Parameter Approximations in Econometrics

ANDREW CHESHER

CeMMAP

The Centre for Microdata Methods and Practice

at the Institute for Fiscal Studies

& University College London

____

NFS Summer Symposium in Econometrics and Statistics

Identification and Inference for Econometric Models

University of California Berkeley, August 2nd 2001

Parameter Approximations

- Consider parameter approximations to a DGP - to a distribution function (or functional of it).
- Accurate for parameter values, λ, close to an interesting value λ^*.
- λ^* is interesting because at $\lambda = \lambda^*$ an unknown element determining the DGP vanishes.
- A parameter approximation is useful if:
  - economic theory is silent about the form of the unknown element,
  - the unknown element does not appear in the approximation.

Parameter Approximations

- In this paper the unknown element is the distribution function of an unobservable variate.
- Three examples:
  - Models of choice with taste variation.
  - Covariate measurement error models.
  - Models with endogeneity.
- Alternative ways of treating the unknown element:
  - Remove it, e.g. by conditioning.
  - Provide an (arbitrary) parametric specification of it.
  - Nonparametrically estimate it.

Example 1: heterogeneity

- Let Y given X and U have distribution function
$$P[Y \leq y|X = x, U = u] = F_{Y|XU}(y|x, u; \theta)$$
- There are realisations of Y and X, but not U.
- The unknown element is the distribution function of U, $F_U(u)$.
- Realisations of Y and X are informative about the conditional DF of Y given $X = x$,
$$F_{Y|X}(y|x; \theta, F_U(\cdot)) = \int F_{Y|XU}(y|x, u; \theta) dF_U(u).$$

Example 2: covariate measurement error

- Let Y given X and U have conditional DF
$$F_{Y|XU}(y|x, u; \theta) = F_{Y|X}(y|x; \theta)$$
- Let $Z = X + \Lambda U$ where U is independent of X, at $\Lambda = 0$ no measurement error.
- The unknown element is the distribution function of U. The pdf of X is $f_X(x)$, also unknown.
- Realisations of only Y and Z are available.
- Realisations of Y and Z are informative about the joint DF of Y and Z,
$$F_{YZ}(y, z; \theta, \Lambda, f_X(\cdot), F_U(\cdot))$$
$$= \int F_{Y|XU}(y|z - \Lambda u; \theta) f_X(z - \Lambda u) dF_U(u)$$

Example 3: endogeneity (1)

- Continuously distributed Y_1 and Y_2 are determined by
$$\begin{aligned} Y_1 &= h_1(Y_2, \varepsilon + \lambda\nu) \\ Y_2 &= h_2(\nu) \end{aligned}$$
where ν and ε are mutually independent.
- Example: Y_1 = wages, Y_2 = schooling, ε is wage heterogeneity and ν = ability. There may be exogenous X as well.
- Policy to change Y_2 exogenously requires knowledge of
$$\beta(y_2, \omega) = \frac{\partial}{\partial a} h_1(a, b)|_{a=y_2, b=\omega}$$

Example 3: endogeneity (2)

- At $\lambda = \lambda^* = 0$ there is NO endogeneity
$$Y_1 = h_1(Y_2, \varepsilon)$$
- If $h_1(Y_2, \varepsilon)$ is monotonic increasing in ε for all Y_2, ε then
$$Q_{Y_1|Y_2}(\tau, y_2) = h_1(y_2, Q_\varepsilon(\tau))$$
  - $Q_\varepsilon(\tau)$ is the τ-quantile of ε
  - $Q_{Y_1|Y_2}(\tau, y_2)$ is the conditional τ-quantile of Y_1 given $Y_2 = y_2$.
- Therefore the function of policy interest can be estimated nonparametrically (Chaudhuri (1991)):
$$\beta(y_2, Q_\varepsilon(\tau)) = \nabla_{y_2} Q_{Y_1|Y_2}(\tau, y_2)$$
- When $\lambda \neq 0$, how is the quantile derivative related to the function of policy interest?

Uses of parameter approximations

- Understanding impact on DGP of departing from $\lambda = \lambda^*$.
- Understanding impact on estimators of $\lambda \neq \lambda^*$ - local specification analysis.
- Tool for developing tests of $H_0 : \lambda = \lambda^*$ - specification tests.
- Tool for studying sensitivity of inference to $\lambda \neq \lambda^*$.
- Tool for constructing "approximately consistent" estimators.

Related work (1)

- Small variance approximations used by Kadane (1971) to compare properties of econometric estimators.
- Rothenberg's (1971) discussion of local identifiability considers small parameter variations around a value which is identifiable.
- Local to unity parameter approximations in time series models are used to approximate sampling distributions of estimators, Ahtola and Tiao (1984), Phillips (1987).
- Local specification analysis (Kiefer and Skoog, (1984)) employs small parameter approximations.

Related work (2)

- Cox (1983), Chesher (1984), Freidlin & Wentzell (1984), Jorgenson (1987), employ small variance approximations in models of overdispersion.
- Carrol, Ruppert, Stefanski and co-authors have made extensive use in measurement error models. Focus on estimation not DGPs.
- Chesher and Schluter (2001), Chesher Dumangane and Smith (2001) use small parameter approximations to study the impact of measurement error on poverty and inequality measures and on event histories.
- Sweeting (1992) develops a general parameter-asymptotic limiting distribution theory for estimators.

Plan of this presentation

- Development of parameter approximations.
- Example: - discrete choice with taste variation.
- Regularising parameter approximations.
- Specification testing: random versus fixed parameters.
- Generic effects of measurement error on quantile regressions: sensitivity analysis.
- Generic effects of weak endogeneity.

"A careful econometrician, armed with a little statistical theory, a modest computer, and a lot of common sense, can always find reasonable approximations for a given inference problem."

T.J. Rothenberg (1984)

Developing parameter approximations: heterogeneity (1)

- We require an approximation to $F_{Y|X}$:
$$F_{Y|X}(y|x;\theta) = \int F_{Y|XU}(y, x, u, \theta) dF_U(u)$$
- Write the DF conditional on X and U as
$$F_{Y|XU}(y, x, u, \theta) = G(y, x, \Lambda u, \theta)$$
Λ is lower triangular, $k \times k$ with elements λ_{ij}. Normalise $V[U] = I_k$.
- Let $\Lambda\Lambda' = \Sigma = [\sigma_{st}] = Var[\Lambda U]$.
- Derivatives of G with respect to elements of $v = \Lambda u$ are G_i, G_{ij} and so forth. $[G_{ij}] = G^{vv}$.

