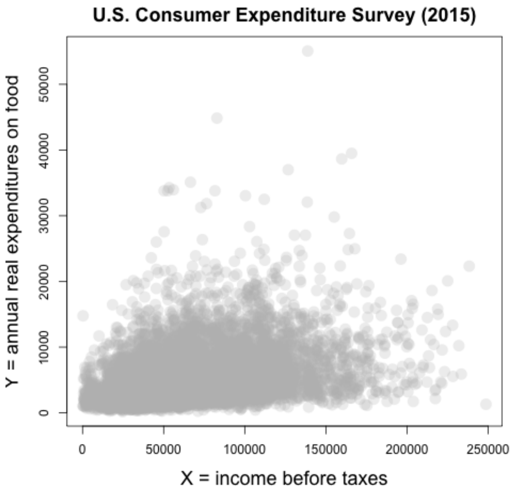
Erasmus University Rotterdam

A Level-Set Method for Multiple-Output (Panel) Quantiles*

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

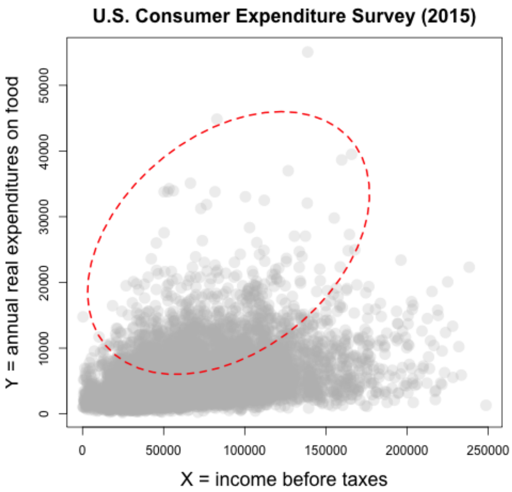
*Joint work with Annika Camehl and Dennis Fok

Example 1: Heterogeneity in Household Consumption Patterns



- $\{y_n, x_n\}_{n=1}^{N}$ and $N = 29,988$
- $y_n = \mu + x_n'\beta + u_n$

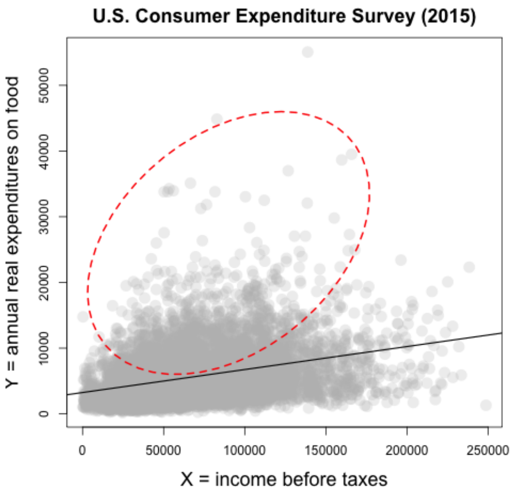
Example 1: Heterogeneity in Household Consumption Patterns



- $\{y_n, x_n\}_{n=1}^N$ and $N = 29,988$
- $y_n = \mu + x_n'\beta + u_n$

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

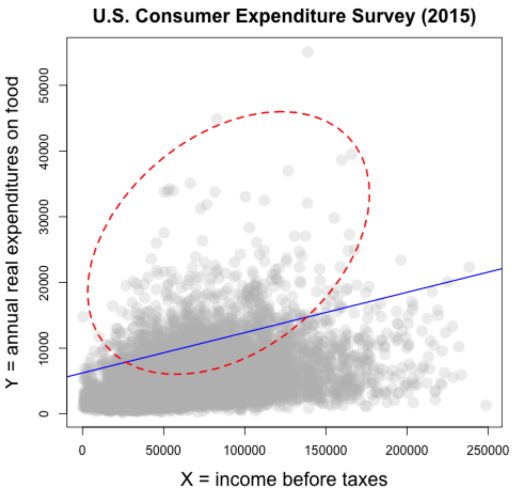
Example 1: Heterogeneity in Household Consumption Patterns



Mean regression

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

Example 1: Heterogeneity in Household Consumption Patterns

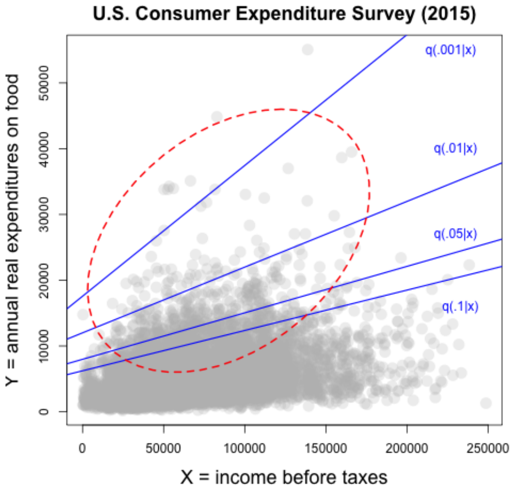


Quantile regression

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

Provides insights into the effects of covariates that are missed with mean regression!

Example 1: Heterogeneity in Household Consumption Patterns



Quantile regression

$$\left\{q(\alpha_l|x_n) = \mu_{(\alpha_l)} + x_n'\beta_{(\alpha_l)}\right\}_{l=1}^{L}$$

Provides a comprehensive picture of the conditional response distribution!

