\((2^{1/2}a,1)\) random variable over its support is 1. Using the change of variable 
\(y = x \left(\frac{\beta J}{I}\right)^{1/2}\), it must be the case that 

\[
\mu(\omega) = (\frac{I}{2\pi J})^{1/2} \int_{-\infty}^{\infty} \exp(-\frac{y^2 I}{2\beta J}) \prod_i \exp((y + \beta h_i) \omega_i) dy / Z_I \tag{A.5}
\]

where

\[
Z_I = (\frac{I}{2\pi J})^{1/2} \int_{-\infty}^{\infty} \exp(-\frac{y^2 I}{2\beta J}) \prod_i M(y + \beta h_i) dy \tag{A.6}
\]

and

\[
M(s) = \exp(s) + \exp(-s) \tag{A.7}
\]

Notice that

\[
\int_{-\infty}^{\infty} \exp(-\frac{y^2 I}{2\beta J}) \prod_i M(y + \beta h_i) dy = \int_{-\infty}^{\infty} \exp(IH_I(y)) dy \tag{A.8}
\]

where

\[
H_I(y) = \frac{1}{I} \sum_i \ln(\exp(v_i(1)) + \exp(v_i(1))) \tag{A.9}
\]

and

\[
v_i(\omega_i) = \beta h_i \omega_i + y \omega_i - \frac{y^2}{2\beta J} \tag{A.10}
\]

Notice that if \(h_i = h \forall i\), then \(H_I(y)\) does not depend on \(I\).

It is shown rigorously by Amaro de Matos and Perez (1991) that as \(I\to\infty\),
integrals of the form

\[
\int_{-\infty}^{\infty} \exp(IH_I(y)) dy \tag{A.11}
\]

101
"pack" all mass onto the global maximizing point

\[ y^* = \text{argmax}_y(H(y)) \]  (A.12)

where

\[ H(y) = \int_{-\infty}^{\infty} \ln(\exp(\beta h + y - \frac{y^2}{2\beta J}) + \exp(-\beta h - y - \frac{y^2}{2\beta J}))dF(h) \]  (A.13)

and \( F(h) \) is the cumulative distribution function of \( h \). One therefore expects (and can prove) that

\[ y_I^* = \text{argmax}_y(H_I(y)) \Rightarrow y^* = \text{argmax}_y(H(y)) \]  (A.14)

Simple algebra reveals that the first order condition for the maximum of \( H(\beta Jm) \) over \( m \) is

\[ m = \int_{-\infty}^{\infty} \tanh(\beta h + \beta Jm)dF(h) \]  (A.15)

When \( h_i = h \ \forall \ i \), this equation also holds for the expected value of each individual \( i \) and hence the sample average by symmetry, which gives us Theorem 4.

Finally, this result suggests that if we replace the integral over \( y \) in eq. (A.5) with a Dirac delta function whose mass is at \( y^* \), we can obtain an approximate probability for the system of the form

\[ \mu(\omega) = \left( \frac{I}{2\pi\beta J} \right)^{1/2} \exp\left( -\frac{y^*^2 I}{2\beta J} \right) \Pi \exp((y^* + \beta h_i)\omega_i)/Z_I \]  (A.16)

where \( Z_I \) is a normalizing constant.

Let $E(\omega_i)$ denote the expectation of $\omega_i$ with respect to the probability measure (A.2). In order to determine the behavior of sample averages as $I \to \infty$, we again consider the case where $h_i = h \; \forall \; i$. Notice that the argument of the previous section implies that

$$\lim_{I \to \infty} E(\omega_i) = \frac{\lim_{I \to \infty} \int_{-\infty}^{\infty} \exp(IH_1(y))G_1(y)dy}{\lim_{I \to \infty} \int_{-\infty}^{\infty} \exp(IH_1(y))dy}$$  \hspace{1cm} (A.17)$$

where

$$H_1(y) = \ln(M(\beta h + y)) - \frac{y^2}{2\beta J}$$  \hspace{1cm} (A.18)$$

and

$$G_1(y) = \frac{\exp(\beta h + y) - \exp(-\beta h - y)}{M(\beta h + y)} = \frac{M'(\beta h + y)}{M(\beta h + y)}$$  \hspace{1cm} (A.19)$$

