Ma and Koenker, 'Demystification of the Magic Formula'

$$\begin{aligned} Y_{i1} &= Y_{i2}\alpha_1 + x_i^T\alpha_2 + \nu_{i1} + \lambda\nu_{i2} && (2.1) \\ Y_{i2} &= z_i\beta_1 + x_i^T\beta_2 + \nu_{i2} && (2.2) \end{aligned}$$

For simplicity, Y_{i2} and z_i are scalars; ($\nu' s$ independent of (z,x)).

Thus, $\nu_{i2} = Y_{i2} - z_i\beta_1 - x_i^T\beta_2$ and

$$Y_{i1} = Y_{i2}\alpha_1 + x_i^T\alpha_2 + \nu_{i1} + \lambda(Y_{i2} - z_i\beta_1 - x_i^T\beta_2),$$

which gives rise, after collecting terms, to the 'hybrid' form:

$Y_1 = Y_2(\alpha_1 + \lambda) - z\beta_1\lambda + X(\alpha_2 - \lambda\beta_2) + \nu_1$, or $Y_1 = W\delta$, with

$W = (Y_2 \vdots z \vdots X)$, $\delta = (\alpha_1 + \lambda, -\beta_1\lambda, \alpha_2 - \lambda\beta_2)$

Demystification (2)

Expressing the hybrid form as $Y_1 = W\delta$, with $W = (Y_2 \vdots z \vdots X)$, $\delta = (\alpha_1 + \lambda, -\beta_1\lambda, \alpha_2 - \lambda\beta_2)$

So $\alpha_1 = \delta_1 + \frac{\delta_2}{\beta_1} = (\alpha_1 + \lambda) + \frac{-\beta_1\lambda}{\beta_1} = \alpha_1$

It turns out that the 2SLS estimate of α_1, say $\hat{\alpha}_1$, is $\hat{\alpha}_1 = \hat{\delta}_1 + \frac{\hat{\delta}_2}{\hat{\beta}_1}$, (where $\hat{\delta}$ is OLS on 'hybrid'.)

Demystification (3)

You could do it with quantiles:

$$Q_1(\tau_1|Y_2,x,z) = Y_2(\alpha_1+\lambda) - z\beta_1\lambda + x^T(\alpha_2 - \lambda\beta_2) + F_1^{-1}(\tau_1)$$

$$Q_2(\tau_2|x,z) = z\beta_1 + x^T\beta_2 + F_2^{-1}(\tau_2)$$

Provided $\nabla_z Q_2(\tau_2|x,z) = \beta_1 \neq 0$, following Chesher (2003):

$$\alpha_1 = \nabla_{Y_2} Q_1(\tau_1|Y_2,x,z) + \frac{\nabla_z Q_1(\tau_1|Y_2,x,z)}{\nabla_z Q_2(\tau_2|x,z)}$$

$$\alpha_2 = \nabla_x Q_1(\tau_1|Y_2,x,z) - \frac{\nabla_z Q_1(\tau_1|Y_2,x,z)}{\nabla_z Q_2(\tau_2|x,z)} \nabla_x Q_2(\tau_2|x,z)$$

Discussion: The ‘unusual’ thing here is that α_1 and α_2 are not dependent on τ_1, τ_2 : *because*, that is *assumed* by the form of the model (2.1-2.2). ‘Location shift’. Our ‘thought experiments’ should be about ‘altering Y_2’s distribution not its value.’

Going Beyond Location Shift

$$
\begin{aligned}
Y_{i1} &= Y_{i2}\alpha_1 + x_i^T\alpha_2 + \delta Y_{i2}(\nu_{i1} + \lambda\nu_{i2}) && (2.6)\\
Y_{i2} &= z_i\beta_1 + x_i^T\beta_2 + \gamma z_i\nu_{i2} && (2.7)
\end{aligned}
$$