Developing parameter approximations: heterogeneity (1)

- Expand G with remainder term R_1

$$G(y,x,\Lambda u;\theta) =$$

$$G(y,x,0;\theta) + \sum_{i,j} \lambda_{ij} u_j G_i(y,x,0;\theta) + \sum_{i,j,k,l} \frac{1}{2} \lambda_{ij}\lambda_{kl} u_j u_l G_{ik}(y,x,0;\theta) + R_1$$

- Integrate term by term, $F_{Y|X} = \int G \times dF_U(u)$, use $E[U]=0$, $V[U]=I_k$, $\sum_j \lambda_{ij}\lambda_{kj} = \sigma_{ik}$,

$$\begin{aligned} F_{Y|X} &\simeq G(y,x,0;\theta) + \frac{1}{2}\sum_{i,j,k,l} \lambda_{ij}\lambda_{kj} G_{ik}(y,x,0;\theta) \\ &= G(y,x,0;\theta) + \frac{1}{2}\sum_{i,k} \sigma_{ik} G_{ik}(y,x,0;\theta) \\ &= G(y,x,0;\theta) + \frac{1}{2}\,\text{trace}\left(G^{vv}(y,x,0;\theta)\Sigma\right). \end{aligned}$$

and $G(y,x,0;\theta)$, is $F_{Y|XU}(y|X=x,U=0,\theta)$

Developing parameter approximations: heterogeneity (1)

- The remainder term, R_1 can be written, with $\Lambda = [\lambda_{ij}]$, and $\|\Lambda^*\| < \|\Lambda\|$:

$$R_1 = \frac{1}{6}\sum_{i,j,k,l,m,n} \lambda_{ij}\lambda_{kl}\lambda_{mn} u_j u_l u_n G_{ikm}(y,x,\Lambda^* u;\theta)$$

Suppose $\exists$ finite valued $M(y,x,\theta)$ and C such that, $\forall\ v = \{v_s\}_{s=1}^S$, and $\forall\ i,\ k$, and m,

$$\sup_{i,k,m} \left| \frac{\partial^3}{\partial v_i \partial v_k \partial v_m} G(y,x,v;\theta) \right| \le M(y,x,\theta)$$

$$E\left[\left|U_i U_j U_k\right|\right] < C.$$

Then the remainder term R_2 has the property

$$|R_2| \le$$

$$\frac{1}{6} M(y,x;\theta) \sum_{i,j,k,l,m,n} \lambda_{ij}\lambda_{kl}\lambda_{mn} \int \left|u_j u_l u_n\right| dF_U(u)$$

$$\le \frac{1}{6} M(y,x;\theta) C \sum_{i,j,k,l,m,n} \lambda_{ij}\lambda_{kl}\lambda_{mn}$$

Example: Mixed Multinomial Logit Model (1)

Probability of choice $i \in \{1,\dots I\}$ conditional on $X = x$ is

$$P[i|x] = \int \frac{\exp(x'\beta_i + u_i)}{\sum_{j=1}^I \exp(x'\beta_j + u_j)} dF_U(u)$$

where $u = \{u\}_{i=1}^I$ is a vector of unobserved variates, assumed independent of X.

We have the small variance approximation (Chesher and Santos-Silva (2001))

$$g(i|x;\beta,\Omega) = \frac{\exp(x_i'\beta + \sum_{s=1}^I \sum_{t=s}^I{}_{s,t\neq i^*}\ \omega_{st} z_i^{st}(x;\beta))}{\sum_{j=1}^I \exp(x_j'\beta + \sum_{s=1}^I \sum_{t=s}^I{}_{s,t\neq i^*}\ \omega_{st} z_j^{st}(x;\beta))}$$

in which i^* identifies a base alternative relative to which the u's are measured,

$$\omega_{st} = Cov[u_s - u_{i^*}, u_t - u_{i^*}]$$

Example: Mixed Multinomial Logit Model - small parameter approximation

$$g(i|x;\beta,\Omega) = \frac{\exp(x_i'\beta + \sum_{s=1}^I \sum_{t=s}^I{}_{s,t\neq i^*}\ \omega_{st} z_i^{st}(x;\beta))}{\sum_{j=1}^I \exp(x_j'\beta + \sum_{s=1}^I \sum_{t=s}^I{}_{s,t\neq i^*}\ \omega_{st} z_j^{st}(x;\beta))}$$

where

$$z_i^{st}(x;\beta) = \begin{cases} 0 & i = i^* \\ \frac{1}{2} - p(s|x;\beta) & i \neq i^*,\quad s = t,\quad i = s \\ 0 & i \neq i^*,\quad s = t \quad i \neq s \\ -p(t|x;\beta) & i \neq i^*,\quad s \neq t \quad i = s \\ -p(s|x;\beta) & i \neq i^*,\quad s \neq t \quad i = t \\ 0 & i \neq i^*,\quad s \neq t \quad i \neq s \quad i \neq t \end{cases}$$

and

$$p(u|x,\beta) = \frac{\exp(x'\beta_u)}{\sum_{j=1}^I \exp(x'\beta_j)}$$

Regularising parameter approximations (1)

- It may be useful to have probabilities $\in [0,1]$, summing to 1, densities positive, probability mass exactly 1.

- Consider a 1st order "raw" approximation to a density $f(y;\lambda)$ with $\lambda^* = 0$:

$$f^R(y;\lambda) = f(y;0) + \lambda' g(y).$$

- With $f(y,0) > 0$, $h(\cdot) > 0$, twice differentiable, $h(1) = \nabla h(1) = 1$

$$\begin{aligned} f^R(y;\lambda) &= f(y;0) \times \left(1 + \lambda' \frac{g(y)}{f(y;0)}\right) \\ &= f(y;0) \times h\left(1 + \lambda' \frac{g(y)}{f(y;0)}\right) + o(\lambda) \end{aligned}$$

which is necessarily positive, and f^R is correct to $O(\lambda)$.