We employ LaPlace's method (Kac (1968) pg. 248 or Ellis (1985), pp. 38, 50-51) to obtain the limiting values of these integrals. As described above, intuitively, all mass in these integrals gets packed onto the global maximizer $y^*$. We restate the following useful result which is proven in Murray (1984) pg. 34.

**Approximation Theorem.**

Let $H(t)$ be a function on the interval $(a,b)$ which takes a global maximum at a point $\alpha$ in the interval and let $H(t)$ be smooth enough to possess a second-order Taylor expansion at point $\alpha$ with $H''(\alpha) < 0$. Let $G(t)$ denote a continuous
function. Then

\[ \int_{-\infty}^{\infty} G(t) \exp(I H(t)) dt = \]

\[ \exp(I H(\alpha)) G(\alpha) \left( \frac{-2\pi}{IH''(\alpha)} \right)^{1/2} + O(I^{-3/2}) \]  \hspace{1cm} (A.20)

This formula states, in a precise way, the sense in which the mass of the integral piles up at the maximizer \( \alpha \) as \( I \to \infty \). Using this formula, letting \( \alpha = y \),

\[ \frac{\int_{-\infty}^{\infty} \exp(I H_1(y)) G_1(y) dy}{\int_{-\infty}^{\infty} \exp(I H_1(y)) dy} = \]

\[ \frac{\exp(I H_1(\alpha)) G_1(\alpha) \left( \frac{-2\pi}{IH''_1(\alpha)} \right)^{1/2} + O(I^{-3/2})}{\exp(I H_1(\alpha)) \left( \frac{-2\pi}{IH''_1(\alpha)} \right)^{1/2} + O(I^{-3/2})} \]  \hspace{1cm} (A.21)

which is easily seen to converge to \( G_1(\alpha) \) as \( I \to \infty \). Hence we have

\[ \lim_{I \to \infty} E(\omega_i) = G_1(y^*) = m^* = \frac{\exp(\beta h + y^*) - \exp(-\beta h - y^*)}{M(\beta h + y^*)} = \]

\[ \frac{M'(\beta h + y^*)}{M(\beta h + y^*)} = \tanh(\beta h + y^*) = \tanh(\beta h + \beta Jm^*) \]  \hspace{1cm} (A.22)

where \( y^* = \beta Jm^* \).

The problem that \( m^* \) solves appears mysterious at first glance. However, there is an interesting connection between our solution to the behavior of a social planner and the maximization of social surplus as analyzed in McFadden (1981). Following McFadden, social surplus will equal \( \sum (u(\omega_i; X_i) - \frac{1}{2}(\omega_i - \bar{\omega})^2) \). If all agents have common characteristics \( X_i \), then following (A.2), the probability of
the social surplus can be expressed as a function of \( G(\omega) = \sum_i h\omega_i + \frac{1}{2}\beta i (\sum_i \omega_i)^2 \). Then it can be shown (Brock (1993)) that
\[
\beta(\lim_{I \to \infty} E(\max_{\omega} I^{-1}G(\omega))) = \lim_{I \to \infty} (I^{-1} \ln Z_I) = \\
\max_y \ln(\exp(-\frac{y^2}{2\beta I})M(\beta h + y)) = \\
\max_m \ln(\exp(-\frac{(\beta Jm)^2}{2\beta I})M(\beta h + \beta Jm))
\]

(A.23)

As would be expected, one maximizes a notion of social welfare in the large economy limit in order to find the socially optimal states.

