Online Appendix to "Has Moral Hazard Become a More Important Factor in Managerial Compensation?"

George-Levi Gayle and Robert A. Miller

Tepper School of Business, Carnegie Mellon University

August 2008

Abstract

In this online appendix we formally show that the model in the main text is identified, describe the empirical implementation of our estimation technique, and derive the asymptotic covariance of our estimator.

I. Identification

The identification of this model is an application of Gayle and Miller (2008). For the reasons given in the text, we proceed as if true compensation, w_t, and excess returns, x_t, are observed for the purposes of establishing identification of the other parameters. Identification of the remaining parameters, namely the risk-aversion parameter (ρ), tastes for shirking over diligence (α_2/α_1), tastes for diligence over the value of quitting (α_2/α_0), and the signalling function ($g(x)$) proceeds in two steps. First, we prove (α_2/α_1), (α_2/α_0), and $g(x)$ are identified if ρ is known. Then we give sufficient conditions for identifying ρ.

Defining $v_t(x, \rho)$ as

$$v_t(x, \rho) \equiv \left(\frac{\alpha_0}{\alpha_2}\right)^{1/(b_t - 1)} \exp\left[\frac{\rho w_t(x)}{b_{t+1}}\right], \tag{1}$$

it follows from the optimal contract for diligent work,

$$w_t = \frac{b_{t+1}}{\rho(b_t - 1)} \ln\left(\frac{\alpha_2}{\alpha_0}\right) + \frac{b_{t+1}}{\rho} \ln\left[1 + \eta_t \left(\frac{\alpha_2}{\alpha_1}\right)^{1/(b_t-1)} - \eta_t g(x_t)\right], \tag{2}$$

that for a given value for ρ, a transformation of the optimal compensation, depending only on (observed) bond prices, is a linear mapping of $g(x)$. Namely,

$$v_t(x, \rho) = 1 + \eta_t \left[(\alpha_2/\alpha_1)^{1/(b_t-1)} - g(x)\right]. \tag{3}$$

So, if the values of the intercept and the slope of the mapping could be found, and the value of ρ were known, then $g(x)$ could be simply determined. Taking the expectation with respect to x conditional on the price of bonds at time t yields

$$E\left[v_t(x, \rho)\right] = 1 + \eta_t(\alpha_2/\alpha_1)^{1/(b_t-1)} - \eta_t \equiv \underline{v}_t(\rho). \tag{4}$$

We now impose a regularity condition on $g(x)$, satisfied by our parameterization, that says $g(x) \to 0$ as $x \to \infty$. Intuitively this condition states that the shareholders attach negligible probability to a manager shirking if the firm's excess returns are extraordinarily high. The condition implies

$$\lim_{x\to\infty} v_t(x, \rho) = 1 + \eta_t(\alpha_2/\alpha_1)^{1/(b_t-1)} \equiv \overline{v}_t(\rho). \tag{5}$$

Solving for the signaling function $g(x)$, the nonpecuniary benefit ratio (α_2/α_1), and the tastes for participation (α_2/α_0) given ρ, using equations (3), (4), and (5), proves the following.

Proposition 1. *For any $\rho^* > 0$,*

$$\begin{aligned}
\frac{\alpha_2^*}{\alpha_0^*} &= \left[E_t\left\{\exp\left[-\frac{\rho^* w_t(x)}{b_{t+1}}\right]\right\}\right]^{1-b_t} \\
\frac{\alpha_2^*}{\alpha_1^*} &= \left(\frac{\overline{v}_t(\rho^*)-1}{\overline{v}_t(\rho^*)-\underline{v}_t(\rho^*)}\right)^{b_t-1} \\
g^*(x) &= \frac{\overline{v}_t(\rho^*)-v_t(x,\rho^*)}{\overline{v}_t(\rho^*)-\underline{v}_t(\rho^*)}.
\end{aligned}$$

Proof of Proposition 1. The expression for α_2^*/α_0^* follows directly from rearranging the participation constraint (7). Subtracting equation (5) from (3), we obtain

$$\eta_t g(x) = \overline{v}_t(\rho^*) - v_t(x,\rho^*).$$

Subtracting equation (4) from (5) we obtain

$$\eta_t = \overline{v}_t(\rho^*) - \underline{v}_t(\rho^*). \tag{6}$$

Substituting for η_t using (6) in the previous equation and making $g(x)$ the subject of the resulting equation yields:

$$g(x) = \frac{\overline{v}_t(\rho^*)-v_t(x,\rho^*)}{\overline{v}_t(\rho^*)-\underline{v}_t(\rho^*)}.$$

