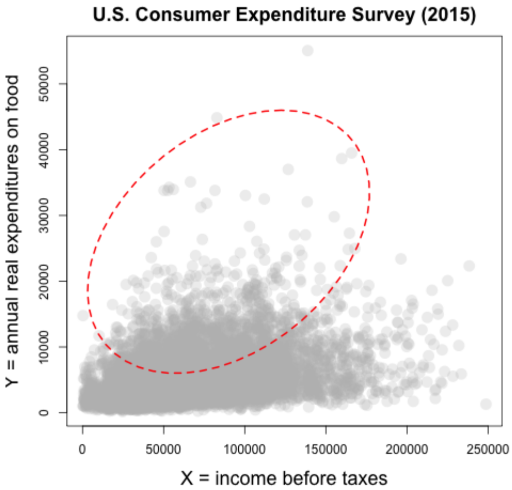
Erasmus University Rotterdam

Multivariate quantile regression using superlevel sets of conditional densities*

Kathrin Gruber
Department of Econometrics
Erasmus School of Economics
gruber@ese.eur.nl

*Joint work with Annika Camehl and Dennis Fok

Example: Heterogeneity in Household Consumption Patterns

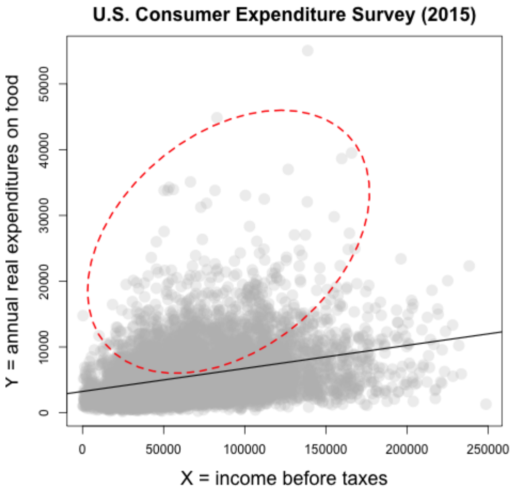


$\{y_n, x_n\}_{n=1}^{N}$ and $N = 29,988$

$y_n = \mu + x_n'\beta + u_n$

Assume, our interest is in the "top" (i.e., 10%) household segment.

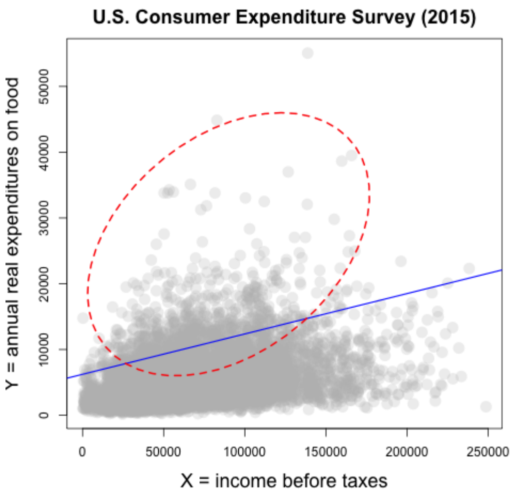
Example: Heterogeneity in Household Consumption Patterns



(Conditional) mean regression

$$\overline{m}(x_n) = \mu + x_n'\beta$$

Example: Heterogeneity in Household Consumption Patterns

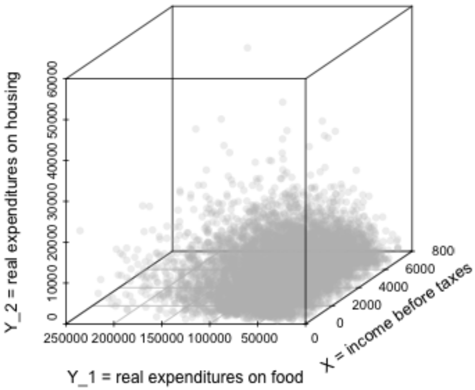


(Conditional) quantile regression

$$q(.1|x_n) = \mu_{(.1)} + x_n' \beta_{(.1)}$$

Example: Heterogeneity in Household Consumption Patterns

that is: $\boldsymbol{y}_n = (y_{n1}, \ldots, y_{nK})'$



$$\boldsymbol{q}(\alpha|\boldsymbol{x}_n) = \boldsymbol{\mu}_{(\alpha)} + \boldsymbol{B}_{(\alpha)}\boldsymbol{x}_n$$

where $\boldsymbol{\mu}_{(\alpha)}$ is $K \times 1$ and $\boldsymbol{B}_{(\alpha)}$ is $K \times G$

$\Rightarrow$ Seemingly unrelated regression, simultaneous equations, VAR ...