Now that the expected value for each choice has been analyzed, we can consider laws of large numbers for data generated in this environment. First, we consider the sample mean, \( \bar{\omega}_I = I^{-1} \sum \omega_i \). Notice that the limiting behavior of the sample mean in distribution (\( \Rightarrow d \)) can be inferred from weak convergence (\( \Rightarrow w \)) since weak convergence necessarily implies convergence in distribution (see Lukacs (1975) pg. 9 for a typical proof.) By Tchebychev’s inequality,
\[
\mu(|\bar{\omega}_I - m^*| \geq \epsilon) \leq \frac{\text{Var}(\bar{\omega}_I - m^*)}{\epsilon^2}
\]

(A.24)

so it is sufficient to prove \( \lim_{I \to \infty} \text{Var}(\bar{\omega}_I - m^*) = 0 \). To do this, it is sufficient to show that \( I^{-2} \sum_i \omega_i \sum_j \omega_j \Rightarrow \text{w} m^{*2} \). However, this can be verified (after considerable algebra) by computing \( I^{-2} \sum_i \omega_i \sum_j \omega_j \) directly and using LaPlace’s method as employed in Murray above to verify that \( I^{-2} \sum_i \omega_i \sum_j \omega_j \Rightarrow \text{w} \tanh(\beta h + y^*)^2 = m^{*2} \). This proves Theorem 12.

iii. Maximum likelihood theory

Consider \( g = 1...G \) distinct neighborhoods with observations \( X_{i, g}, i = 1...I \)
and \( \bar{\omega}_g = I^{-1} \sum_i \omega_{i,g} \) available for each \( g \). Define the likelihood function for the data from these neighborhoods as \( \prod_g \mu(\omega_g) \) where \( \omega_g = (\omega_{1,g}, \ldots, \omega_{I,g}) \). When choices are consistent with the solution to a social planners problem, the likelihood function within each neighborhood will have the form

\[
\mu(\omega_g) \sim \exp \left( \sum_i \left( \frac{1}{2} c' X_{i,g} + \frac{I}{I} (\sum_j \omega_{j,g}) \right) \omega_{i,g} \right)
\]  

(A.25)

which can be rewritten as

\[
\left( \frac{I}{2 \pi \beta J} \right)^{1/2} \int_{-\infty}^{\infty} \exp \left( -\frac{\omega_g^2 I J}{2} \right) \prod_g \exp \left( \frac{1}{2} c' X_{i,g} + J \bar{\omega}_g \right) \omega_{i,g} \, d\omega_g
\]  

(A.26)

Define the parameter vector \( \theta = (c, J) \). One can consider the mean log likelihood over all observations

\[
\frac{1}{GI} \sum_g \ln \mu(\omega_g).
\]  

(A.27)

For large \( I \), and letting \( F(x) = \frac{\exp(x)}{1 + \exp(x)} \) the density of this likelihood will approximately equal

\[
\frac{1}{G} \sum_g \frac{1}{I} \sum_i \left( \frac{1 + \omega_{i,g}}{2} \ln \left( F(c' X_{i,g} + 2J \mu_{I,g}) \right) + \frac{1 - \omega_{i,g}}{2} \ln \left( F(- (c' X_{i,g} + 2J \mu_{I,g})) \right) \right)
\]  

(A.28)

where

\[
\mu_{I,g} = \arg \max \left( H_{I,g}(\mu) \right)
\]  

(A.29)

and

\[
H_{I,g}(\mu) =
\]

106
\[-\frac{\mu^2 J}{2} + \frac{1}{N} \sum_i \ln(\exp(\frac{1}{2}c^i X_{i,g} + J\mu)) + \exp(-\frac{1}{2}c^i X_{i,g} - J\mu) \]  

(A.30)

Note that $H_{I,g}$ converges to

\[
H_g(\mu) =
\]

\[-\frac{\mu^2 J}{2} + \int \ln(\exp(\frac{1}{2}c^i X_{i,g} + J\mu)) + \exp(-\frac{1}{2}c^i X_{i,g} - J\mu))dF_g(X_{i,g}) \]  

