Appendix 1

Grossman (1972): On the Concept of Health Capital and the Demand for Health

Wage Effects

To obtain the wage elasticities of medical care and the time spent producing health, three equations must be partially differentiated with respect to the wage. These equations are the gross investment production function and the two first-order conditions for cost minimization:

$$I(M,TH;E) = Mg(t;E) = (\tilde{H}+\delta)H,$$
$$W' = \pi g',$$
$$P = \pi(g - tg').$$

Since I is linear homogenous in M and TH,

$$\frac{\partial(g-tg')}{\partial M} = -\frac{t\partial(g-tg')}{\partial TH},$$
$$\frac{\partial g'}{\partial TH} = -\frac{1}{t}\frac{\partial(g-tg')}{\partial TH},$$
$$\sigma_p = \frac{(g-tg')g'}{I\{[\partial(g-tg')]/\partial TH\}}.$$

Therefore, the following relationships hold:

$$\frac{\partial(g-tg')}{\partial M} = -\frac{t(g-tg')g'}{I\sigma_p},$$
$$\frac{\partial g'}{\partial TH} = -\frac{1}{t}\frac{(g-tg')g'}{I\sigma_p}, \tag{A1}$$

$$\frac{\partial(g - tg')}{\partial TH} = \frac{(g - tg')g'}{lq_p}.$$

Carrying out the differentiation, one gets

$$g'\frac{dTH}{dW} + (g - tg')\frac{dM}{dW} = -\frac{H(H+\delta)\varepsilon}{\pi}\left(\frac{d\pi}{dW} - \frac{\pi}{W}\right),$$

$$1 = g'\frac{d\pi}{dW} + \pi\left(\frac{\partial g'}{\partial TH}\frac{dTH}{dW} + \frac{\partial g'}{\partial M}\frac{dM}{dW}\right),$$

$$0 = (g - tg')\frac{d\pi}{dW} + \pi\left[\frac{\partial(g - tg')}{\partial TH}\frac{dTH}{dW} + \frac{\partial(g - tg')}{\partial M}\frac{dM}{dW}\right].$$

Using the cost-minimization conditions and (A1) and rearranging terms, one has

$$l\varepsilon\frac{d\pi}{dW} + W\frac{dTH}{dW} + P\frac{dM}{dW} = \frac{l\varepsilon\pi}{W},$$

$$l\sigma_p\frac{d\pi}{dW} - \frac{1}{t}P\frac{dTH}{dW} + P\frac{dM}{dW} = l\frac{\pi}{W}\sigma_p, \qquad \text{(A2)}$$

$$l\sigma_p\frac{d\pi}{dW} + W\frac{dTH}{dW} - tW\frac{dM}{dW} = 0.$$

Since (A2) is a system of three equations in three unknowns – dTH/dW, dM/dW, and $d\pi/dW$ - Cramer's rule can be applied to solve for, say, dM/dW :

$$\frac{dM}{dW} = \frac{\begin{vmatrix} l\varepsilon + W & +\dfrac{l\varepsilon\pi}{W} \\ l\sigma_p - \left(\dfrac{1}{t}\right)P & +l\left(\dfrac{\pi}{W}\right)\sigma_p \\ l\sigma_p + W & -0 \end{vmatrix}}{\begin{vmatrix} l\varepsilon + W & +P \\ l\sigma_p - \left(\dfrac{1}{t}\right)P & +P \\ l\sigma_p + W & -tW \end{vmatrix}}.$$

The determinant in the denominator reduces to $(I\sigma_p \pi^2 I^2)/THM$. The determinant in the numerator is

$$\frac{I\sigma_p}{THM}\left(I\pi\sigma_p THM + I\pi\varepsilon \frac{P}{W} M^2 \right).$$

Therefore,

$$\frac{dM}{dW} = \frac{THM}{I\pi}\left(\sigma_p + \frac{\varepsilon PM}{WTH} \right).$$

