times ((0, 1), B(0, 1), \mathrm{Leb})$, where $((0, 1), B(0, 1), \mathrm{Leb})$ is the canonical probability space, consisting of the unit segment of the real line equipped with Borel sets and the Lebesgue measure. We need this for measure-theoretic reasons, in order to claim the required strong representation.

(N) *F_U has a density f_U with respect to the Lebesgue measure on $\mathbb{R}^d$ with a convex support set $\mathcal{U}$.*

The distribution F_U describes a reference distribution for a vector of latent variables U, taking values in $\mathbb{R}^d$, that we would like to link to Y via a strong representation of the form mentioned in the introduction. This vector will be one of many random vectors V having a distribution function F_U, but there will only be one $V = U$, in the sense specified below, that will provide the required strong representation. The leading cases for the reference distribution F_U include:

- the standard uniform distribution on the unit d-dimensional cube, $U(0,1)^d$, and
- the standard normal distribution $N(0, I_d)$ over $\mathbb{R}^d$.

Our goal here is to create a deterministic mapping that transforms a random vector U with distribution F_U into Y such that Y conditional on Z has the conditional distribution $F_{Y|Z}$. That is, we want to have a strong representation property like (1.2) that we stated in the introduction. Moreover, we would like this transform to have *a monotonicity property*, as in the scalar case. Specifically, in the vector case we require this transform to be a *gradient of a convex function*, which is a plausible generalization of monotonicity from the scalar case. Indeed, in the scalar case the requirement that the transform is the gradient of a convex map reduces to the requirement that the transform is non-decreasing. We shall refer to the resulting transform as the *conditional vector quantile function* (CVQF). The following theorem shows that such map *exists* and is *uniquely determined* by the stated requirements.

Theorem 2.1 (Conditional Vector Quantiles as Conditional Brenier Maps). *Suppose condition (N) holds.*

(i) There exists a measurable map $(u,z) \mapsto Q_{Y|Z}(u,z)$ from $\mathcal{UZ}$ to $\mathbb{R}^d$, such that for each z in $\mathcal{Z}$, the map $u \mapsto Q_{Y|Z}(u,z)$ is the unique (F_U-almost everywhere) gradient of convex function such that, whenever $V \sim F_U$, the random vector $Q_{Y|Z}(V,z)$ has the distribution function $F_{Y|Z}(\cdot, z)$, that is,

$$F_{Y|Z}(y,z) = \int 1\{Q_{Y|Z}(u,z) \leq y\} F_U(du), \quad \text{for all } y \in \mathbb{R}^d. \tag{2.1}$$

(ii) Moreover, there exists a random variable V such that P*-almost surely*

$$Y = Q_{Y|Z}(U,Z), \text{ and } U \mid Z \sim F_U. \tag{2.2}$$

The theorem is our *first main result* that we announced in the introduction. It should be noted that the theorem does not require Y to have an absolutely continuous distribution, it holds for discrete and mixed outcome variables; only the reference distribution for the latent variable U is assumed to be continuous. It is also noteworthy that in the classical case of Y and U being *scalars* we recover the classical conditional quantile function as well as the strong representation formula based on this function (Matzkin 2003, Koenker, 2005). Regarding the proof, the first assertion of the theorem is a consequence of fundamental results due to McCann (1995) (as, e.g, stated in Villani (2003), Theorem 2.32) who in turn refined the fundamental results of Brenier (1991). These results were obtained in the case without conditioning. The second assertion is a consequence of Dudley-Philipp (1983) result on abstract couplings in Polish spaces.

Remark 2.1 (**Monotonicity**). The transform $(u, z) \mapsto (Q_{Y|Z}(u, z), z)$ has the following monotonicity property:

$$(Q_{Y|Z}(u, z) - Q_{Y|Z}(\overline{u}, z))^\top (u - \overline{u}) \geq 0 \quad \forall u, \overline{u} \in \mathcal{U}, \forall z \in \mathcal{Z}. \tag{2.3}$$

■

Remark 2.2 (**Uniqueness**). In part (i) of the theorem, $u \mapsto Q_{Y|Z}(u, z)$ is equal to a gradient of some convex function $u \mapsto \varphi(u, z)$ for F_U-almost every value of $u \in \mathcal{U}$ and it is unique in the sense that any other map with the same properties will agree with it F_U-almost everywhere. In general, the gradient $u \mapsto \nabla_u \varphi(u, z)$ exists F_U-almost everywhere, and the set of points $\mathcal{U}_e$ where it does not is negligible. Hence the map $u \mapsto Q_{Y|Z}(u, z)$ is still definable at each $u_e \in \mathcal{U}_e$ from the gradient values $\varphi(u, z)$ on $u \in \mathcal{U} \setminus \mathcal{U}_e$, by defining it at each u_e as a smallest-norm element of $\{v \in \mathbb{R}^d : \exists u_k \in \mathcal{U} \setminus \mathcal{U}_e : u_k \to u_e, \nabla_u \varphi(u_k, z) \to v\}$.
■

Let us assume further that the following condition holds:

(C) *For each $z \in \mathcal{Z}$, the distribution $F_{Y|Z}(\cdot, z)$ admits a density $f_{Y|Z}(\cdot, z)$ with respect to the Lebesgue measure on $\mathbb{R}^d$.*

Under this condition we can recover U uniquely in the following sense:

Theorem 2.2 (**Inverse Conditional Vector Quantiles**). *Suppose conditions (N) and (C) holds.*

Then there exists a measurable map $(y, z) \mapsto Q^{-1}_{Y|Z}(y, z)$, mapping $\mathcal{YZ}$ to $\mathbb{R}^d$, such that for each z in $\mathcal{Z}$, the map $y \mapsto Q^{-1}_{Y|Z}(y, z)$ is the inverse of $u \mapsto Q_{Y|Z}(u, z)$ in the sense that:

$$Q^{-1}_{Y|Z}(Q_{Y|Z}(u, z), z) = u,$$

for almost all u under F_U. Furthermore, we can construct U in (2.2) as follows,

$$U = Q^{-1}_{Y|Z}(Y,Z), \text{ and } U \mid Z \sim F_U. \tag{2.4}$$

It is also of interest to state a further implication, which occurs under (N) and (C).

Corollary 2.1 (**Conditional Forward and Backward Monge-Ampère Equations**). *Assume that conditions (N) and (C) hold and, further, that the map $u \mapsto Q_{Y|Z}(u,z)$ is continuously differentiable and injective for each $z \in \mathcal{Z}$. Under this condition, the following conditional forward Monge-Ampère equation holds for all $(u,z) \in \mathcal{UZ}$:*

$$f_U(u) = f_{Y|Z}(Q_{Y|Z}(u,z),z)\, det[\mathrm{D}_u Q_{Y|Z}(u,z)] = \int \delta(u - Q^{-1}_{Y|Z}(y,z)) f_{Y|Z}(y,z)dy, \tag{2.5}$$

where δ is the Dirac delta function in $\mathbb{R}^d$ and $\mathrm{D}_u = \partial/\partial u^\top$. Reversing the roles of U and Y, we also have the following conditional backward Monge-Ampère equation holds for all $(u,z) \in \mathcal{YZ}$:

$$f_{Y|Z}(y,z) = f_U(Q^{-1}_{Y|Z}(y,z))\, det[\mathrm{D}_y Q^{-1}_{Y|Z}(y,z)] = \int \delta(y - Q_{Y|Z}(u,z)) f_U(u)du. \tag{2.6}$$

The latter expression is useful for linking the conditional density function to the conditional vector quantile function and for setting up *maximum likelihood estimation* of conditional quantile functions, which we comment on in the concluding section. Equations (2.5) and (2.6) are *partial differential equations* of the Monge-Ampère type, carrying an additional index $z \in \mathcal{Z}$. These equations could be used directly to solve for conditional vector quantiles given conditional densities. In the next section we describe a variational approach to recovering conditional vector quantiles.

2.2. **Conditional Vector Quantiles as Optimal Transport.**

Under additional moment assumptions, the CVQF can be characterized and even defined as solutions to a regression version of the Monge-Kantorovich-Brenier's optimal transportation problem or, equivalently, a conditional correlation maximization problem.

We assume that the following conditions hold:

(M) The second moment of Y and the second moment of U are finite, namely

$$\int\int \|y\|^2 F_{YZ}(dy,dz) < \infty \text{ and } \int \|u\|^2 F_U(du) < \infty.$$

We consider the following optimal transportation problem with conditional independence constraints:

$$\min_V \{\mathrm{E}\|Y - V\|^2 : V \mid Z \sim F_U\}, \tag{2.7}$$

where the minimum is taken over all random vectors V defined on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Under condition (M) we will see that a *solution* exists and is given by $V = U$ constructed in the previous section.

Remark 2.3 (Matching Factor Interpretation for Latent Vector U). The variational formulation (2.7) immediately provides a useful interpretation for U – they can be thought as latent factors, independent of each other and explanatory variables Z and having a prescribed marginal distribution F_U, and that best explain the variation in Y. Therefore, the conditional vector quantile model (2.2) provides a non-linear latent factor model for Y with factors U solving the matching problem (2.7). This interpretation suggests that this model may be useful in applications which require measurement of multidimensional unobserved factors, for example, cognitive ability, persistence, and various other latent propensities; see, for example, Cunha, Heckman, Schennach (2010) . ■

The problem (2.7) is the conditional version of the classical Monge-Kantorovich problem with Brenier's quadratic costs, which was solved by Brenier in considerable generality in the unconditional case. In the unconditional case, the canonical Monge problem is to transport a pile of coal with mass distributed across production locations via F_U into a pile of coal with mass distributed across consumption locations via F_Y, and it can be rewritten in terms of random variables V and Y. We are seeking to match Y with a version of V that is closest in mean squared sense subject to V having a prescribed distribution. Our conditional version above (2.7) imposes the additional conditional independence constraint $V \mid Z \sim F_U$.

The problem above is equivalent to covariance maximization problem subject to the prescribed conditional independence and distribution constraints:

$$\max_V \{\mathrm{E}(V^\top Y) : \quad V \mid Z \sim F_U\}, \tag{2.8}$$

where the maximum is taken over all random vectors V defined on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$. This type of problem will be convenient for us, as it most directly connects to convex analysis and leads to a convenient dual program. This form also connects to unconditional multivariate quantile maps defined Ekeland, Galichon, Henry (2010), who employed them for purposes of risk analysis. Here, we pursue the conditional version and connect these to the conditional Brenier functions defined in the previous section[4].

The dual program to (2.8) can be stated as:

$$\min_{(\psi,\varphi)} \mathrm{E}(\varphi(V,Z) + \psi(Y,Z)) \ : \ \varphi(u,z) + \psi(y,z) \geq u^\top y, \text{ for all } (z,y,u) \in \mathcal{Z} \times \mathbb{R}^{2d}, \tag{2.9}$$

[4]In the unconditional context, the previous section is noteworthy. Indeed, as in the seminal work of McCann (1995), it leads to a more general definition of vector/multivariate quantiles in that it does not require moment assumptions.

where V is any vector such that $V \mid Z \sim F_U$, and minimization is performed over Borel maps $(y,z) \mapsto \psi(y,z)$ from $\mathcal{Z} \times \mathbb{R}^d$ to $\mathbb{R} \cup \{+\infty\}$ and $(u,z) \mapsto \varphi(z,u)$ from $\mathcal{Z} \times \mathbb{R}^d$ to $\mathbb{R} \cup \{+\infty\}$, where $y \mapsto \psi(y,z)$ and $u \mapsto \varphi(u,z)$ are lower-semicontinuous for each value $z \in \mathcal{Z}$.

Theorem 2.3 (**Conditional Vector Quantiles as Optimal Transport**). *Suppose conditions (N), (C), and (M) hold.*

(i) There exists a pair of maps $(u,z) \mapsto \varphi(u,z)$ and $(y,z) \mapsto \psi(y,z) = \varphi^*(y,z)$, each mapping from $\mathbb{R}^d \times \mathcal{Z}$ to $\mathbb{R}$, that solve the problem (2.9). For each $z \in \mathcal{Z}$, the maps $u \mapsto \varphi(u,z)$ and $y \mapsto \varphi^*(y,z)$ are convex and are Legendre transforms of each other:

$$\varphi(u,z) = \sup_{y \in \mathbb{R}^d} \{u^\top y - \varphi^*(y,z)\}, \quad \varphi^*(y,z) = \sup_{u \in \mathbb{R}^d} \{u^\top y - \varphi(u,z)\},$$

for all $(u,z) \in \mathcal{UZ}$ and $(y,z) \in \mathcal{YZ}$.