(A.31)

so that under regularity conditions such as those described in Newey and McFadden (1994) pg. 2121 it must be the case that

\[
\mu_{l,g} \Rightarrow \omega g =\arg\max(H_G(\mu))
\]  

(A.32)

Notice that the naive estimator introduced in Section 3 inserts $\omega_g$ in place of $\mu_{l,g}$ in the sample likelihood (A.28) above and selects $\theta$ to maximize the modified sample log likelihood function. This means that the naive estimator does not allow the data to directly address the possibility of discontinuous neighborhood responses because the standard maximum likelihood theory in logistic models yields a strictly concave optimization problem. Hence the optimization problem will be continuous in parameters such as the distribution function of individual characteristics, $F_g(X_{i,g})$.

This suggests that one might wish to modify this log likelihood by adding a penalty function of the form

\[
\frac{A}{G} \sum_g (\omega_g - \mu_{l,g})^2
\]  

(A.33)

$A = 0$ will correspond to the naive estimator. Intuitively, as $A$ increases the penalty will push the parameter estimates towards those of the complete estimator, i.e. one which accounts for the relationship between the neighborhood
characteristics and neighborhood mean behavior.

2. Proof of Theorem 5.

For a given parameter set \((k, c, d, J)\), assume by way of contradiction that there exists an alternative \((\bar{k}, \bar{c}, \bar{d}, \bar{J})\) such that on \(\text{supp}(X, Y, m^e)\) we have

\[
(k - \bar{k}) + (c' - \bar{c}')X_i + (d' - \bar{d}')Y_{n(i)} + (J - \bar{J})m^e_{n(i)} = 0
\]

(A.34)

and

\[
m^e_{n(i)} = m_{n(i)} = \int \omega_i dF(\omega_i | k + c'X_i + d'Y_{n(i)} + Jm_{n(i)})dF_X | Y_{n(i)} = \\
\int \omega_i dF(\omega_i | \bar{k} + \bar{c}'X_i + \bar{d}'Y_{n(i)} + \bar{J}m_{n(i)})dF_X | Y_{n(i)}.
\]

(A.35)

Notice the Proposition is true if it is the case that \(J - \bar{J}\) is zero. Otherwise, \(X_i\) and \(Y_{n(i)}\) would lie in a proper linear subspace of \(R^{r+s}\) which violates Assumption \(i\). Equation (A.34) implies that for elements of \(\text{supp}(X, Y, m^e)\), conditional on \(Y_{n(i)}\)

\[
(c' - \bar{c}')X_i = \rho(Y_{n(i)})
\]

(A.36)

where \(\rho(Y_{n(i)}) = -(k - \bar{k}) - (d' - \bar{d}')Y_{n(i)} - (J - \bar{J})m^e_{n(i)}\). (A.36) must hold for all neighborhoods, including \(n_0\) as described in Assumption \(iv\). of the Theorem. This would mean that, conditional on \(Y_{n_0}\), and given that \(X_i\) cannot contain a constant by Assumption \(iii\), that \(X_i\) is contained in a proper linear subspace of \(R^r\) and therefore violates the Assumption \(iv\). of the Proposition. Hence, \(c\) is identified.

Given identification of \(c\), (A.34) now implies, if \(J \neq \bar{J}\), that \(m^e_{n(i)}\) is a linear function of \(Y_{n(i)}\), unless \((d' - \bar{d}')\) and/or \(m^e_{n(i)}\) is always equal to zero. The
latter is ruled out by Assumption vi. Linear dependence of $m^e_{n(i)}$ on $\gamma_{n(i)}$ when $(d' - \bar{d}') \neq 0$ contradicts the combination of the requirement that support of $m^e_{n(i)}$ is $[-1, 1]$ with Assumption v., that the support of each component of $\gamma_{n(i)}$ is unbounded, since $\gamma_{n(i)}$ can, if it is unbounded, assume values with positive probability that violate the bounds on $m^e_{n(i)}$. So, $J$ is identified. If $J$ is identified and $(d' - \bar{d}') \neq 0$, then (A.34) requires that