In elasticity notation, this becomes

$$\ell_{M,W} = (1-K)\varepsilon + K\sigma_p. \tag{A3}$$

Along similar lines, one easily shows that

$$\ell_{TH,W} = (1-K)(\varepsilon - \sigma_p). \tag{A4}$$

The Role of Human Capital

To convert the change in productivity due to a shift in human capital into a change in average or marginal cost, let the percentage changes in the marginal products of medical care and own time for a one-unit change in human capital be given by

$$\frac{\partial(g - tg')}{\partial E}\frac{1}{g - tg'} = \frac{g\hat{g} - tg'\hat{g}'}{g - tg'},$$

$$\frac{\partial g'}{\partial E}\frac{1}{g'} = \hat{g}'.$$

If a shift in human capital were "factor neutral," the percentage changes in these two marginal products would be equal:

$$\hat{g} = \frac{g\hat{g} - tg'\hat{g'}}{g - tg'},$$

or

$$\hat{g'} = \hat{g} = r_H. \tag{A5}$$

The average cost of gross investment in health is defined as

$$\pi = (PM + WTH)I^{-1} = (P + Wt)g^{-1}.$$

Given factor neutrality,

$$\frac{d\pi}{dE}\frac{1}{\pi} = -\hat{g} = -r_H. \tag{A6}$$

This coincides with the percentage change in marginal cost, since

$$\pi = P(g - tg')^{-1},$$

and

$$\frac{d\pi}{dE}\frac{1}{\pi} = -\left(\frac{g\hat{g} - tg'\hat{g'}}{g - tg'}\right) = -\hat{g'} = -\hat{g} = -r_H. \tag{A7}$$

Taking the *total* derivative of E in the gross investment function, one computes this parameter thus:

$$\frac{dI}{dE}\frac{1}{I} = M\frac{(g - tg')}{I}\hat{M} + \frac{THg'}{I}\hat{TH} + r_H.$$

Since $\hat{M} = \hat{TH}$ and $\hat{H} = \hat{I}$, the last equation can be rewritten as

$$\hat{H} = \hat{M} + r_H.$$

Solving for $\hat{M}$ and noting that $\hat{H} = r_H \varepsilon$, one gets

$$\hat{M} = r_H(\varepsilon - 1). \tag{A8}$$

Appendix 2

Muurinen(1982): Demand for Health

The Comparative Static Predictions:

Age Effects

$$[U_s/\lambda + Y_s]\phi' = [\delta(t, X)]C. \qquad \text{(A9)}$$

Equation (A9) in logarithms is

$$\ln[U_s/\lambda + Y_s] + \ln\phi' = \ln\delta(t, X) + \ln C. \qquad \text{(A10)}$$

The differentiation of this with respect to time yields

$$(\partial\ln\phi'/\partial\ln K^h)(\partial\ln K^h/\partial t) = (\partial\ln\delta/\partial t) + (\partial\ln C/\partial t). \qquad \text{(A11)}$$

If, following Grossman, $\partial\ln\phi'/\partial\ln K^h$ is denoted by $-1/\varepsilon$, negative of the inverse of the marginal efficiency of health capital, and a tilde above the variable and t as its subscript used to denote its percentage change over time, (A11) becomes

$$-(1/\varepsilon)\tilde{K}^h_t = \tilde{\delta}_t + \tilde{C}_t. \qquad \text{(A12)}$$

Since it is assumed that C is dependent on time only through the marginal productivity of medical care, it follows that [from $C = (P^M - Y_M)/f$]

$$\dot{C} = C(-\dot{f}/f), \qquad \text{(A13)}$$

or

$$\dot{C}/C = -\tilde{f}_t. \qquad \text{(A14)}$$

If the productivity of medical care decreases with age, $\tilde{f}_t$ is negative and $\tilde{C}_t$ positive. It is further assumed that

$$\dot{C}/C = q. \tag{A15}$$

Therefore, (A12) becomes

$$K_t^h = -\varepsilon[\delta_t + q], \tag{A16}$$

whereas

$$\dot{K}^h(t) = f(t)M(t) - \delta[t, X(t)]K^h(t), \tag{A17}$$

To derive the effect of age on medical care, equation (A17) is solved for M. In logarithmic form this can be written as

$$\ln M = \ln K^h + \ln(\dot{K}^h / K^h + \delta) - \ln f. \tag{A18}$$

Differentiating this with respect to time, taking into account equations (A14) and (A15), and assuming that δ_t is constant, one arrives at equation:

$$M_t = \frac{\delta\,\delta_t(1-\varepsilon) + \varepsilon^2\,\delta_t(\delta_t + q)}{\delta - \varepsilon(\delta_t + q)} + q(1-\varepsilon). \tag{A19}$$

From (A16), the denominator of (A19) equals $\delta + K_t^h$. Multiplying through this expression by K^h gives

$$K^h\delta + \dot{K}^h, \tag{A20}$$

which equals gross investment in health. Since this is positive in the main model, the denominator of (A19) is always positive.