(iii) We can take the gradient $(u,z) \mapsto \nabla_u \varphi(u,z)$ of $(u,z) \mapsto \varphi(u,z)$ as the conditional vector quantile function, namely, for each $z \in \mathcal{Z}$, $Q_{Y|Z}(u,z) = \nabla_u \varphi(u,z)$ for almost every value u under F_U.

(iv) We can take the gradient $(y,z) \mapsto \nabla_y \varphi^*(y,z)$ of $(y,z) \mapsto \varphi^*(y,z)$ as the conditional inverse vector quantile function, namely, for each $z \in \mathcal{Z}$, $Q^{-1}_{Y|Z}(y,z) = \nabla_y \varphi(z,y)$ for almost every value y under $F_{Y|Z}(\cdot, z)$.

(v) The vector $U = Q^{-1}_{Y|Z}(Y,Z)$ is a solution to the primal problem (2.8) and is unique in the sense that any other solution U^* obeys $U^* = U$ almost surely under P. *The primal (2.8) and dual (2.9) have the same value.*

(vi) The maps $u \mapsto \nabla_u \varphi(u,z)$ and $y \mapsto \nabla \varphi^*_y(y,z)$ are inverses of each other: for each $z \in \mathcal{Z}$, and for almost every u under F_U and almost every y under $F_{Y|Z}(\cdot, z)$

$$\nabla_y \varphi^*(\nabla_u \varphi(u,z), z) = u, \quad \nabla_u \varphi(\nabla_y \varphi^*(y,z), z) = y.$$

This theorem provides a number of analytical properties, formalizing the variational interpretation of conditional vector quantiles, providing the potential functions $(u,z) \mapsto \varphi(u,z)$ and $(y,z) \mapsto \varphi^*(y,z)$, which are mutual Legendre transforms, and whose gradients are the conditional vector quantile functions and its inverse. Intellectually this problem is a conditional generalization of the fundamental results by Brenier as presented in Villani (2005), Theorem 2.12.

Example 2.1 (**Conditional Normal Vector Quantiles**). Here we consider the normal conditional vector quantiles. Consider the case where

$$Y \mid Z \sim N(\mu(Z), \Omega(Z)).$$

Here $z \mapsto \mu(z)$ is the conditional mean function and $z \mapsto \Omega(z)$ is a conditional variance function such that $\Omega(z) > 0$ (in the sense of symmetric matrices) for each $z \in \mathcal{Z}$ with $\mathrm{E}\|\Omega(Z)\| + \mathrm{E}\|\mu(Z)\|^2 < \infty$. The reference distribution is given by $U \mid Z \sim N(0, I)$. Then we have the following conditional vector quantile model:

$$
\begin{aligned}
Y &= \mu(Z) + \Omega^{1/2}(Z)U, \\
U &= \Omega^{-1/2}(Z)(Y - \mu(Z)).
\end{aligned}
$$

Here we have the following conditional potential functions

$$
\begin{aligned}
\varphi(u, z) &= \mu(z)^\top u + \frac{1}{2} u^\top \Omega^{1/2}(z) u, \\
\psi(y, z) &= \frac{1}{2}(y - \mu(z))^\top \Omega^{-1/2}(z)(y - \mu(z)),
\end{aligned}
$$

which are mutual Legendre transforms, and the following conditional vector quantile functions:

$$
\begin{aligned}
Q_{Y|Z}(u, z) &= \nabla_u \varphi(u, z) = \mu(z) + \Omega^{1/2}(z) u, \\
Q^{-1}_{Y|Z}(y, z) &= \nabla_y \psi(y, z) = \Omega^{-1/2}(z)(y - \mu(z)).
\end{aligned}
$$

It is not easy but also not difficult to show that $V = U$ solves the covariance maximization problem (2.8). An interesting feature of this example is that the conditional vector quantile functions are *linear*. ■

3. Vector Quantile Regression

3.1. **Linear Formulation.** Here we use the following notation:

- We let $X = f(Z)$ denote a vector of regressors formed as transformations of Z, such that the first component of X is 1 (intercept term in the model) and such that conditioning on X is equivalent to conditioning on Z. The dimension of X is denoted by p and we shall denote $X = (1, X_{-1})$ with $X_{-1} \in \mathbb{R}^{p-1}$.

In practice, X would often consist of a constant and some polynomial or spline transformations of Z as well as their interactions. Note that conditioning on X is equivalent to conditioning on Z if, for example, a component of X contains a one-to-one transform of Z.

Denote by F_X the distribution function of X and $F_{UX} = F_U F_X$. Let $\mathcal{X}$ denote the support of F_X and $\mathcal{UX}$ the support of F_{UX}. We define *linear vector quantile regression model* (VQRM) as the following linear model of CVQF.

(L) The following linearity condition holds:

$$Y = Q_{Y|X}(U, X) = \beta_0(U)^\top X, \quad U \mid X \sim F_U,$$

where $u \mapsto \beta_0(u)$ is a map from $\mathcal{U}$ to the set $\mathcal{M}_{p\times d}$ of $p \times d$ matrices such that $u \mapsto \beta_0(u)^\top x$ is a monotone, smooth map, in the sense of being a gradient of a convex function:

$$\beta_0(u)^\top x = \nabla_u \Phi_x(u), \quad \Phi_x(u) := (B_0(u)^\top x), \text{ for all } (u, x) \in \mathcal{U}\mathcal{X},$$

where $u \mapsto B_0(u)$ is C^1 map from $\mathcal{U}$ to $\mathbb{R}^d$, and $u \mapsto B_0(u)^\top x$ is a strictly convex map from $\mathcal{U}$ to $\mathbb{R}$.

The parameter $\beta(u)$ is indexed by the quantile index $u \in \mathcal{U}$ and is a $d\times p$ *matrix* of quantile regression coefficients. Of course in the scalar case, when $d = 1$, this matrix reduces to a vector of quantile regression coefficients. This model is a natural analog of the classical QR for scalar Y where the similar regression representation holds. One example where condition (L) holds was Example 2.1, describing the conditional normal vector regression. It is of interest to specify other examples where condition (L) holds or provides a plausible approximation.

Example 3.1 (Saturated Specification)**.** The regressors $X = f(Z)$ with $\mathrm{E}\|f(Z)\|^2 < \infty$ are *saturated* with respect to Z, if, for any $g \in L^2(F_Z)$, we have $g(Z) = X^\top \alpha_g$. In this case the linear functional form (L) is not a restriction. For $p < \infty$ this can occur if and only if Z takes on a finite set of values $\mathcal{Z} = \{z_1, \ldots, z_p\}$, in which case we can write:

$$Q_{Y|X}(u, X) = \sum_{j=1}^{p} Q_{Y|Z}(u, z_j) 1(Z = z_j) =: B_0(u)^\top X,$$

$$B_0(u) := \begin{pmatrix} Q_{Y|Z}(u, z_1)^\top \\ \vdots \\ Q_{Y|Z}(u, z_p)^\top \end{pmatrix}, \quad X := \begin{pmatrix} 1(Z = z_1) \\ \vdots \\ 1(Z = z_p) \end{pmatrix}.$$

Here the problem is equivalent to considering p unconditional vector quantiles in populations corresponding to $Z = z_1, \ldots, Z = z_p$. ■

The rationale for using linear forms is two-fold – one is convenience of estimation and representation of functions and another one is approximation property. We can approximate a smooth convex potential by a smooth linear potential, as the following example illustrates for a particular approximation method.

Example 3.2 (Linear Approximation)**.** Let $(u, z) \mapsto \varphi(u, z)$ be of class C^a with $a > 1$ on the support $(u, z) \in \mathcal{U}\mathcal{Z} = [0, 1]^{d+k}$. Consider a trigonometric tensor product basis of functions $\{(u,z) \mapsto q_j(u)f_l(z), j \in \mathbb{N}, l \in \mathbb{N}\}$ in $L^2[0,1]^{d+k}$. Then there exists a JL vector $(\gamma_{jl} : j \in \{1,...,J\}, l \in \{1,...,L\})$ such that the linear map:

$$(u,z) \mapsto \Phi^{JL}(u,z) := \sum_{j=1}^{J}\sum_{l=1}^{L} \gamma_{jl} q_j(u) f_l(z) =: B_0^L(u)^\top f^L(z),$$

where $B_0^L(u) = (\sum_{j=1}^J \gamma_{jl} q_j(u), l \in \{1,...,L\})$ and $f^L(z) = (f_l(z), l \in \{1,...,L\})$, provides uniformly consistent approximation of the potential and its derivative:

$$\lim_{J,L\to\infty} \sup_{(u,z)\in\mathcal{U}\mathcal{Z}} \Big(|\varphi(u,z) - \Phi^{JL}(u,z)| + \|\nabla_u \varphi(u,z) - \nabla_u \Phi^{JL}(u,z)\| \Big) = 0.$$

■

The approximation property provides a rationale for the linear specification (1.3). If the linear specification does not hold we say that the model is *misspecified*. If the model is flexible enough, then the approximation error is small, and we effectively ignore the error when assuming (1.3). However, when constructing a sensible estimator we must allow the possibility that the model is misspecified, which means we can't really force (1.3) onto data. Our proposal for estimation presented next does not force (1.3) onto data, but if (1.3) is true, then the true conditional vector quantile function would be recovered perfectly as a result (in population).

3.2. **Linear Program for VQR.** Our approach to multivariate quantile regression is based on the multivariate extension of the covariance maximization problem with a mean-independence constraint:

$$\max_V \{\mathrm{E}(V^\top Y) : \ V \sim F_U, \ \mathrm{E}(X \mid V) = \mathrm{E}(X)\}. \tag{3.1}$$

Note that the constraint condition is a relaxed form of the previous independence condition.

Remark 3.1. The new condition $V \sim F_U$, $\mathrm{E}(X \mid V) = \mathrm{E}(X)$ is weaker than $V \mid X \sim F_U$, but the two conditions coincide if X is saturated relative to Z, as in Example 3.1, in which case $\mathrm{E}(g(Z)V) = \mathrm{E}X'\alpha_g V = \mathrm{E}(X'\alpha_g)\mathrm{E}(V) = \mathrm{E}g(Z)\mathrm{E}V$ for every $g \in L^2(F_Z)$. More generally, this example suggests that the richer X is, the closer the mean-independence condition becomes to the conditional independence. ■

The relaxed condition is sufficient to guarantee that the solution exists not only when (L) holds, but more generally when the following quasi-linear assumption holds.

(QL) We have a quasi-linear representation a.s.

$$Y = \beta(\widetilde{U})^\top X, \quad \widetilde{U} \sim F_U, \quad \mathrm{E}(X \mid \widetilde{U}) = \mathrm{E}(X),$$

where $u \mapsto \beta(u)$ is a map from $\mathcal{U}$ to the set $\mathcal{M}_{p\times d}$ of $p \times d$ matrices such that $u \mapsto \beta(u)^\top x$ is a gradient of convex function for each $x \in \mathcal{X}$ and a.e. $u \in \mathcal{U}$:

$$\beta(u)^\top x = \nabla_u \Phi_x(u), \quad \Phi_x(u) := (B(u)^\top x),$$

where $u \mapsto B(u)$ is C^1 map from $\mathcal{U}$ to $\mathbb{R}^d$, and $u \mapsto B(u)^\top x$ is a strictly convex map from $\mathcal{U}$ to $\mathbb{R}$.

This condition allows for a degree of misspecification, which allows for a latent factor representation where the latent factor obeys the relaxed independence constraints.

Theorem 3.1. *Suppose conditions (M), (N) , (C), and (QL) hold.*

(i) The random vector $\widetilde{U}$ entering the quasi-linear representation (QL) solves (3.1).

(ii) The quasi-linear representation is unique a.s. that is if we also have $Y = \overline{\beta}(\overline{U})^\top X$ with $\overline{U} \sim F_U, \mathrm{E}(X \mid \overline{U}) = \mathrm{E}X$, $u \mapsto X^\top \overline{\beta}(u)$ is a gradient of a strictly convex function in $u \in \mathcal{U}$ a.s., then $\overline{U} = \widetilde{U}$ and $X^\top \beta(\widetilde{U}) = X^\top \overline{\beta}(\widetilde{U})$ a.s.

(iii) Under condition (L) and assuming that $\mathrm{E}(XX^\top)$ has full rank, $\widetilde{U} = U$ a.s. and U solves (3.1). Moreover, $\beta_0(U) = \beta(U)$ a.s.

The last assertion is important – it says that if (L) holds, then the linear program with the *relaxed* independence constraint will find the true linear vector quantile regression in the population.