Finally, making (α_2/α_1) the subject of equation (4) and then substituting for η_t using (6), we obtain

$$\frac{\alpha_2^*}{\alpha_1^*} = \left(\frac{\overline{v}_t(\rho^*)-1}{\eta_t}\right)^{b_t-1} = \left(\frac{\overline{v}_t(\rho^*)-1}{\overline{v}_t(\rho^*)-\underline{v}_t(\rho^*)}\right)^{b_t-1},$$

as required.

Proposition 1 establishes that if ρ^* is known then (α_2^*/α_1^*), (α_2^*/α_0^*), and $g^*(x)$ are identified since they can be written as a mapping of the data, because consistent estimates of the mappings $\overline{v}_t(\rho)$ and $\underline{v}_t(\rho)$ can be obtained from the data. See Gayle and Miller (2008) for details on constructing nonparametric consistent estimates of these quantities. A natural place to begin investigating the identification of ρ^* is the participation constraint. When $(\alpha_2/\alpha_0) > 1$, meaning the nonpecuniary benefits of working do not fully compensate the manager for the total benefits of his alternative, and thus expected compensation is positive, the data imply a lower bound for the risk-aversion parameter, ρ. To picture this, define the mapping

$$\psi_t(\rho) \equiv E_t\left[\exp\left(-\frac{\rho w_t(x)}{b_{t+1}}\right)\right].$$

From its definition, $\psi_t(0) = 1$, while the assumption above implies

$$\psi_t'(0) = \frac{\partial}{\partial\rho} E\left[\exp\left(-\frac{\rho w_t(x)}{b_{t+1}}\right)\right]_{\rho=0} = -E\left[\frac{w_t(x)}{b_{t+1}}\right] < 0.$$

Also $\psi_t(\rho)$ is convex in ρ because

$$\frac{\partial^2}{\partial\rho^2}\left[\exp\left(-\frac{\rho w_t(x)}{b_{t+1}}\right)\right] = \left(\frac{w_t(x)}{b_{t+1}}\right)^2 \exp\left(-\frac{\rho w_t(x)}{b_{t+1}}\right) > 0$$

and the expectations operator preserves convexity. Assuming $\alpha_2/\alpha_0 > 1$, it now follows that $\psi_t(\rho)$ crosses the unit level from below just once at say ρ_t, which implies $\psi_t(\rho) > 1$ for all $\rho > \rho_t$. This rules out the possibility that $\rho^* \leq \rho_t$. Intuitively, the participation equation is satisfied by different combinations of ρ and α_2/α_0 satisfying $\rho > \rho_t$ and $\alpha_2/\alpha_0 = \psi_t(\rho)^{1-b_t}$ as we see in Figure 1.

Along this line, as ρ increases, the person becomes more risk averse, the expected utility from $w_t(x)$ declines along with its certainty equivalent, but this is just offset by nonpecuniary amenities from the job. Consequently an observer with cross-sectional data on a homogeneous set of firms and managerial compensation paid out in that period cannot distinguish between a sample of managers with a high risk tolerance and unpleasant working conditions, versus a sample with lower tolerance but more nonpecuniary benefits. The remaining parameters are then inferred from the value ascribed to ρ, the slope of the contract with respect to abnormal returns determining $g(x)$ and thence the probability distribution of abnormal returns under shirking.