Applications

Study of heterogeneity in treatment participation
(e.g., Abadie, Angrist & Imbens, 2002; Athey & Imbens 2006; Firpo, 2007; Chernozhukov & Hansen, 2013)

Value-at-risk (tail value) measurement
(Chernozukov & Umantsev, 2001; Engle & Manganelli, 2004)

Panel model formulations to control for individual and time-specific effects (Koenker, 2004; Galvao, 2011; Arellano & Bonhomme, 2017)

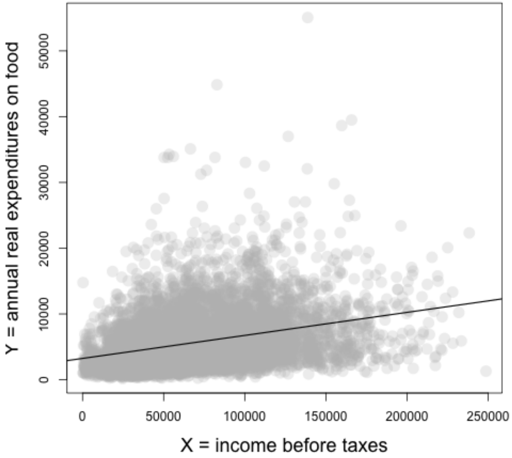
Exploration of multiple-output and functional responses
(Hallin, Paindaveine & Siman, 2010; Wei 2008; Carlier, Chernozhukov & Galichon, 2016, Hallin & Siman, 2016)

This Talk

- The quantile is defined as a property of an (estimated) conditional multivariate density.
- This so-called super-level set (quantifies the uncertainty in a (set of) response variable(s) conditional on other response variables) enables a clear probabilistic interpretation and enjoys favorable quantile properties.
- Multivariate as well as univariate probabilistic formulations of (linear and non-linear) cross-sectional and panel regression quantiles are obtained in a comprehensive framework.

Regression Quantiles in a Nutshell

(Conditional) mean regression

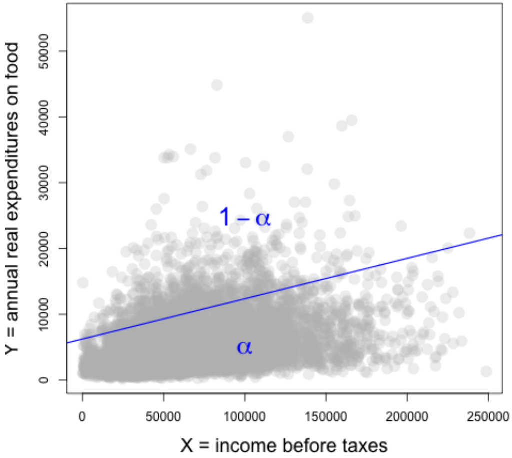


$$\underset{\mu,\beta}{\operatorname{argmin}} \sum_{n=1}^{N} |y_n - \overline{m}(x_n)|^2$$

$\Rightarrow$ solved with numerical linear algebra (i.e., OLS).

Regression Quantiles in a Nutshell

(Conditional) quantile regression



$$\underset{\mu_{(\alpha)},\beta_{(\alpha)}}{\text{argmin}} \sum_{n=1}^{N} \rho_\alpha |y_n - q(\alpha|x_n)|$$

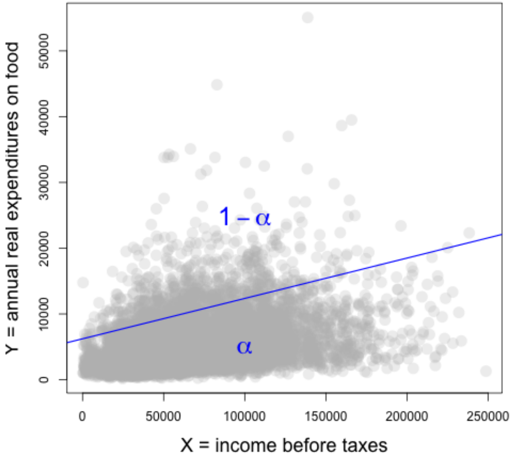
for $0 < \alpha < 1$, and asymmetric weight ("check") function:

$$\rho_\alpha(u_n) = \begin{cases} \alpha u_n, & u_n > 0 \\ -(1-\alpha)u_n, & u_n \leq 0 \end{cases}$$

$\Rightarrow$ solved with linear programming (see, Koenker & Bassett, 1978).

Regression Quantiles in a Nutshell

(Conditional) quantile regression



$$\rho_\alpha(u_n) = u_n(\alpha I_{(u_n>0)} - (1-\alpha) I_{(u_n \leq 0)})$$

equivalent to:

$$u_n \overset{iid}{\sim} \mathcal{AL}(0, \sigma, \alpha)$$

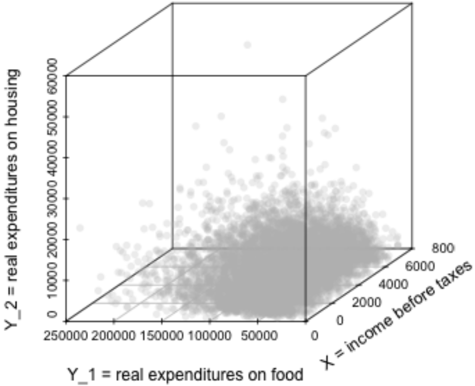
with asymmetry parameter $0 < \alpha < 1$

Example

$\Rightarrow$ ML (and Bayesian) inference is straightforward (see, Koenker & Machado, 1999; Yu & Moyeed, 2001).

Regression Quantiles in a Nutshell

Multiple-outputs: $\boldsymbol{y}_n = (y_{n1}, \ldots, y_{nK})'$



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

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!

Multivariate Quantiles

Attempt 1: Conditioning

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

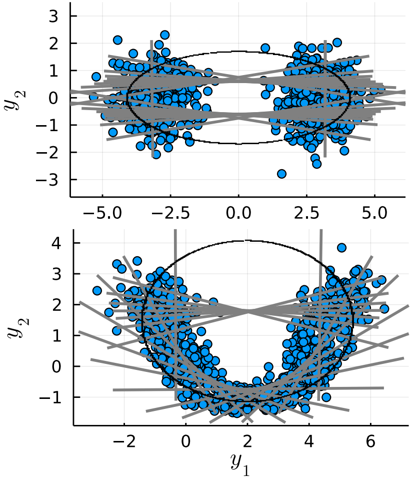
- Input-space augmentation, assumes all "regressors" are fixed!