- Here $\lim_{||\lambda|| \to 0} (o(\lambda)/||\lambda||) = 0$.

Regularising parameter approximations (2)

- A proper approximation:

$$f^P(y;\lambda) = C(\lambda)^{-1} f(y;0) \times h\left(1 + \lambda' \frac{g(y)}{f(y;0)}\right)$$

where

$$C(\lambda) = \int f(y;0) h\left(1 + \lambda' \frac{g(y)}{f(y;0)}\right) dy$$

- But this is only correct if $C(\lambda) = 1 + o(\lambda)$.

- It is:

$$C(0) = \int f(y;0) dy = 1$$

$$\nabla C(\lambda)|_{\lambda=0} = \int g(y) dy = \int \nabla_\lambda f(y;\lambda)|_{\lambda=0}\, dy$$

$$= \nabla_\lambda \int f(y;\lambda) dy \Big|_{\lambda=0} = 0$$

Specification tests

- Score tests of $H_0 : \lambda = 0$ are specification tests to detect appearance of the unknown element.

- There is the proper approximate log likelihood function for N independent realisations of Y, $\{Y_n\}_{n=1}^N$,

$$l^A = \sum_{n=1}^{N} -\log C(\lambda) + \log f(Y_n;0) + \sum_{n=1}^{N} \log\left(h\left(1+\lambda'\frac{g(Y_n)}{f(Y_n;0)}\right)\right)$$

- The approximate score for λ at $\lambda = 0$ is

$$S^A = \sum_{n=1}^{N} \frac{g(Y_n)}{f(Y_n;0)}.$$

Example: random parameters

- In the heterogeneity example, let $\theta = \bar{\theta} + \Lambda u$

- Write the conditional DF of Y given X and U as $F(y, x, \bar{\theta} + \Lambda u)$.

- A test of $H_0 : \Lambda = 0$ is a test of a fixed parameter model against a random parameter alternative and

$$g(Y_n) = \nabla_{\theta\theta'} f(Y_n, X_n, \bar{\theta}).$$

- The score for Λ is therefore

$$S^A = \sum_{n=1}^{N} \frac{\nabla_{\theta\theta'} f(Y_n, X_n, \bar{\theta})}{f(Y_n; X_n, \bar{\theta})}$$

$$= \sum_{n=1}^{N} \nabla_{\theta\theta'} \log f(Y_n, X_n, \bar{\theta})$$

$$+ \sum_{n=1}^{N} \nabla_{\theta} \log f(Y_n, X_n, \bar{\theta}) \nabla_{\theta'} \log f(Y_n, X_n, \bar{\theta})$$

Measurement error and quantile regression (1)

- The τ-quantile of Y given $X = x$ is the QRF: $Q_{Y|X}(\tau, x)$, defined implicitly by

$$F_{Y|X}(Q_{Y|X}(\tau, x)|x) = \tau.$$

- Let $Z = X + \Lambda U$ be measurement error contaminated X. Realisations of Y and Z are informative about the τ-quantile of Y given $Z = z$ is $Q_{Y|Z}(\tau, z)$, defined implicitly by

$$F_{Y|Z}(Q_{Y|Z}(\tau, z)|z) = \tau.$$

- Write the conditional quantile conditional on Z as $Q_{Y|Z}(\tau, z; \Sigma)$ where $\Sigma = Var[\Lambda U]$ and

$$Q_{Y|X}(\tau, z) = Q_{Y|Z}(\tau, z; 0)$$

and develop a Taylor series approximation

$$Q_{Y|Z}(\tau, z; \Sigma) = Q_{Y|Z}(\tau, z; 0)$$

$$+ \sum_{i,j} \sigma_{ij} \frac{\partial}{\partial \sigma_{ij}} Q_{Y|Z}(\tau, z; \Sigma)\Big|_{\Sigma=0} + o(\Sigma)$$

Measurement error and quantile regression (2)

- To develop an expression for $\frac{\partial}{\partial \sigma_{ij}} Q_{Y|Z}(\tau, z; \Sigma)\Big|_{\Sigma=0}$ use the following approximation to $F_{Y|Z}(y|z)$ (Chesher (1991))

$$F_{Y|Z}(y|z) = F^A_{Y|Z}(y|z) + o(\Sigma)$$

$$F^A_{Y|Z}(y|z) = F_{Y|X}(y|z)$$

$$+ \sum_{i,j} \sigma_{ij} \left(F^i_{Y|X}(y|z) g^j_X(z) + \frac{1}{2} F^{ij}_{Y|X}(y|z)\right)$$

where for example

$$F^{ij}_{Y|X}(y|z) = \frac{\partial^2}{\partial x_i \partial x_j} F_{Y|X}(y|x)\Big|_{x=z}.$$

- The function $g_X(\cdot)$, is the log probability density function of X,

$$g_X(z) = \log f_X(x)$$

with derivatives as follows.

$$g^j_X(z) = \frac{\partial}{\partial x_j} g_X(x)\Big|_{x=z}$$

Measurement error and quantile regression (3)

- The approximate error contaminated QRF is

$$Q_{Y|Z}(\tau, z) = Q_{Y|X}(\tau, z) -$$

$$\sum_{i,\tau} \sigma_{ij} \frac{F^i_{Y|X}(Q_{Y|Z}|z) g^j_X(z) + \frac{1}{2} F^{ij}_{Y|X}(Q_{Y|Z}|z)}{F^Y_{Y|X}(Q_{Y|Z}|z)} + o(\Sigma)$$

- In terms of $Q_{Y|X}(\tau, z)$.

$$Q_{Y|Z}(\tau, z) = Q_{Y|X}(\tau, z) +$$

$$\sum_{i,j} \sigma_{ij} \left(Q^i_{Y|X}(\tau, z) g^j_X(z) + \frac{1}{2} Q^{ij}_{Y|X}(\tau, z)\right)$$

$$-\frac{1}{2} \frac{1}{Q^\tau_{Y|X}(\tau, z)} \sum_{i,j} \sigma_{ij} Q^{\tau i}_{Y|X}(\tau, z) Q^j_{Y|X}(\tau, z)$$

$$-\frac{1}{2} \frac{1}{Q^\tau_{Y|X}(\tau, z)} \sum_{i,j} \sigma_{ij} Q^{\tau j}_{Y|X}(\tau, z) Q^i_{Y|X}(\tau, z)$$

$$+\frac{1}{2} \frac{Q^{\tau\tau}_{Y|X}(\tau, z)}{Q^\tau_{Y|X}(\tau, z)^2} \sum_{i,j} \sigma_{ij} Q^i_{Y|X}(\tau, z) Q^j_{Y|X}(\tau, z) + o(\Sigma)$$