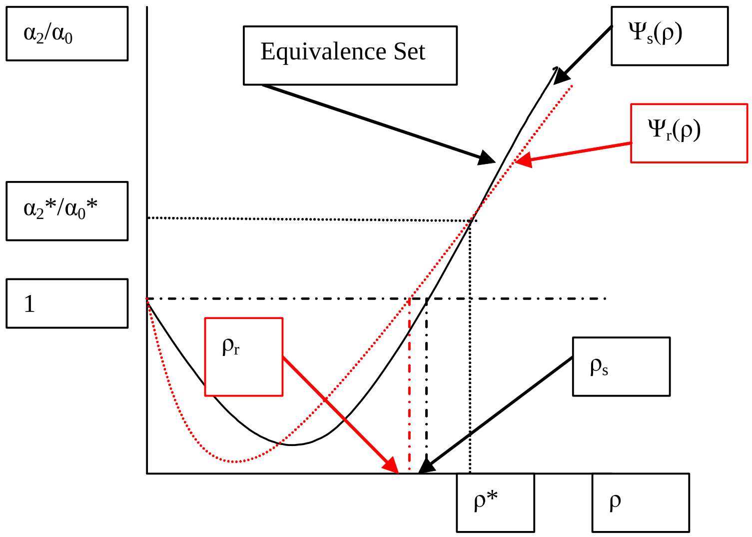

Figure 1: Equivalence Set

Accordingly, we now suppose there are data on at least two states $s \in S$, that is dates with distinct bond prices, or sectors where the nonpecuniary benefits of the job and the alternative opportunities for work are the same. More formally, the two states have different compensation plans $w_r(x)$ and $w_s(x)$ but the same nonpecuniary benefits from diligent work α_2. In this case $w_r(x) \neq w_s(x)$ because the probability density function of abnormal returns from working diligently differs by state, that is $f_{2r}(x) \neq f_{2s}(x)$, or the density from shirking differs, that is $f_{1r}(x) \neq f_{1s}(x)$.

The existence of multiple states provides a means of identifying ρ. Since the participation condition holds for each state $s \in S$ separately, we can in principle solve moment conditions of the form

$$\left[\int \exp\left[-\rho b_{r+1}^{-1} w_r(x)\right] f_{2r}(x)\mathrm{d}x\right]^{\varkappa(r)} = \left[\int \exp\left[-\rho b_{s+1}^{-1} w_r(x)\right] f_{2s}(x)\mathrm{d}x\right]^{\varkappa(s)}$$

in ρ, where $\varkappa(r) = \varkappa(s) = 1$ when they are sectors and $\varkappa(r) = 1 - b_r$ and $\varkappa(s) = 1 - b_s$ when they are dates. Figure 1 illustrates how identification would be achieved with two states, ρ^* determined by a unique intersection of $\psi_s(\rho)$ with $\psi_t(\rho)$. Although there may be multiple roots in ρ to the equations defined by the separate states $r \in S$ and $s \in S$, if there is a unique root common to all possible pairs, then ρ is identified.

II. Empirical Implementation and Standard Errors

In the old sample and the new restricted sample, the data are ordered by $n \in \{1, .., N\}$, where each observation refers to a firm–year vector of variables, including compensation paid to the three top executives, the abnormal return, the number of employees, the asset-to-equity ratio, GDP that year, the bond price in the current year (denoted b_n), the bond price the following year (denoted b_{1n}), and sector dummy variables.

A. Stage Zero

Recall that

$$x_n = \pi_n - \pi - z_n \gamma,$$

where γ is a 2×1 vector and z_{nt} is a 1×2 vector of sector-specific constants and GDP, and that x is estimate as the residual of the regression of z_n on $\pi_n - \pi$. Let $\gamma^{(N)}$ denote the estimate of γ from that regression. For each sector, we estimate the lower bound of the excess return distribution as

$$\psi^{(N)}\left(\gamma^{(N)}\right) = \min_{\{1,..,N\}} \left\{\pi_n - \pi - z_n \gamma^{(N)}\right\}. \quad (7)$$

Let γ_0 denote the true value of γ in the population. Note that if γ_0 were known then

$$\psi^{(N)}(\gamma_0) = \min_{\{1,..,N\}} \{\pi_n - \pi - z_n \gamma_0\}$$

would be a super-consistent estimate of $\psi_0(\gamma_0)$, the true value of ψiin the population. However, since we are using $\gamma^{(N)}$ instead of γ_0, the following Lemma establishes that it is only $\sqrt{N}$ consistent and gives its asymptotic variance.

Lemma 2. *Under standard regularity conditions,*

$$\sqrt{N}\left(\psi^{(N)}\left(\gamma^{(N)}\right)-\psi_0(\gamma_0)\right) \Rightarrow N(0,\mathrm{var}(\psi_0)),$$

where $\mathrm{var}(\psi_0)=\sigma^2(z)\underline{z}(z'z)^{-1}\underline{z}'$, $\sigma^2(z)=\mathrm{var}(x|z)$ *and* $\underline{z}$ *is the value of* z *at the minimum of* x.