Multivariate Quantiles

Attempt 2: Directional

- Convex intersection of α-quantile halfspaces for different (Koenker-Bassett) regression hyperplanes (Hallin, Paindaveine & Siman, 2010).

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



Multivariate Quantiles

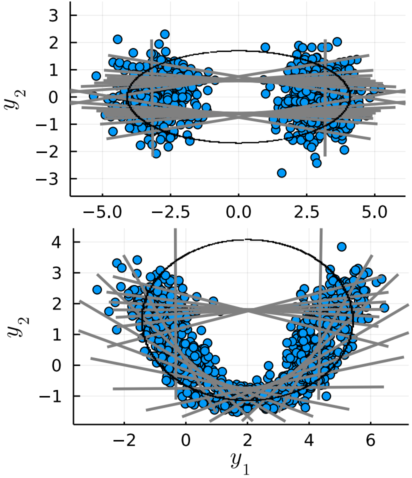
Attempt 2: Directional

- Convex intersection of α-quantile halfspaces for different (Koenker-Bassett) regression hyperplanes (Hallin, Paindaveine & Siman, 2010).

Attempt 3: Direct

- Find an ellipsoid around a (determined) center with α-probability mass (e.g., Hallin & Siman, 2016).

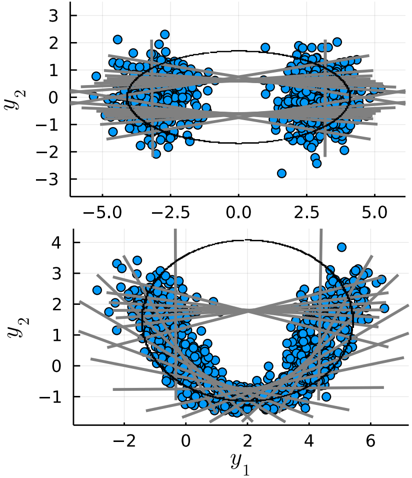
Areas within the black lines (countours) give the 80%-**elliptical quantile**.



Multivariate Quantiles

- The (directional) quantile contours are not guaranteed to cover α.
- The quantile regions can cover large parts with little to no probability mass.
- The definitions cannot easily be extended (i.e., to more than two outputs/ inputs).

Interpretability and practical use?



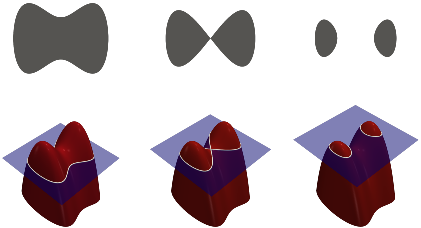
Multivariate Quantiles

Level set

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

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

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



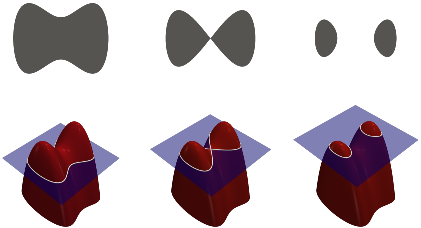
Multivariate Quantiles

Super-level set

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

for threshold $d > 0$, 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 d.



Multivariate Quantiles

Super-level set

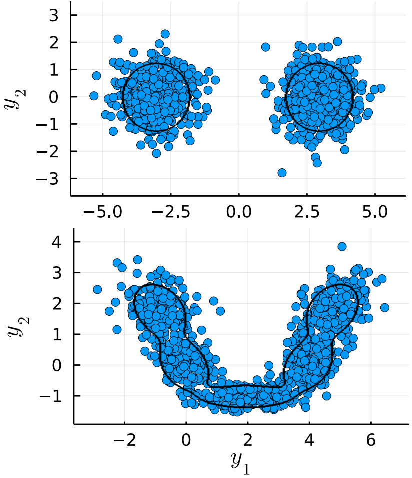
$$\mathcal{L}(f; d) = \left\{ \boldsymbol{y}_n \in \mathbb{R}^K : f(\boldsymbol{y}_n) \geq d \right\}$$

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

Super-level set "quantile"

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

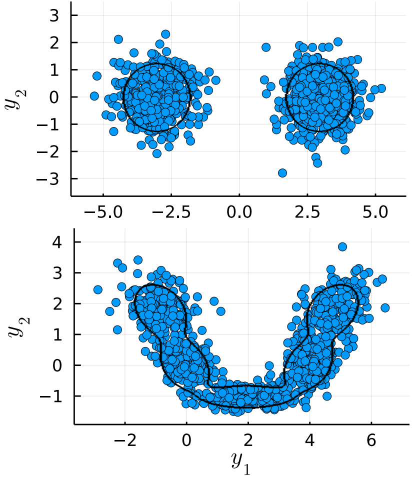
Areas within the lines (quantile contours) correspond to 80% probability mass.



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 inputs/ outputs are straightforward (Camehl, Fok & Gruber, 2024).

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



Implementation

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

$$\kappa_m \geq 0, \sum_{m=1}^{M} \kappa_m = 1$$

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

- Straightforward incorporation of prior (shape and center) information.
- No particular residual distribution assumption.
- Computationally cheap and data driven bandwith parameter selection.