3.3. **Dual Program for Linear VQR.** As explained in details in the appendix, the dual of (3.1) reads

$$\inf_{(\psi,b)} \mathrm{E}(\psi(X,Y)) + \mathrm{E}b(V)^\top \mathrm{E}(X) \;:\; \psi(x,y) + b(u)^\top x \geq u^\top y, \; \forall\; (y,x,u) \in \mathcal{Y}\mathcal{X}\mathcal{U}, \tag{3.2}$$

where $V \sim F_U$, where the infimum is taken over all continuous functions $(y,x) \mapsto \psi(y,x)$, mapping $\mathcal{Y}\mathcal{X}$ to $\mathbb{R}$ and $u \mapsto b(u)$ mapping $\mathcal{U}$ to $\mathbb{R}$, such that $\mathrm{E}(\psi(X,Y))$ and $\mathrm{E}b(V)$ are finite.

Since for fixed b, the largest ψ which satisfies the pointwise constraint in (3.2) is given by

$$\psi(x,y) := \sup_{u \in \mathcal{U}} \{u^\top y - b(u)^\top x\},$$

one may equivalently rewrite (3.2) as the minimization over continuous b of

$$\int \sup_{u\in\mathcal{U}}\{u^\top y - B(u)^\top x\} F_{Y|X}(dx,dy) + \int b(u)^\top E(X) F_U(du).$$

By standard arguments, the infimum over continuous functions coincides with the one over smooth or simply integrable functions.

Theorem 3.2. *Under (M) and (QL), we have that the optimal solution to the dual is given by functions:*

$$\psi(x,y) = \sup_{u\in\mathcal{U}}\{u^\top y - B(u)^\top x\}, \quad b(u) = B(u).$$

This result can be recognized as a consequence of strong duality of the linear programming.

3.4. **The dual and the primal relations in general.** In this subsection we do not assume condition (QL) and wish to study which information (3.1) can give regarding the dependence of X and Y. Once again, a good starting point is duality. Without (QL), the existence of optimal functions ψ and B is *not* obvious, and is proven under the following assumptions:

(G) The support of $W = (X_{-1}, Y)$, say $\mathcal{W}$, is a closure of an open bounded convex subset of $\mathbb{R}^{p-1+d}$, the density f_W of W is uniformly bounded from above and does not vanish anywhere on the interior of $\mathcal{W}$. The set $\mathcal{U}$ is a closure of an open bounded convex subset of $\mathbb{R}^d$, and the density f_U is strictly positive over $\mathcal{U}$.

Theorem 3.3. *Suppose that condition (G) holds.*

(i) The dual problem (3.2) admits at least a solution (ψ, B) such that

$$\psi(x,y) = \sup_{u\in\mathcal{U}}\{u^\top y - B(u)^\top x\}.$$

(ii) Moreover, under the same condition, the map

$$(x,y) \mapsto \psi(x,y),$$

is differentiable for F_{YX}-almost every $(y,x) \in \mathcal{Y}\mathcal{X}$.

The first result is a consequence of a non-trivial argument given in the proof. The second is a consequence of generalized envelope theorems by Milgrom and Segal (2002).

We next discuss the implications of the existence of the solution to a dual program and of smoothness conditions of maps ψ and B.

Suppose that $\widetilde{U}$ solves (3.1) and (ψ, B) solves its dual (3.2). Recall that, without loss of generality, we can take ψ as a function given by

$$\psi(x,y) = \sup_{u\in\mathcal{U}}\{u^\top y - B(u)^\top x\}. \tag{3.3}$$

Observe that $(x,y) \mapsto \psi(x,y)$ is convex in both arguments. The primal-dual relations give

$$\psi(x,y) + B(u)^\top x \geq u^\top y \text{ for all } (x,y,u) \in \mathcal{X}\mathcal{Y}\mathcal{U},$$

and almost-surely

$$\psi(X,Y) + B(\widetilde{U})^\top X = \widetilde{U}^\top Y.$$

Since ψ is convex and is given by (3.3), it follows that we have:

$$(-B(\widetilde{U}), \widetilde{U}) \in \partial\psi(X,Y), \text{ or, equivalently, } (X,Y) \in \partial\psi^*(-B(\widetilde{U}), \widetilde{U}).$$

If $(x,y) \mapsto \psi(x,y)$ is differentiable in both arguments for F_{YX}-a.e. $(x,y) \in \mathcal{Y}\mathcal{X}$, then we have that almost-surely:

$$\widetilde{U} = \nabla_y\psi(X,Y),\ -B(\widetilde{U}) = \nabla_x\psi(X,Y).$$

Problems (3.1) and (3.2) have thus enabled us to find:

- $\widetilde{U} \sim F_U :\quad \mathrm{E}(X \mid \widetilde{U}) = 0$ and ψ convex,
- such that $(X,Y) \in \partial\psi^*(-b(\widetilde{U}), \widetilde{U})$.

Quasi-linear specification of vector quantile regression rather asks whether we can write $Y = \nabla_u(B(\widetilde{U})^\top X) := \nabla_u\Phi_X(\widetilde{U})$ with $u \mapsto \Phi_x(u) := B(\widetilde{u})^\top x$ is convex in u for fixed x. We now wish to explain that smoothness of $u \mapsto B(\widetilde{u})$ is tightly related to quasi-linearity.

Note that since ψ is smooth, and if B is smooth, then ψ solves the vectorial Hamilton-Jacobi equation:

$$\nabla_x\psi(x,y) + B(\nabla_y\psi(x,y)) = 0. \tag{3.4}$$

In addition, if B are smooth then, by the envelope theorem,

$$Y = \nabla_u B(\widetilde{U})^\top X = \nabla_u\Phi_X(\widetilde{U}).$$

We then see that φ and B are consistent with vector quantile regression estimation, but we are still short of the convexity requirement. The lemma below shows that convexity holds over support of $\widetilde{U}$ given $X = x$. However, this convexity is weaker than condition (QL), which imposes convexity over entire $u \in \mathcal{U}$, the marginal support of $\widetilde{U}$.

Lemma 3.1. *Suppose that the dual program admits a solution, then*

$$\Phi_X(\widetilde{U}) = \Phi_X^{**}(\widetilde{U})$$

P-almost surely, where Φ_x^{**} denotes the convex envelope of Φ_x.

3.5. **Connecting to Scalar Quantile Regression.** We consider the connection to the canonical, scalar quantile regression problem, where Y is scalar and for each probability index t, the linear functional form $x \mapsto x^\top \beta(t)$ is used. Koenker and Bassett define linear quantile regression as $X^\top \beta(t)$ with $\beta(t)$ solving the minimization problem

$$\beta(t) \in \arg\min_{\beta \in \mathbb{R}^p} \mathrm{E}(\rho_t(Y - X^\top \beta)), \tag{3.5}$$

where the loss function ρ_t is given by $\rho_t(z) := tz_- + (1-t)z_+$ with z_- and z_+ denoting as usual the negative and positive parts of z. The above formulation makes sense and $\beta(t)$ is unique under the following simplified conditions:

(QR) $\mathrm{E}|Y| < \infty$, $(y, x) \mapsto f_{Y|X}(y, x)$ is uniformly continuous, and $\mathrm{E}(wXX^\top)$ is positive-definite, for $w = f_{Y|X}(X'\beta(t), X)$.

For further use, note that (3.5) can be conveniently rewritten as

$$\min_{\beta \in \mathbb{R}^p} \{\mathrm{E}(Y - X^\top \beta)_+ + (1-t)\mathrm{E}X^\top \beta\}. \tag{3.6}$$

Koenker and Bassett showed that this convex program admits as dual formulation

$$\max\{\mathrm{E}(A_t Y) \ : \ A_t \in [0,1], \ \mathrm{E}(A_t X) = (1-t)\mathrm{E}X\}. \tag{3.7}$$

An optimal $\beta = \beta(t)$ for (3.6) and an optimal rank-score variable A_t in (3.7) may be taken to be

$$A_t = 1(Y > X^\top \beta(t)), \tag{3.8}$$

and thus the constraint $\mathrm{E}(A_t X) = (1-t)\mathrm{E}X$ reads:

$$\mathrm{E}(1(Y > X^\top \beta(t))X) = (1-t)\mathrm{E}X. \tag{3.9}$$

which simply are the first-order conditions for (3.6).

We say that the specification of quantile regression is quasi-linear if

$$t \mapsto x^\top \beta(t) \text{ is increasing on } (0,1). \tag{3.10}$$

Define the rank variable $\widetilde{U} = \int_0^1 A_t dt$, then under this assumption we have that

$$A_t = 1(\widetilde{U} > t),$$

and the first-order conditions imply that for each $t \in (0,1)$

$$\mathrm{E}1(\widetilde{U} \geq t) = (1-t), \quad \mathrm{E}1(\widetilde{U} \geq t)X = (1-t)\mathrm{E}X.$$

The first property implies that $\widetilde{U} \sim U(0,1)$ and the second property can be easily shown to imply the *mean-independence* condition:

$$\mathrm{E}(X \mid \widetilde{U}) = \mathrm{E}X.$$

Thus quantile regression *naturally leads* to the mean-independence condition and the quasi-linear latent factor model. This is the reason we used mean-independence condition as a starting point in formulating the vector quantile regression. Moreover, in both vector and scalar cases, we have that, when the conditional quantile function is linear (not just quasi-linear), the quasi-linear representation coincides with the linear representation and $\widetilde{U}$ becomes fully independent of X.

The following result explains the connection more formally.

Theorem 3.4 (**Connection to Scalar Quantile Regression**). *Suppose that (QR) hods.*

(i) If (3.10) holds, then for $\widetilde{U} = \int_0^1 A_t dt$ we have the quasi-linear model holding

$$Y = X^\top \beta(\widetilde{U}) \text{ a.s., } \widetilde{U} \sim U(0,1) \text{ and } \mathrm{E}(X \mid \widetilde{U}) = \mathrm{E}(X).$$

Moreover, $\widetilde{U}$ solves the correlation maximization problem with a mean-independence constraint:

$$\max\{\mathrm{E}(VY) : \ V \sim U(0,1), \ \mathrm{E}(X \mid V) = \mathrm{E}(X)\}. \tag{3.11}$$

(ii) The quasi-linear representation above is unique almost surely. That is, if we also have $Y = \overline{\beta}(\overline{U})^\top X$ with $\overline{U} \sim U(0,1), \mathrm{E}(X \mid \overline{U}) = \mathrm{E}X$, $u \mapsto X'\overline{\beta}(u)$ is increasing in $u \in (0,1)$ a.s., then $\widetilde{U} = \overline{U}$ and $X'\beta(\widetilde{U}) = X'\overline{\beta}(\overline{U})$ a.s.

(iii) Consequently, if the conditional quantile function is linear, namely $Q_{Y|X}(u) = X^\top \beta_0(u)$, so that $Y = X^\top \beta_0(U)$, then the latent factors in the quasi-linear and linear specifications coincide, namely $U = \widetilde{U}$, and so do the model coefficients, namely $\beta_0(U) = \beta(U)$.

4. Implementation of Vector Quantile Regression

In practice to implement computation or estimation, we shall have to discretize the problems. In this section, we shall specialize the previous results to the case when the distribution ν of (X, Y), and the distribution μ of U, are discrete. The primarily case we have in mind is the case when an empirical sample of the distribution of (X, Y) is observed, and when the distribution of U is approximated on a grid, but our setting allows to cover more general cases, hence we shall allow for nonuniform weights ν_j of observations (x_j, y_j), as well as irregular grids for the sample points of u_k and nonuniform weights μ_k.

The observations are response variables $y_j \in \mathbb{R}^d$ and regressors $x_j \in \mathbb{R}^p$ $(1 \leq j \leq n)$. It is assumed that the constant is included as a regressor, thus one assumes that the first entry of x_j is equal to one, that is $x_{j1} = 1$. Let Y be the $n \times d$ matrix of row vectors y_j and X the $n \times r$ matrix of row vectors x_j. By assumption, the first column of X is thus made of ones. Let ν be a $n \times 1$ matrix such that ν_j is the probability of observation (X_j, Y_j) (hence

$\nu_j \geq 0$ and $\sum_j \nu_j = 1$). Let m be the number of points of U. Let U be a $m \times d$ matrix, where the ith row stands for vector $u_i \in \mathbb{R}^d$. Let μ be a $m \times 1$ matrix such that μ_i is the probability weight of u_i (hence $\mu_i \geq 0$ and $\sum_k \mu_k = 1$).