The conditional (Koenker-Bassett) quantile concept is not easy to extend!

This Talk

- The quantile is defined as a property of an (estimated) conditional multivariate density.
- This so-called super-level set enables a clear probabilistic interpretation and enjoys favorable quantile properties.
- Linear and non-linear multivariate as well as univariate regression quantiles are obtained in a comprehensive (fully) Bayesian framework.

Multivariate Quantiles

Attempt 1: Conditional

$$\boldsymbol{q}(\alpha) = [q_{y_{n1}}(\alpha|\boldsymbol{y}_{n(-1)}), \ldots, q_{y_{nK}}(\alpha|\boldsymbol{y}_{n(-K)})]'$$

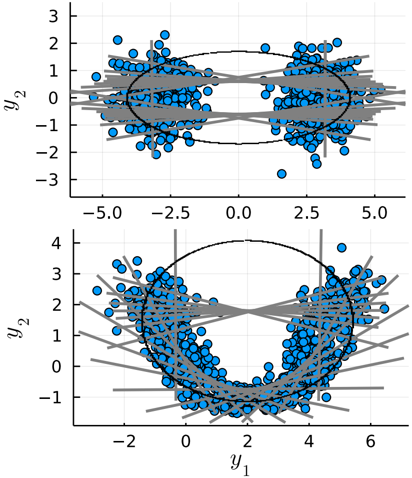
- Input space augmentation.
- Assumes all "regressors" are fixed!

Multivariate Quantiles

Attempt 2: Directional

- Convex intersection of α-quantile halfspaces for different (Koenker-Bassett) regression hyperplanes (Hallin, Paindaveine & Siman, 2010).
- The (directional) quantile contours are not guaranteed to cover α.

Areas within the gray lines (contours) give the 80%-**directional quantile** (20 directions).

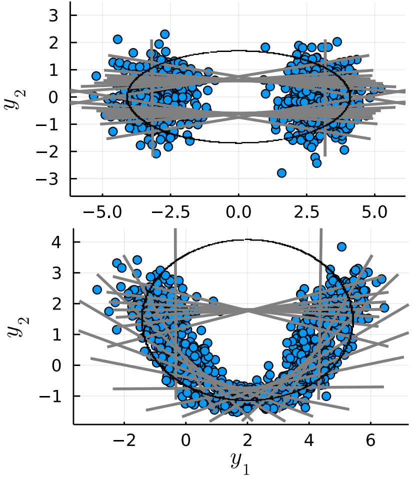


Multivariate Quantiles

Attempt 3: Direct

- Find an ellipsoid around a (determined) center with α-probability mass (e.g., Hallin & Siman, 2016).
- The quantile regions can cover large parts with little to no probability mass.

Areas within the black lines (countours) give the 80%-**elliptical quantile**.



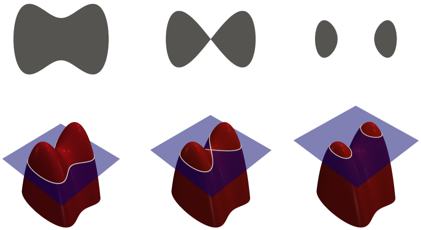
Multivariate Quantiles

Level set

$$\mathcal{L}(f;t) = \left\{ \boldsymbol{y}_n \in \mathbb{R}^K : f(\boldsymbol{y}_n) = t \right\}$$

$\Rightarrow$ Cross-section of $f(\cdot)$ at a given (constant) value t (Osher & Sethian, 1988).

Level sets of a bivariate bimodal distribution for three different values of t.



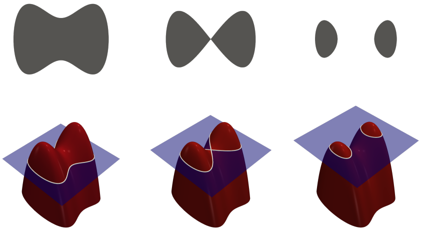
Multivariate Quantiles

Super-level set

$$\mathcal{L}(f;t) = \left\{ \boldsymbol{y}_n \in \mathbb{R}^K : f(\boldsymbol{y}_n) \geq t \right\}$$

for threshold $t > 0$, gives the highest density region for $f(\cdot)$ (see, e.g., Hartigan, 1987).

Level sets of a bivariate bimodal distribution for three different values of t.