Measurement error and QRFs: one covariate

- Consider the case with a SINGLE covariate.

$$Q_{Y|Z}(\tau, z) = Q_{Y|X}(\tau, z)$$

$$+ \sigma^2 Q^x_{Y|X}(\tau, z) g^x_X(z) + \frac{\sigma^2}{2} Q^{xx}_{Y|X}(\tau, z)$$

$$-\sigma^2 \frac{Q^{\tau x}_{Y|X}(\tau, z) Q^x_{Y|X}(\tau, z)}{Q^\tau_{Y|X}(\tau, z)}$$

$$+\frac{\sigma^2}{2} \frac{Q^{\tau\tau}_{Y|X}(\tau, z) Q^x_{Y|X}(\tau, z)^2}{Q^\tau_{Y|X}(\tau, z)^2} + o(\sigma^2)$$

- Derivatives here are e.g.,

$$Q^\tau_{Y|X}(\tau, z) = \nabla_\tau Q_{Y|X}(\tau, z)$$

$$Q^x_{Y|X}(\tau, z) = \nabla_x Q_{Y|X}(\tau, x)\Big|_{x=z}$$

- Derivatives of $\sigma^2 Q_{Y|Z}(\tau, z)$ can replace derivatives of $\sigma^2 Q_{Y|X}(\tau, z)$ without disturbing the order of the approximation.

Measurement error and parallel QRFs

- Parallel conditional quantiles:
$$Q_X(\tau, x) = a(\tau) + b(x)$$
arise when Y is a location shift of a random variable $W \perp X$,
$$Y = b(X) + W.$$

- With $Q_W(\tau) = a(\tau)$ denoting the τ-quantile of W,
$$Q_X(\tau, x) = Q_W(\tau) + b(x).$$

- In this case $Q_X^{\tau x}(\tau, z) = 0$ and the approximation is
$$\begin{aligned} Q_Z(\tau, z) &= a(\tau) + b(z) \\ &\quad + \sigma^2 b^x(z) g_X^x(z) \\ &\quad + \frac{\sigma^2}{2} b^{xx}(z) \\ &\quad + \frac{\sigma^2}{2} \frac{a^{\tau\tau}(\tau) b^x(z)^2}{a^\tau(\tau)^2} + o(\sigma^2) \end{aligned}$$

"Sometimes, under some circumstances, asymptotic arguments lead to good approximations. Often they do not." T.J. Rothenberg (1984)

——

Accuracy of approximate QRFs (1)

- The approximation is EXACT for the fully Gaussian model, apart from vertical location of the QRFs.

- Consider numerical calculations with exponential power (EP) distributions.
$$\begin{aligned} Y &= \beta_0 + \beta_1 X + \sigma_W W \\ Z &= X + \sigma U \end{aligned}$$
W and U (mean 0, variance 1) and X (mean 0, variance 3) are independent EP variates with shape parameters: γ_W, γ_X, γ_U.

Accuracy of approximate QRFs (2)

- Exponential power distributed S with
$$E[S] = \mu, \quad Var[S] = \sigma^2$$
$\gamma \in (-1, 1)$ has pdf.
$$f_S(s) = A \exp\left(-B \left|\frac{s-\mu}{\sigma}\right|^{\frac{2}{1+\gamma}}\right)$$

- A and B are functions of γ and σ^2.

- At $\gamma = 0$, S is Gaussian.

- At $\gamma = 1$, S is Laplace.

- As $\gamma \to -1$, $S \to Unif[\mu - 3^{\frac{1}{2}}\sigma, \mu + 3^{\frac{1}{2}}\sigma]$

Sensitivity analysis for QRFs

- Suppose a parametric error free QRF is specified - e.g. linear
$$Q_X(\tau, x) = \beta_0 + \beta_1 x + \sigma_W Q_W(\tau)$$
where $Q_W(\tau)$ is the τ-quantile of $W \perp X$.

- There is the approximation
$$\begin{aligned} \tilde{Q}_Z(\tau, z) &= \beta_0^*(\tau) + \beta_1 \left(z + \sigma^2 g_Z^x(z)\right) \\ \beta_0^*(\tau) &= \beta_0 + \sigma_W Q_W(\tau) - \frac{\sigma^2}{2\sigma_W} \beta_1^2 g_W^w(Q_W(\tau)) \end{aligned}$$

- $g_Z^x(z)$ is the derivative of the log density of Z.

- For any value (chosen/estimated) of σ^2 we can estimate using $\hat{g}_X^x(z)$).

- Expect plim $\hat{\tilde{\beta}}_1 - \beta_1 = o(\sigma^2)$.

Sensitivity analysis for QRFs: Monte Carlo

- The error free QRF is
$$Q_X(\tau, x) = \beta_0 + \beta_1 x + \sigma_W Q_W(\tau)$$
$$\beta_0 = 0, \beta_1 = 1, \sigma_W = 1$$
$$E[W] = E[V] = 0 \quad Var[W] = Var[V] = 1$$
$$E[X] = 0 \quad Var[X] = 3$$

- W, X and V are EP variates with shape parameters $\gamma_W, \gamma_X, \gamma_V \in \{-0.5, 0, +0.5\}$.

- $R^2 = 0.75$. Mean regression attenuation is 25%.

- Sample size 400. 2000 Monte Carlo replications.

- Examine σ^2 known and estimated. $g_Z^x(z)$ known and (sieve) estimated.

Exponential series estimation of $g_Z^x(z)$

- Use the exponential series density estimator of Barron and Sheu (1991).

- The data are mapped by affine transformation onto the unit interval.