One looks for a $m \times n$ matrix π such that π_{ij} is the joint probability of (u_i, X_j, Y_j) which maximizes

$$\mathrm{E}_\pi \left[U^\top Y \right] = \sum_{ij} \pi_{ij} y_j^\top u_i = \mathrm{Tr}(U' \pi Y)$$

subject to constraint $(X, Y) \sim \nu$, which rewrites $\pi' 1_m = \nu$, where 1_m is a $m \times 1$ vector of ones, and subject to constraints $U \sim \mu$ and $\mathrm{E}\left[X \mid U\right] = \mathrm{E}\left[X\right]$, which rewrites $\pi X = \mu \mathrm{E}[X]$, where $\overline{x} = \nu' X$ is a $1 \times r$ vector.

Hence, the μ-QR problem in the discrete case is given in its primal form by

$$\begin{aligned} &\max_{\pi \geq 0} \mathrm{Tr}\left(U' \pi Y\right) \\ &\pi' 1_m = \nu \;\; [\psi] \\ &\pi X = \mu \mathrm{E}[X] \;\; [b] \end{aligned} \tag{4.1}$$

where the square brackets show the associated Lagrange multipliers and in its dual form by

$$\begin{aligned} &\min_{\psi, b} \psi' \nu + \mathrm{E}[X] b' \mu \\ &\psi 1_m' + X b' \geq Y U' \;\; \left[\pi'\right] \end{aligned} \tag{4.2}$$

where ψ is a $n \times 1$ vector, and b is a $m \times r$ matrix.

Problems (4.1) and (4.2) are two linear programming problems dual to each other. However, in order to implement them on standard numerical analysis softwares such as R or Matlab coupled with a Linear Programming software such as Gurobi, we need to convert matrices into vectors. This is done using the *vec* operation, which is such that if A is a $p \times q$ matrix, $vec(A)$ is a $pq \times 1$ matrix such that $vec\,(A)_{i+p(j-1)} = A_{ij}$. The use of the Kronecker product will also greatly facilitate computations. Recall that if A is a $p \times q$ matrix and B is a $p' \times q'$ matrix, then the Kronecker product $A \otimes B$ is the $pp' \times qq'$ matrix such that for all relevant choices of indices i, j, k, l,

$$(A \otimes B)_{i+p(k-1), j+q(l-1)} = A_{ij} B_{kl}.$$

The fundamental property linking Kronecker products and the *vec* operator is

$$vec\left(B X A^T\right) = (A \otimes B)\, vec\,(X)\,.$$

Introduce $vec\pi = vec\,(\pi)$, the optimization variable of the "vectorized problem". Note that the variable $vec\pi$ is a $mn \times 1$ vector. The objective function rewrites $\mathrm{Tr}\left(U' \pi Y\right) =$

$vec\pi' vec\,(UY')$; as for the constraints, $vec\,(1_m'\pi) = (I_n \otimes 1_m')\, vec\pi$ is a $n \times 1$ vector; and $vec\,(\pi X) = (X' \otimes I_m)\, vec\pi$ is a $mr \times 1$ vector. Thus we can rewrite the linear VQR program as:

$$\begin{aligned} &\max_{vec\pi \geq 0} vec\,(UY')'\, vec\pi \\ &(I_n \otimes 1_m')\, vec\pi = vec\,(\nu') \\ &(X' \otimes I_m)\, vec\pi = vec\,(\mu \mathrm{E}[X]) \end{aligned} \tag{4.3}$$

which is a LP problem with mn variables and $mr+n$ constraints. The constraints $(I_n \otimes 1_m')$ and $(X' \otimes I_m)$ are very sparse, which can be taken advantage of from a computational point of view.

5. Empirical Illustration

We demonstrate the use of the approach on a classical application of Quantile Regression since Koenker and Bassett (1982): Engel's (1857) data on household expenditures, including 199 Belgian working-class households surveyed by Ducpetiaux in 1855, and 36 observations from all over Europe surveyed by Le Play on the same year. Due to the univariate nature of classical QR, Koenker and Bassett limited their focus on the regression of food expenditure over total income. But in fact, as seen in Figure 1 Engel's dataset is richer and classifies household expenses in nine broad categories: 1. Food; 2. Clothing; 3. Housing; 4. Heating and lighting; 5. Tools; 6. Education; 7. Safety; 8. Medical care; and 9. Services. This allows us to have a multivariate dependent variable. While we could in principle have $d = 9$, we focus for illustrative purposes on a two-dimensional dependent variable ($d = 2$), and we choose to take Y_1 as food expenditure (category #1) and Y_2 as housing and domestic fuel expenditure (category #2 plus category #4). We take $X = (X_1, X_2)$ with $X_1 = 1$ and X_2=total expenditure as an explanatory variable. Descriptive statistics are offered in Table 1.

Table 1. Descriptive statistics

	Minimum	**Maximum**	**Median**	**Average**
Food	242,32	2032,68	582,54	624,15
Shelter+Fuel	11,17	660,24	113,36	136,62
Clothing	5,00	520,00	111,76	135,54
Else	0,00	1184,40	39,50	69,21
Total	377,06	4957,83	883,99	982,47

11

A. Budgets der Ausgaben mitgetheilt von Ducpetiaux.

Localitäten. — Bezeichnung der Arbeiter.	Kategorie der Arbeiter-Familien.	Ausgaben für: 1. Nahrung.	2. Kleidung.	3. Wohnung.*	4. Heizung und Beleuchtung.	5. Geräthe etc.	6. Erziehung etc.	7. öffentliche Sicherheit etc.	8. Gesundheitspflege etc.	9. persönliche Dienstleistungen.	Ueberhaupt.
		Fr.	Fr.	Fr.	Fr.	Fr.	Fr.	Fr.	Fr.	Fr.	Fr.
I. Provinz Brabant.											
lt Nivelles:											
Tagelöhner	1	576,28	129,34	78,32	46,80	.	.	.	5,20	.	835,94
Weber	2	631,80	83,00	79,80	73,84	.	5,00	.	.	.	873,74
Schieferbrecher	3	608,64	134,00	108,00	78,00	.	5,00	.	17,80	.	951,44

FIGURE 1. An exerpt from Engel's [26] compilation of Ducpetiaux's data on household expenditure. The total expendable income is broken down into nine categories: 1 Food (Nahrung) 2. Clothing (Kleidung) 3. Housing (Wohnung) 4. Heating and Lighting (Heizung und Beleuchtung) 5. Tools (Geräthe) 6. Education (Erziehung) 7. Public safety (Öffentliche Sicherheit) 8. Health (Gesundheitspflege) 9. Personal service (persönliche Dienstleistungen).

Our first focus is the comparison between the classical quantile regression (in red) and vector quantile regression (in green) when $d = 1$. Figure 2 compares these curves, both Y_1 (left panel) and for Y_2 (right panel) as dependent variables, and for five quartiles of income (0%, 25%, 50%, 75%, 100%). In each case, the curves exhibit very little difference between classical quantile regression and vector quantile regression.

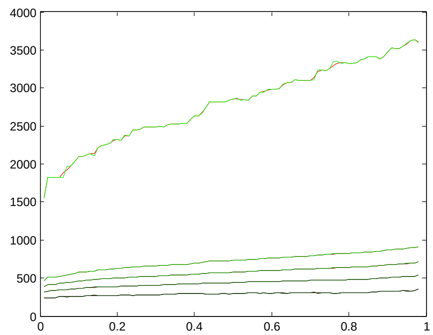
The representation obtained by the classical quantile regression taking in turns Y_1 and Y_2 as the dependent variable gives

$$Y_1 = \beta_1 (U)^\top X \text{ and } Y_2 = \beta_2 (U)^\top X \tag{5.1}$$

with $U \sim \mathcal{U}([0,1])$. In contrast, the two-dimensional vector quantile regression with $Y = (Y_1, Y_2)$ as a dependent variable yields a representation

$$Y_1 = \frac{\partial b}{\partial u_1} (U_1, U_2)^\top X \text{ and } Y_2 = \frac{\partial b}{\partial u_2} (U_1, U_2)^\top X \tag{5.2}$$

where $(U_1, U_2) \sim F_U = U([0,1]^2)$. U_1 and U_2 have an interesting interpretation. U_1 is a *propensity for food expenditure*, while U_2 is a *propensity for domestic (housing and heating) expenditure*. The quantity $U(x, y) = \mathbb{E}[U|X = x, Y = y]$ is a measure of joint propensity of observation $Y = y$ conditional on $X = x$.



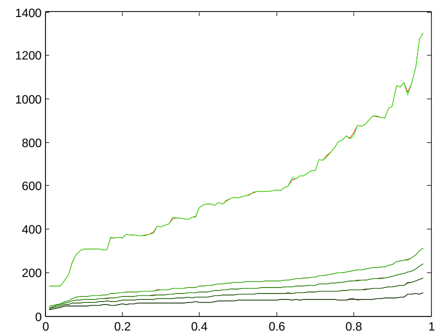


FIGURE 2. Classical quantile regression (red) and one-dimensional vector quantile regression (green) with income as explanatory variable and with: (i) Food expenditure as dependent variable (Left) and (ii) Housing expenditure as dependent variable (Right). The maps $u \to \beta(u)^\top x$ are plotted for five values of income x (quartiles).

In Figure 3, we set $x = (1,\ 883.99)$, where $x_2 = 883.99$ is the median value of the total expenditure X_2, and we draw the corresponding representations (5.1) and (5.2).

Under specification $Y_j = \sum_i \beta_{ij}(U) X_i$, thus, as $X_1 = 1$, $Y_j = \alpha_j(U) + \sum_{i\neq 1} \beta_{ij}(U)(X_i - \overline{x}_i)$, where

$$\alpha_j(u) = \beta_{1j}(u) + \sum_{i\neq 1} \beta_{ij}(u)\,\overline{x}_i$$

which we denote

$$Y = \alpha(U) + \beta_{-1}(U)^\top (X - \overline{x}).$$

Hence, under specification, $\alpha(u) = \mathbb{E}[Y|X = \overline{x}, U = u]$. Figure 4 provides a plot of $u \to \alpha_1(u)$ (top left), $u \to \alpha_2(u)$ (bottom left), $u \to \beta_{21}(u)$ (top right), and $u \to \beta_{22}(u)$ (bottom right).

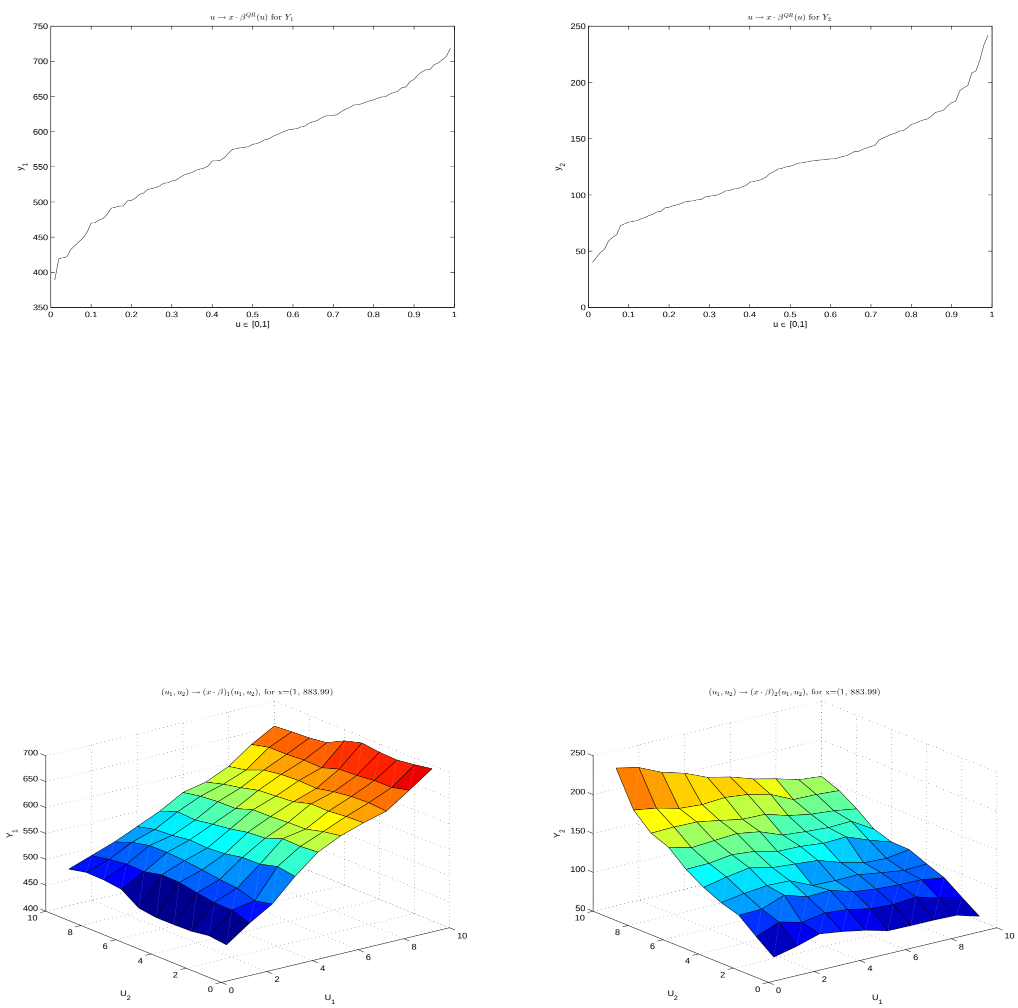


FIGURE 3. Predicted outcome conditional on total expenditure equal to median value, that is $X_2 = 883.99$. Left: food expenditure, Right: housing expenditure. Top: as predicted by classical quantile regression. Bottom: as predicted by vector quantile regression.

