Econometric Theory, **31**, 2015, 1102–1116.
doi:10.1017/S0266466614000711

WHAT DO QUANTILE REGRESSIONS IDENTIFY FOR GENERAL STRUCTURAL FUNCTIONS?

YUYA SASAKI
Johns Hopkins University

This paper shows what quantile regressions identify for general structural functions. Under fairly mild conditions, the quantile partial derivative identifies a weighted average of heterogeneous structural partial effects among the subpopulation of individuals at the conditional quantile of interest. This result justifies the use of quantile regressions as means of measuring heterogeneous causal effects for a general class of structural functions with multiple unobservables.

1. INTRODUCTION

Since the seminal work by Koenker and Bassett (1978), quantile regressions have been widely used to model unobserved heterogeneity in econometric analyses—see surveys by Koenker and Hallock (2001) and Koenker (2005). Quantile partial derivatives (QPDs) are often interpreted as heterogeneous structural partial effects, e.g., random coefficients under the comonotonicity (Koenker, 2005, Sect. 2.6). When relevant assumptions fail, however, quantile regressions may misspecify true structural functions in several ways.

A general class of structural models consists of nonparametric functions g of the form

$$Y = g(X, U),$$

where X is an observed input, U is an M-dimensional vector of unobserved inputs, and Y is an observed output or outcome. The quantile regression of Y on X is the nonstructural function $Q_{Y|X}$ defined by

$$Q_{Y|X}(\tau \mid x) = \inf\{y \mid F_{Y|X}(y \mid x) \geqslant \tau\}.$$

While g takes M dimensions of unobserved factors, $Q_{Y|X}$ takes only a scalar latent variable τ. If (i) X is exogenous and (ii) g is monotone with respect to a

I would like to thank the co-editor, Victor Chernozhukov, three anonymous referees, Andrew Chesher, Stefan Hoderlein, Roger Koenker, Blaise Melly, and Richard Spady for helpful comments and suggestions. All errors are mine. Address correspondence to Yuya Sasaki, Johns Hopkins University, Department of Economics, 440 Mergenthaler Hall, 3400 N. Charles St., Baltimore, MD 21218, USA; e-mail: sasaki@jhu.edu.

 © Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

scalar U (i.e., $M = 1$), then the quantile regression $Q_{Y|X}$ can be used to represent the structural function g.

Several approaches are suggested to relax condition (i).[1] On the other hand, condition (ii) is considered to be essential for the representation of g by $Q_{Y|X}$. Without the monotonicity assumption, the (local) slope parameter of quantile regressions fails to coincide with the true structural partial effect because individuals' ranks τ can arbitrarily change through a counterfactual (infinitesimal) shift in x, regardless of whether X is exogenous.

In applications, however, economic models often involve nonscalar unobserved factors U and/or nonmonotonic dependence of Y on U. Unobserved skills, abilities, or capabilities U are considered to take multiple dimensions, e.g., "capability vectors" (Sen, 1985) or cognitive and noncognitive abilities including "self-regulation, motivation, time preference, far-sightedness, adventurousness, ..." (Heckman, 2007). These unobserved factors may not be summarized by a scalar index in production of labor outcomes. Nonmonotonicity of Y on U is also natural, as is the case for inverted U-shaped structures of behavioral outcomes that are empirically found in neurology, psychology, and medicine, e.g., the Yerkes–Dodson law states that mental arousal enhances performance up to some point, but its marginal effects are negative thereafter.

Without the monotonicity with respect to a scalar unobserved heterogeneity, the quantile regression parameters generally fail to coincide with structural partial effects. Given the extensive use of quantile regressions in applications and the rich set of econometric methods available in the literature, however, we do not want to give up quantile regressions for presumably nonmonotonic structural functions. Fortunately, quantile regressions still measure important features of the underlying structural functions even if the monotonicity assumption fails, which concerns the main claim of this paper. Hoderlein and Mammen (2007) show that the QPD coincides with the unweighted average of the structural partial effects at the locality defined by (Y, X). This paper investigates this problem further by formally accounting for the expectation taken on a set of measure zero, and shows that the QPD generally coincides with a weighted average of the structural partial effects.

Our claim that the QPD identifies a weighted average parallels the well-known feature of the 2SLS that identifies a weighted average of structural causal effects under various treatment intensities (Angrist and Imbens, 1995). This paper makes similar contributions to the literature by warranting the effectiveness of quantile regressions for inference of structural causal effects. Nonmonotonicity is not the only type of misspecification by quantile regressions. Angrist, Chernozhukov, and Fernández-Val (2006) show that linear quantile regression parameters minimize a weighted mean square error at the population level for potentially nonlinear quantile regressions, which handles another way of misspecification. By showing what nonlinear quantile regressions identify for general structural functions without monotonicity, this paper complements their work so we can understand quantile regressions in a more general structural framework by both results combined together.

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

2. AN INTUITIVE ILLUSTRATION

In this section, we use heuristic arguments to illustrate how the local slope of a quantile regression can be decomposed into a weighted average of structural partial effects within a specifically characterized subpopulation. It will be followed up with formal theoretical discussions in the next section. We first define the objects of interest and fix their notations below.

DEFINITION 1 (Structural Partial Effect). *Given a structural function g, the structural partial effect at (x,u) is defined by $\beta(x,u) := \frac{\partial}{\partial x} g(x,u)$.*

DEFINITION 2 (Quantile Partial Derivative). *Given a quantile regression $Q_{Y|X}$, the QPD at x is defined by $\frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x)$ for each quantile $\tau \in (0,1)$.*

These definitions presume differentiability, but we postpone discussions of assumptions until Section 3. The QPD is the local slope parameter of a quantile regression. Hence, the QPD at (τ, x) can be consistently estimated by the standard local linear quantile smoothers, such as $\hat{\beta}$ in the solution $(\hat{\alpha}, \hat{\beta})$ to $\min_{\alpha,\beta} \sum_i \rho_\tau(Y_i - \alpha - \beta X_i) K\left(\frac{X_i - x}{h}\right)$, where ρ_τ is the check function, K is a kernel function, and h is a bandwidth parameter. Identification and estimation strategies under endogeneity can be found in the references listed in Footnote 1.

The major source of discrepancy between the structural function g and the quantile regression $Q_{Y|X}$ lies in the different dimensions of unobserved factors, U and τ. Geometry of the set of U constructed by the structural function g plays an important role in our analysis. We define the lower contour set of U evaluated by $g(x, \cdot)$ below a given level of y as follows:

$$V(y,x) = \{u \in \mathbb{R}^M \mid g(x,u) \leqslant y\}.$$

Its boundary is denoted by $\partial V(y,x)$.