Multivariate Quantiles

Super-level set

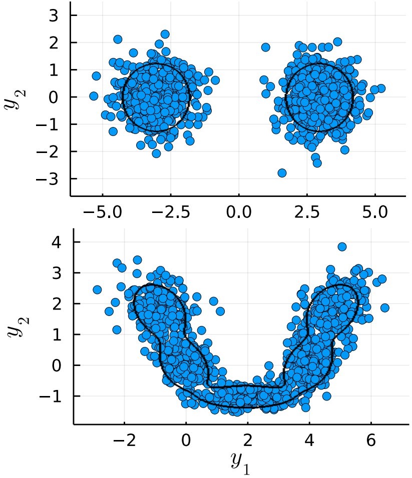
$$\mathcal{L}(f;t) = \left\{ \boldsymbol{y}_n \in \mathbb{R}^K : f(\boldsymbol{y}_n) \geq t \right\}$$

for threshold $t > 0$, gives the highest density region for $f(\cdot)$ (see, e.g., Hartigan, 1987).

Super-level set quantile

$$\boldsymbol{q}(\alpha) = \mathcal{L}(f; t^\star_\alpha),$$
$$t^\star_\alpha = \sup \left\{ Pr(\boldsymbol{y}_n \in \mathcal{L}(f;t)) \geq \alpha \right\}$$

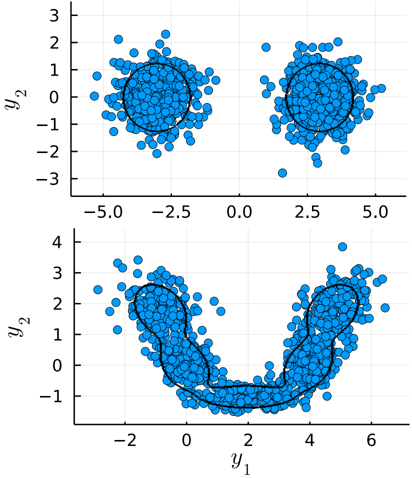
Areas within the lines (quantile contours) correspond to 80% probability mass.



Multivariate Quantiles

- Supports a clear probabilistic interpretation (in terms of α).
- (Flexible) quantile regions cover areas with high probability mass.
- Extensions to more than two outputs are straightforward.

Areas within the lines (quantile contours) correspond to 80% probability mass.



Super-Level Set Quantiles vs. HPD Sets

Highest posterior density set

- Operationalizes uncertainty of a model parameter for a (typically) univariate and unimodal posterior distribution (i.e., an interval, Box & Tiao, 1965).

Super-level set

- Quantifies uncertainty in a (set of) response variable(s), conditional on other response variables, for arbitrarily shaped joint posterior distributions (i.e., an interval or a set of intervals).

Let's go (fully) Bayesian

(Overfitted) Finite Gaussian Mixture Model:

$$f\Big(\boldsymbol{y}_n|\boldsymbol{x}_n\Big) = \sum_{m=1}^{M} \kappa_m \phi\Big(\boldsymbol{g}_m(\boldsymbol{x}_n), \boldsymbol{\Sigma}_m\Big),$$

$$\text{where } \boldsymbol{g}_m(\boldsymbol{x}_n) = \boldsymbol{\mu}_m + \boldsymbol{B}_m \boldsymbol{x}_n,$$

$$\boldsymbol{\kappa}|\{\bar{\rho}_m\} \sim \mathcal{D}\Big(\bar{\rho}_1, \cdots, \bar{\rho}_M\Big),$$

$$\bar{\rho}_m \sim \mathcal{G}\Big(\underline{a}_1, 1/(\underline{a}_2 M)\Big).$$

with M comparatively large (Nobile & Fearnside, 2007; Rousseau & Mengersen, 2011) and a Shrinkage Prior on $\phi(\boldsymbol{g}_m(\boldsymbol{x}_n), \boldsymbol{\Sigma}_m)$ (Malsiner-Walli, Frühwirth-Schnatter & Grün, 2016).