- The unknown density of z is specified as
$$f_Z(z) \propto f_Z^0(z) \exp\left(\sum_{j=1}^m \theta_j h_j(z)\right)$$
where $f_Z^0(z) = 1$ is the uniform kernel density on $[0, 1]$ and the $h_j(\cdot)$ is the jth order Legendre polynomial.

- Estimate θ_j's by ML ($m = 8$).

- The estimated log density derivative is simply
$$\hat{g}_Z^x(z) = \sum_{j=1}^m \hat{\theta}_j \nabla_z h_j(z)$$

Weak endogeneity (1)

- Continuously distributed Y_1 and Y_2 are determined by
$$\begin{aligned} Y_1 &= h_1(Y_2, \varepsilon + \lambda\nu) \\ Y_2 &= h_2(\nu) \end{aligned}$$
where ν and ε are mutually independent.

- Example: Y_1 = wages, Y_2 = schooling, ε as wage heterogeneity and ν = ability. There may be exogenous X as well.

- Implementation of policy to change Y_2 exogenously requires knowledge of
$$\beta(y_2, \omega) = \frac{\partial}{\partial a} h_1(a, b)|_{a=y_2, b=\omega}$$

- At $\lambda = 0$, there is no endogeneity and
$$\beta(y_2, Q_\varepsilon(\tau)) = \nabla_{y_2} Q_{Y_1|Y_2}(\tau, y_2)$$

- What is $\nabla_{y_2} Q_{Y_1|Y_2}(\tau, y_2)$ when $\lambda \neq 0$?

Weak endogeneity (2)

- Assume $h_1(\cdot,\cdot)$ is monotonic increasing in 2nd argument, $h_2(\cdot)$ monotonic increasing. There is an inverse function
$$g_2(Y_2) = \nu$$

- We have
$$\begin{aligned} Y_1 &= h_1(Y_2, \varepsilon + \lambda\nu) \\ Y_2 &= h_2(\nu) \end{aligned}$$
and so at any $Y_2 = y_2$
$$Y_1 = h_1(y_2, \varepsilon + \lambda g_2(y_2))$$
monotonicity implies
$$Q_{Y_1|Y_2}(\tau, y_2; \lambda) = h_1(y_2, Q_\varepsilon(\tau) + \lambda g_2(y_2))$$
$$\nabla_{y_2} Q_{Y_1|Y_2}(\tau, y_2; \lambda) = \nabla_1 h_1(y_2, Q_\varepsilon(\tau) + \lambda g_2(y_2))$$
$$+\lambda \nabla_{y_2} g_2(y_2) \nabla_2 h_1(y_2, Q_\varepsilon(\tau) + \lambda g_2(y_2))$$

Weak endogeneity (3)

- The approximate y_2 derivative of the conditional quantile with endogeneity ($\lambda \neq 0$) is
$$\nabla_{y_2} Q_{Y_1|Y_2}(\tau, y_2; \lambda) = \nabla_{y_2} Q_{Y_1|Y_2}(\tau, y_2; 0)$$
$$+\lambda g_2(y_2) \nabla_2 h_1(y_2, Q_\varepsilon(\tau))$$
$$+\lambda \nabla_{y_2} g_2(y_2) \nabla_2 \nabla_1 h_1(y_2, Q_\varepsilon(\tau))$$
$$+o(\lambda)$$
where $\nabla_i h_1(\cdot,\cdot)$ signifies the derivative of $h_1(\cdot,\cdot)$ with respect to its ith argument.

- Easier to interpret (and use) when expressed in terms of quantiles. Note:
$$\nabla_2 h_1(y_2, Q_\varepsilon(\tau)) = \frac{\nabla_\tau Q_{Y_1|Y_2}(\tau, y_2; 0)}{\nabla_\tau Q_\varepsilon(\tau)}$$
$$\nabla_2 \nabla_1 h_1(y_2, Q_\varepsilon(\tau)) = \frac{\nabla_{y_2} \nabla_\tau Q_{Y_1|Y_2}(\tau, y_2; 0)}{\nabla_\tau Q_\varepsilon(\tau)}$$

Weak endogeneity (4)

- After manipulation
$$\nabla_{y_2} Q_{Y_1|Y_2}(\tau, y_2; \lambda) = \nabla_{y_2} Q_{Y_1|Y_2}(\tau, y_2; 0)$$
$$+\lambda^+ g_2(y_2) \nabla_\tau Q_{Y_1|Y_2}(\tau, y_2; \lambda)$$
$$+\lambda^+ \nabla_{y_2} g_2(y_2) \nabla_{y_2} \nabla_\tau Q_{Y_1|Y_2}(\tau, y_2; \lambda)$$
$$+o(\lambda)$$
where
$$\lambda^+ = \frac{\lambda}{\nabla_\tau Q_\varepsilon(\tau)} = \lambda f_\varepsilon(Q_\varepsilon(\tau))$$

- $Q_{Y_1|Y_2}(\tau, y_2; \lambda)$ and its derivatives can be estimated nonparametrically, as can $g_2(y_2)$. Conduct sensitivity analysis by considering variations in λ^+ in
$$\hat\beta(y_2, Q_\varepsilon(\tau)) = \nabla_{y_2} \hat Q_{Y_1|Y_2}(\tau, y_2; \lambda)$$
$$-\lambda^+ \hat g_2(y_2) \nabla_\tau \hat Q_{Y_1|Y_2}(\tau, y_2; \lambda)$$
$$-\lambda^+ \nabla_{y_2} \hat g_2(y_2) \nabla_{y_2} \nabla_\tau \hat Q_{Y_1|Y_2}(\tau, y_2; \lambda)$$

Concluding remarks

- Parameter approximations to DGPs can eliminate elements about which economic theory is silent. Can be used to:
  - characterise the impact of local departures from DGPs in which the unknown element is absent,
  - assess the impact of such local departures on inference when the unknown element is ignored ,
  - develop specification tests to detect the presence of the unknown element,
  - produce locally consistent estimates of parameters without specifying the unknown element.

- Other applications: local to vanishing sample selection, non-compliance, stochastic volatility...