Implementation

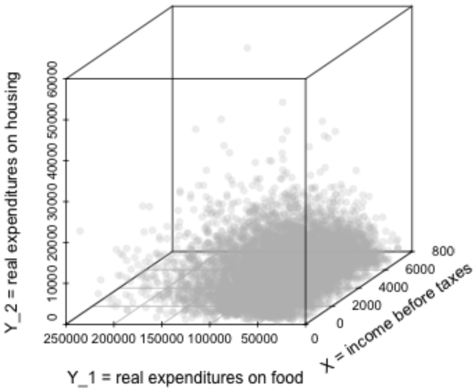
$$\boldsymbol{\mu}_m^{\mathcal{K}|\mathcal{C}}(\boldsymbol{y}_\mathcal{C}, \boldsymbol{x}) = \boldsymbol{g}_{m,\mathcal{K}}(\boldsymbol{x}) + \boldsymbol{\Sigma}_{m,\mathcal{K},\mathcal{C}} \boldsymbol{\Sigma}_{m,\mathcal{C},\mathcal{C}}^{-1} \Big(\boldsymbol{y}_\mathcal{C} - \boldsymbol{g}_{m,\mathcal{C}}(\boldsymbol{x})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

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

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

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

Heterogeneity in Household Consumption Patterns (cont.)



Overfitted Finite Gaussian Mixture Model

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

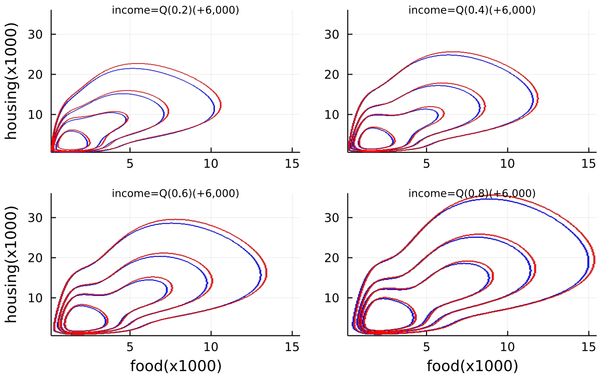
MCMC samples

- 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 $\alpha \in \{.2, .4, .6, .8\}$.

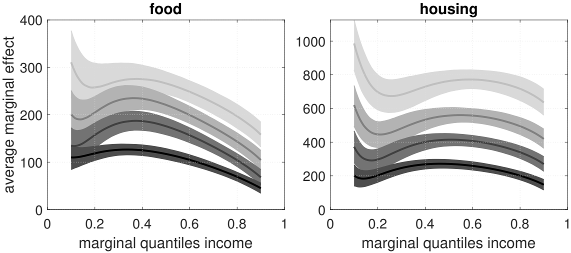
Red lines corr. to an income increase of $6,000. QTE



- $.2$ 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.)

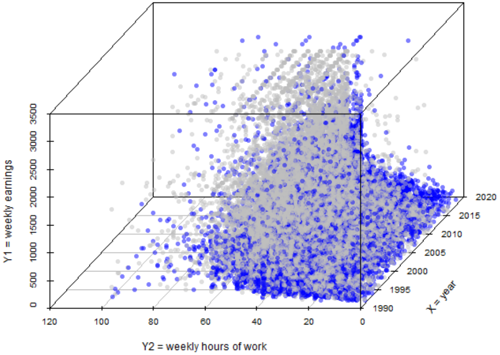
Quantile-varying marginal effects conditional on income.
Shaded areas give the 90% C.I. for four income α-levels.



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

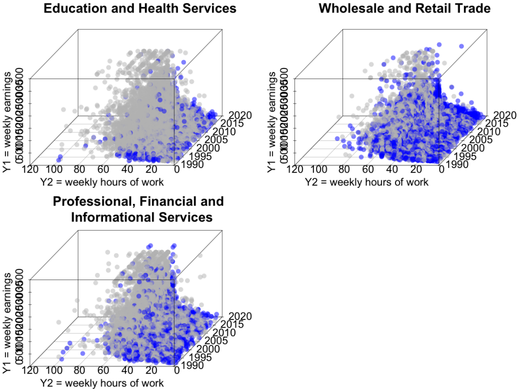
Example 2: Labor Market at Risk

Random sample of the current population survey (high-skilled workers in gray and less-skilled workers in blue)



Example 2: Labor Market at Risk

Random sample of the current population survey (high-skilled workers in gray and less-skilled workers in blue)



- Labor composition across industries and changes in productivity (co-movement structure).
- Time-varying quantile over (multivariate) conditional distribution.

Multivariate Panel Regression

$$y_{k,nt} = \mathbf{a}_k \mathbf{x}_{nt} + \mathbf{b}_{k,n} \mathbf{z}_{nt} + u_{k,nt},$$

with $(\text{Var}(\mathbf{b}_{k,n}) = \sigma^2 \mathbf{I})$ and $\text{E}(u_{k,nt}) = 0$, $\text{Var}(u_{k,nt}) = (\sigma^2_{kk})_n$, $\text{Cov}(u_{k,nt}, u_{j,nt}) = (\sigma_{kj})_{k \neq j,n}$ and $\text{Cov}(u_{k,nt}, \mathbf{b}_{k,n}) = \mathbf{0}$.

Reparameterization

$$\boldsymbol{\beta}_{k,n} \mathbf{r}_{k,nt} + \mathbf{a}_k \mathbf{x}_{nt} + \mathbf{b}_{k,n} \mathbf{z}_{nt}$$

where $\mathbf{r}_{k,nt} = (\mathbf{y}_t)^{k,n} - (\mathbf{A})^{k,n} \mathbf{x}_t - (\mathbf{B})^{k,n} \mathbf{z}_t$, $\mathbf{A}$ are the (stacked) population effects and $\mathbf{B}$ are the (stacked) individual deviations.

"equation-by-equation representation"

Time-Varying Super-level Set

$$q(\alpha) = \mathcal{L}(f; d_\alpha^* | \boldsymbol{w}_{nt}),$$

$$d_\alpha^* = \sup\{d : \Pr[\boldsymbol{y}_{nt} \in \mathcal{L}(d|\boldsymbol{w}_{nt})] \geq \alpha\}$$

where $\boldsymbol{w}_{nt}$ is the conditioning set at time point t (incl. $\boldsymbol{y}_{\mathcal{C},nt}$, $\boldsymbol{x}_{nt}$, $\boldsymbol{z}_{nt}$ and $\boldsymbol{r}_{nt}$).