$$ (d' - \bar{d}') \gamma_{n(i)} = -(k - \bar{k}) \quad \text{(A.37)} $$

for all $\gamma_{n(i)} \in \text{supp}(\gamma_{n(i)})$. This implies, since by Assumption iii. $\gamma_{n(i)}$ does not contain a constant, that $\text{supp}(\gamma_{n(i)})$ is contained in a proper linear subspace of $R^k$, which contradicts condition ii. of the Theorem. Therefore, $d' = \bar{d}'$. This immediately implies that $k = \bar{k}$ and the Theorem is verified.


As is done in the text, $A$ denotes the parameter set $(k, c, d, J)$ and the conditional mean function is $H = k + c \gamma_i + d' \gamma_{n(i)} + J G(m)$. To verify the theorem, it is necessary to show that the components of the gradient vector

$$ d_A H = \frac{\partial H}{\partial A} + \frac{\partial H}{\partial m} \frac{\partial m}{\partial A} \quad \text{(A.38)} $$

define a linearly independent collection of functions of $\gamma_i$ and $\gamma_{n(i)}$ on $\text{supp}(\gamma_i, \gamma_{n(i)})$. Differentiation implies the following, which we will use.

$$ \frac{\partial H}{\partial A} = (1, \gamma_i, \gamma_{n(i)} G(m)) \quad \text{(A.39)} $$

$$ \frac{\partial H}{\partial m} = J(1 + \xi \frac{d G(m)}{m}) \quad \text{(A.40)} $$
Since $J \neq 1$ and $g$ is $C^2$, the neighborhood $N_\epsilon$ can always be chosen so that an implicit function $m(\overline{X}_{n(i)};A,\xi)$ exists. Also, define the function $J(m,\xi) = J(1 + \xi \frac{dg(m)}{m})$.

Rewrite the gradient as

$$d_A H = \frac{1}{1 - J(m,\xi)}(1, X_i + J(m,\xi)(\overline{X}_{n(i)} - X_i), \overline{X}_{n(i)}, m + \xi g(m)) \quad (A.41)$$

If $\xi$ is close enough to zero, $J(m,\xi)$ cannot equal 1 since $J \neq 1$ by Assumption iii. This is a vector proportional to the form $v = (1, v_2(\overline{X}_i, \overline{X}_{n(i)}), v_3(\overline{X}_{n(i)}), v_4(\overline{X}_{n(i)}))$. Notice that we have eliminated $m$ since its implicit function solution makes it a function of $\overline{X}_{n(i)}$. In order to show linear independence, we must verify that

$$a_1 + a_2 v_2(\overline{X}_i, \overline{X}_{n(i)}) + a_3 v_3(\overline{X}_{n(i)}) + a_4 v_4(\overline{X}_{n(i)}) = 0 \quad (A.42)$$

implies that $a_1 = a_2 = a_3 = a_4 = 0$.

Since only $v_2$ depends on $X_i$, (A.42) can only hold if $a_2 = 0$; otherwise Assumption ii. would be violated. Further, if $a_4 = 0$, then Assumption i. is violated. This is true because $v_3(\overline{X}_{n(i)})$ is proportional to $\overline{X}_{n(i)}$. We can therefore, without loss of generality assume $a_4 = -1$.