Blank page

Table 1: Means and standard deviations of QRF slope estimates ignoring measurement error

			$\gamma_V = -0.5$		$\gamma_V = 0.0$		$\gamma_V = +0.5$	
τ	γ_Y	γ_X	mean	s.d.	mean	s.d.	mean	s.d.
0.50	-0.5	-0.5	.738	.029	.755	.031	.772	.033
		0.0	.734	.031	.750	.033	.769	.034
		+0.5	.728	.034	.744	.035	.761	.038
	0.0	-0.5	.736	.030	.755	.031	.774	.032
		0.0	.732	.031	0.750	.033	.771	.034
		+0.5	.725	.034	.743	.035	.763	.035
	+0.5	-0.5	.736	.028	.756	.029	.778	.032
		0.0	.730	.030	.750	.032	.772	.033
		+0.5	.723	.033	.743	.034	.764	.037
0.75	-0.5	-0.5	.746	.034	.753	.034	.764	.036
		0.0	.742	.034	.750	.036	.761	.037
		+0.5	.739	.038	.747	.038	.757	.040
	0.0	-0.5	.746	.033	.752	.034	.763	.036
		0.0	.743	.034	.750	.036	.761	.037
		+0.5	.740	.036	.745	.038	.756	.039
	+0.5	-0.5	.747	.032	.753	.034	.763	.036
		0.0	.743	.034	.750	.035	.760	.037
		+0.5	.739	.036	.746	.038	.756	.039
0.90	-0.5	-0.5	.765	.042	.748	.044	.736	.044
		0.0	.766	.043	.750	.044	.740	.046
		+0.5	.769	.045	.754	.047	.743	.048
	0.0	-0.5	.766	.043	.747	.043	.735	.047
		0.0	.768	.044	.750	.044	.738	.047
		+0.5	.770	.045	.752	.045	.744	.048
	+0.5	-0.5	.770	.045	.746	.044	.733	.046
		0.0	.771	.044	.750	.045	.737	.047
		+0.5	.773	.046	.754	.046	.742	.048

Table 2: Means and standard deviations of measurement error corrected QRF slope estimates with σ^2 known and $g_Z^x(\cdot)$ known

			$\gamma_V = -0.5$		$\gamma_V = 0.0$		$\gamma_V = +0.5$	
τ	γ_Y	γ_X	mean	s.d.	mean	s.d.	mean	s.d.
0.50	-0.5	-0.5	0.989	.040	1.011	.040	1.028	.042
		0.0	0.978	.042	1.000	.044	1.026	.046
		+0.5	0.972	.043	0.996	.046	1.021	.050
	0.0	-0.5	0.986	.041	1.010	.040	1.031	.041
		0.0	0.976	.041	1.000	.043	1.028	.045
		+0.5	0.970	.044	0.995	.046	1.024	.047
	+0.5	-0.5	0.987	.039	1.013	.038	1.036	.041
		0.0	0.974	.040	1.000	.043	1.030	.044
		+0.5	0.966	.042	0.995	.044	1.025	.048
0.75	-0.5	-0.5	0.994	.045	1.007	.044	1.018	.046
		0.0	0.989	.046	1.000	.047	1.015	.050
		+0.5	0.988	.049	0.998	.050	1.011	.053
	0.0	-0.5	0.992	.044	1.005	.044	1.018	.046
		0.0	0.990	.046	1.000	.048	1.014	.049
		+0.5	0.988	.047	0.996	.049	1.013	.052
	+0.5	-0.5	0.993	.044	1.005	.044	1.018	.046
		0.0	0.991	.046	1.000	.047	1.014	.049
		+0.5	0.989	.047	0.997	.049	1.012	.052
0.90	-0.5	-0.5	1.004	.056	0.994	.058	0.984	.058
		0.0	1.020	.058	1.000	.059	0.986	.062
		+0.5	1.029	.058	1.005	.062	0.984	.064
	0.0	-0.5	1.005	.056	0.990	.057	0.982	.059
		0.0	1.023	.059	1.000	.059	0.984	.062
		+0.5	1.032	.059	1.003	.059	0.986	.063
	+0.5	-0.5	1.007	.059	0.988	.059	0.978	.059
		0.0	1.026	.059	1.001	.059	0.981	.062
		+0.5	1.036	.059	1.004	.059	0.984	.065

Table 3: Means and standard deviations of measurement error corrected QRF slope estimates with σ^2 unknown and $g_Z^x(\cdot)$ known

			$\gamma_V = -0.5$		$\gamma_V = 0.0$		$\gamma_V = +0.5$	
τ	γ_Y	γ_X	mean	s.d.	mean	s.d.	mean	s.d.
0.50	-0.5	-0.5	0.870	0.107	1.024	.127	1.087	.130
		0.0	-	-	-	-	-	-
		+0.5	1.117	.168	1.017	.161	0.910	.149
	0.0	-0.5	0.867	.106	1.023	.122	1.095	.129
		0.0	-	-	-	-	-	-
		+0.5	1.123	.160	1.018	.161	0.909	.152
	+0.5	-0.5	0.874	.105	1.029	.120	1.101	.128
		0.0	-	-	-	-	-	-
		+0.5	1.122	.164	1.020	.158	0.908	.152
0.75	-0.5	-0.5	0.892	.121	1.013	.137	1.074	.142
		0.0	-	-	-	-	-	–
		+0.5	1.106	.180	1.008	.180	0.899	.161
	0.0	-0.5	0.888	.119	1.017	.133	1.078	.146
		0.0	-	-	-	-	-	-
		+0.5	1.098	.170	1.004	.175	0.903	.161
	+0.5	-0.5	0.890	.116	1.013	.136	1.073	.144
		0.0	-	-	-	-	-	-
		+0.5	1.102	.178	1.011	.170	0.903	.162
0.90	-0.5	-0.5	0.933	.152	0.988	.181	1.015	.188
		0.0	-	-	-	-	-	-
		+0.5	1.077	.218	0.988	.216	0.880	.194
	0.0	-0.5	0.931	.158	0.993	.169	1.020	.192
		0.0	-	-	-	-	-	-
		+0.5	1.066	.227	0.980	.221	0.886	.194
	+0.5	-0.5	0.934	.158	0.981	.182	1.013	.196
		0.0	-	-	-	-	-	-
		+0.5	1.064	.227	0.987	.217	0.887	.201