If the time preference and interest rates differ, and if Y_s is equal to zero, differentiating the logarithmic form of

$$\left(1/\lambda e^{(r-\rho)t}\right)I_s\phi' = \delta(t,X)C', \tag{A21}$$

with respect to time gives

$$(r-\rho)-(1/\varepsilon)\dot{K}_t^h = \tilde{\delta}_t + q. \tag{A22}$$

Education Effects

Education effects on health are derived in the same way as age effects taking into account the fact that C is not dependent on education in the present model. Its effects on medical care are found out by differentiating (A18) with respect to education. This gives

$$\tilde{M}_E = \tilde{K}_E^h + (1/(\tilde{K}_t^h+\delta))\partial \dot{K}_t^h/\partial E + \partial\delta/\partial E, \tag{A23}$$

where $\tilde{M}_E = (dM/dE)(1/M)$, etc.

$$\tilde{K}_E^h = -\varepsilon\tilde{\delta}_E. \tag{A24}$$

Assuming $\tilde{\delta}_t$ is independent of education, and taking into account (A24) and (A16), equation (A25) is obtained:

$$\tilde{M}_E = \frac{\delta\tilde{\delta}_E(1-\varepsilon)+\varepsilon^2\tilde{\delta}_E(\tilde{\delta}_t+q)}{\delta-\varepsilon(\tilde{\delta}_t+q)}. \tag{A25}$$

Wealth Effects

Equation (A9) can be rewritten as

$$(U_s + \lambda Y_s)\phi' = \lambda[\delta(t, X)]C. \tag{A26}$$

Taking logarithms of (A26) and differentiating the result with respect to λ, gives

$$Y_s/(U_s + \lambda Y_s) - (1/\varepsilon)K_\lambda^h = 1/\lambda. \tag{A27}$$

Denoting $(dK^h/d\lambda)(\lambda/K^h)$ by $e_{K^h,\lambda}$ and $U_s/(U_s + \lambda Y_s)$ by m,

$$1 - m - (1/\varepsilon)e_{K^h,\lambda} = 1, \tag{A28}$$

which gives

$$e_{K^h,\lambda} = -\varepsilon m, \tag{A29}$$

Differentiating (A18) with respect to λ, and taking into account that both δ and $\tilde{\delta}_t$ are independent of λ,

$$e_{M,\lambda} = e_{K^h,\lambda}. \tag{A30}$$

where $e_{M,\lambda}$ is the elasticity of medical care with respect to λ.

Appendix 3

1.) Johansen's Procedure:

Johansen (1988) sets his analysis within the following framework. Begin by defining a general polynomial distributed lag model of a vector of variables x as

$$X_t = \pi_1 X_{t-1} + \ldots + \pi_1 X_{t-1} + \varepsilon_t \qquad t = 1,\ldots,T \qquad \text{(A31)}$$

where X_t is a vector of N variables of interest; π_i are NXN coefficient matrices, and ε_t is an independently identically distributed N-dimensional vector with zero mean and covariance matrix Ω. Within this framework the long run, or cointegrating matrix is given by

$$I - \pi_1 - \pi_2 \ldots - \pi_k = \pi \qquad \text{(A32)}$$

where I is the identity matrix.

π will therefore be an NXN matrix. The number, r, of distinct cointegrating vectors, which exists between the variables of X, will be given by the rank of π. In general, if X consists of variables which must be differenced once in order to be stationary [integrated of order one of I(1)] then, at most, r must be equal to N – 1, so that $r \leq N - 1$. Now we define two matrices α, β both of which are N x r such that

$$\pi = \alpha\beta' \qquad \text{(A33)}$$

So the rows of β form the r distinct cointegrating vectors.