Quantile regressions describe statistical relationships between y and x while keeping the probability τ fixed. Therefore, the key concept we use to understand quantile regressions is the "law of probability conservation." For ease of illustration, consider the case of $M = 2$ where all of X, U_1, and U_2 are positive factors of Y in a structural function g.

We can readily see under this setting that $V(y, x + \Delta x) \subset V(y,x)$ holds for infinitesimal $\Delta x > 0$. Figure 1(a) illustrates such a case, where the solid and dotted curves draw $\partial V(y,x)$ and $\partial V(y, x + \Delta x)$, respectively. Integrating the density f_U over the area between the two curves yields the probability loss that results from this increase Δx in x. The quantity $\langle \vec{\omega}(u), \vec{n}(u) \rangle f_U(u)$ measures the marginal local probability gain at u, where $\vec{\omega}(u)$ denotes the rate of change in the boundary point $u \in \partial V(y,x)$ and $\vec{n}(u)$ denotes the unit outward normal vector of $\partial V(y,x)$ at u. Because $\langle \vec{\omega}(u), \vec{n}(u) \rangle < 0$ is graphically the case at each $u \in \partial V(y,x)$ in Figure 1(a), the local probability gains are negative, i.e., an increase in x induces probability losses.

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

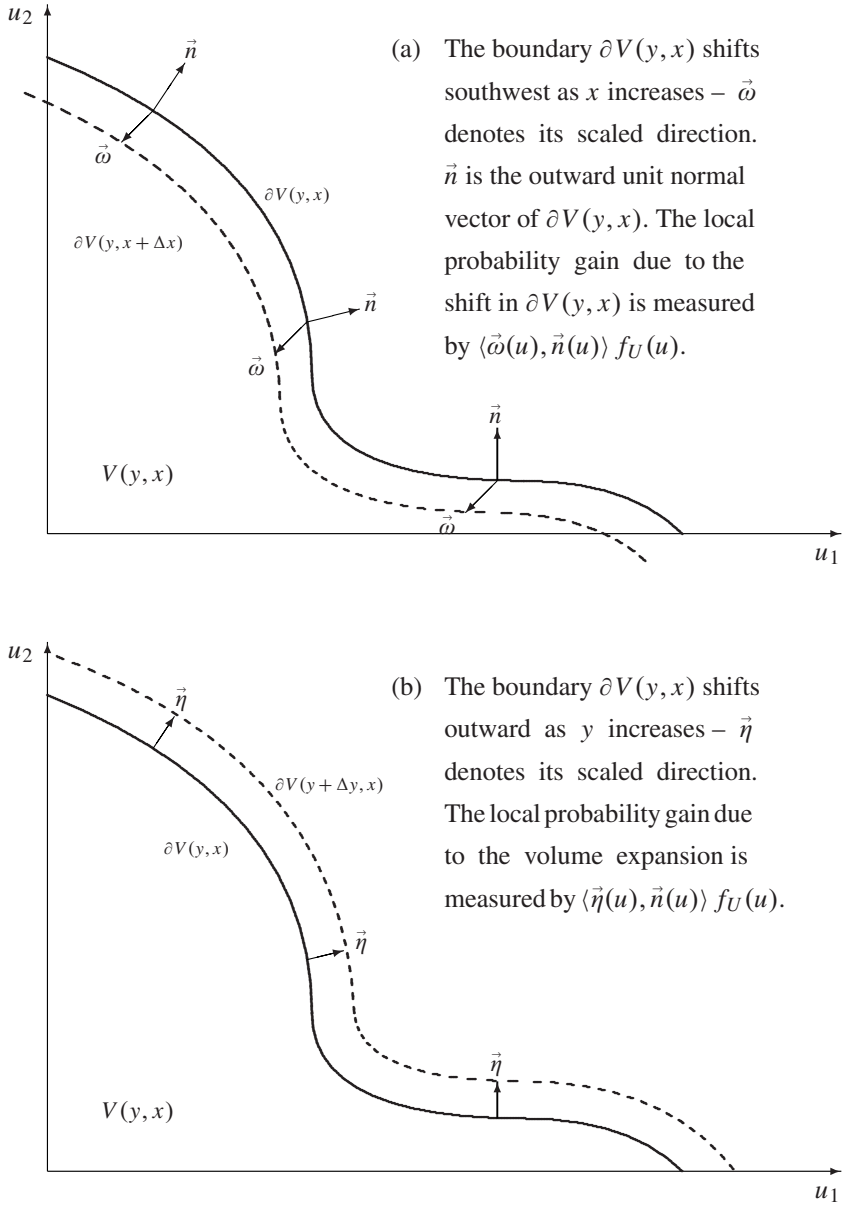


FIGURE 1. (a) A shift in the volume $V(y,x)$ in response to a marginal increase in x and (b) an expansion of the volume $V(y,x)$ in response to a marginal increase in y. The law of probability conservation requires that the probability loss in (a) must be compensated for by the probability gain in (b).

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

In order to satisfy the law of probability conservation, the probability loss must be compensated for by appropriately expanding the lower contour set $V(y,x)$. From its definition, we can see that increasing y expands the volume, and $V(y+\Delta y,x) \supset V(y,x)$ holds for $\Delta y > 0$, as illustrated in Figure 1(b). Integrating the density f_U over the area between the two curves yields the probability gain which results from this increase in y. The marginal local probability gain at u is measured by $\langle \vec{\eta}(u), \vec{n}(u) \rangle f_U(u)$, where $\vec{\eta}(u)$ denotes the rate of change in the boundary point $u \in \partial V(y,x)$ through this expansion of the area of $V(y,x)$.