In the previous model (2.1–2.2) the "δY_{i2}" of the 1st eqn was "1", as was the γz_i of the 2nd eqn. If $\delta \neq 0$, $\gamma \neq 0$, you can write this as:

$$
\begin{aligned}
Y_{i1} &= Y_{i2}(\alpha_1 + \delta\nu_{i1} - \frac{\delta\beta_1\lambda}{\gamma}) + x_i^T\alpha_2 + \frac{Y_{i2}^2}{z_i}(\frac{\delta\lambda}{\gamma}) - \frac{Y_{i2}x_i^T}{z_i}(\frac{\delta\lambda\beta_2}{\gamma})\\
Y_{i2} &= z_i(\beta_1 + \gamma\nu_{i2}) + x_i^T\beta_2
\end{aligned}
$$

and we have conditional quantile functions

$$
\begin{aligned}
Q_1(\tau_1|Y_{i2}, x_i, z_i) &= Y_{i2}\theta_1(\tau_1) + x_i^T\theta_2(\tau_1) + \frac{Y_{i2}^2}{z_i}\theta_3(\tau_1) - \frac{Y_{i2}x_i^T}{z_i}\theta_4(\tau_1)\\
Q_2(\tau_2|x, z) &= z\beta_1(\tau_2) + x^T\beta_2(\tau_2) && (2.8\text{–}2.9)
\end{aligned}
$$

Follow the Coefficients

$$
\begin{aligned}
Y_{i1} &= Y_{i2}(\alpha_1 + \delta\nu_{i1} - \frac{\delta\beta_1\lambda}{\gamma}) + x_i^T\alpha_2 + \frac{Y_{i2}^2}{z_i}(\frac{\delta\lambda}{\gamma}) - \frac{Y_{i2}x_i^T}{z_i}(\frac{\delta\lambda\beta_2}{\gamma}) \\
Y_{i2} &= z_i(\beta_1 + \gamma\nu_{i2}) + x_i^T\beta_2
\end{aligned}
$$

$$
\begin{aligned}
Q_1(\tau_1|Y_{i2}, x_i, z_i) &= Y_{i2}\theta_1(\tau_1) + x_i^T\theta_2(\tau_1) + \frac{Y_{i2}^2}{z_i}\theta_3(\tau_1) - \frac{Y_{i2}x_i^T}{z_i}\theta_4(\tau_1) \\
Q_2(\tau_2|x, z) &= z\beta_1(\tau_2) + x^T\beta_2(\tau_2) \qquad \text{(2.8–2.9)}
\end{aligned}
$$

So:

$$
\begin{aligned}
\theta_1(\tau_1) &= (\alpha_1 + \delta\nu_{i1} - \frac{\delta\beta_1\lambda}{\gamma}) = (\alpha_1 + \delta F_1^{-1}(\tau_1) - \frac{\delta\beta_1\lambda}{\gamma}) \\
\theta_2(\tau_1) &= \alpha_2, \quad \theta_3(\tau_1) = (\frac{\delta\lambda}{\gamma}), \quad \theta_4(\tau_1) = -(\frac{\delta\lambda\beta_2}{\gamma}) \\
\beta_1(\tau_2) &= (\beta_1 + \gamma\nu_{i2}) = \beta_1 + \gamma F_2^{-1}(\tau_2), \quad \beta_2(\tau_2) = \beta_2
\end{aligned}
$$

Follow the Coefficients (2)

$$
\begin{aligned}
\theta_1(\tau_1) &= (\alpha_1 + \delta\nu_{i1} - \frac{\delta\beta_1\lambda}{\gamma}) = (\alpha_1 + \delta F_1^{-1}(\tau_1) - \frac{\delta\beta_1\lambda}{\gamma}) \\
\theta_2(\tau_1) &= \alpha_2, \quad \theta_3(\tau_1) = (\frac{\delta\lambda}{\gamma}), \quad \theta_4(\tau_1) = -(\frac{\delta\lambda\beta_2}{\gamma}) \\
\beta_1(\tau_2) &= (\beta_1 + \gamma\nu_{i2}) = \beta_1 + \gamma F_2^{-1}(\tau_2), \quad \beta_2(\tau_2) = \beta_2
\end{aligned}
$$

Some of these depend on τ_1 and τ_2, and some (most) do not. (Note that θ_1 'depends' on τ_2 through $\beta_1 = \beta_1(\tau_2) - \gamma F_2^{-1}(\tau_2)$.) You can write the model as:

$$
\begin{aligned}
Q_1(\tau_1|Q_2(\tau_2|x,z),x,z) &= Q_2(\tau_2|x,z)(\alpha_1 + \delta(F_1^{-1}(\tau_1) + \lambda F_2^{-1}(\tau_2))) \quad (***) \\
&\quad + x_i^T\alpha_2 \\
Q_2(\tau_2|x,z) &= z(\beta_1 + \gamma F_2^{-1}(\tau_2)) + x^T\beta_2
\end{aligned}
$$

Discussion on ‘Following the Coefficients’

The structural effect we seek can be obtained by (1) applying the magic formula to the system (2.8–2.9) or (2) by ‘direct inspection’ of (***) as

$$\pi(\tau_1, \tau_2) = (\alpha_1 + \delta(F_1^{-1}(\tau_1) + \lambda F_2^{-1}(\tau_2)))$$

From p.6: "It should be emphasized, however, that direct estimation of (2.8) and (2.9) without imposing the nonlinear restrictions implied by the model would fail to provide the exact cancellation of the exact calculation. Consequently, evaluating the *estimated* structural effect would produce an expression depending on the exogenous variables x and z, and we would need some scheme to average over the covariate space to obtain the structural effect."