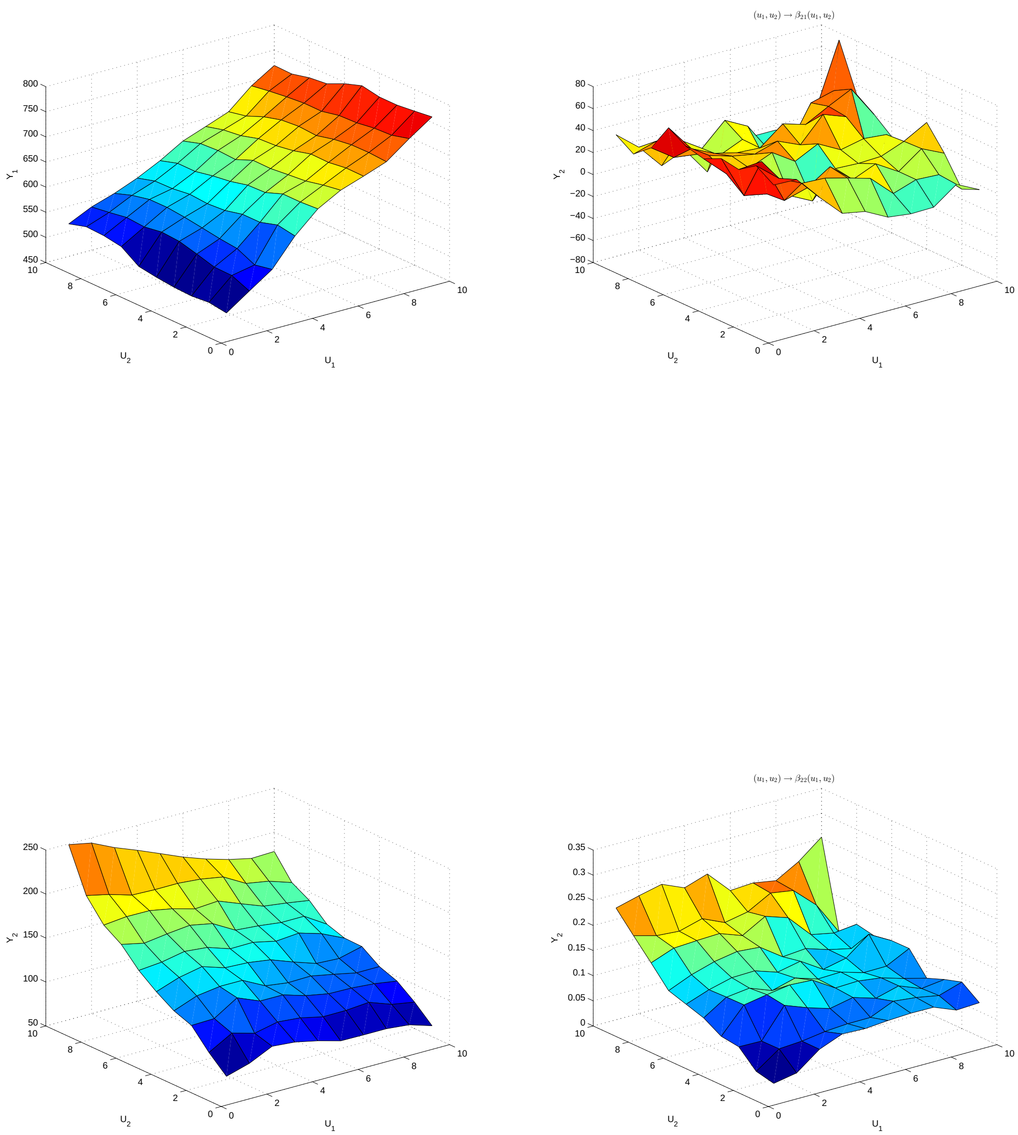


FIGURE 4. Plots of $u \to \alpha(u) + \beta_2(u)$. Left: α ; right: β_2. Top: first dimension; bottom: second dimension.

Appendix

Appendix A. Proofs for Section 2

A.1. **Proof of Theorem 2.1.** The first assertion of the theorem is a consequence of the refined version of Brenier's theorem given by McCann (1995) (as, e.g, stated in Villani (2003), Theorem 2.32), which we apply for each $z \in \mathcal{Z}$. In particular, this implies that for each $z \in \mathcal{Z}$, the map $u \mapsto Q_{Y|Z}(u, z)$ is measurable.

Next we note that $(Q_{Y|Z}(V, Z), Z)$ is a proper random vector, hence a measurable map from $(\Omega, \mathcal{A})$ to $(\mathbb{R}^{d+k}, \mathcal{B}(\mathbb{R}^{d+k}))$. For any rectangle $A \times B \subset \mathbb{R}^{d+k}$:

$$\begin{aligned} \mathrm{P}((Y, Z) \in A \times B) &= \int_B \left[\int_A F_{Y|Z}(dy, z) \right] F_Z(dz) && \text{(A.1)} \\ &= \int_B \left[\int 1\{(Q_{Y|Z}(u, z) \in A\} F_U(du) \right] dF_Z(dz) && \text{(A.2)} \\ &= \mathrm{P}((Q_{Y|Z}(V, Z), Z) \in A \times B), && \text{(A.3)} \end{aligned}$$

where penultimate equality follows from the previous paragraph. Since measure over rectangles uniquely pins down the probability measure on all Borel sets via Caratheodory's extension theorem, it follows that the law of $(Q_{Y|Z}(V, Z), Z)$ is properly defined on all Borel sets and is equal to that of (Y, Z).

Note that since the above argument works for every (V, Z) with support on $\mathcal{UZ}$, we can select (V, Z) such that the sigma-field generated by these variables coincides with the Borel sigma field $\mathcal{B}(\mathcal{UZ})$ on $\mathcal{UZ}$ generated from the open sets of the topology defined relative to $\mathbb{R}^{d+k}$. Hence it must be that the map $(u, z) \mapsto (Q_{Y|Z}(u, z), z)$ is measurable.

To show the second assertion we invoke Dudley-Phillip's (1983) coupling result given in their Lemma 2.11.

Lemma A.1 (Dudley-Phillip's coupling)**.** *Let S and T be Polish spaces and Q a law on $S \times T$, with marginal law μ on S. Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space and J a random variable on Ω with values in S and $J \sim \mu$. Assume there is a random variable W on Ω, independent of J, with values in a Polish space R and law ν on R having no atoms. Then there exists a random variable I on Ω with values in T such that $(J, I) \sim Q$.*

First we recall that our probability space has the form:

$$(\Omega, \mathcal{A}, \mathrm{P}) = (\Omega_0, \mathcal{A}_0, \mathrm{P}_0) \times (\Omega_1, \mathcal{A}_1, \mathrm{P}_1) \times ((0, 1), B(0, 1), \mathrm{Leb}),$$

where $(0, 1), B(0, 1), \mathrm{Leb})$ is the canonical probability space, consisting of the unit segment of the real line equipped with Borel sets and the Lebesgue measure. We use this canonical space to carry W, which is independent of any other random variables appearing below, and which has the uniform distribution on $R = [0,1]$. The space $R = [0,1]$ is Polish and the distribution of W has no atoms.

Next we apply the lemma to $J = (Y, Z)$ to show existence of $I = U$, where both J and I live on the probability space $(\Omega, \mathcal{A}, \mathrm{P})$ and that obeys the second assertion of the theorem. The variable J takes values in the Polish space $S = \mathbb{R}^d \times \mathbb{R}^k$, and the variable I takes values in the Polish space $T = \mathbb{R}^d$.

Next we describe a law Q on $S \times T$ by defining a triple (Y^*, Z^*, U^*) that lives on a suitable probability space. We consider a random vector Z^* with distribution F_Z, a random vector $U^* \sim F_U$, independently distributed of Z^*, and $Y^* = Q_{Y|Z}(U^*, Z^*)$ uniquely determined by the pair (U^*, Z^*), which completely characterizes the law Q of (Y^*, Z^*, U^*). In particular, the triple obeys $Z^* \sim F_Z$, $U^*|Z^* \sim F_U$ and $Y^* \mid Z^* = z \sim F_{Y|Z}(\cdot, z)$. Moreover, the set $\{(y^*, z^*, u^*) : \|y^* - Q_{Y|Z}(u, z^*)\| = 0\} \subset S \times T$ is assigned probability mass 1 under Q.

By the lemma quoted above, given J, there exists an $I = U$, such that $(J, I) \sim Q$, but this implies that $U|Z \sim F_U$ and that $\|Y - Q_{Y|Z}(U, Z)\| = 0$ with probability 1 under P. ■

A.2. **Proof of Theorem 2.2.** We condition on $Z = z$. By reversing the roles of V and Y, we can apply Theorem 2.1 to claim that there exists a map $y \mapsto Q^{-1}_{Y|Z}(y, z)$ with the properties stated in the theorem such that $Q^{-1}_{Y|Z}(Y, z)$ has distribution function F_U, conditional on $Z = z$. Hence for any test function $\xi : \mathbb{R}^d \to \mathbb{R}$ such that $\xi \in C_b(\mathbb{R}^d)$ we have

$$\int \xi(Q^{-1}_{Y|Z}(Q_{Y|Z}(u, z), z)) F_U(du) = \int \xi(u) F_U(du).$$

This implies that for F_U-almost every u, we have

$$Q^{-1}_{Y|Z}(Q_{Y|Z}(u, z), z) = u.$$

Hence P-almost surely

$$Q^{-1}_{Y|Z}(Y, Z) = Q^{-1}_{Y|Z}(Q_{Y|Z}(U, Z), Z) = U.$$

Thus we can set $U = Q^{-1}_{Y|Z}(Y, Z)$ P-almost surely in Theorem 2.1. ■

A.3. **Proof of Theorem 2.3.** The result follows from Villani (2005), Theorem 2.12. ■

Appendix B. Proofs for Section 3

B.1. **Proof of Theorem 3.1.** We first establish part(i). We have a.s.

$$Y = \nabla \Phi_X(\widetilde{U}), \text{ with } \Phi_X(u) = B(u)^\top X.$$

For any $V \sim F_U$ such that $\mathrm{E}(X|V) = E(X)$, and $\Phi_x^*(y) := \sup_{v\in\mathcal{U}}\{v^\top y - \Phi_x(v)\}$, we have

$$\mathrm{E}[\Phi_X(V) + \Phi_X^*(Y)] = \mathrm{E}B(V)^\top E(X) + \mathrm{E}\Phi_X^*(Y) := M,$$

where M depends on V only through F_U. We have by Young's inequality

$$V^\top Y \leq \Phi_X(V) + \Phi_X^*(Y).$$

but $Y = \nabla\Phi_X(\widetilde{U})$ a.s. implies that a.s.

$$\widetilde{U}^\top Y = \Phi_X(\widetilde{U}) + \Phi_X^*(Y),$$

so taking expectations gives

$$\mathrm{E}V^\top Y \leq M = \mathrm{E}\widetilde{U}^\top Y, \quad \forall V \sim F_U : \mathrm{E}(X|V) = E(X),$$

which yields the desired conclusion.