Proof. Define

$$\psi^{(N)}(\gamma)=\min_{\{1,..,N\}}\{\pi_n-\pi-z_n\gamma\} \text{ for any } \gamma\in\mathbb{R}$$

and

$$\psi_0(\gamma)=p\lim_{N\to\infty}\psi^{(N)}(\gamma) \text{ for any } \gamma\in\mathbb{R}.$$

Next note

$$\psi^{(N)}(\gamma^{(N)})-\psi_0(\gamma_0)=\psi^{(N)}(\gamma^{(N)})-\psi^{(N)}(\gamma_0)+\psi^{(N)}(\gamma^0)-\psi_0(\gamma_0). \tag{8}$$

Since $\gamma^{(N)}$ is a $\sqrt{N}$-consistent estimator of γ_0,

$$\psi^{(N)}(\gamma^{(N)})-\psi^{(N)}(\gamma_0)=O_p(N^{-\frac{1}{2}}),$$

and since $\psi^{(N)}(\gamma_0)$ is a super-consistent estimator of $\psi_0(\gamma_0)$,

$$\psi^{(N)}(\gamma_0)-\psi_0(\gamma_0)=O_p(N^{-1}).$$

Therefore

$$\begin{aligned}\psi^{(N)}\left(\gamma^{(N)}\right)-\psi_0(\gamma_0) &= O_p\left(N^{-\frac{1}{2}}\right)+O_p\left(N^{-1}\right)\\ &= O_p\left(\max\left\{N^{-\frac{1}{2}},N^{-1}\right\}\right)\\ &= O_p\left(N^{-\frac{1}{2}}\right). \qquad (9)\end{aligned}$$

Hence $\psi^{(N)}\left(\gamma^{(N)}\right)$ is $\sqrt{N}$ consistent, which implies that $\sqrt{N}\left(\psi^{(N)}\left(\gamma^{(N)}\right) - \psi_0(\gamma_0)\right) \Rightarrow N(0, \mathrm{var}(\psi_0))$. The variance formula for $\mathrm{var}(\psi_0)$ follows from the asymptotic variance of $\psi_0\left(\gamma^{(N)}\right)$.

B. Stage One

In order to take into account the pre-estimation in x, we now make its dependence on γ explicit by defining

$$x_n(\gamma) \equiv \pi_n - \pi - z_n\gamma.$$

For each sector, the log likelihood of observing $x_n(\gamma)$ is given by

$$(10) \quad l(\psi_0(\gamma_0), x(\gamma_0), \sigma) = \log\sigma + \ln\Phi\left[\frac{\mu(\psi_0(\gamma_0), \sigma) - \psi_0(\gamma_0)}{\sigma}\right] + \frac{[x(\gamma_0) - \mu(\psi_0(\gamma_0), \sigma)]^2}{2\sigma^2},$$

where $\mu(\psi_0(\gamma_0), \sigma)$ is defined as the implicit solution in μ_2 of the following equation.

$$(11) \qquad \mu_2 + \sigma\frac{\phi\left[\frac{\mu_2 - \psi_0(\gamma_0)}{\sigma}\right]}{\Phi\left[\frac{\mu_2 - \psi_0(\gamma_0)}{\sigma}\right]} = 0$$

Let $S(\psi_0(\gamma_0), x(\gamma_0), \sigma)$, the score, be the derivative of $l(\psi_0(\gamma_0), x(\gamma_0), \sigma)$ with respect to σ and define

$$(12) \qquad h_0(\psi_0(\gamma_0), x(\gamma_0), \sigma) = \begin{bmatrix} S(\psi_0(\gamma_0), x(\gamma_0), \sigma) \\ z'x(\gamma_0) \end{bmatrix}$$

to be the 3×1 vector of moment condition with

$$(13) \qquad E[h_0(\psi_0(\gamma_0), x(\gamma_0), \sigma)] = 0.$$

Define

$$G_\sigma = E\left[\frac{\partial^2 l(\psi_0(\gamma_0), x(\gamma_0), \sigma)}{\partial\sigma\partial\sigma}\right],$$

$$G_\gamma = E\left[\frac{\partial S(\psi_0(\gamma_0), x(\gamma_0), \sigma)}{\partial \gamma}\right],$$

$$S(z) = S(\psi_0(\gamma_0), x(\gamma_0), \sigma),$$

$$D = -E[z'z],$$

and

$$\varphi(z) = E[z'z]^{-1} z' x(\gamma_0).$$

Under standard regularity conditions,*

$$\sqrt{N}(\sigma^N - \sigma) \Longrightarrow N(0, V(\sigma)),$$

where

$$V(\sigma) = G_\sigma^{-1} E[\{S(z) + G_\gamma \varphi(z)\}\{S(z) + G_\gamma \varphi(z)\}'] G_\sigma^{-1\prime}.$$

This follows directly from Theorem 6.1 of Whitney K. Newey and Daniel McFadden (1994).