Implementation

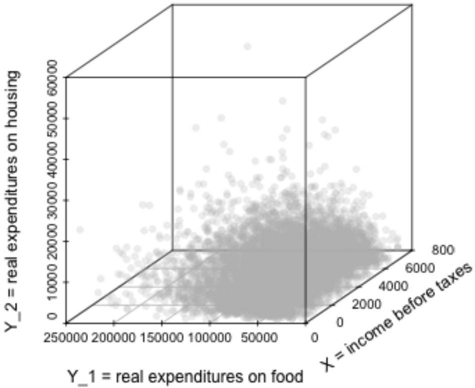
$$\boldsymbol{\mu}_m^{\mathcal{K}|\mathcal{C}}(\boldsymbol{y}_\mathcal{C}, \boldsymbol{x}) = \boldsymbol{g}_{m,\mathcal{K}}(\boldsymbol{x}) + \boldsymbol{\Sigma}_{m,\mathcal{K},\mathcal{C}} \boldsymbol{\Sigma}_{m,\mathcal{C},\mathcal{C}}^{-1} \Big(\boldsymbol{y}_\mathcal{C} - \boldsymbol{g}_{m,\mathcal{C}}(\boldsymbol{x}) \Big),$$

$$\boldsymbol{\Sigma}_m^{\mathcal{K}|\mathcal{C}} = \boldsymbol{\Sigma}_{m,\mathcal{K},\mathcal{K}} - \boldsymbol{\Sigma}_{m,\mathcal{K},\mathcal{C}} \boldsymbol{\Sigma}_{m,\mathcal{C},\mathcal{C}}^{-1} \boldsymbol{\Sigma}_{m,\mathcal{C},\mathcal{K}} \ ,$$

$$\omega_m^\mathcal{C}(\boldsymbol{y}_\mathcal{C}, \boldsymbol{x}) = \frac{\kappa_m \phi\Big(\boldsymbol{g}_{m,\mathcal{C}}(\boldsymbol{x}), \boldsymbol{\Sigma}_{m,\mathcal{C},\mathcal{C}}\Big)}{\sum_{l=1}^M \kappa_l \phi\Big(\boldsymbol{g}_{l,\mathcal{C}}(\boldsymbol{x}), \boldsymbol{\Sigma}_{l,\mathcal{C},\mathcal{C}}\Big)} \ ,$$

... serve as inputs to the Level-set algorithm to compute $\tilde{\boldsymbol{q}}(\alpha)$.

Heterogeneity in Household Consumption Patterns (cont.)



Hyperparameters

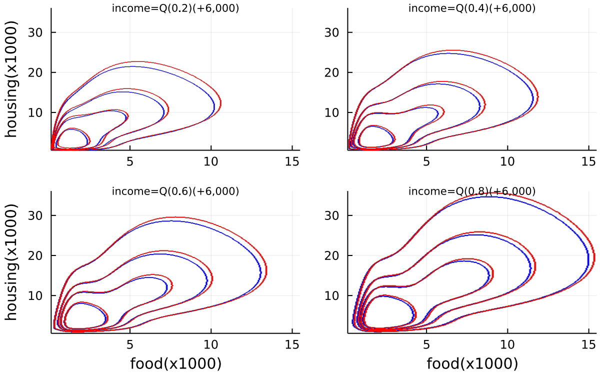
- $M = 5$
- $\underline{a}_1 = 10$, $\underline{a}_2 = 40$ (Dirichlet prior)
- $\underline{b}_1 = .5$, $\underline{b}_2 = .5$ (Gamma prior)

MCMC samples

- Effective: 200,000 / 40 (Burn-in: 400,000)

Heterogeneity in Household Consumption Patterns (cont.)

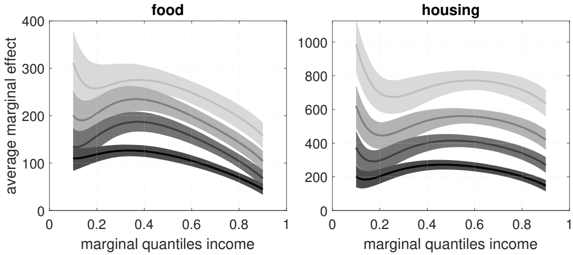
Bivariate quantiles for food and housing conditional on four different income levels. Blue lines corr. to $\alpha \in \{.2, .4, .6, .8\}$. Red lines corr. to an income increase of $6,000. QTE



- $.2$ expenditure quantile does not react considerably.
- $.8$ expenditure ($.2$ income) quantile substantially increases spending.
- No clear substitution patterns between food and housing.

Heterogeneity in Household Consumption Patterns (cont.)

Quantile-varying marginal effects conditional on income.
Shaded areas give the 90% C.I. for four income α-levels.



- Low-income quantile households dedicate most of the additional income to food and shelter.
- Highest-income quantile households hardly increase spendings at all.