Finally, we decompose these marginal probability changes into structural components. For the relation $0 = g(x,u) - y$ that holds on the boundary $\partial V(y,x)$ under regularity conditions, the vectors $\vec{n}(u)$, $\vec{\omega}(u)$, and $\vec{\eta}(u)$ can be written in terms of the structure as follows.

$$\vec{n}(u) = \left(\frac{\partial g(x,u)/\partial u_1}{\|\nabla_u g(x,u)\|}, \frac{\partial g(x,u)/\partial u_2}{\|\nabla_u g(x,u)\|} \right)',$$

$$\vec{\omega}(u) = \left(\frac{-\beta(x,u)}{\partial g(x,u)/\partial u_1}, \frac{-\beta(x,u)}{\partial g(x,u)/\partial u_2} \right)', \text{and}$$

$$\vec{\eta}(u) = \left(\frac{1}{\partial g(x,u)/\partial u_1}, \frac{1}{\partial g(x,u)/\partial u_2} \right)'.$$

Therefore, the marginal local probability loss and the marginal local probability gain are expressed in terms of the structure as follows.

$$\text{Marginal local probability loss:} \quad -\langle \vec{\omega}(u), \vec{n}(u) \rangle f_U(u) = 2 \frac{\beta(x,u)}{\|\nabla_u g(x,u)\|} f_U(u)$$

and

$$\text{Marginal local probability gain:} \quad \langle \vec{\eta}(u), \vec{n}(u) \rangle f_U(u) = 2 \frac{1}{\|\nabla_u g(x,u)\|} f_U(u).$$

The QPD is the marginal response of y to x while maintaining the probability balance between the total loss and the total gain. Therefore, the relation

"Total y-marginal probability gain $\times$ QPD = Total x-marginal probability loss"

must be satisfied under the law of probability conservation, and we thus obtain the QPD decomposition formula:

$$\begin{aligned} \frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x) &= \frac{\text{Total } x\text{-marginal probability loss}}{\text{Total } y\text{-marginal probability gain}} \\ &= \frac{\int_{\partial V(y,x)} \frac{\beta(x,u)}{\|\nabla_u g(x,u)\|} f_U(u) du}{\int_{\partial V(y,x)} \frac{1}{\|\nabla_u g(x,u)\|} f_U(u) du}. \end{aligned}$$

This heuristically derived equality shows that the QPD can be interpreted as a weighted average of the true structural partial effects $\beta(x,u)$, where the weights are proportional to $f_U(u)/\|\nabla_u g(x,u)\|$. Therefore, albeit the unknown weights,

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

the local slope parameter of the quantile regression can be used as a sensible measure of the true structural causal effects, even if the monotonicity assumption fails.

This informal result has a couple of problems. First, it was not logically derived, and the assumptions required for its derivation remain vague. Second, the integral is not precisely defined, and the assumptions for integrability are not stated. The next section presents a formal derivation of this heuristically discovered characterization of the QPD.

3. THE RESULT

We introduce the algebra on $\partial V(y,x)$ defined by $\mathcal{B}(y,x) := \{S \cap \partial V(y,x) \mid S \in \mathcal{B}(\mathbb{R}^M)\}$, where $\mathcal{B}(\mathbb{R}^M)$ is the Borel σ-algebra. The boundary $\partial V(y,x)$ itself is a Borel set, because it is closed regardless of whether $g(x, \cdot)$ is continuous. This fact implies that $\mathcal{B}(y,x) \subset \mathcal{B}(\mathbb{R}^M)$ holds, i.e., every element $S \in \mathcal{B}(y,x)$ is also a Borel set. Let m^M denote the Lebesgue measure on $\mathbb{R}^M$, and let H^{M-1} denote the $(M-1)$-dimensional Hausdorff measure restricted to $(\partial V(y,x), \mathcal{B}(y,x))$—see Section A.1 in the Appendix for its definition. The velocities of the boundary $\partial V(y,x)$ at u with respect to a change in y and a change in x are denoted by $\partial v(y,x;u)/\partial y$ and $\partial v(y,x;u)/\partial x$, respectively.

With Σ denoting an $(M-1)$-dimensional rectangle, the boundary $\partial V(\cdot,\cdot)$ can be represented by a map $\Sigma \times \mathcal{Y} \times \mathcal{X} \to \mathbb{R}^M$, i.e., the $(M-1)$-dimensional manifold $\partial V(y,x)$ can be parameterized through a map $\pi_{(y,x)} : \Sigma \to \partial V(y,x)$ for each $(y,x) \in \mathcal{Y} \times \mathcal{X}$. Furthermore, the velocity $\partial v(y,\cdot\,;\cdot)/\partial x$ of the boundary with respect to a change in x with y fixed can be represented by a map $\mathcal{X} \to L^1(\Sigma)$ for each $y \in \mathcal{Y}$, provided that the function $\partial v(y,x\,;\cdot)/\partial x : \partial V(y,x) \to \mathbb{R}^M$ is absolutely integrable with respect to the parameterization over Σ for each $(y,x) \in \mathcal{Y} \times \mathcal{X}$. Similarly, $\partial v(\cdot,x\,;\cdot)/\partial y$ can be represented by a map $\mathcal{Y} \to L^1(\Sigma)$ for each $x \in \mathcal{X}$ provided absolute integrability.

The following assumption establishes conditions on the structure $(g, F_{U|X})$ under which the QPD can be decomposed into a weighted average of structural partial effects $\beta(x,u)$.

Assumption 1. (i) g is continuously differentiable. (ii) $\|\nabla_u g(x, \cdot)\| \neq 0$ on $\partial V(y,x)$. (iii) The conditional distribution of U given X is absolutely continuous with respect to m^M, and $f_{U|X} \in C^1(\mathcal{X}; L^1(\mathbb{R}^M))$ holds. (iv) $\int_{\partial V(y,x)} f_{U|X}(u \mid x) dH^{M-1}(u) > 0$. (v) $\partial V(y,\cdot) \in C^1(\Sigma \times \mathcal{X}; \mathbb{R}^M)$ holds for each $y \in \mathcal{Y}$ and $\partial V(\cdot,x) \in C^1(\Sigma \times \mathcal{Y}; \mathbb{R}^M)$ holds for each $x \in \mathcal{X}$. (vi) $\partial v(y,\cdot\,;\cdot)/\partial x \in C^1(\mathcal{X}; L^1(\Sigma))$ holds for each $y \in \mathcal{Y}$ and $\partial v(\cdot,x\,;\cdot)/\partial y \in C^1(\mathcal{Y}; L^1(\Sigma))$ holds for each $x \in \mathcal{X}$.

Part (i) requires that the structural function g is smooth. Part (ii) requires that unobserved heterogeneity U is a nontrivial factor of the production g in the locality (y,x) of interest. This assumption ensures that the contour set

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

$\{u \in \mathbb{R}^M \mid g(x,u) = y\}$ does not consist of critical points of $g(x,\cdot)$. Part (iii) requires that unobserved heterogeneity is continuously distributed. The absolute continuity essentially rules out point masses, such as discretely supported categories, of unobservables. This assumption admits the conditional density (which we denote by $f_{U|X}$), but the conditional density function $f_{U|X}(\cdot \mid x)$ itself need not be continuous. While part (iii) is stated in terms of the conditional distribution, it can of course be stated in terms of the marginal distribution when X is exogenous, which is the case we consider mostly in the rest of the paper. Part (iv) requires that individuals with unobserved factors U residing on the boundary $\partial V(y,x)$ are substantially present with respect to the measure H^{M-1}. Part (v) requires that the boundary $\partial V(y,x)$ is smooth. Furthermore, part (vi) requires that movements of the boundary with changes in y and x are also smooth.