C. Stage Two-Estimation and Standard Error

Having obtained estimates of the coefficients σ and $\psi_0(\gamma^0)$, which determine the probability density function for abnormal returns, $f_2(x)$, we estimated the remaining parameters $\theta \equiv (\rho, u_1, a_1, a_2, \xi)$ from orthogonality conditions derived from the participation and incentive-compatibility constraints, along with the score of the optimal contract's likelihood function in a generalized method-of-moments procedure, after substituting our estimate for σ, $\psi_0(\gamma^0)$, and $x(\gamma^0)$ obtained in the first step. Let the true value of θ be denoted by $\theta^o \equiv (\rho^o, u_1^o, a_1^o, a_2^o, \xi^o)$.

The first vector of orthogonality conditions is constructed from the participation constraints (a vector of three executives) of the form

(14) $$h_{1n}(\boldsymbol{\theta}) = \exp[-b_{1n}^{-1}(\rho \widetilde{w}_n + \xi)] - (a_2' z_n)^{1/(1-b_n)}.$$

The distributional assumptions on ε_n imply

$$E\left\{\exp[-b_{1n}^{-1}(\rho^o\widetilde{w}_n+\xi)]\big|\,w_n,b_{1n}\right\}=\exp\left[-b_{1n}^{-1}(\rho^o w_n)\right]. \tag{15}$$

Because the participation equation is met with equality under the optimal contract, it follows that

$$E[h_{1n}(\theta^o)]=0. \tag{16}$$

The second vector of orthogonality conditions is based on the incentive-compatibility constraint. Define the vector

$$h_{2n}(\boldsymbol{\theta},x_n(\gamma),\sigma,\psi(\gamma))=\exp\left[-b_{1n}^{-1}(\rho\widetilde{w}_n+\xi)\right]\left[\frac{f_1(x_n(\gamma),\sigma,\psi(\gamma))}{f_2(x_n(\gamma),\boldsymbol{\theta},\sigma,\psi(\gamma))}-(a_1'z_n)^{1/(b_n-1)}\right]. \tag{17}$$

The incentive-compatibility constraint is also met with equality under the optimal contract, when the parameters are set to their true values, implying

$$E\left[h_{2n}(\boldsymbol{\theta}^0,x_n(\gamma^0),\sigma^0,\psi(\gamma^0))\right]=0, \tag{18}$$

where $(\sigma^o,\psi(\gamma^o),\gamma^0)$ are the true values of (σ,ψ,γ).

The final set of orthogonality conditions comes from the properties of the optimal contract. According to definition of ε, the observed compensation can be written as

$$\widetilde{w}_n=\frac{b_{1n}}{\rho(b_n-1)}\ln(a_2'z_n)+\frac{b_{1n}}{\rho}\ln\left[1+\eta_n(a_1'z_n)^{\frac{1}{(b_n-1)}}-\eta_n\frac{f_1(x_n(\gamma),\sigma,\psi(\gamma))}{f_2(x_n(\gamma),\boldsymbol{\theta},\sigma,\psi(\gamma))}\right]+\varepsilon_n, \tag{19}$$

where η_n is the unique, strictly positive solution to the following equation in η.

$$\int[\eta(a_1'z_n)^{1/(b_n-1)}-\eta\frac{f_1(x_n(\gamma),\sigma,\psi(\gamma))}{f_2(x_n(\gamma),\boldsymbol{\theta},\sigma,\psi(\gamma))}+1]^{-1}f_2(x_n(\gamma),\boldsymbol{\theta},\sigma,\psi(\gamma))\mathrm{d}x=1 \tag{20}$$