The 'Panoramic View'

Given the separate contributions of $F_1^{-1}(\tau_1)$ and $F_2^{-1}(\tau_2)$, it is clear that $\pi(\tau_1, \tau_2)$ reflects not only the fact that the stochastic effect of Y_2 on Y_1 arises from two distinct sources, but also provides structural insight into how these sources are related. Suppose we fix τ_1 so v_1 is fixed at its τ_1 quantile, changes in τ_2 in $\pi_1(\tau_1, \tau_2)$ reflect how the distribution of v_2 affects the τ_1 quantile of the response Y_1. On the other hand, if we fix τ_2, and allow τ_1 to change, this sheds light on how the τ_2 quantile of Y_2 influences the whole distribution of the response Y_1. By considering variation in both τ_1 and τ_2 we obtain a panoramic view of the stochastic relationship between Y_2 and Y_1.

Estimation of structural quantile treatment effects

Exposition of Ma and Koenker pp. 7–10. Two methods: ‘weighted averaged derivative’ and 'control variate'. We start with the parametric recursive structural system:

$$\begin{aligned} Y_{i1} &= \varphi_1(Y_{i2}, x_i, \nu_{i1}, \nu_{i2}; \alpha) \qquad (2.10) \\ Y_{i2} &= \varphi_2(z_i, x_i, \nu_{i2}; \beta) \qquad (2.11) \end{aligned}$$

so the inverse function

$$\nu_{i2} = \tilde{\varphi}_2(Y_{i2}, z_i, x_i; \beta), \qquad \text{exists and}$$

$$Y_{i1} = \varphi_1(Y_{i2}, x_i, \nu_{i1}, \tilde{\varphi}_2(Y_{i2}, z_i, x_i; \beta); \alpha)$$

Write the conditional quantile functions:

$$\begin{aligned} Q_1(\tau_1 | Y_{i2}, x_i, z_i) &= h_1(Y_{i2}, x_i, z_i; \theta) \\ Q_2(\tau_2 | z_i, x_i) &= h_2(z_i, x_i; \beta) \end{aligned}$$

(Rhetorical question: why θ not α ?)

Weighted Average Derivatives

$$\begin{aligned} Q_1(\tau_1|Y_{i2}, x_i, z_i) &= h_1(Y_{i2}, x_i, z_i; \theta) \\ Q_2(\tau_2|z_i, x_i) &= h_2(z_i, x_i; \beta) \end{aligned}$$

$$\hat{\theta}(\tau_1) = \underset{\theta \in \Theta}{argmin} \sum_{i=1}^{n} \sigma_{i1} \rho_{\tau_1}(Y_{i1} - h_1(Y_{i2}, x_i, z_i; \theta)) \tag{2.12}$$

$$\hat{\beta}(\tau_2) = \underset{\beta \in B}{argmin} \sum_{i=1}^{n} \sigma_{i2} \rho_{\tau_2}(Y_{i2} - h_2(z_i, x_i; \beta)) \tag{2.13}$$

Goal is to estimate the weighted average quantile treatment effect:

$$\pi_1(\tau_1, \tau_2) = \int \left\{ \nabla_y Q_1(\tau_1|y_i, x_i, z_i) + \frac{\nabla_z Q_1(\tau_1|y_i, x_i, z_i)}{\nabla_z Q_2(\tau_2|x_i, z_i)} \right\} w(x_i, z_i) \mathrm{d}x \mathrm{d}z$$

There is also a weighted average exogenous variable effect:

$$\pi_2(\tau_1, \tau_2) = \int \left\{ \nabla_x Q_1(\tau_1 | y_i, x_i, z_i) - \frac{\nabla_z Q_1(\tau_1 | y_i, x_i, z_i)}{\nabla_z Q_2(\tau_2 | x_i, z_i)} \nabla_x Q_2(\tau_2 | x_i, z_i) \right\} \times w(x_i, z_i) \mathrm{d}x \mathrm{d}z$$