Under this set of regularity conditions, the following lemma suggests how to decompose the QPD into heterogeneous structural partial effects $\beta(x,u)$.

LEMMA 1 (Structural Decomposition of the QPD). *If Assumption 1 is true at* (y,x), $\tau := F_{Y|X}(y \mid x) \in (0,1)$, *and* $f_{U|X}(\cdot \mid \cdot)$ *is continuously differentiable, then the QPD* $\frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x)$ *exists and*

$$\frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x) = \frac{\int_{\partial V(y,x)} \frac{\beta(x,u)}{\|\nabla_u g(x,u)\|} \frac{f_{U|X}(u|x) \cdot M\pi^{(M-1)/2}}{2^{M-1}\Gamma(\frac{M+1}{2})} dH^{M-1}(u) - \int_{V(y,x)} \frac{\partial}{\partial x} f_{U|X}(u \mid x) dm^M(u)}{\int_{\partial V(y,x)} \frac{1}{\|\nabla_u g(x,u)\|} \frac{f_{U|X}(u|x) \cdot M\pi^{(M-1)/2}}{2^{M-1}\Gamma(\frac{M+1}{2})} dH^{M-1}(u)}$$

holds, where Γ *denotes the gamma function. In particular, when* X *is exogenous, we have*

$$\frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x) = \frac{\int_{\partial V(y,x)} \frac{\beta(x,u)}{\|\nabla_u g(x,u)\|} f_U(u) dH^{M-1}(u)}{\int_{\partial V(y,x)} \frac{1}{\|\nabla_u g(x,u)\|} f_U(u) dH^{M-1}(u)}. \tag{3.1}$$

Because quantile regressions describe relationships between x and y while keeping the probability τ fixed, the key idea behind this result is "the law of probability conservation," as informally discussed in Section 2. The numerator in the first equation measures the net probability loss induced by the surface outflow of the probability minus the probability gain within the volume as a result of an infinitesimal change in x, accounting for endogeneity. The denominator measures the surface inflow of probability as a result of an infinitesimal change in y. The ratio thus yields the marginal relation between x and y while compensating for the two net probability changes in order to conserve probability.

The first equation is apparently complicated since it accounts for endogeneity. If X is exogenous, then the conditional density $f_{U|X=x}$ reduces to the marginal density f_U, and the second term in the numerator can be eliminated so that the QPD simplifies to (3.1). The integrals in these equations are well-defined, because the Hausdorff measure is a metric outer measure and hence every Borel set $S \in \mathcal{B}(y,x) \subset \mathcal{B}(\mathbb{R}^M)$ is (H^{M-1})-measurable. While the integrals are defined,

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

whether or not they exist (i.e., are finite) may be another issue. Provided that they exist, this lemma shows that the QPD is decomposed into a weighted average of the true structural partial effects $\beta(x,u)$ on the boundary $\partial V(y,x)$ of the lower contour set $V(y,x)$.

In order to make a formal interpretation of the decomposition formula (3.1), we next discuss conditions for the integrability. Consider the function $\mu_{y,x} : \mathcal{B}(y,x) \to \mathbb{R}$ defined by

$$\mu_{y,x}(S) := \frac{\int_S \frac{1}{\|\nabla_u g(x,u)\|} f_U(u) dH^{M-1}(u)}{\int_{\partial V(y,x)} \frac{1}{\|\nabla_u g(x,u)\|} f_U(u) dH^{M-1}(u)} \qquad \text{for all } S \in \mathcal{B}(y,x).$$

Integrability allows this expression to be defined. The following assumption guarantees that those integrals exist,[2] $\mu_{y,x}$ is indeed defined, $\mu_{y,x}$ is a probability measure in particular, and thus the integrals can be interpreted as moments.

Assumption 2. Let $\gamma(x,u) := \|\nabla_u g(x,u)\|^{-1}$. There exist values $p \geqslant 1$ and $q \geqslant 1$ satisfying $p^{-1} + q^{-1} = 1$ such that $\|\gamma(x, \cdot)\|_{L^p(\partial V(y,x), H^{M-1})} < \infty$ and $\|f_U\|_{L^q(\partial V(y,x), H^{M-1})} < \infty$ hold.

This assumption requires that $\gamma(x, \cdot)$ and f_U reside in dual function spaces of each other. Lower-level assumptions can be stated with common choices of p and q, such as $(p,q) = (1,\infty), (2,2)$, and $(\infty,1)$. Imposing this assumption with $(p,q) = (\infty,1)$ is perhaps the most natural because we can then view $f_U \Big/ \int_{\partial V(y,x)} f_U dH^{M-1}$ as a "conditional density" and $\|\nabla_u g(x, \cdot)\|^{-1}$ as uniformly bounded weights.

LEMMA 2 (Weights). *If Assumption 2 is true at (y,x), then $\mu_{y,x}$ is defined and is a probability measure on $(\partial V(y,x), \mathcal{B}(y,x))$.*

By this lemma, Assumption 2 allows us to rewrite the decomposition formula (3.1) derived in Lemma 1 into the moment expression:

$$\frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x) = \int \beta(x,u) d\mu_{y,x}(u) = \mathbb{E}_{\mu_{y,x}}[\beta(x,U)]. \tag{3.2}$$

This equation formalizes our interpretation of it as a weighted average of the heterogeneous structural partial effects $\beta(x,u)$. The QPD is the mean of the true structural partial effects $\beta(x,U)$ with respect to the probability measure $\mu_{y,x}$ on $(\partial V(y,x), \mathcal{B}(y,x))$. The weights of this measure are proportional to $f_U / \|\nabla_u g(x, \cdot)\|$, but not in general to the actual density f_U of the unobservables.

Because $\mu_{y,x}$ is restrictively defined, the mean is taken only over the subpopulation of individuals with $u \in \partial V(y,x)$. Policy makers would perhaps ask which individuals constitute this subpopulation for which the weighted average structural partial effects are measured by the QPD. The next lemma characterizes this subpopulation under the assumption of sufficient smoothness and variability of $g(x, \cdot)$.