We next establish part(ii). We can argue similarly to above to show that

$$Y = \overline{\beta}(\overline{U})^\top X = \nabla\overline{\Phi}_X(\overline{U}), \text{ for } \Phi_X(u) = \overline{B}(u)^\top X,$$

and that for $\overline{\Phi}_x^*(y) := \sup_{v\in\mathcal{U}}\{v^\top y - \Phi_x(v)\}$ we have a.s.

$$\widetilde{U}^\top Y = \Phi_X(\widetilde{U}) + \Phi_X^*(Y).$$

Using the fact that $\widetilde{U} \sim \overline{U}$ and the fact that mean-independence gives $\mathrm{E}(B(\widetilde{U})^\top X) = \mathrm{E}(B(\overline{U})^\top X) = \mathrm{E}B(\widetilde{U})\mathrm{E}(X)$, we have

$$\mathrm{E}(\widetilde{U}Y) = \mathrm{E}(\psi(X,Y) + B(\widetilde{U})^\top X) = \mathrm{E}(\psi(X,Y) + B(\overline{U})^\top X) \geq \mathrm{E}(\overline{U}Y)$$

but reversing the role of U and $\overline{U}$, we also have $\mathrm{E}(UY) \leq \mathrm{E}(\overline{U}Y)$ and then

$$\mathrm{E}(\overline{U}Y) = \mathrm{E}(\psi(X,Y) + B(\overline{U})^\top X)$$

so that, thanks to inequality

$$\psi(x,y) + B(u)^\top x \geq u^\top y, \ \forall(u,x,y) \in \mathcal{U}\mathcal{X}\mathcal{Y},$$

we have

$$\psi(X,Y) + B(\overline{U})^\top X = \overline{U}^\top Y, \ \text{ a.s. },$$

which means that $\overline{U}$ solves $\max_{u\in\mathcal{U}}\{u^\top Y - B(u)^\top X\}$ which, by strict concavity admits $\widetilde{U}$ as unique solution. This proves that $\widetilde{U} = \overline{U}$ and thus a.s. we have

$$(\overline{\beta}(\widetilde{U}) - \beta(\widetilde{U}))^\top X = 0.$$

The part (iii) is a consequence of part (i). Note that by part (ii) we have that $\widetilde{U} = U$ a.s. and $(\beta(U) - \beta_0(U))^\top X = 0$ a.s. Since U and X are independent, we have that, for $e_1, ..., e_p$ denoting vectors of the canonical basis in $\mathbb{R}^p$:

$$\begin{aligned}0 &= \mathrm{E}\left(e_j^\top(\beta(U) - \beta_0(U))^\top XX^\top(\beta(U) - \beta_0(U))e_j\right)\\ &= \mathrm{E}\left(e_j^\top(\beta(U) - \beta_0(U))^\top \mathrm{E}XX^\top(\beta(U) - \beta_0(U))e_j\right)\\ &\geq \mathrm{mineg}(\mathrm{E}XX^\top)\mathrm{E}\left(\|(\beta(U) - \beta_0(U))e_j\|^2\right).\end{aligned}$$

Since $\mathrm{E}XX^\top$ has full rank this implies that $\mathrm{E}\|(\beta(U) - \beta_0(U))e_j\|^2 = 0$ for each j, which implies the rest of the claim. ■

B.2. **Proof of Theorem 3.2.** We have that any feasible pair (ψ, b) obeys the constraint

$$\psi(x, y) + b(u)^\top x \geq u^\top y, \quad \forall (y, x, u) \in \mathcal{YXU}.$$

Let $\widetilde{U} \sim F_U : \mathrm{E}(X \mid U) = \mathrm{E}(X)$ be the solution to the primal program. Then for any feasible pair (ψ, b) we have:

$$\mathrm{E}\psi(X, Y) + \mathrm{E}b(\widetilde{U})^\top \mathrm{E}X = \mathrm{E}\psi(X, Y) + \mathrm{E}b(\widetilde{U})^\top X \geq \mathrm{E}Y^\top \widetilde{U}.$$

Moreover, the last inequality holds as equality holds if

$$\psi(x, y) = \sup_{u \in \mathcal{U}}\{u^\top y - B(u)^\top x\}, \quad b(u) = B(u), \tag{B.1}$$

which is a feasible pair by (QL). In particular, as noted in the proof of the previous theorem, we have that

$$\psi(X, Y) + b(\widetilde{U})^\top X = Y^\top \widetilde{U}$$

It follows that $\mathrm{E}Y^\top \widetilde{U}$ is the optimal value and it is attained by the pair (B.1). ■

B.3. **Proof of Theorem 3.3.** To show part (i). We define notations. Recall that we consider the dual problem:

$$\inf_{(\psi,b)} \mathrm{E}(\psi(X, Y)) + \mathrm{E}b(V)^\top \mathrm{E}(X) \ : \ \psi(x, y) + b(u)^\top x \geq u^\top y, \ \forall \ (y, x, u) \in \mathcal{YXU}, \tag{B.2}$$

where $V \sim F_U$, where the infimum is taken over all continuous functions.

Write $X = (1, \widetilde{X}^\top)^\top$, where $\widetilde{X}$ denotes the non-constant component of vector X. Let f_W denote the joint density of $(\widetilde{X}, Y)$ with support $W = \mathcal{W}$, and F_U the distribution of U with support set $\mathcal{U}$. Assume without loss of generality that

$$\mathrm{E}(\widetilde{X}) = 0.$$

Let $\mathcal{W}_o$ denote the interior of $\mathcal{W}$. We also partition

$$b(u) = (\varphi(u), v(u)^\top)^\top,$$

with $u \mapsto \varphi(u)$ mapping $\mathcal{U}$ to $\mathbb{R}$, corresponding to the coefficient in front of the constant.

Let us denote by $(0, \overline{y})$ the mean of f_W:

$$\int_{\mathcal{W}_o} \widetilde{x}\, F_W(d\widetilde{x}, dy) = 0,\ \int_{\mathcal{W}_o} y\, F_W(d\widetilde{x}, dy) =: \overline{y}.$$

Observe that $(0, \overline{y}) \in \mathcal{W}_o$, since, otherwise, by convexity, F_W would be supported on $\partial\mathcal{W}_o$ which would contradict our assumption that $f_W \in L^\infty(\mathcal{W}_o)$.

With this notation, we wish to prove the existence of optimal potentials for our dual problem:

$$\inf_{\psi,\varphi,v} \int_{\mathcal{W}_o} \psi(\widetilde{x}, y) F_W(d\widetilde{x}, dy) + \int_{\mathcal{U}} \varphi(u) F_U(du) \tag{B.3}$$

subject to the pointwise constraint that

$$\psi(\widetilde{x}, y) + \varphi(u) \geq u^\top y - v(u)^\top \widetilde{x},\ (\widetilde{x}, y) \in \mathcal{W},\ u \in \mathcal{U}. \tag{B.4}$$

Of course, we can take ψ that satisfies

$$\psi(\widetilde{x}, y) := \sup_{u \in \mathcal{U}} \{u^\top y - v(u)^\top \widetilde{x} - \varphi(u)\}$$

so that ψ can be chosen convex and 1 Lipschitz with respect to y. In particular, we have

$$\psi(\widetilde{x}, \overline{y}) - \|y - \overline{y}\| \leq \psi(\widetilde{x}, y) \leq \psi(\widetilde{x}, \overline{y}) + \|y - \overline{y}\|. \tag{B.5}$$

The problem being invariant by the transform $(\psi, \varphi) \to (\psi + C, \psi - C)$, for C being an arbitrary constant, we can add as a normalization the condition that

$$\psi(0, \overline{y}) = 0. \tag{B.6}$$

This normalization and the constraint (B.4) imply that

$$\varphi(t) \geq t^\top \overline{y} - \psi(0, \overline{y}) \geq -\|t\| \|\overline{y}\| \geq -C\|\overline{y}\|. \tag{B.7}$$

We note that there is one extra invariance of the problem: if one adds an affine term $q^\top \widetilde{x}$ to ψ this does not change the cost and neither does it affect the constraint, provided one modifies b accordingly by subtracting to it the constant vector q. Take then q in the subdifferential of $\widetilde{x} \mapsto \psi(\widetilde{x}, \overline{y})$ at 0 and change ψ into $\psi - q^\top \widetilde{x}$, we obtain a new potential with the same properties as above and with the additional property that $\psi(\cdot, \overline{y})$ is minimal at $\widetilde{x} = 0$, and thus $\psi(\widetilde{x}, \overline{y}) \geq 0$, together with (B.5) this gives the lower bound

$$\psi(\widetilde{x}, y) \geq -\|y - \overline{y}\| \geq -C \tag{B.8}$$

where the bound comes from the boundedness of $\mathcal{W}_o$. From now on, C will denote a generic constant maybe changing from one line to another.

Now take a minimizing sequence $(\psi_n, \varphi_n, v_n) \in C(\mathcal{W}, \mathbb{R}) \times C(\mathcal{U}, \mathbb{R}) \times C(\mathcal{U}, \mathbb{R}^N)$ where for each n, ψ_n has been chosen with the same properties as above. Since φ_n and ψ_n are bounded from below ($\varphi_n \geq -C\|\overline{y}\|$ and $\psi_n \geq C$) and since the sequence is minimizing, we deduce immediately that ψ_n and φ_n are bounded sequences in L^1. Let $z = (\widetilde{x}, y) \in \mathcal{W}_o$ and $r > 0$ be such that the distance between z and the complement of $\mathcal{W}_o$ is at least $2r$, so that $B_r(z)$ is a ball that is at least at distance r from $\partial\mathcal{W}_o$, by assumption there is an $\alpha_r > 0$ such that $f_W \geq \alpha_r$ on $B_r(z)$. We then deduce from the convexity of ψ_n:

$$C \leq \psi_n(z) \leq \frac{\int_{B_r(z)} \psi_n}{|B_r(z)|} \leq \frac{\int_{B_r(z)} |\psi_n|}{|B_r(z)|} \leq \frac{\|\psi_n\|_{L^1(f_W)}}{|B_r(z)|\alpha_r},$$

where $|B_r(z)|$ denotes the volume of $B_r(z)$ with respect to the Lebesgue measure, so that ψ_n is actually bounded in L^∞_{loc} and by convexity, we also have

$$\|\nabla\psi_n\|_{L^\infty(B_r(z))} \leq \frac{2}{R-r}\|\psi_n\|_{L^\infty(B_R(z))}$$

whenever $R > r$ and $B_R(z) \subset \mathcal{W}_o$ (see for instance Lemma 5.1 in [12] for a proof of such bounds). We can thus conclude that ψ_n is also locally uniformly Lipschitz. Therefore, thanks to Ascoli's theorem, we can assume, taking a subsequence if necessary, that ψ_n converges locally uniformly to some potential ψ.

Let us now prove that v_n is bounded in L^1, for this take $r > 0$ such that $B_{2r}(0, \overline{y})$ is included in $\mathcal{W}_o$. For every $\widetilde{x} \in B_r(0)$, any $t \in \mathcal{U}$ and any n we then have

$$-v_n(t)^\top \widetilde{x} \leq \varphi_n(t) - t^\top \overline{y} + \|\psi_n\|_{L^\infty(B_r(0,\overline{y}))} \leq C + \varphi_n(t)$$

maximizing in $\widetilde{x} \in B_r(0)$ immediately gives

$$\|v_n(t)\|r \leq C + \varphi_n(t).$$

From which we deduce that v_n is bounded in L^1 since φ_n is.

From Komlos theorem (see [39]), we may find a subsequence such that the Cesaro means

$$\frac{1}{n}\sum_{k=1}^{n} \varphi_k, \ \frac{1}{n}\sum_{k=1}^{n} v_k$$

converge a.e. respectively to some φ and b. Clearly ψ, φ and b satisfy the linear constraint (B.4), and since the sequence of Cesaro means $(\psi'_n, \phi'_n, v'_n) := n^{-1}\sum_{k=1}^{n}(\psi_k, \phi_k, v_k)$ is also minimizing, we deduce from Fatous' Lemma

$$\begin{aligned}
&\int_{\mathcal{W}_o} \psi(\widetilde{x}, y) F_W(d\widetilde{x}, dy) + \int_{\mathcal{U}} \varphi(u) F_U(du) \\
&\leq \liminf_n \int_{\mathcal{W}_o} \psi'_n(\widetilde{x}, y) F_W(d\widetilde{x}, dy) + \int_{\mathcal{U}} \varphi'_n(u) F_U(du) = \inf(B.3)
\end{aligned}$$

which ends the existence proof.