Looking at the two formulas (here's the weighted average quantile treatment effect again),

$$\pi_1(\tau_1, \tau_2) = \int \left\{ \nabla_y Q_1(\tau_1 | y_i, x_i, z_i) + \frac{\nabla_z Q_1(\tau_1 | y_i, x_i, z_i)}{\nabla_z Q_2(\tau_2 | x_i, z_i)} \right\} w(x_i, z_i) \mathrm{d}x \mathrm{d}z$$

we can see how to estimate these by taking averages over the sample of the corresponding estimated quantities.

A Control Variate Estimator

Back to the 2SLS example at the beginning. Recall, that the usual way to do 2SLS is to replace Y_2 by $\hat{Y}_2$ and 'run the regression'. But we could instead compute $\hat{\nu}_2 = Y_2 - \hat{Y}_2$ and include it in the regression while leaving Y_2 (not $\hat{Y}_2$) on the right hand side: the resulting estimates of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ would be the same as obtained by 2SLS. This result generalizes.

The control variate is going to be $\nu_2(\tau_2) = \nu_2 - F_2^{-1}(\tau_2)$ or for observation i: $\nu_{i2}(\tau_2) = \nu_{i2} - F_2^{-1}(\tau_2)$ (A little difficulty in this notation: $\nu_{i2}(\tau_2)$ is different from ν_{i2}.) Write the system of equations conditioning on the control variate:

$$\begin{aligned} Q_1(\tau_1|Y_{i2}, x_i, z_i, \nu_{i2}(\tau_2)) &= g_1(Y_{i2}, x_i, \nu_{i2}(\tau_2); \alpha(\tau_1, \tau_2)) \\ Q_2(\tau_2|z_i, x_i) &= g_2(z_i, x_i; \beta(\tau_2)) \end{aligned}$$

A Control Variate Estimator (2)

Notice that in the above system $Q_2(\tau_2|z_i, x_i) = g_2(z_i, x_i; \beta(\tau_2))$, that $g_2 = h_2$ in $Q_2(\tau_2|z_i, x_i) = h_2(z_i, x_i; \beta)$. Solving

$$\hat{\beta}(\tau_2) = \underset{\beta \in B}{argmin} \sum_{i=1}^{n} \sigma_{i2} \rho_{\tau_2}(Y_{i2} - g_2(z_i, x_i; \beta)),$$

we can invert to obtain

$$\begin{aligned} \nu_2 &= \tilde{\varphi}(Y_2, z, x; \beta), \qquad so \\ F_2^{-1}(\tau_2) &= \tilde{\varphi}(g_2(z, x; \beta), z, x; \beta) \end{aligned}$$

So the control variate can be obtained as

$$\begin{aligned} \hat{\nu}_{i2}(\tau_2) &= \tilde{\varphi}(Y_{i2}, z_i, x_i; \hat{\beta}) - \tilde{\varphi}(g_2(z_i, x_i; \hat{\beta}), z_i, x_i; \hat{\beta}) \\ &= \hat{\nu}_2 - \hat{F}_2^{-1}(\tau_2) \end{aligned}$$

Control Variate Estimator: Commentary and Conclusion

Note that the above procedure is valid regardless of the dimension of z_i, so as long as the model is identifiable $\hat{v}_{i2}(\tau_2)$ incorporates information on all of the available instruments. But it does so in a much more parsimonious fashion than by introducing z_i directly into what we have referred to in the introduction to Section 2 as the hybrid form of the first structural equation obtained by substituting for v_{i2}.

Once $\hat{v}_{i2}(\tau_2)$ is available, we can estimate the parameters of the first structural equation by reexpressing φ_1 as

$$g_1(Y_{i2}, x_i, \hat{v}_{i2}(\tau_2); a) = \varphi_1(Y_{i2}, x_i, F_1^{-1}(\tau_1), \hat{v}_{i2}(\tau_2); \alpha)$$

absorbing $F_1^{-1}(\tau_1)$ into the new parameter vector a, and solving

$$\hat{\alpha}(\tau_1, \tau_2) = \underset{a \in \mathscr{A}}{\operatorname{argmin}} \sum_{i=1}^{n} \sigma_{i1} \rho_{\tau_1}(Y_{i1} - g_1(Y_{i2}, x_i, \hat{v}_{i2}(\tau_2); a)).$$

Note that the coefficient of $\hat{v}_{i2}(\tau_2)$ may be interpreted as the degree of endogeneity, and a test for the local endogeneity could thus be formed based on the significance of the estimate for this coefficient.