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

LEMMA 3 (Subpopulation). *If Assumption 1 (i) and (ii) are true at* (y, x)*, then* $\partial V(y, x) = \{u \in \mathbb{R}^M \mid g(x, u) = y\}$ *holds.*

In fact, this lemma is used in the proof of Lemma 1, which may appear to reverse the order of presentation, but they are logically coherent because the assumptions used for Lemma 1 encompass the assumptions used for Lemma 3. This lemma shows that the subpopulation $\partial V(y, x)$ of interest is the set of individuals residing in the "isoquantile"

$$\{u \in \mathbb{R}^M \mid g(x, u) = y\}$$

of $\tau = F_{Y|X}(y \mid x)$. Summarizing Lemmas 1–3 together, we obtain the following main result of this paper.

THEOREM 1 (Causal Interpretation of the QPD). *Suppose that Assumptions 1 and 2 are true at* (y, x) *and that* X *is exogenous. Then the QPD at* $\tau = F_{Y|X}(y \mid x) \in (0, 1)$ *is the mean structural partial effect:*

$$\frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x) = \mathbb{E}_{\mu_{y,x}}[\beta(x, U)],$$

where the mean is taken with respect to the probability measure $\mu_{y,x}$ *whose weights are proportional to*

$$\frac{f_U}{\|\nabla_u g(x, \cdot)\|}, \tag{3.3}$$

and the mean is taken over the subpopulation of individuals in the isoquantile $\{u \in \mathbb{R}^M \mid g(x, u) = y\}$ *of* $\tau = F_{Y|X}(y \mid x)$.

While the QPD does not exactly coincide with a single structural partial effect $\beta(x, u)$ in general, it is still a sensible measure of heterogeneous structural partial effects. Because the weights are positive on the support of the true underlying distribution F_U under Assumption 1, the QPD gives a strict convex combination of structural partial effects $\beta(x, u)$ within the isoquantile.

The weights (3.3) take the form of the ratio of the density of U to the gradient norm of the structural function with respect to u. In addition to the underlying density f_U, heavier weights are assigned to the locality of u at which the unobserved heterogeneity marginally matters less in the structural function g. Although the weight itself is not identified from data, it can be related to observed data. Under Assumption 1, similar calculations to the ones produced in the proof of Lemma 1 yield

$$\int_{\partial V(y,x)} \frac{f_U(u)}{\|\nabla_u g(x, u)\|} d\tilde{H}^{M-1}(m) = f_{Y|X=x}(y),$$

where $\tilde{H}^{M-1} := \frac{M\pi^{(M-1)/2}}{2^{M-1}\Gamma\left(\frac{M+1}{2}\right)} H^{M-1}$ is a rescaled Hausdorff measure. This equality shows that the weights integrate with respect to $\tilde{H}^{M-1}$ into the observed

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

conditional density $f_{Y|X=x}$. This relation is analogous to the fact that weights of Angrist et al. (2006) can be written as an integral of the observed conditional density $f_{Y|X=x}$.

In order to emphasize the distinction between weighted and unweighted averages, consider a simple example of structural functions that entail nonconstant weights in the structural decomposition of the QPD. Consider the structural function defined by

$$g(x,u_1,u_2) = \left[u_1^{\rho}+u_2^{\rho}\right]^{\frac{1}{\rho}}\cdot x \qquad \text{for all } x \geqslant 0 \text{ and } (u_1,u_2)\in \text{supp}(U)\subset \mathbb{R}^2_+.$$

Given the observed factor X fixed, this structure exhibits a constant elasticity of substitution (CES) between the unobserved factors U_1 and U_2. Some calculations show $\|\nabla_u g(x,u)\| = \left[u_1^{2\rho-2}+u_2^{2\rho-2}\right]^{\frac{1}{2}}\cdot x^{\rho}\cdot y^{1-\rho}$ where $y = g(x,u_1,u_2)$. Consequently, the weight factor $1/\|\nabla_u g(x,\cdot)\|$ is generally nonconstant in (u_1,u_2) on $\partial V(y,x) = \left\{u\in \text{supp}(U)\subset \mathbb{R}^2_+ \,\middle|\, \left[u_1^{\rho}+u_2^{\rho}\right]^{\frac{1}{\rho}} = \frac{y}{x}\right\}$ for fixed (y,x), and thus the QPD generally fails to be the unweighted average of the structural partial effects. However, there is a point of ρ at which the weight factor reduces to a constant, namely $1/\|\nabla_u g(x,\cdot)\| \equiv \sqrt{2}\cdot x$ whenever $\rho = 1$. Note that $\rho = 1$ has the unobserved factors U_1 and U_2 as perfect substitutes of each other. At any other value of ρ, these factors are complementary, and the weight factor $1/\|\nabla_u g(x,u)\|$ is not constant in u. This inspection suggests that the structural complementarity among multiple unobservables is the source of nonconstant weights in the QPD decomposition formula.

The lack of complementarity among unobservables is the case with the linear random coefficient models:

$$g(x,u_1,u_2) = u_1 + u_2 x. \tag{3.4}$$

It is well known that the random coefficients correspond to the QPDs under the assumption of "comonotonicity"—see Koenker (2005, Sect. 2.6). The main theorem of this paper provides a generalization of this random-coefficient interpretation of quantile regressions for the case in which the comonotonicity assumption fails. Given that the weight factor $1/\|\nabla_u g(x,u)\|$ is constant in u under the model (3.4), Theorem 1 implies that the QPD, $\frac{\partial}{\partial x}Q_{Y|X}(\tau \mid x)$, identifies the *unweighted* conditional mean of the random coefficient, i.e., $\mathbb{E}_{\mu_{y,x}}[U_2]$, where $\mu_{y,x}$ is the weight-free measure over the subpopulation at the τ-th isoquantile.

4. SUMMARY

The QPD $\frac{\partial}{\partial x}Q_{Y|X}(\tau \mid x)$ measures a weighted average of structural partial effects $\beta(x,u)$ within the subpopulation of individuals in the isoquantile $\{u\in\mathbb{R}^M \mid g(x,u) = y\}$, where the weights are proportional to $f_U/\|\nabla_u g(x,\cdot)\|$. This result implies that quantile regression parameters are still useful for inference of

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

heterogeneous causal effects even if the quantile regression equipped with only a scalar latent variable misspecifies the true structural function which entails potentially multidimensional unobserved explanatory variables. In particular, the QPD identifies the unweighted conditional mean of the random coefficient in linear random coefficient models without the comonotonicity assumption. This paper contributes to the literature of empirical economics by warranting that quantile regressions are effective statistical tools to analyze general structural functions with multiple unobservables. One need not invoke the monotonicity assumption in order to justify his/her use of quantile regressions if only weighted averages of structural causal effects are investigated.