The condition for linear independence can now be written as

$$m(\overline{X}_{n(i)};A,\xi) + \xi g(m(\overline{X}_{n(i)};A,\xi)) = a_1 + a_3 \overline{X}_{n(i)} \quad (A.43)$$

We pair this with the self-consistency condition written as

$$m(\overline{X}_{n(i)};A,\xi) =
$$

$$k + (c' + d') \overline{X}_{n(i)} + J(m(\overline{X}_{n(i)};A,\xi)) + \xi g(m(\overline{X}_{n(i)};A,\xi)). \quad (A.44)$$

We will verify that (A.43) and (A.44) lead to a contradiction when $\frac{dg}{dm}$ differs
across any two \( m \) values, say \( m_1 \) and \( m_2 \). Since at least two such values must exist by Assumption \( iv \), this will complete the proof.

On the open set \( O \) described by Assumption \( iv \), we can differentiate both these equations with respect to \( \overline{X}_{n(i)} \), obtaining

\[
(1 + \xi \frac{dm(\overline{X}_{n(i)}, A, \xi)}{dm}) \frac{dm(\overline{X}_{n(i)}, A, \xi)}{d\overline{X}_{n(i)}} = a_3 \tag{A.45}
\]

and

\[
\frac{dm(\overline{X}_{n(i)}, A, \xi)}{d\overline{X}_{n(i)}} \left( 1 - J(1 + \xi \frac{dg(m(\overline{X}_{n(i)}, A, \xi))}{dm}) \right) = (c' + d') \tag{A.46}
\]

Equating \( \frac{dm(\overline{X}_{n(i)}, A, \xi)}{d\overline{X}_{n(i)}} \) across these expressions yields

\[
(1 - J(1 + \xi \frac{dg(m(\overline{X}_{n(i)}, A, \xi))}{dm}))a_3 - (c + d)(1 + \xi \frac{dg(m(\overline{X}_{n(i)}, A, \xi))}{dm}) = 0 \tag{A.47}
\]

or

\[
\xi \frac{dg(m(\overline{X}_{n(i)}, A, \xi))}{dm} \left( (c + d) + Ja_3 \right) + ((c + d) + Ja_3 - a_3) = 0 \tag{A.48}
\]

Recall that \( \xi \) and \( \frac{dg}{dm} \) are scalars, whereas \( c, d, \) and \( a_3 \) are \( r \times 1 \) vectors. By construction of \( g \), we have the existence of two values of \( m \), call them \( m_1 \) and \( m_2 \) such that in the population data, \( \frac{dg}{dm} \) differs across them. Applying this component by component to (A.47), one can show that this implies that \( (c + d) + Ja_3 = 0 \). By (A.48), this means that \( a_3 = 0 \). But from (A.49), this would imply that

\[
m(\overline{X}_{n(i)}, A, \xi) + \xi g(m(\overline{X}_{n(i)}, A, \xi)) = a_1. \tag{A.49}
\]
But this would contradict the part of Assumption iv. that $m_{n(i)}$ in the data is nonconstant. Therefore, the model and assumptions described by the Theorem require that the components of the gradient (A.38) are linearly independent when $\xi \neq 0$. Notice that when $\xi = 0$, the gradient will not be of full rank, because $m(\overline{X}_{n(i)}, A, 0)$ is linear in $\overline{X}_{n(i)}$. Hence the local nonidentification of the linear-in-means model can be perturbed away by a $C^2$-small change from $Jm$ to $Jm + \xi g(m)$, which completes the proof.
Bibliography


113


Bénabou, R., (1996a), "Equity and Efficiency in Human Capital Investment:


116


Coleman, J., E. Campbell, J. Hobson, J. McPartland, A. Mood, F. Weinfeld, and


Labov, W., (1972a), Sociolinguistic Patterns, Philadelphia, University of Pennsylvania Press.

Labov, W., (1972b), Language in the Inner City, Philadelphia, University of Pennsylvania Press.

Opportunities in Baltimore: Early Evidence," working paper, Joint Center for Poverty Research, Northwestern University.

Lancaster, T., (1990), The Analysis of Transition Data, New York: Cambridge University Press.


127


Spitzer, F., (1971b), Random Fields and Interacting Particle Systems, Providence: Mathematical Association of America.