- (Asymptotic) conditional coverage with exact finite-sample marginal properties.
- Multivariate prediction regions that contain the true outcome with high probability and without requiring strong distributional assumptions (Shafer & Vovk., 2008).

$$f\left(c_{nt,m}, y_{k,nt}\right) = \prod_m \omega_m^{c_{nt,m}} \mathcal{N}\left(\boldsymbol{\beta}_{k,n,m} \mathbf{r}_{k,nt} + \mathbf{a}_{k,m} \mathbf{x}_{nt} + \mathbf{b}_{k,n,m} \mathbf{z}_{nt}, v_{k,n,m}^{-1}\right)^{c_{nt,m}}$$

where $c_{nt,m} = 1$ if obs. t of unit n belongs to component m, $c_{nt,m} = 0$ otherwise.

Variational Inference

$$q(\mathbf{c}, \boldsymbol{\theta}) = \prod_m q(\omega_m) \prod_n q(c_{nt,m}) \prod_k q(\mathbf{a}_{k,m}) q(\mathbf{b}_{k,n,m}) q(\boldsymbol{\beta}_{k,n,m}) q(v_{k,n,m}^{-1}),$$

where $\boldsymbol{\theta} = [\boldsymbol{\omega}, \mathbf{a}, \mathbf{b}, \boldsymbol{\beta}, \mathbf{V}]$, and $q(\cdot)$ are computable densities.
(Iteratively) optimize the variational parameters to find the (best) approximate distribution:

$$q^{\star}(\mathbf{c}, \boldsymbol{\theta}) = \mathsf{E}_{q(\mathbf{c},\boldsymbol{\theta})}\left[\ln p(\mathbf{c}, \mathbf{y}_t) - \ln q(\mathbf{c}, \boldsymbol{\theta})\right].$$

Optimal Variational Densities

Implementation

Super-level set posterior for observation set $\mathcal{D}_{nt} = \{\boldsymbol{y}_{nt}, \boldsymbol{w}_{nt}\}$:

$$p(\boldsymbol{y}_{nt}|\mathcal{D}_{nt}) = \Pr(\boldsymbol{y}_{nt} \in \mathcal{L}(d^*_\alpha|\boldsymbol{w}_t)|\mathcal{D}_{nt}).$$

Compute (posterior) update given proposal $\boldsymbol{y}^\star_{nt}$:

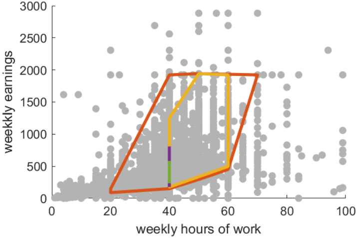
$$\mathsf{E}(f(\boldsymbol{y}^\star_{nt})|\mathcal{D}_{nt}) = \phi(a^\star_m)$$
$$\mathsf{Var}(f(\boldsymbol{y}^\star_{nt})|\mathcal{D}_{nt}) = \phi(a^\star_m) - \phi(a^\star_m)^2 - 2T(\phi(a^\star_m), c^\star_m)$$

where $a^\star_m = \frac{\boldsymbol{\mu}_{nt,m}(\boldsymbol{y}^\star_{nt}|\mathcal{D}_{nt})}{\sqrt{1+\boldsymbol{\Sigma}_{n,m}(\boldsymbol{y}^\star_{nt}|\mathcal{D}_{nt})}}$, $c^\star = \frac{1}{\sqrt{1+2\boldsymbol{\Sigma}_{n,m}(\boldsymbol{y}^\star_{nt}|\mathcal{D}_{nt})}}$,
$\boldsymbol{\mu}_{nt,m} = \boldsymbol{\beta}_{n,m}\mathbf{r}_{nt} + \mathbf{a}_m\mathbf{x}_{nt} + \mathbf{b}_{n,m}\mathbf{z}_{nt}$, $\boldsymbol{\Sigma}_{n,m} = \mathbf{V}_{n,m}$ and $T(\cdot)$ is Owen's T function ("look-ahead aquisition").

Heterogeneity in Labor Markets across Industries

(a) Manufacturing

(b) Education and Health

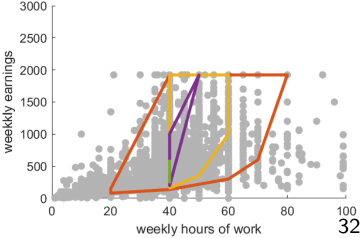


(c) Agri., Mining, Transport.

(d) Wholesale and Retail Trade

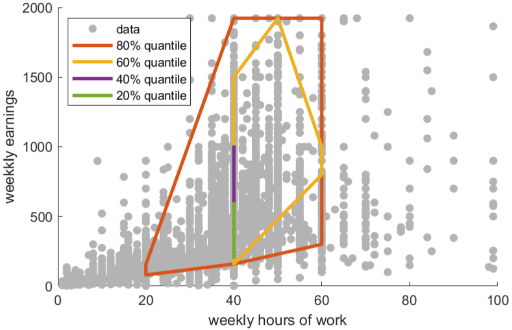
(e) Professional Services

(f) Financial, Info. Services



Heterogeneity in Labor Markets across Time

(g) year 2008



(h) year 2020

Summary & Outlook

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.

- Probabilistic formulation of the multivariate dynamic panel quantile model accounts for (time-varying) dependencies and heterogeneity.
- Variational inference and look-ahead strategy enable scalability to high-dimensions.

Many thanks for your attention!