Johansen then demonstrates the following theorem.

Theorem: The maximum likelihood estimate of the space spanned by β is the space spanned by the r canonical variates corresponding to the r largest squared canonical correlations between the residuals of X_{t-k} and ΔX_t correlated for the effect of the lagged differences of the X process. The likelihood ratio test statistic for the hypothesis that there are at most r cointegrating vectors is

$$-2\ \ln Q = -T \sum_{i=r+1}^{N} \ln(1 - \hat{\lambda}_i) \tag{A34}$$

where $\hat{\lambda}_{r+1} \ldots \hat{\lambda}_N$ are the (N − r) smallest squared canonical correlations. Johansen then goes on to demonstrate the properties of the maximum likelihood estimates and, more importantly, he shows that the likelihood ratio test has an asymptotic distribution, which is a function of an (N − r) dimensional Brownian motion, which is independent of any nuisance parameters. This means that a set of critical values can be tabulated, which will be correct for all models. He demonstrates that the space spanned by β is consistently estimated by the space spanned by $\hat{\beta}$.

In order to implement this theorem we begin by reparameterising (A31) into the Error Correction Model:

$$\Delta X_t = \Gamma_i \Delta X_{t-1} + \ldots + \Gamma_{k-1} \Delta X_{t-k+1} + \Gamma_K X_{t-k} + \varepsilon_t \tag{A35}$$

where

$$\Gamma_i = -I + \pi_1 + \ldots \pi_i; \qquad I = 1 \ldots k$$

The equilibrium matrix π is now clearly identified as $-\Gamma_k$.

Johansen's suggested procedure begins by regressing ΔX_t on the lagged differences of ΔX_t which yields a set of residuals R_{ot}. We then regress X_{t-k} on the lagged differences ΔX_{t-j}, which yields residuals R_{kt}. The likelihood functions, in terms of α, β and Ω is then proportional to

$$L(\alpha,\beta,\Omega) = |\ \Omega\ |^{-T/2} \exp[-1/2\sum_{t=1}^{T} R_{ot} + \alpha\beta' R_{kt})' \qquad \text{(A36)}$$

$$\Omega^{-1}(R_{ot} + \alpha\beta' R_{kt})]$$

If β were fixed we could maximize over α and Ω by a regression of R_{ot} on $-\beta' R_{kt}$ which gives

$$\hat{\alpha}(\beta) = -S_{ok}\beta(\beta' S_{kk}\beta)^{-1} \qquad \text{(A37)}$$

and

$$\hat{\Omega}(\beta) = S_{oo} - S_{ok}\beta(\beta' S_{kk}\beta)^{-1}\beta' S_{ko} \qquad \text{(A38)}$$

where

$$S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R_{jt}' \qquad \text{i, j} = 0,\text{k}$$

So maximizing the likelihood function may be reduced to minimizing

$$|\ S_{oo} - S_{ok}\beta(\beta' S_{kk}\beta)^{-1}\beta' S_{ko}\ | \qquad \text{(A39)}$$

It may be shown that (A39) will be minimized when

$$|\ \beta' S_{kk}\beta - \beta' S_{ko} S S_{oo}^{-1} S_{ok}\beta\ | / |\ \beta' S_{kk}\beta\ | \qquad \text{(A40)}$$

Attains a minimum with respect to β.

We now define a diagonal matrix D which consists of the ordered eigen values $\lambda_1 > \ldots > \lambda_N$ of $(S_{ko}S_{oo}^{-1}S_{ok})$ with respect to S_{kk}. That is λ_i satisfies

$$| \lambda S_{kk} - S_{ko}S_{oo}^{-1}S_{ok} | = 0 \qquad \text{(A41)}$$

Define E to be the corresponding matrix of eigen vectors so that

$$S_{kk}ED = S_{ko}S_{oo}^{-1}S_{ok}E \qquad \text{(A42)}$$

where we normalize E such that $E'S_{kk}E = I$.