Denoting the density of $\widetilde{w}_n$ conditional on z_n and x_n as $f_{\theta,\sigma,\psi,\gamma}(\widetilde{w}_n\,|z_n,x_n)$, we can write the score with respect to θ for the likelihood of observing $\widetilde{w}_n$ as

$$h_{3n}(\boldsymbol{\theta},x_n(\gamma),\sigma,\psi(\gamma)) = \nabla_\theta \ln f_{\theta,s,\psi,\gamma}(\widetilde{w}_n\,|z_n,x_n). \tag{21}$$

From the definition of a score,

$$E\left[h_{3n}(\boldsymbol{\theta}^0,x_n(\gamma^0),\sigma^0,\psi(\gamma^0))\right] = 0. \tag{22}$$

Our estimator for θ was found by forming a $q\times 1$ vector function $h_{4n}(\boldsymbol{\theta},x_n(\gamma),\sigma,\psi(\gamma))$ from $h_{1n}(\boldsymbol{\theta}), h_{2n}(\boldsymbol{\theta},x_n(\gamma),\sigma,\psi(\gamma))$ and $h_{3n}(\boldsymbol{\theta},x_n(\gamma),\sigma,\psi(\gamma))$ and minimizing

$$\left[\frac{1}{N}\sum_{n=1} h_{4n}(\boldsymbol{\theta},x_n(\gamma^N),\sigma^N,\psi(\gamma^N))\right]' A_N \left[\frac{1}{N}\sum_{n=1} h_{4n}(\boldsymbol{\theta},x_n(\gamma^N),\sigma^N,\psi(\gamma^N))\right] \tag{23}$$

with respect to $\boldsymbol{\theta}$ subject to equation(20) which defines η_n, where A_N, which is a $q\times q$ matrix converging to some constant nonsingular matrix A, and the estimators $(\sigma^{(N)},\psi(\gamma^{(N)}),x_n(\gamma^N))$ come from the first two steps.

Let

$$h_{4n}(\boldsymbol{\theta},\boldsymbol{\theta}_1) = h_{4n}(\boldsymbol{\theta},x_n(\gamma),\sigma,\psi(\gamma)),$$

where $\boldsymbol{\theta}_1 = (\gamma,\sigma)'$. Next, define

$$G_{\boldsymbol{\theta}} = E\left[\nabla_{\boldsymbol{\theta}} h_{4n}(\boldsymbol{\theta},\boldsymbol{\theta}_1)\right],$$

$$G_{\boldsymbol{\theta}_1} = E\left[\nabla_{\boldsymbol{\theta}_1} h_{4n}(\boldsymbol{\theta},\boldsymbol{\theta}_1)\right],$$

$$h_4(z) = h_{4n}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}_1^0\right),$$

$$M = E\left[\nabla_{\boldsymbol{\theta}_1} h_0(\boldsymbol{\theta}_1^0)\right],$$

and

$$\varphi_1(z) = -M^{-1} h_0 \left(\boldsymbol{\theta}_1^0\right).$$

Under standard regularity conditions

$$\sqrt{N}(\boldsymbol{\theta}^N - \boldsymbol{\theta}^0) \Longrightarrow N(0, V_1),$$

where

$$V_1 = (G'_{\boldsymbol{\theta}} A G_\theta)^{-1} E[\{G'_{\boldsymbol{\theta}} A h_4(z) + G'_{\boldsymbol{\theta}} A G_{\boldsymbol{\theta}_1} \varphi_1(z)\}\{G'_{\boldsymbol{\theta}} A h_4(z) + G'_{\boldsymbol{\theta}} A G_{\boldsymbol{\theta}_1} \varphi_1(z)\}'](G'_{\boldsymbol{\theta}} A G_\theta)^{-1}.$$

The result follows from applying Theorem 6.1 of Whitney Newey and Daniel McFadden (1994) to moments based on the first-order conditions of the minimization problem (23). In our application:

$$\text{plim} A_N = E\{h_4(z) h_4(z)'\}^{-1}.$$

REFERENCES

Gayle, George-Levi, and Robert A. Miller. 2008. "Identifying and Testing Generalized Moral Hazard Models of Managerial Compensation." Tepper school of Business, Carnegie Mellon University.

Newey, Whitney K., and Daniel McFadden. 1994. "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics,* ed. R. F. Engle and D. L. McFadden. 2111–245. Amsterdam: North-Holland.

Notes

*See Newey and McFadden (1994) for examples of these regularity conditions.