References

Arellano M, Bonhomme S (2017). Nonlinear panel data estimation via quantile regression. *Econometric Journal,* 19:C61– 94.

Abadie A, Angrist J, Imbens G (2002). Instrumental variables estimates of subsidized training on the quantile of trainee earnings. *Econometrica*, 70:91–117.

Athey S, Imbens G (2006). Identification and inference in nonlinear difference- in-differences models. *Econometrica*, 74: 431–497.

Balandat M, Karrer B, Jiang D, Daulton S, Letham B, Wilson A G, and Bakshy E (2020). Botorch: A framework for efficient Monte-Carlo Bayesian optimization. *Advances in Neural Information Processing Systems* 33, 21524–21538.

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

Camehl, A., D. Fok, and K. Gruber (2024). On superlevel sets of conditional densities and multivariate quantile regression. *Journal of Econometrics*, 105807.

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

Chernozhukov V, Umantsev L (2001). Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics*, 26: 271–293.

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

References

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

Firpo S (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75:259–76

Galvao A F (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, 164(1):142 – 157.

Hallin M, Paindaveine D, Siman M (2010). Multivariate quantiles and muliple- 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.

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

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

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

References

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.

Shafer G, Vovk V (2008). A tutorial on conformal prediction. *Journal of Machine Learning Research, 9*:371–421.

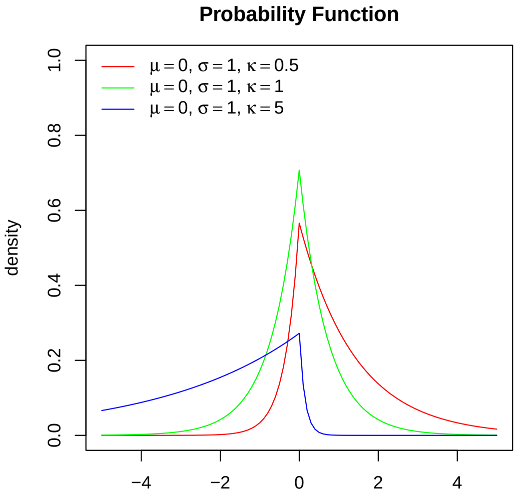
Taddy M A, Kottas A (2010). A Bayesian nonparametric approach to inference for quantile regression. *Journal of Business & Economic Statistics*, 28: 357–369.

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.

Yau, C, C. Holmes (2011). Hierarchical Bayesian nonparametric mixture models for clustering with variable relevance determination. *Bayesian Analysis* 6 (2), 329.

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

Asymmetric Laplace Distribution



(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).

Prior Distributions

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

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

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

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

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

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)

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

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

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

Sampling Algorithm

- Simulate mixture parameters conditional on $\boldsymbol{z}_n$ $(n = 1, \ldots, N,\ m = 1, \ldots, M)$:
  - Sample $\{\kappa_m\}$ from $\mathcal{D}(\bar{\rho}_1, \ldots, \bar{\rho}_M)$ 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 $(n = 1, \ldots, N)$:
  - $\pi_m \equiv \Pr[z_n = m | \boldsymbol{y}_m, \boldsymbol{\kappa}, \boldsymbol{\mu}, \boldsymbol{B}, \boldsymbol{\Sigma}] \propto \kappa_m \phi\big(\boldsymbol{y}_n; g_m(\boldsymbol{x}_n), \boldsymbol{\Sigma}_m\big)$.
  - Sample $\{z_n\}$ from $\mathcal{M}(\pi_1, \ldots, \pi_M)$.
- 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 $\overline{\boldsymbol{V}}_0 = \text{diag}(R_1^2 \lambda_1, \ldots, R_K^2 \lambda_K)$.

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

Simulation Exercise

1. Multivariate Gaussian
2. Multivariate Student-t
3. Multivariate log-Gaussian
4. Conditional heteroskedasticity
5. Multivariate Gaussian mixture

$1,000$ data sets with a sample size of $10,000$ for each DGP.

Hyperparameters

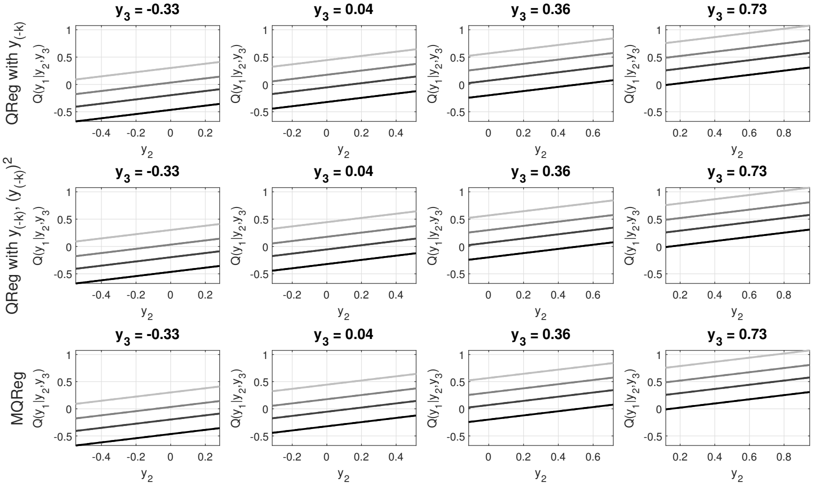
- $M = 5$
- $\underline{a}_1 = 10$, $\underline{a}_2 = 40$ (Dirichlet prior)
- $\underline{b}_1 = .5$, $\underline{b}_2 = .5$ (Gamma prior)

MCMC samples

- Effective: 50,000 / 10 (Burn-in: 10,000)