Properties of the Estimators: Asymptopia

The weights σ_{ij}, $j = 1, 2$ that appear above in the various estimating expressions are optimally chosen to be $f(\xi_{ij})$, a density at the quantile of the disturbance in question. (In particular $\xi_{i1} = Q_1(\tau_1 | Y_{i2}, x_i, z_i, \nu_{i2}(\tau_2))$, $\xi_{i2} = Q_2(\tau_2 | z_i, x_i)$.)

When this is done the CV estimator is more efficient than the WAD estimator. Monte Carlos support this.

Marginal vs. Joint Independence Models

Following the discussion (p.10) in Chesher, "Identification of Non–Additive Structural Functions", 2005 World Congress paper. If we assume

$$Y_1 = h(Y_2, X, U_1) \qquad\qquad U_1 \perp X$$

we have 'marginal independence.' The notation here includes in X covariates which locally or globally have no effect on Y_1. (In other contexts these are notationally distinguished and call 'instruments'.) Chesher's paper outlines properties that are generic to these models; on the whole, they are more difficult to work with than models that additionally supply an equation for Y_2 :

$$Y_2 = g(X, U_2) \qquad\qquad U_2 \perp X,$$

and which replace $U_1 \perp X$ with $U_1 \perp X | U_2$ which in turn implies $(U_1, U_2) \perp X$, whence the name 'joint independence.'

IV for Quantile Regression Models with 'Marginal Independence': Chernozhukov and Hansen (2005, 2006)

The first of these papers (a (long) note in *Econometrica*) proposes a model that exhibits marginal independence and compares it with Chesher (2003) and Imbens and Newey (2003), characterized as 'control function' approaches. The discussion in Section 3, p. 254 should be read in conjunction with Chesher's World Congress discussion.

The 2006 *J.Econometrics* paper explicitly discusses estimation. We focus exclusively on the estimation algorithm given in Section 3.2, 'An instrumental variable quantile regression process and an analogy with two stage least squares', starting on p.9 of the proofs.

Chernozhukov and Hansen's IVQR (CH–IVQR)

Basic model is written:

$$q(d, x, \tau) = d'\alpha(\tau) + x'\beta(\tau) \tag{3.4}$$

where d is a dim(α) vector of treatment variables (possibly interacted with covariates), x covariates. Define 'the weighted quantile regression function'

$$Q_n(\tau, \alpha, \beta, \gamma) \equiv \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - D_i'\alpha - X_i'\beta - \hat{\Phi}_i(\tau)'\gamma) \cdot \hat{V}_i(\tau)$$

where $\hat{\Phi}_i(\tau) \equiv \hat{\Phi}(\tau, X_i, Z_i)$ is a dim(α) vector of (transformations of) instruments, and $\hat{V}_i(\tau) \equiv \hat{V}(\tau, X_i, Z_i)$ a positive weight function. Let $\|x\|_A$ be the norm of a vector x using metric A (e.g. $x'Ax$ where A is positive definite). Then:

$$\hat{\alpha}(\tau) = arg \inf_{\alpha \in A} \|\hat{\gamma}(\alpha, \tau)\|_{A(\tau)}, \quad where \tag{3.5}$$

$$(\hat{\beta}(\alpha, \tau), \hat{\gamma}(\alpha, \tau)) = arg \inf_{\beta, \gamma} Q_n(\tau, \alpha, \beta, \gamma) \tag{3.6}$$

‘Practical’ CH–IVQR

The foregoing amounts to:

1. For a given τ, define a grid of value $\{\alpha_j, j = 1, ..., J\}$ and run the ordinary quantile regression of $Y_i - D_i'\alpha_j$ on X_i and $\hat{\Phi}_i(\tau)$ to obtain coefficients $\hat{\beta}(\alpha_j, \tau), \hat{\gamma}(\alpha_j, \tau)$.
2. Choose $\hat{\alpha}(\tau)$ as that value among $\{\alpha_j, j = 1, ..., J\}$ that makes $\|\hat{\gamma}(\alpha_j, \tau)\|_{A(\tau)}$ as close to zero as possible. Choose $\hat{\beta}(\tau)$ that corresponds to α_j.

(Discussion–why this works.)