Conclusion

Super-level sets provide a coherent framework for multivariate and univariate conditional as well as marginal quantiles:

- (1) no quantile crossing, (2) flexible quantile contours with exact probability coverage, (3) easy to extend quantile concept.

The overfitted GMM allows for straightforward incorporation of prior information regarding shapes and centers:

- enables a data driven bandwith parameter selection without unpleasant computational features (slow convergence, long runtimes; Polonik, 1997).
- makes no particular residual distribution assumption (see, e.g., Sriram, Ramamoorthi & Ghosh (2013) on the invalidity of the $\mathcal{AL}$-likelihood).

References

Brown P J, Griffin J E (2010). Inference with Normal-Gamma prior distributions in regression problems. *Bayesian Analysis*, 5: 171–188.

Carlier G, Chernozhukov V, Galichon A (2016). Vector quantile regression: an optimal transport approach. *The Annals of Statistics*, 44: 1165–1192.

Chernozhukov V, Hansen C (2013). Quantile models with endogeneity. *Annual Review of Economics*, 5:57–81

Doksum K (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *The Annals of Statistics*, 2: 267–277.

Hallin M, Paindaveine D, Siman M (2010). Multivariate quantiles and multiple-output regression quantiles: from L1 optimization to halfspace depth. *The Annals of Statistics*, 38: 635–703.

Hallin M, Siman M (2016). Elliptical multiple-output quantile regression and convex optimization. *Statistics & Probability Letters*, 109: 232–237.

Hartigan J A (1987). Estimation of a convex density contour in two dimensions. *Journal of the American Statistical Association*, 82: 267–270.

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

Koenker R, Machado J (1999). Goodness of fit and related inference processes for quantile regression, *Journal of the American Statistical Association*, 94: 1296-309.

Malsiner-Walli G, Frühwirth-Schnatter S, Grün B (2016). Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and Computing*, 26: 303–324.

References

Polonik W (1995). Measuring mass concentrations and estimating density contour clusters-an excess mass approach. *The Annals of Statistics*, 855–881.

Nobile A, Fearnside A T (2007). Bayesian finite mixtures with an unknown number of components: The allocation sampler. *Statistics and Computing*, 17: 147–162.

Rousseau J, Mengersen K (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73: 689–710.

Wei Y (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *Journal of the American Statistical Association*, 103:, 397–409.

Yu K, Moyeed R A (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54: 437-447.

Prior Distributions

$$\boldsymbol{\mu}_m | \bar{\boldsymbol{v}}_0, \overline{\boldsymbol{V}}_0 \sim \mathcal{N}(\bar{\boldsymbol{v}}_0, \overline{\boldsymbol{V}}_0),$$

$$\bar{\boldsymbol{v}}_0 \sim \mathcal{N}(\underline{\boldsymbol{v}}, \underline{\boldsymbol{V}}),$$

$$\underline{\boldsymbol{v}} = median(\boldsymbol{y}_n), \underline{\boldsymbol{V}}^{-1} = \mathbf{0}$$

$$\overline{\boldsymbol{V}}_0 = \text{diag}(R_1^2 \lambda_1, \dots, R_K^2 \lambda_K),$$

$$\lambda_k \sim \mathcal{G}(\underline{b}_1, 1/\underline{b}_2).$$

with response variable-specific value ranges $\{R_k\}$ and local shrinkage factors $\{\lambda_k\}$ $(\underline{b}_1, \underline{b}_2 > 0)$. (see, Brown & Griffin, 2010)

$$\text{vec}(\boldsymbol{B}_m) \sim \mathcal{N}(\boldsymbol{c}_0, \boldsymbol{C}_0)$$

$$\boldsymbol{\Sigma}_m \sim \mathcal{IW}(\boldsymbol{S}_0, s_0)$$

where $\boldsymbol{S}_0 = \boldsymbol{I}$ and $s_0 > 2 + K$.