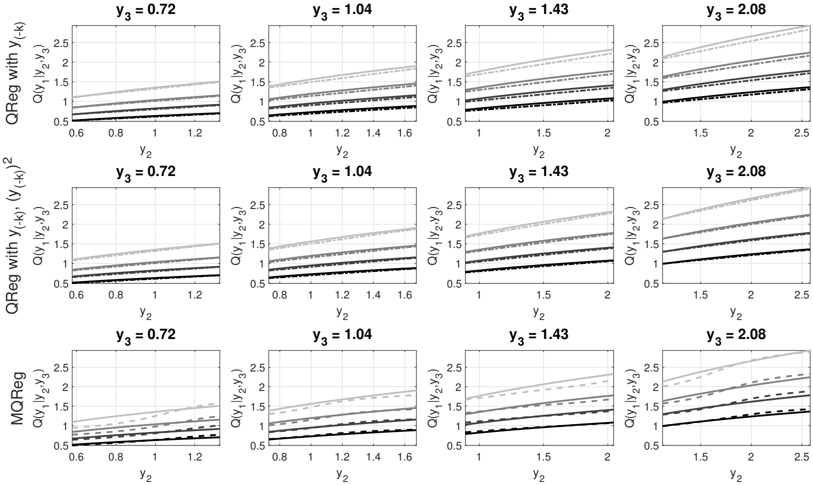
Simulation Exercise: Multivariate Gaussian

Estimated (dashed lines) vs. true (solid lines) conditonal quantiles for $\alpha \in \{.2, .4, .6, .8\}$



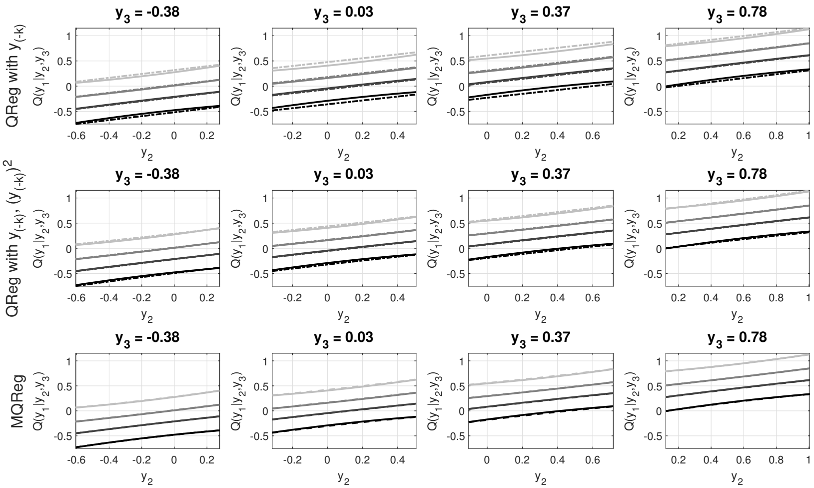
Simulation Exercise: Multivariate Student-t

Estimated (dashed lines) vs. true (solid lines) conditonal quantiles for $\alpha \in \{.2, .4, .6, .8\}$



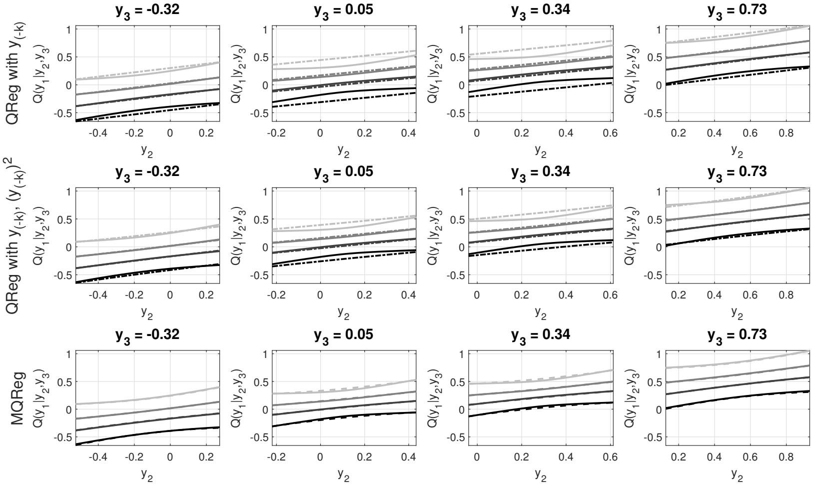
Simulation Exercise: Multivariatae log-Gaussian

Estimated (dashed lines) vs. true (solid lines) conditonal quantiles for $\alpha \in \{.2, .4, .6, .8\}$



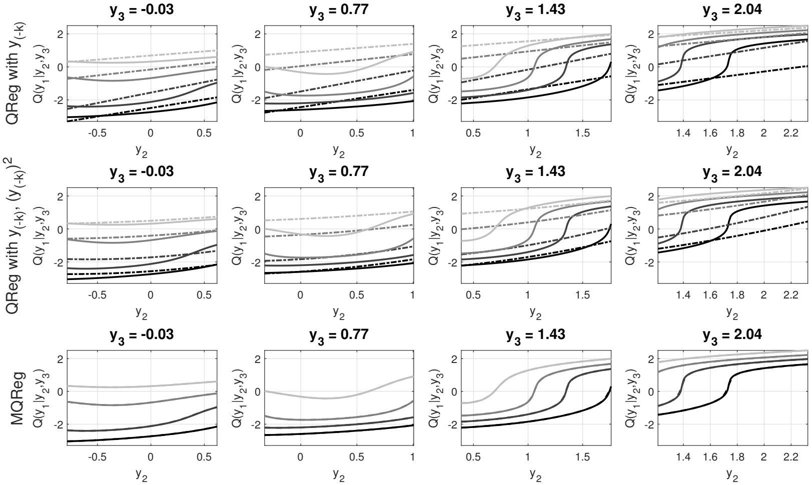
Simulation Exercise: Conditional Heteroskedasticty

Estimated (dashed lines) vs. true (solid lines) conditonal quantiles for $\alpha \in \{.2, .4, .6, .8\}$