NOTES

1. Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007), Imbens and Newey (2009), d'Haultfoeuille and Février (2011), and Torgovitsky (2011) propose IV-based approaches to identify the potential quantile regression $Q_{Y|X}$ when X is endogenous in data. Also related is Chesher (2003), but he explicitly identifies the partial effect for a structural function g with $M > 1$, instead of $Q_{Y|X}$, under a recursive model. Furthermore, Schennach (2008) shows identification of quantile regressions with measurement errors using IVs.

2. Assumption 2 guarantees that the weights $f_U / \|\nabla_u g(x, \cdot)\|$ are absolutely integrable on the boundary $\partial V(y, x)$, and hence the integral in the QPD decomposition formula (3.1) exists under Assumptions 1 and 2. Specifically, the weights $f_U / \|\nabla_u g(x, \cdot)\|$ are shown to belong to $L^1(\partial V(y, x), H^{M-1})$ under Assumption 2 in the proof of Lemma 2, and the structural partial effects $\beta(x, \cdot)$ belong to $L^\infty(\partial V(y, x), H^{M-1})$ under Assumption 1. Hence, it follows that their product is integrable on the boundary $(\partial V(y, x), H^{M-1})$.

3. A set $S \subset \mathbb{R}^M$ is $(M-1)$-rectifiable if there exists a Lipschitz mapping of a bounded measurable set $S' \subset \mathbb{R}^{M-1}$ to S.

REFERENCES

Angrist, J., V. Chernozhukov, & I. Fernández-Val (2006) Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica* 74(2), 539–563.

Angrist, J.D. & G.W. Imbens (1995) Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90(430), 431–442.

Chernozhukov, V. & C. Hansen (2005) An IV model of quantile treatment effects. *Econometrica* 73(1), 245–261.

Chernozhukov, V., G.W. Imbens, & W.K. Newey (2007) Instrumental variable estimation of nonseparable models. *Journal of Econometrics* 139(1), 4–14.

Chesher, A. (2003) Identification in nonseparable models. *Econometrica* 71(5), 1405–1441.

d'Haultfoeuille, X. & P. Février (2011) Identification of Nonseparable Models with Endogeneity and Discrete Instruments. Working paper, CREST.

Heckman, J.J. (2007) The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences* 104(33), 13250–13255.

Hoderlein, S. & E. Mammen (2007) Identification of marginal effects in nonseparable models without monotonicity. *Econometrica* 75(5), 1513–1518.

Imbens, G.W. & W.K. Newey (2009) Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77(5), 1481–1512.

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

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

Koenker, R. & G. Bassett, Jr. (1978) Regression quantiles. *Econometrica* 46(1), 33–50.
Koenker, R. & K.F. Hallock (2001) Quantile regression. *Journal of Economic Perspectives* 15(4), 143–156.
Lukeš, J. & J. Malý (2005) *Measure and Integration*. Matfyz Press.
Padula, M. (2011) *Asymptotic Stability of Steady Compressive Fluids*. Springer.
Rogers, C.A. & K. Falconer (1998) *Hausdorff Measures*. Cambridge University Press.
Schennach, S.M. (2008) Quantile regression with mismeasured covariates. *Econometric Theory* 24(4), 1010–1043.
Sen, A. (1985) A sociological approach to the measurement of poverty: A reply to professor Peter Townsend. *Oxford Economic Papers* 37(4), 669–676.
Torgovitsky, A. (2011) Identification and Estimation of Nonparametric Quantile Regressions with Endogeneity. Working paper, Northwestern University.

MATHEMATICAL APPENDIX

A.1. Notes on Differentiation under Integral Sign

This appendix section introduces an auxiliary lemma. Let $V(t) \subset \mathbb{R}^M$ denote a volume which may vary with time $t \in (0,T) \subset \mathbb{R}$, and let $\partial V(t)$ denote its surface. The velocity of the surface at $z \in \mathbb{R}^M$ is denoted by $\partial v(z,t)/\partial t$, and the unit outward normal vector of the surface at z is denoted by $\nu(z,t)$. If Σ denotes an $(M-1)$-dimensional rectangle, then we can represent the moving surface $\partial V(\cdot)$ as a map $\Sigma \times (0,T) \to \mathbb{R}^M$, i.e., for each $t \in (0,T)$ the $(M-1)$-dimensional manifold $\partial V(t)$ can be parameterized through $\pi_t : \Sigma \to \partial V(t)$ by elements in the rectangle $\Sigma \subset \mathbb{R}^{M-1}$. Furthermore, the velocity $\partial v/\partial t$ of the moving surface can be represented by a map $(0,T) \to L^1(\Sigma)$, provided that the function $\partial v(\cdot,t)/\partial t : \partial V(t) \to \mathbb{R}^M$ is absolutely integrable with respect to the parameterization over Σ for each $t \in (0,T)$, i.e., the composition $\Sigma \overset{\pi_t}{\to} \partial V(t) \overset{\partial v(\cdot,t)/\partial t}{\longrightarrow} \mathbb{R}^M$ is $L^1(\Sigma)$ for each $t \in (0,T)$. With these representations, Padula (2011, p. 12, Lem. 1.3.2) provides the following statement.