Table 4: Means and standard deviations of measurement error corrected QRF slope estimates with σ^2 known and $g_Z^x(\cdot)$ estimated

			$\gamma_V = -0.5$		$\gamma_V = 0.0$		$\gamma_V = +0.5$	
τ	γ_Y	γ_X	mean	s.d.	mean	s.d.	mean	s.d.
0.50	-0.5	-0.5	0.979	.048	1.002	.049	1.024	.052
		0.0	0.972	.047	0.994	.050	1.021	.052
		+0.5	0.968	.047	0.991	.051	1.017	.056
	0.0	-0.5	0.977	.049	1.003	.049	1.027	.051
		0.0	0.969	.046	0.994	.049	1.024	.052
		+0.5	0.965	.048	0.991	.051	1.020	.052
	+0.5	-0.5	0.978	.048	1.005	.047	1.032	.051
		0.0	0.968	.047	0.993	.049	1.024	.052
		+0.5	0.963	.046	0.992	.049	1.021	.055
0.75	-0.5	-0.5	0.986	.053	0.999	.053	1.015	.055
		0.0	0.984	.051	0.993	.053	1.012	.056
		+0.5	0.986	.052	0.994	.054	1.008	.060
	0.0	-0.5	0.984	.052	0.999	.052	1.016	.055
		0.0	0.985	.051	0.993	.053	1.011	.057
		+0.5	0.986	.051	0.994	.053	1.009	.057
	+0.5	-0.5	0.986	.052	0.997	.052	1.015	.054
		0.0	0.986	.051	0.994	.054	1.010	.057
		+0.5	0.985	.051	0.994	.053	1.008	.057
0.90	-0.5	-0.5	0.999	.063	0.987	.064	0.979	.067
		0.0	1.015	.064	0.994	.064	0.983	.067
		+0.5	1.027	.063	1.003	.065	0.983	.068
	0.0	-0.5	0.999	.064	0.985	.064	0.977	.068
		0.0	1.019	.063	0.992	.064	0.980	.068
		+0.5	1.029	.061	1.002	.064	0.984	.067
	+0.5	-0.5	1.003	.064	0.983	.063	0.975	.067
		0.0	1.021	.064	0.997	.066	0.977	.069
		+0.5	1.032	.063	1.002	.063	0.982	.069

Table 5: Means and standard deviations of measurement error corrected QRF slope estimates with σ^2 unknown and $g_Z^x(\cdot)$ estimated

			$\gamma_V = -0.5$		$\gamma_V = 0.0$		$\gamma_V = +0.5$	
τ	γ_Y	γ_X	mean	s.d.	mean	s.d.	mean	s.d.
0.50	-0.5	-0.5	0.820	.102	0.903	.136	0.972	.169
		0.0	-	-	-	-	-	-
		+0.5	0.944	.182	0.907	.170	0.863	.148
	0.0	-0.5	0.818	.101	0.906	.137	0.974	.173
		0.0	-	-	-	-	-	-
		+0.5	0.947	.181	0.904	.153	0.865	.156
	+0.5	-0.5	0.817	.097	0.908	.128	0.976	.188
		0.0	-	-	-	-	-	-
		+0.5	0.950	.172	0.906	.150	0.862	.147
0.75	-0.5	-0.5	0.835	.118	0.900	.152	0.958	.183
		0.0	-	-	-	-	-	–
		+0.5	0.940	.187	0.902	.180	0.845	.162
	0.0	-0.5	0.830	.116	0.903	.151	0.955	.187
		0.0	-	-	-	-	-	-
		+0.5	0.939	.187	0.888	.175	0.853	.180
	+0.5	-0.5	0.830	.117	0.896	.136	0.949	.196
		0.0	-	-	-	-	-	-
		+0.5	0.941	.178	0.896	.165	0.845	.168
0.90	-0.5	-0.5	0.856	.163	0.884	.173	0.906	.220
		0.0	-	-	-	-	-	-
		+0.5	0.939	.214	0.888	.212	0.824	.199
	0.0	-0.5	0.859	.158	0.883	.193	0.902	.222
		0.0	-	-	-	-	-	-
		+0.5	0.933	.218	0.878	.219	0.829	.214
	+0.5	-0.5	0.857	.155	0.878	.173	0.898	.235
		0.0	-	-	-	-	-	-
		+0.5	0.933	.214	0.883	.203	0.823	.206

Figure 1: Exact and approximate τ-QRFs: $\tau \in \{0.5, 0.75, 0.9\}$, $\gamma_Y = +0.5$