Simulation Exercise: Multivariate Gaussian Mixture

Estimated (dashed lines) vs. true (solid lines) conditonal quantiles for $\alpha \in \{.2, .4, .6, .8\}$



Finite Gaussian Mixture (Panel) Model

$$f\left(\mathbf{y}_t | \mathbf{x}_t, \mathbf{z}_t\right) \sim \sum_{m=1}^{M} \omega_m \phi\left(\mathbf{A}_m \mathbf{x}_t + \mathbf{B}_m \mathbf{z}_t, \boldsymbol{\Lambda}_m\right),$$

with $\boldsymbol{\Lambda}_m = \boldsymbol{\Sigma}_m^{-1}$ (such that $\mathrm{E}(\mathbf{u}_{nt}\mathbf{u}'_{nt}) = \boldsymbol{\Sigma}_{nn,m}$ and $\mathrm{E}(\mathbf{u}_{nt}\mathbf{u}'_{it}) = \boldsymbol{\Sigma}_{ni,m}$ with $\boldsymbol{\Sigma}_{nn,m} \neq 0$, $\boldsymbol{\Sigma}_{ni,m} \neq 0$ for $n \neq i$).

For $\boldsymbol{\Lambda}_m = \mathbf{L}'_m \mathbf{V}_m \mathbf{L}_m$ with lower triangular matrix $\mathbf{L}_m = \mathbf{I} - \tilde{\mathbf{A}}_m$ and diagonal matrix $\mathbf{V}_m$:

$$(\mathbf{I} - \tilde{\mathbf{A}}_m)\mathbf{y}_t = (\mathbf{I} - \tilde{\mathbf{A}}_m)\left[\mathbf{A}_m \mathbf{x}_t + \mathbf{B}_m \mathbf{z}_t\right] + (\mathbf{I} - \tilde{\mathbf{A}}_m)\mathbf{u}_t$$

$$\mathbf{y}_t = \tilde{\mathbf{A}}_m\left[\mathbf{y}_t - \mathbf{A}_m \mathbf{x}_t - \mathbf{B}_m \mathbf{z}_t\right] + \mathbf{A}_m \mathbf{x}_t + \mathbf{B}_m \mathbf{z}_t + \mathbf{e}_t$$

and $\mathrm{E}[\mathbf{e}_t \mathbf{e}'_t] = \mathbf{V}_m^{-1}$.

Prior Distributions

$$\boldsymbol{\omega} | \{\rho_m\} \sim \mathcal{D}(\rho_1, \ldots, \rho_M)$$

$$\mathbf{a}_{k,m} \sim \mathcal{N}(\mathbf{0}, \tau_k^a \mathbf{V}^a)$$

$$\mathbf{b}_{k,n,m} \sim \mathcal{N}(\mathbf{0}, \tau_k^b \mathbf{V}_n^b)$$

$$\boldsymbol{\beta}_{k,n,m} \sim \mathcal{N}(\mathbf{0}, \tau_k^\beta \mathbf{I})$$

$$\tau_k^a \sim \mathcal{IG}(s^a, \tau_k^a),$$

$$\tau_k^b \sim \mathcal{IG}(s^b, \tau_k^b),$$

with s^a, $\tau_k^a > 0$, $s^a/\tau_k^a < \infty$, and s^b, $\tau_k^b > 0$, $s^b/\tau_k^b < \infty$.

$$v_{k,n,m} \sim \mathcal{G}(s, \lambda_k),$$

with $s = 1/2$ and $\lambda_k > 0$.

Optimal Variational Densities

$$q^{\star}(\mathbf{c}) = \mathsf{Cat}\left(\mathbf{c}|\boldsymbol{\omega}\right)$$

with responsibilities (i.e., normalized allocation probabilities) $w_{nt,m}$:

$$\begin{aligned} w_{nt,m} = & \, \mathsf{E}_{q(\boldsymbol{\omega})}[\ln(\omega_m)] + \frac{1}{2}\mathsf{E}_{q(\mathbf{V})}[\ln(|\mathbf{V}_{n,m}|)] - \frac{K}{2}\ln(2\pi) \\ & - \frac{1}{2}\mathsf{E}_{q(\boldsymbol{\beta})q(\mathbf{a})q(\mathbf{b})q(\mathbf{V})}\left[\left(\mathbf{y}_{nt} - \boldsymbol{\mu}_{n,m}\right)' \mathbf{V}_{n,m}\left(\mathbf{y}_{nt} - \boldsymbol{\mu}_{n,m}\right)\right]. \end{aligned}$$

where $\boldsymbol{\mu}_{n,m} = \boldsymbol{\beta}_{n,m}\mathbf{r}_{nt} + \mathbf{a}_m\mathbf{x}_{nt} + \mathbf{b}_{n,m}\mathbf{z}_{nt}$.

$$\begin{aligned} q^{\star}\left(\boldsymbol{\theta}|\mathbf{c}\right) = \mathcal{D}\left(\boldsymbol{\omega}|\left\{\rho_m\right\}\right)\prod_m & \mathcal{N}\left(\boldsymbol{\beta}_{k,n,m}|\boldsymbol{\mu}^{\boldsymbol{\beta}}_{n,m,k}, \Sigma^{\boldsymbol{\beta}}_{n,m,k}\right)\mathcal{N}\left(\mathbf{a}_{k,m}|\boldsymbol{\mu}^{\mathbf{a}}_{k,m}, \Sigma^{\mathbf{a}}_{k,m}\right) \\ & \mathcal{N}\left(\mathbf{b}_{k,n,m}|\boldsymbol{\mu}^{\mathbf{b}}_{k,n,m}\Sigma^{\mathbf{b}}_{k,n,m}\right)\mathcal{G}\left(\upsilon_{k,n,m}|s, \lambda_k\right), \end{aligned}$$

Back