LEMMA 4. *Suppose that* $\partial V \in C^1(\Sigma \times (0,T); \mathbb{R}^M)$ *and* $\partial v/\partial t \in C^1((0,T); L^1(\Sigma))$. *If* $h \in C^1((0,T); L^1(\mathbb{R}^M))$, *then we have*

$$\frac{d}{dt}\int_{V(t)} h(z,t)\, d\mathcal{V}(z) = \int_{V(t)} \frac{\partial h(z,t)}{\partial t}\, d\mathcal{V}(z) + \oint_{\partial V(t)} h(z,t)\left\langle \frac{\partial v(z,t)}{\partial t}, \nu(z,t)\right\rangle d\mathcal{S}(z). \tag{A.1}$$

The formula involves two different integral symbols, $\int$ and $\oint$. They are defined with respect to distinct measures, $\mathcal{V}$ and $\mathcal{S}$. Since this distinctness is one of the most important aspects of this paper, we make it and the definitions of these measures more explicit. The first integral $\int$ is usually understood as the Riemann integral on $\mathbb{R}^M$, but is more generally defined with respect to the Lebesgue measure m^M on $\mathbb{R}^M$. Hence, we treat $\mathcal{V}$ as m^M. On the other hand, the second integral $\oint$ is the surface integral, and is defined with respect to an $(M-1)$-dimensional measure on $\mathbb{R}^M$ through parameterization. An $(M-1)$-dimensional measure on $\mathbb{R}^M$ need not be unique in general, but all

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

$(M-1)$- dimensional measures on $\mathbb{R}^M$ coincide on the σ-algebra generated by $(M-1)$-rectifiable sets.[3] Among many unweighted $(M-1)$-dimensional measures, there exist ones that are defined on all Borel subsets of $\mathbb{R}^M$. One of them is the $(M-1)$-dimensional Hausdorff measure obtained by the restriction of the function $H^{M-1}: 2^{\mathbb{R}^M} \to \mathbb{R}$ defined by $H^{M-1}(S) = \sup_{\delta>0} \inf\left\{\sum_{i=1}^{\infty} (\mathrm{diam} U_i)^{M-1} \mid \cup_{i=1}^{\infty} U_i \supset S,\ \mathrm{diam} U_i < \delta\right\}$ to the collection of Carathéodory-measurable sets. Since the $(M-1)$-dimensional Hausdorff measure does not evaluate the $(M-1)$-dimensional unit square to be one, we usually normalize it by the constant $H^{M-1}([0,1]^{M-1} \times \{0\}) \equiv \left(\frac{4}{\pi}\right)^{\frac{M-1}{2}} \Gamma\left(1+\frac{M-1}{2}\right)$ to obtain the normalized $(M-1)$-dimensional Hausdorff measure. Hence, we treat $\mathcal{S}$ as $\frac{\pi^{(M-1)/2}}{2^{M-1}\cdot\Gamma(\frac{M+1}{2})} H^{M-1}$. (Let us remark that this particular choice of $\mathcal{S}$ is not crucial for our result, but we make such a concrete choice in order to avoid mathematical ambiguity in our arguments.) See Rogers and Falconer (1998) and Lukeš and Malý (2005) for comprehensive reviews of these concepts. With these technical auxiliaries, we informatively write the standard formula (A.1) to our goal as follows:

$$\frac{d}{dt}\int_{V(t)} h(z,t)\, dm^M(z) = \int_{V(t)} \frac{\partial h(z,t)}{\partial t}\, dm^M(z) + \int_{\partial V(t)} \left\langle \frac{\partial v(z,t)}{\partial t}, v(z,t) \right\rangle \frac{h(z,t)\cdot \pi^{(M-1)/2}}{2^{M-1}\cdot\Gamma(\frac{M+1}{2})}\, dH^{M-1}(z).$$

A.2. Proof of Lemma 1

Proof. By Assumption 1(iii), we can write

$$\mathcal{T}(y,x) := F_{Y|X}(y \mid x) = P_{U|X=x}(V(y,x)) = \int_{V(y,x)} f_{U|X=x} dm^M(u).$$

Let $\partial v(y,x;u)/\partial x$ be the velocity of the boundary $\partial V(y,x)$ with respect to a change in x at u, and let $v(y,x;u)$ be the unit outward normal of the surface $\partial V(y,x)$ at u. By the definition of $V(y,x)$ satisfying the relation $0 = g(x,u) - y$ on its boundary following Lemma 3, these two vectors are explicitly written under Assumption 1(i) and (ii) as follows:

$$\frac{\partial v(y,x;u)}{\partial x} = \left(\frac{-\beta(x,u)}{\partial g(x,u)/\partial u_1}, \cdots, \frac{-\beta(x,u)}{\partial g(x,u)/\partial u_M}\right)'$$
$$v(y,x;u) = \frac{\nabla_u g(x,u)}{\|\nabla_u g(x,u)\|}.$$

We adapt the current setting to Lemma 4 by letting x and u play the roles of t and z, respectively, for a fixed y. Our notations $\partial V(y,\cdot)$, $\partial v(y,\cdot;\cdot)/\partial x$, and $f_{U|X}(\cdot \mid \cdot)$ respectively correspond to $\partial V(\cdot)$, $\partial v(\cdot,\cdot)/\partial t$, and $h(\cdot,\cdot)$ in terms of the notations of Lemma 4. With these notational translations, Assumption 1(iii), (v), and (vi) directly satisfies the conditions of Lemma 4, and hence we obtain

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

$$\begin{aligned}\frac{\partial}{\partial x}\mathcal{T}(y,x) &= \int_{V(y,x)} \frac{\partial}{\partial x} f_{U|X}(u \mid x) dm^M(u) \\ &\quad + \int_{\partial V(y,x)} \left\langle \frac{\partial v(y,x;u)}{\partial x},\ v(y,x;u) \right\rangle \frac{f_{U|X}(u \mid x) \cdot \pi^{(M-1)/2}}{2^{M-1}\Gamma(\frac{M+1}{2})} dH^{M-1}(u) \\ &= \int_{V(y,x)} \frac{\partial}{\partial x} f_{U|X}(u \mid x) dm^M(u) \\ &\quad - \int_{\partial V(y,x)} \frac{\beta(x,u)}{\|\nabla_u g(x,u)\|} \frac{f_{U|X}(u \mid x) \cdot M\pi^{(M-1)/2}}{2^{M-1}\Gamma(\frac{M+1}{2})} dH^{M-1}(u).\end{aligned}$$

Also, since

$$\frac{\partial v(y,x;u)}{\partial y} = \left(\frac{1}{\partial g(x,u)/\partial u_1}, \cdots, \frac{1}{\partial g(x,u)/\partial u_M} \right)'$$

holds under Assumption 1(i) and (ii) by the definition of $V(y,x)$ satisfying the relation $0 = g(x,u) - y$ on its boundary by Lemma 3,

$$\left\langle \frac{\partial v(y,x;\cdot)}{\partial y},\ v(y,x;\cdot) \right\rangle = \frac{M}{\|\nabla_u g(x,\cdot)\|}.$$