To show part (ii), we note that a.e. apply the generalized envelope theorem by Milgrom and Segal (2002). Alternatively this can demonstrated from Rademacher's theorem. ∎

B.4. **Proof of Theorem 3.4.** Obviously $A_t = 1 \Rightarrow \widetilde{U} \geq t$, and $\widetilde{U} > t \Rightarrow A_t = 1$. Hence $\mathrm{P}(\widetilde{U} \geq t) \geq \mathrm{P}(A_t = 1) = \mathrm{P}(Y > \beta(t)^\top X) = (1-t)$ and $\mathrm{P}(\widetilde{U} > t) \leq \mathrm{P}(A_t = 1) = (1-t)$ which proves that $\widetilde{U}$ is uniformly distributed and $\{\widetilde{U} > t\}$ coincides with $\{\widetilde{U}_t = 1\}$ a.s. We thus have $\mathrm{E}(X1\{\widetilde{U} > t\}) = \mathrm{E}(XA_t) = \mathrm{E}X(1-t) = \mathrm{E}X\mathrm{E}A_t$, with standard approximation argument we deduce that $\mathrm{E}(Xf(\widetilde{U})) = \mathrm{E}X\mathrm{E}f(\widetilde{U})$ for every $f \in C([0,1],\mathbb{R})$ which means that $E(X \mid \widetilde{U}) = E(X)$.

As already observed $\widetilde{U} > t$ implies that $Y > \beta(t)^\top X$ in particular $Y \geq \beta(\widetilde{U} - \delta)^\top X$ for $\delta > 0$, letting $\delta \to 0^+$ and using the a.e. continuity of $u \mapsto \beta(u)$ we get $Y \geq \beta(\widetilde{U})^\top X$. The converse inequality is obtained similarly by remaking that $\widetilde{U} < t$ implies that $Y \leq \beta(t)^\top X$.

Let us now prove that $\widetilde{U}$ solves (3.11). Take V uniformly distributed and mean-independent from X and set $V_t := 1\{V > t\}$, we then have $\mathrm{E}(XV_t) = 0$, $\mathrm{E}(V_t) = (1-t)$ but since A_t solves (3.7) we have $\mathrm{E}(V_tY) \leq \mathrm{E}(A_tY)$. Observing that $V = \int_0^1 V_t dt$ and integrating the previous inequality with respect to t gives $\mathrm{E}(VY) \leq \mathrm{E}(UY)$ so that $\widetilde{U}$ solves (3.11).

Next we show part(ii). Let us define for every $t \in [0,1]$ $B(t) := \int_0^t \beta(s)ds$. Let us also define for (x,y) in $\mathbb{R}^{N+1}$:

$$\psi(x,y) := \max_{t\in[0,1]} \{ty - B(t)^\top x\}$$

thanks to monotonicity condition, the maximization program above is strictly concave in t for every y and each $x \in X$. We then note that

$$Y = \beta(\widetilde{U})^\top X = \nabla B(\widetilde{U})^\top X \text{ a.s.}$$

exactly is the first-order condition for the above maximization problem when $(x,y) = (X,Y)$. In other words, we have

$$\psi(x,y) + B(t)^\top x \geq ty, \ \forall (t,x,y) \in [0,1] \times \mathcal{X} \times \mathbb{R} \tag{B.9}$$

with an equality holding a.s. for $(x,y,t) = (X,Y,\widetilde{U})$, i.e.

$$\psi(X,Y) + B(\widetilde{U})^\top X = UY, \quad \text{a.s.} \tag{B.10}$$

Using the fact that $\widetilde{U} \sim \overline{U}$ and the fact that mean-independence gives $\mathrm{E}(B(\widetilde{U})^\top X) = \mathrm{E}(b(\overline{U})^\top X) = E(X)$, we have

$$\mathrm{E}(UY) = \mathrm{E}(\psi(X,Y) + B(\widetilde{U})^\top X) = \mathrm{E}(\psi(X,Y) + B(\overline{U})^\top X) \geq \mathrm{E}(\overline{U}Y)$$

but reversing the role of $\widetilde{U}$ and $\overline{U}$, we also have $\mathrm{E}(UY) \leq \mathrm{E}(\overline{U}Y)$ and then

$$\mathrm{E}(\overline{U}Y) = \mathrm{E}(\psi(X,Y) + B(\overline{U})^\top X)$$

so that, thanks to inequality (B.9)

$$\psi(X,Y) + B(\overline{U})^\top X = \overline{U}Y, \quad \text{a.s.}$$

which means that $\overline{U}$ solves $\max_{t\in[0,1]}\{tY - \varphi(t) - B(t)^\top X\}$ which, by strict concavity admits $\widetilde{U}$ as unique solution.

Part (iii) is a consequence of Part (ii) and independence of $\widetilde{U}$ and X. Note that by part (ii) we have that $\widetilde{U} = U$ a.s. and that $(\beta(U) - \beta_0(U))^\top X = 0$ a.s. Since U and X are independent, we have that

$$\begin{aligned}
0 &= \mathrm{E}\left((\beta(\widetilde{U}) - \beta_0(U))^\top XX^\top(\beta(U) - \beta_0(U))\right)\\
&= \mathrm{E}\left((\beta(U) - \beta_0(U))^\top \mathrm{E}XX^\top(\beta(U) - \beta_0(U))\right)\\
&\geq \mathrm{mineg}(\mathrm{E}XX^\top)\mathrm{E}\left(\|(\beta(U) - \beta_0(U))\|^2\right).
\end{aligned}$$

Since $\mathrm{E}XX^\top$ has full rank this implies that $\mathrm{E}\|(\beta(U) - \beta_0(U))\|^2 = 0$, which implies the rest of the claim. ■

B.5. **Proof of Lemma 3.1.** . Define $\psi_x(y) := \psi(x,y)$. We then have

$$Y \in \partial\psi_X^*(U), \quad Y \in \partial_u\psi^*(-b(U), U).$$

We have that $\psi_x(y) \leq \Phi_x^*(y)$ by definition, hence

$$\psi_x^*(u) \leq (\Phi_x)^{**}(u) \leq \Phi_x(u)$$

where Φ_x^{**} denotes the convex envelope of Φ_x. The duality relations give

$$\widetilde{U}^\top Y = \psi_X(Y) + \Phi_X(\widetilde{U}) = \Phi_X^*(Y) + \Phi_X(\widetilde{U})$$

and then

$$\Phi_X^{**}(\widetilde{U}) \geq \widetilde{U}^\top Y - \Phi_X^*(Y) = \Phi_X(\widetilde{U}).$$

Hence, $\Phi_X(\widetilde{U}) = \Phi_X^{**}(\widetilde{U})$ almost surely. ■

Appendix C. Additional Results for Section 2 and Section 3: Rigorous Proof of Duality for Conditional Vector Quantiles and Linear Vector Quantile Regression

We claimed in the main text we stated two dual problems for CVQF and VQR without the proof. In this sections we prove that these assertions were rigorous. Here for simplicity of notation we assume that $X = Z$, which entails no loss of generality under our assumption that conditioning on Z and X is equivalent.

We shall write $X = (1, \widetilde{X}^\top)^\top$, where $\widetilde{X}$ denotes the non-constant component of vector X. Let f_W denote the joint density of $(\widetilde{X}, Y)$ with support $W = \mathcal{W}$, and F_U the distribution of U with support set $\mathcal{U}$. Assume without loss of generality that

$$\mathrm{E}(\widetilde{X}) = 0.$$

C.1. **Duality for Conditional Vector Quantiles.** Let k and d be two integers, K_1 be a compact subset of $\mathbb{R}^k$ and K_2 and K_3 be compact subsets of $\mathbb{R}^d$. Let $\nu = F_W$ have support $\mathcal{W} \subseteq K_1 \times K_2$, and we may decompose $\nu = m \otimes \nu^{\widetilde{x}}$ where m denotes the first marginal of ν. We also assume that m is centered i.e.

$$\overline{x} := \int_{K_1} \widetilde{x} m(d\widetilde{x}) = 0$$

Finally, let $\mu = F_U$ have support $\mathcal{U} \subseteq K_3$. We are interested here in rigorous derivation for dual formulations for covariance maximization under an independence and then a mean-independence constraint.

Duality for the independence constraint. First consider the case of an independence constraint:

$$\sup_{\theta \in I(\mu,\nu)} \int_{K_1 \times K_2 \times K_3} u^\top y \, \theta(d\widetilde{x}, dy, du) \tag{C.1}$$

where $I(\mu, \nu)$ consists of the probability measures θ on $K_1 \times K_2 \times K_3$ such that $\Pi_{X,Y\#}\theta = \nu$ and $\Pi_{X,U\#}\theta = m \otimes \mu$, namely that

$$\int_{K_1 \times K_2 \times K_3} \psi(\widetilde{x}, y)\theta(d\widetilde{x}, dy, du) = \int_{K_1 \times K_2} \psi(\widetilde{x}, y)\nu(d\widetilde{x}, dy), \ \forall \psi \in C(K_1 \times K_2),$$

and

$$\int_{K_1 \times K_2 \times K_3} \varphi(\widetilde{x}, u)\theta(d\widetilde{x}, dy, du) = \int_{K_1 \times K_3} \varphi(\widetilde{x}, u)m(d\widetilde{x})\mu(du), \ \forall \varphi \in C(K_1 \times K_3).$$

As already noticed, given a random (X, Y) such that $\mathrm{Law}(X, Y) = \nu$, (C.1) is related to the problem of finding U independent of X and having law μ which is maximally correlated to Y. It is clear that $I(\mu, \nu)$ is a nonempty (take $m \otimes \nu^{\widetilde{x}} \otimes \mu$) convex and weakly $*$ compact set so that (C.1) admits solutions. Let us consider now:

$$\inf_{(\psi,\varphi) \in C(K_1 \times K_2) \times C(K_1 \times K_3)} \int_{K_1 \times K_2} \psi(\widetilde{x}, y)\nu(d\widetilde{x}, dy) + \int_{K_1 \times K_3} \varphi(\widetilde{x}, u)m(d\widetilde{x})\mu(du) \tag{C.2}$$

subject to the constraint

$$\psi(\widetilde{x}, y) + \varphi(\widetilde{x}, u) \geq u^\top y, \ \forall (\widetilde{x}, y, u) \in K_1 \times K_2 \times K_3. \tag{C.3}$$

Then we have

Theorem C.1. *The infimum in (C.2)-(C.3) coincides with the maximum in (C.1). This common value also coincides with the infimum of* $\int_{K_1\times K_2} \psi\nu + \int_{K_1\times K_2} \varphi m\otimes\mu$ *taken over* $L^1(\nu)\times L^1(m\otimes\mu)$ *functions that satisfy the constraint (C.3).*

Proof. Let us rewrite (C.2) in standard convex programming form as :

$$\inf_{(\psi,\varphi)} F(\psi,\varphi)+G(\Lambda(\psi,\varphi))$$

where $F(\psi,\varphi) := \int_{K_1\times K_2}\psi\nu+\int_{K_1\times K_3}\varphi m\otimes\mu$, Λ is the linear continuous map from $C(K_1\times K_2)\times C(K_1\times K_3)$ to $C(K_1\times K_2\times K_3)$ defined by

$$\Lambda(\psi,\varphi)(\widetilde{x},y,u) := \psi(\widetilde{x},y)+\varphi(\widetilde{x},u),\ \forall(\widetilde{x},y,u)\in K_1\times K_2\times K_3$$

and G is defined for $\eta\in C(K_1\times K_2\times K_3)$ by:

$$G(\eta)=\begin{cases} 0 & \text{if } \eta(\widetilde{x},y,u)\geq u^\top y\\ +\infty & \text{otherwise.}\end{cases}$$

It is easy to check that the Fenchel-Rockafellar theorem (see Ekeland and Temam [25]) applies here so that the infimum in (C.2) coincides with

$$\sup_{\theta\in\mathcal{M}(K_1\times K_2\times K_3)} -F^*(\Lambda^*\theta)-G^*(-\theta). \tag{C.4}$$

Direct computations give that

$$-G^*(-\theta)=\begin{cases} \int_{K_1\times K_2\times K_3} u^\top y\ \theta(d\widetilde{x},dy,du) & \text{if } \theta\geq 0\\ -\infty & \text{otherwise.}\end{cases}$$

that $\Lambda^*\theta=(\Pi_{X,Y\#}\theta,\Pi_{X,U\#}\theta)$ and

$$F^*(\Lambda^*\theta)=\begin{cases} 0 & \text{if } (\Pi_{X,Y\#}\theta,\Pi_{X,U\#}\theta)=(\nu,m\otimes\mu)\\ +\infty & \text{otherwise.}\end{cases}$$

This shows that the maximization problem (C.4) is the same as (C.1). Therefore the infimum in (C.2) coincides with the maximum in (C.1). When one relaxes (C.2) to L^1 functions, we obtain a problem whose value is less than that of (C.2) (because minimization is performed over a larger set) and larger than the supremum in (C.1) (direct integration of the inequality constraint), the common value of (C.1) and (C.2)-(C.3) therefore also coincides with the value of the L^1 relaxation of (C.2)-(C.3). ∎

C.2. **Duality for Linear Vector Quantile Regression.** Here we use the notation where we partition

$$b(u) = (\varphi(u), v(u)^\top)^\top,$$

with $u \mapsto \varphi(u)$ mapping $\mathcal{U}$ to $\mathbb{R}$, corresponding to the coefficient in front of the constant. Let $\mathcal{W}_o$ denote the interior of $\mathcal{W}$.