The maximum likelihood estimator of β is now given by the first r rows of E, that is, the first r eigen vectors of $(S_{ko}S_{oo}^{-1}S_{ok})$ with respect to S_{kk}. These are the canonical variates and the corresponding eigen values are the squared canonical correlations of R_{kt} with respect to R_{ot}. These eigen values may then be used in the test proposed in (A34) to test either for the existence of a cointegrating vector r =1 or the number of cointegrating vectors N > r >1.

Johansen (1988) calculates the critical values for the likelihood ratio test for the cases where $m \leq 5$, where m = P – r, and P is the number of variables in the set under consideration and r is the maximum number of cointegrating vectors being tested for.

2.) Unit Root Test Procedure:

Statistic	Test equation	Null hypothesis	Reference/source
ADF $(\bar{\tau}_{\tau})$	$\Delta x_t = \mu + \delta T + \alpha x_{t-1} + \sum_{i=1}^{m} \gamma_i \Delta x_{t-i} + \varepsilon_t$	$\alpha = 0$	Dickey and Fuller (1981)
ADF $(\bar{\tau}_{\mu})$	$\Delta x_t = \mu + xX_{t-1} + \sum_{i=1}^{m} \gamma_i \Delta x_{t-i} + \varepsilon_t$	$\alpha = 0$	Dickey and Fuller (1981)
DP$[t^{*}_{3,n}(3)]$	$\Delta^3 x_t = \mu + \alpha\Delta^2 x_{t-1} + \varepsilon_t$	$\alpha = 0$	Dickey and Pantula (1987)
DP$[t^{*}_{2,n}(3)]$	$\Delta^3 x_t = \mu + \alpha\Delta^2 x_{t-1} + \beta\Delta x_{t-1} + \varepsilon_t$	$\alpha = 0, \beta = 0$	Dickey and Pantula (1987)
DP$[t^{*}_{1,n}(3)]$	$\Delta^3 x_t = \mu + \alpha\Delta^2 x_{t-1} + \beta\Delta x_{t-1} + \gamma x_{t-1} + \varepsilon_t$	$\alpha = 0, \beta = 0, \gamma = 0$	Dickey and Pantula (1987)
$Z(t_{\delta^*})$	$\Delta x_t = \mu^* + \delta^* T + (\alpha^* - 1)x_{t-1} + \varepsilon_t$	$\delta^* = 0$	Perron (1988)
$Z(\Phi_2)$	$\Delta x_t = \mu^* + \delta^* T + (\alpha^* - 1)x_{t-1} + \varepsilon_t$	$\mu^* = \delta^* = 0 and \alpha^* = 1$	Perron (1988)
$Z(\Phi_3)$	$\Delta x_t = \mu^* + \delta^* T + (\alpha^* - 1)x_{t-1} + \varepsilon_t$	$\delta^* = 0 and \alpha^* = 1$	Perron (1988)
$Z(\alpha^*)$	$\Delta x_t = \mu^* + \delta^* T + (\alpha^* - 1)x_{t-1} + \varepsilon_t$	$\alpha^* = 0$	Perron (1988)
$Z(t_{\alpha^*})$	$\Delta x_t = \mu^* + \delta^* T + (\alpha^* - 1)x_{t-1} + \varepsilon_t$	$\alpha^* = 0$	Perron (1988)
$Z(\alpha)$	$\Delta x_t = \mu + (\alpha - 1)x_{t-1} + \varepsilon_t$	$\alpha = 1$	Perron (1988)

$Z(t_\mu)$	$\Delta x_t = \mu + (\alpha - 1)x_{t-1} + \varepsilon_t$	$\alpha = 1$	Perron (1988)
$Z(\Phi_1)$	$\Delta x_t = \mu + (\alpha - 1)x_{t-1} + \varepsilon_t$	$\mu = 0, \alpha = 1$	Perron (1988)
[illegible]	$(x_t - \mu) = \rho(x_{t-1} - \mu) + \varepsilon_t$	$\rho = 1$	Sims (1980)

Notes: All test listed above were conducted, although results of simple Dickey-Fuller DF($\hat{\tau}_\mu$), DF($\hat{\tau}_\tau$), sequential Dickey-Pantula and Sims tests, are not presented in tables appearing in the main text. Results of these tests tended to confirm the findings from other test procedures. All results and regression details from all test equations are available from request.