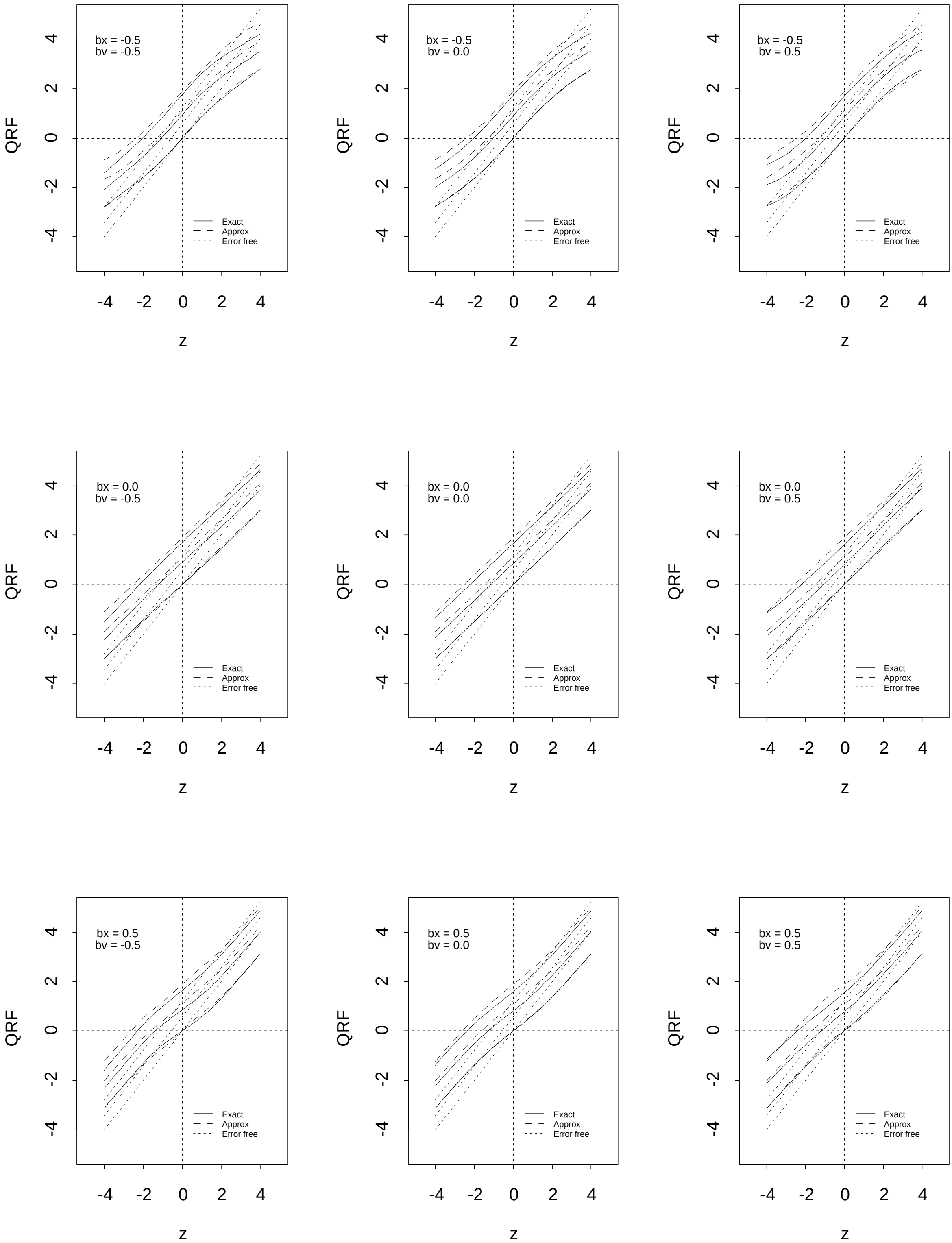

Figure 2: Exact and approximate τ-QRFs: $\tau \in \{0.5, 0.75, 0.9\}$, $\gamma_Y = 0.0$

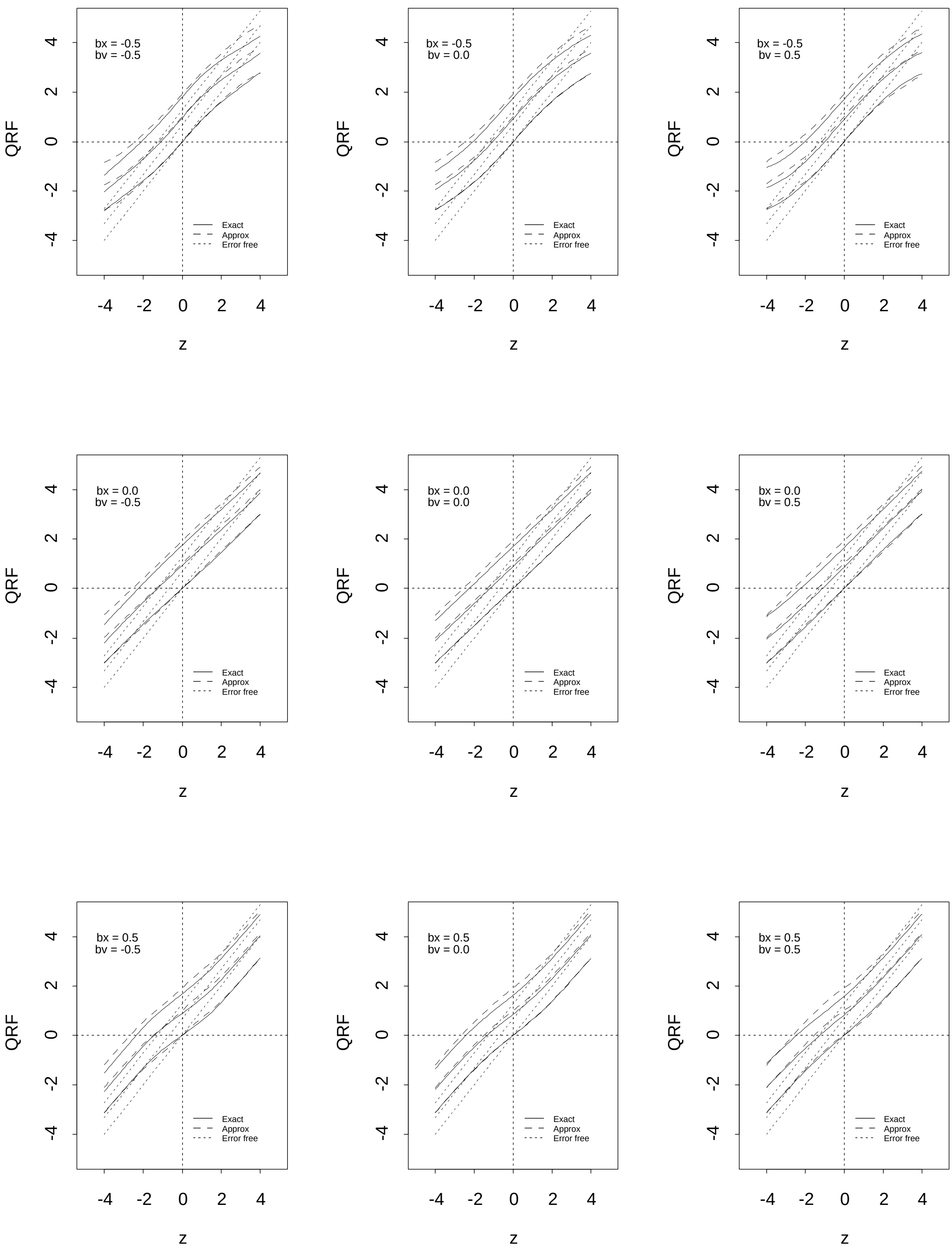

Figure 3: Exact and approximate QRFs: $\tau \in \{0.5, 0.75, 0.9\}$, $\gamma_Y = -0.5$