Sampling Algorithm

- Simulate mixture parameters conditional on $\boldsymbol{z}_n$ $(n = 1, \dots, N,\ m = 1, \dots, M)$:
  - Sample $\{\kappa_m\}$ from $\mathcal{D}(\bar{\rho}_1, \dots, \bar{\rho}_M)$ where $\bar{\rho}_m = \rho_m + N_m$, $N_m = \#\{n : z_n = m\}$.
  - Sample $\{\boldsymbol{\mu}_m\}$ from $\mathcal{N}(\bar{\boldsymbol{v}}_m, \overline{\boldsymbol{V}}_m)$.
  - Sample $\{\boldsymbol{B}_m\}$ from $\mathcal{N}(\boldsymbol{c}_m, \boldsymbol{C}_m)$.
  - Sample $\{\boldsymbol{\Sigma}_m\}$ from $\mathcal{IW}(\boldsymbol{S}_m, s_m)$.
- Sample z_n to classify observations conditional on mixture parameters $(n = 1, \dots, N)$:
  - $\pi_m \equiv \Pr[z_n = m | \boldsymbol{y}_m, \boldsymbol{\kappa}, \boldsymbol{\mu}, \boldsymbol{B}, \boldsymbol{\Sigma}] \propto \kappa_m \phi\big(\boldsymbol{y}_n; g_m(\boldsymbol{x}_n), \boldsymbol{\Sigma}_m\big)$.
  - Sample $\{z_n\}$ from $\mathcal{M}(\pi_1, \dots, \pi_M)$.
- Sample hyperparameters:
  - Sample $\{\bar{\rho}_m\}$ simulatneously via a random walk MH-step with proposal density $\log(\rho_m) \sim \mathcal{N}(\log(\rho_m), s^2_{\rho_m})$ from $p(\bar{\rho}_m | \boldsymbol{\kappa}) \propto p(\boldsymbol{\kappa} | \bar{\rho}_m) p(\bar{\rho}_m)$
  - Sample $\{\lambda_k\}$ from $\mathcal{GIG}(\underline{b}_1 - M/2, 2\underline{b}_2, \delta_k)$ where $\delta_k = \sum_{m=1}^{M} (\mu_{m,k} - \bar{v}_{0,k})^2 / R_k^2$.
  - Sample $\bar{\boldsymbol{v}}_0$ from $\mathcal{N}\big(\sum_{m=1}^{M} \boldsymbol{\mu}_k / M, \overline{\boldsymbol{V}}_0 / M\big)$ with $\overline{\boldsymbol{V}}_0 = \mathrm{diag}(R_1^2 \lambda_1, \dots, R_K^2 \lambda_K)$.

Super-level Set Algorithm

Input : chosen coverage probability α
conditional distribution function $F_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\boldsymbol{y})$
grid boundary probability ϵ
dimension-specific grid point number n_{grid}

Output: actual coverage probability p
numerical quantile $\tilde{Q} = \tilde{Q}_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\alpha)$ of size $n_{\text{grid}}^{|\mathcal{K}|}$

for $k \in \mathcal{K}$ **do**
  grid_k = equally spaced n_{grid} vector with values from $F^{-1}_{Y_k|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\epsilon)$ to $F^{-1}_{Y_k|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(1-\epsilon)$;
  $\tilde{Q}_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\alpha)$ = $|\mathcal{K}|$-dimensional array of zeros

P = empty $|\mathcal{K}|$-dimensional array to hold probabilities per hypercube

for $(i_1 \in 2 : n_{grid}),\ (i_2 \in 2 : n_{grid}),\ \ldots,\ (i_{|\mathcal{K}|} \in 2 : n_{grid})$ **do**
  $P_{i_1,i_2,\ldots,i_{|\mathcal{K}|}} = \Pr[Y_k \in [\text{grid}_{k,i_k-1}, \text{grid}_{k,i_k}]\ \forall k \in \mathcal{K} | \boldsymbol{Y}_{\mathcal{C}} = \boldsymbol{y}_{\mathcal{C}}]$

$p = 0$

while $p < \alpha$ **do**
  $\mathcal{I}$ = set of indices for which P equals $\max\{P\}$
  $p = p + \sum_{i \in \mathcal{I}} P_i$
  for $i \in \mathcal{I}$ **do**
    $\tilde{Q}_i = \alpha$
    $P_i = 0$

Quantile-Specific Measures

The local marginal effect in the α-level quantile of Y_k given $\boldsymbol{Y}_{\mathcal{C}} = \boldsymbol{y}_{\mathcal{C}}$ for a change from $\boldsymbol{x}$ to $\boldsymbol{x} + \Delta_g$ is:

$$\beta^{g}_{k|\mathcal{C}}(\alpha|\boldsymbol{y}_{\mathcal{C}}, \boldsymbol{x}) = Q_{Y_k|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}\Big(\alpha|\boldsymbol{x} + \boldsymbol{\Delta}_g\Big) - Q_{Y_k|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}\Big(\alpha|\boldsymbol{x}\Big),$$

where $\boldsymbol{\Delta}_g$ is a vector with a small value δ_g at position g and zeros elsewhere (see Doksum, 1974). Back