Let us denote by $(0, \overline{y})$ the mean of f_W:

$$\int_{\mathcal{W}_o} \widetilde{x} \, F_W(d\widetilde{x}, dy) = 0, \; \int_{\mathcal{W}_o} y \, F_W(d\widetilde{x}, dy) =: \overline{y}.$$

Let us now consider the mean-independent correlation maximization problem:

$$\sup_{\theta \in MI(\mu,\nu)} \int_{K_1 \times K_2 \times K_3} u^\top y \; \theta(d\widetilde{x}, dy, du) \tag{C.5}$$

where $MI(\mu, \nu)$ consists of the probability measures θ on $K_1 \times K_2 \times K_3$ such that $\Pi_{X,Y\#}\theta = \nu$, $\Pi_{U\#}\theta = \mu$ and according to θ, $\widetilde{x}$ is mean independent of u i.e.

$$\langle \sigma_\theta, v \rangle := \int_{K_1 \times K_2 \times K_3} (v(u)^\top \widetilde{x}) \;\; \theta(d\widetilde{x}, dy, du) = 0, \; \forall \beta \in C(K_3, \mathbb{R}^d). \tag{C.6}$$

Again the constraints are linear so that $MI(\mu, \nu)$ is a nonempty convex and weak $*$ compact set so that the infimum in (C.1) is attained. In probabilistic terms, given $(\widetilde{X}, Y)$ distributed according to ν, the problem above consists in finding U with law μ, mean-independent of $\widetilde{X}$ (i.e. such that $\mathrm{E}(\widetilde{X}|U) = \mathrm{E}(\widetilde{X}) = 0$) and maximally correlated to Y.

We claim that (C.5) is dual to

$$\inf_{(\psi,\varphi,b) \in C(K_1 \times K_2) \times C(K_3) \times C(K_3, \mathbb{R}^d)} \int_{K_1 \times K_2} \psi(\widetilde{x}, y) \nu(d\widetilde{x}, dy) + \int_{K_3} \varphi(u) \mu(du) \tag{C.7}$$

subject to

$$\psi(\widetilde{x}, y) + \varphi(u) + v(u)^\top \widetilde{x} \geq u^\top y, \; \forall (\widetilde{x}, y, u) \in K_1 \times K_2 \times K_3. \tag{C.8}$$

Then we have the following duality result

Theorem C.2. *The infimum in (C.7)-(C.8) coincides with the maximum in (C.5). This common value also coincides with the infimum of* $\int_{K_1 \times K_2} \psi \nu + \int_{K_3} \varphi m \otimes \nu$ *taken over* $L^1(\nu) \times L^1(\mu) \times L^1(\mu, \mathbb{R}^d)$ *functions that satisfy the constraint (C.8).*

Proof. Write (C.7)-(C.8) as

$$\inf_{(\psi,\varphi,b)} F(\psi, \varphi, b) + G(\Lambda(\psi, \varphi, b))$$

where $F(\psi,\varphi,b) := \int_{K_1\times K_2}\psi\nu + \int_{K_3}\varphi\mu$, Λ is the linear continuous map from $C(K_1\times K_2)\times C(K_3)\times C(K_3,\mathbb{R}^d)$ to $C(K_1\times K_2\times K_3)$ defined by

$$\Lambda(\psi,\varphi,b)(\widetilde{x},y,u) := \psi(\widetilde{x},y)+\varphi(u)+v(u)^\top\widetilde{x},\ \forall(\widetilde{x},y,u)\in K_1\times K_2\times K_3$$

and G is as in the proof of theorem C.1. For $\theta\in\mathcal{M}(K_1\times K_2\times K_3)$, one directly checks that $\Lambda^*\theta = (\Pi_{\widetilde{X},Y\,\#}\theta, \Pi_{U\#}\theta, \sigma_\theta)$ (where the vector-valued measure σ_θ is defined as in (C.6)). We then get

$$F^*(\Lambda^*\theta) = \begin{cases} 0 & \text{if } (\Pi_{\widetilde{X},Y\,\#}\theta, \Pi_{U\#}\theta, \sigma_\theta) = (\nu,\mu,0) \\ +\infty & \text{otherwise.} \end{cases}$$

We then argue exactly as in the proof of theorem C.1 to conclude. ■

References

[1] Abadie, A., Angrist J., and Imbens, G. (2002). "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings" *Econometrica* 70 (1).

[2] Abrevaya, J. (2001). "The Effects of Demographics and Maternal Behavior on the Distribution of Birth Outcomes." *Empirical Economics* 26:1, pp. 247–57.

[3] Adrian, T., and Brunnermeier M. (2011). "CoVaR". Working paper.

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

[5] Barlow, R. E., D. J. Bartholomew, J. M. Bremner, and H. D. Brunk (1972). *Statistical inference under order restrictions. The theory and application of isotonic regression.* John Wiley & Sons.

[6] Belloni, A. and R.L. Winkler (2011). "On Multivariate Quantiles Under Partial Order". The Annals of Statistics, Vol. 39 No. 2, 1125–1179

[7] Brenier, Y (1991). "Polar factorization and monotone rearrangement of vector-valued functions," Comm. Pure Appl. Math. 44 **4**, pp. 375–417.

[8] Buchinsky, M. (1994). "Changes in U.S. Wage Structure 1963–1987: An Application of Quantile Regression." Econometrica. March, 62:2, pp. 405–58.

[9] Buchinsky, M. (1997). "The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach." Journal of Applied Econometrics. January/February, 13:1, pp. 1–30.

[10] Carlier, G., Dana, R.-A., and Galichon, A. (2012): "Pareto efficiency for the concave order and multivariate comonotonicity," *Journal of Economic Theory* 147(1), pp. 207–229.

[11] Carlier, G., Chernozhulov, V., and Galichon, A. (2013): "Conditional versions of Brenier's theorem". Working paper.

[12] G. Carlier, A. Galichon, Exponential convergence for a convexifying equation, ESAIM, Control, Optimisation and Calculus of Variations, **18** (3), 611-620 (2012).

[13] Chamberlain, G. (1994). "Quantile Regression, Censoring and the Structure of Wages," in *Advances in Econometrics*. Christopher Sims, ed. New York: Elsevier, pp. 171–209.

[14] Chaudhuri P. (1996): "On a geometric notion of quantiles for multivariate data," *Journal of the American Statistical Association* 91, pp. 862–872.

[15] Chernozhukov, V., Fernandez-Val, I., and A. Galichon (2010a) "Quantile and Probability Curves without Crossing" *Econometrica* 78(3), 1093–1125.

[16] Chernozhukov, V., and Galichon, A. (2013). "The empirical μ-quantile process". In preparation.

[17] Chernozhukov, V., and C. Hansen (2006), "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics* 132(2), 491–525.

[18] Chernozhukov, V. and Umantsev, L. (2001). "Conditional Value-at-Risk: Aspects of Modeling and Estimation." *Empirical Economics* 26:1, pp. 271–92.

[19] Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. "Estimating the technology of cognitive and noncognitive skill formation." Econometrica 78.3 (2010): 883-931

[20] Deaton, A. (1997). *The Analysis of Household Surveys*. Johns Hopkins.

[21] Dette, H., Neumeyer, N., and K. Pilz (2006): "A simple Nonparametric Estimator of a Strictly Monotone Regression Function," *Bernoulli*, 12, no. 3, pp 469—490.

[22] Ducpetiaux, E. (1855), *Budgets économiques des classes ouvrières en Belgique*, Bruxelles.

[23] Dudley, R. M., and Walter Philipp, (1983): "Invariance principles for sums of Banach space valued random elements and empirical processes." *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 62.4: 509-552.

[24] I. Ekeland, A. Galichon, M. Henry, Comonotonic measures of multivariate risks, *Mathematical Finance*, **22** (1), 109-132 (2012).

[25] I. Ekeland, R. Temam, *Convex Analysis and Variational Problems*, Classics in Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, (1999).

[26] Engel, E. (1857), "Die Produktions und Konsumptionsverhältnisse des Königreichs Sachsen," *Zeitschrift des Statistischen Bureaus des Königlich Sächsischen Misisteriums des Innerm,* 8, pp. 1-54.

[27] Engle, R. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles, Journal of Business and Economic Statistics 22: 367–381.

[28] Galichon, A. Henry. M. (2012). "Dual theory of choice with multivariate risks". *Journal of Economic Theory* **47** (4), 1501-516 (2012).

[29] Gosling, A., Machin S., and Meghir, C. (2000). "The Changing Distribution of Male Wages in the U.K." *Review of Economic Studies* 67:4, pp. 635–66.

[30] Hallin, M., Paindaveine, D., and Šiman, M. (2010). "Multivariate quantiles and multiple-output regression quantiles: From L1 optimization to halfspace depth". *Annals of Statistics* 38 (2), pp. 635–669.

[31] He, X. (1997), "Quantile Curves Without Crossing," *American Statistician*, 51, pp. 186–192.

[32] Koenker, R. (2005), *Quantile Regression*. Econometric Society Monograph Series 38, Cambridge University Press.

[33] Koenker, R. (2007). *quantreg: Quantile Regression.* R package version 4.10. `http://www.r-project.org`.

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

[35] Koenker, R. and Bassett, G (1982). "Robust Tests of Heteroscedasticity based on Regression Quantiles," *Econometrica* 50, 43–61.

[36] Koenker, R., and Hallock, K. (2001) *Journal of Economic Perspectives* 15 (4), pp. 143–156.

[37] Koenker, R., P. Ng (2005), "Inequality constrained quantile regression." *Sankhyā* 67, no. 2, 418–440.

[38] Koltchinskii, V. (1997). "M-Estimation, Convexity and Quantiles". *Annals of Statistics* 25, No. 2, pp. 435-477.

[39] J. Komlos, A generalization of a problem of Steinhaus, *Acta Mathematica Academiae Scientiarum Hungaricae*, **18** (1–2), 1967, pp. 217–229.

[40] Kong, L. and Mizera, I. (2010): Quantile tomography: using quantiles with multivariate data. arXiv:0805.0056v1

[41] Matzkin, R. (2003) "Nonparametric estimation of nonadditive random functions." Econometrica 71.5: 1339-1375.

[42] Milgrom, P., & Segal, I. (2002). "Envelope theorems for arbitrary choice sets." Econometrica, 70(2), 583-601.

[43] McCann, R.J. (1995) "Existence and uniqueness of monotone measure-preserving maps" *Duke Mathematical Journal* 80 (2), 309-324

[44] Laine, B. (2001). "Depth contours as multivariate quantiles: a directional approach." Master's thesis, Université Libre de Bruxelles.

[45] Le Play, F. (1855). *Les Ouvriers Européens. Etudes sur les traveaux, la vie domestique et la condition morale des populations ouvrières de l'Europe*. Paris.

[46] Poterba, J. and Rueben, K. (1995). "The Distribution of Public Sector Wage Premia: New Evidence Using Quantile Regression Methods." NBER Working Paper No. 4734.

[47] R Development Core Team (2007): *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`.

[48] J.V.Ryff , Measure preserving Transformations and Rearrangements, *J. Math. Anal. and Applications* **31** (1970), 449–458.

[49] Serfling, R. (2004). "Nonparametric multivariate descriptive measures based on spatial quantiles". *Journal of Statistical Planning and Inference* 123, pp. 259-278.

[50] Tukey, J. (1975). *Mathematics and the picturing of data.* In Proc. 1975 International Congress of Mathematicians, Vancouver 523–531. Montreal: Canadian Mathematical Congress.

[51] Villani, C. (2003), *Topics in Optimal Transportation*, Providence: American Mathematical Society.

[52] C. Villani, *Optimal transport: Old and New*, Grundlehren der mathematischen Wissenschaften, Springer-Verlag, Heidelberg, 2009.

CEREMADE, UMR CNRS 7534, Université Paris IX Dauphine, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE

E-mail address: carlier@ceremade.dauphine.fr

Department of Economics, MIT, 50 Memorial Drive, E52-361B, Cambridge, MA 02142, USA

E-mail address: vchern@mit.edu

Sciences Po, Economics Department, 27, rue Saint-Guillaume, 75007 Paris, FRANCE

E-mail address: alfred.galichon.@sciences-po.fr.