We now adapt the current setting to Lemma 4 by letting y and u play the roles of t and z, respectively, for a fixed x. Our notations $\partial V(\cdot,x)$ and $\partial v(\cdot,x;\cdot)/\partial y$ respectively correspond to $\partial V(\cdot)$ and $\partial v(\cdot,\cdot)/\partial t$ in terms of the notations of Lemma 4. Furthermore, since $f_{U|X}$ does not depend on y, h in the notations of Lemma 4 is constant and thus $\partial h/\partial t = 0$. With these notational translations, Assumption 1 (iii), (v), and (vi) directly satisfies the conditions of Lemma 4, and hence we obtain

$$\begin{aligned}\frac{\partial}{\partial y}\mathcal{T}(y,x) &= \int_{\partial V(y,x)} \left\langle \frac{\partial v(y,x;u)}{\partial y},\ v(y,x;u) \right\rangle \frac{f_{U|X}(u \mid x) \cdot \pi^{(M-1)/2}}{2^{M-1}\Gamma(\frac{M+1}{2})} dH^{M-1}(u) \\ &= \int_{\partial V(y,x)} \frac{1}{\|\nabla_u g(x,u)\|} \frac{f_{U|X}(u \mid x) \cdot M\pi^{(M-1)/2}}{2^{M-1}\Gamma(\frac{M+1}{2})} dH^{M-1}(u).\end{aligned}$$

With this expression, observe that Assumption 1(iv) implies that

$$\frac{\partial}{\partial y}\mathcal{T}(y,x) > 0.$$

Because the derivative of the quantile regression is the derivative dy/dx along the fixed quantile $\mathcal{T}(y,x) = \tau$ by definition, we can use Assumption 1, the continuous differentiability of $f_{U|X}(u \mid \cdot)$ assumed in the lemma, and the fact that $\frac{\partial}{\partial y}\mathcal{T}(y,x) \neq 0$ claimed above to deduce

$$\frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x) = -\frac{\frac{\partial}{\partial x}\mathcal{T}(y,x)}{\frac{\partial}{\partial y}\mathcal{T}(y,x)}.$$

Therefore, by substituting the structural expressions for $\frac{\partial}{\partial x}\mathcal{T}(y,x)$ and $\frac{\partial}{\partial y}\mathcal{T}(y,x)$ obtained above, we can write the QPD as

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press

$$\frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x)$$
$$= \frac{\int_{\partial V(y,x)} \frac{\beta(x,u)}{\|\nabla_u g(x,u)\|} \frac{f_{U|X}(u|x)\cdot M\pi^{(M-1)/2}}{2^{M-1}\Gamma(\frac{M+1}{2})} dH^{M-1}(u) - \int_{V(y,x)} \frac{\partial}{\partial x} f_{U|X}(u \mid x) dm^M(u)}{\int_{\partial V(y,x)} \frac{1}{\|\nabla_u g(x,u)\|} \frac{f_{U|X}(u|x)\cdot M\pi^{(M-1)/2}}{2^{M-1}\Gamma(\frac{M+1}{2})} dH^{M-1}(u)}.$$

Finally, in the case of $X \perp\!\!\!\perp U$, set $f_{U|X=x}$ to f_U so we obtain the simplified QPD decomposition formula

$$\frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x) = \frac{\int_{\partial V(y,x)} \frac{\beta(x,u)}{\|\nabla_u g(x,u)\|} f_U(u) dH^{M-1}(u)}{\int_{\partial V(y,x)} \frac{1}{\|\nabla_u g(x,u)\|} f_U(u) dH^{M-1}(u)}$$

as claimed. ■

A.3. Proof of Lemma 2

Proof. Under Assumption 2, the Hölder's inequality yields

$$\int_{\partial V(y,x)} \left| \frac{1}{\|\nabla_u g(x,u)\|} f_U(u) \right| dH^{M-1}(u) \leqslant \|\gamma(x,u)\|_{L^p(\partial V(y,x), H^{M-1})} \cdot \|f_U\|_{L^q(\partial V(y,x), H^{M-1})} < \infty,$$

showing that $\mu_{y,x}$ is defined. The null equality $\mu_{y,x}(\emptyset) = 0$ holds by the fact that $H^{M-1}(\emptyset) = 0$ for H^{M-1} being a measure. The sigma additivity of $\mu_{y,x}$ also follows from that of H^{M-1}, i.e., for any countable sequence $S_1, S_2, \ldots \in \mathcal{B}(y,x)$ of disjoint sets, we have

$$\mu_{y,x}(\cup_i S_i) = \frac{\int_{\cup_i S_i} \frac{1}{\|\nabla_u g(x,u)\|} f_U(u) dH^{M-1}(u)}{\int_{\partial V(y,x)} \frac{1}{\|\nabla_u g(x,u)\|} f_U(u) dH^{M-1}(u)} = \frac{\sum_i \int_{S_i} \frac{1}{\|\nabla_u g(x,u)\|} f_U(u) dH^{M-1}(u)}{\int_{\partial V(y,x)} \frac{1}{\|\nabla_u g(x,u)\|} f_U(u) dH^{M-1}(u)} = \sum_i \mu_{y,x}(S_i).$$

Therefore, $\mu_{y,x}$ is a signed measure on $(\partial V(y,x), \mathcal{B}(y,x))$. Because $f_U \geqslant 0$ is true as a density function and H^{M-1} is a positive measure as a metric outer measure, we have $\mu_{y,x}(S) \geqslant 0$ for all $S \in \mathcal{B}(y,x)$. This shows that $\mu_{y,x}$ is a positive measure on $(\partial V(y,x), \mathcal{B}(y,x))$. Lastly, $\mu_{y,x}(\partial V(y,x)) = 1$ clearly holds by the definition of $\mu_{y,x}$. Therefore, $\mu_{y,x}$ is a probability measure on $(\partial V(y,x), \mathcal{B}(y,x))$. ■

A.4. Proof of Lemma 3

Proof. Because Assumption 1(i) implies that $g(x, \cdot)$ is continuous, we have $V(y,x) = \overline{V}(y,x)$. But then, $\partial V(y,x) = \overline{V}(y,x) \setminus \overset{\circ}{V}(y,x) = V(y,x) \setminus \overset{\circ}{V}(y,x) \subset \{u \in \mathbb{R}^M \mid g(x,u) = y\}$ follows. Conversely, let $u' \in \{u \in \mathbb{R}^M \mid g(x,u) = y\}$. For any $\epsilon > 0$, there exist u'' and u''' such that $g(x,u'') < y$, $g(x,u''') > y$, $\|u' - u''\| < \epsilon$, and $\|u' - u'''\| < \epsilon$ by Assumption 1(ii). Therefore, $u' \in \partial V(y,x)$. ■

https://doi.org/10.1017/S0266466614000711 Published online by Cambridge University Press