Madansky Method Based on Replacement Functions

Extract from:

Efficiency Units, Elementary Hedonic Models (Gorman and Lancaster) With and Without Bundling Restrictions

James J. Heckman
University of Chicago

Econ 312, Spring 2023

- How to estimate the skill prices across sectors when there are unobserved skill prices?
- How to test equality of skill prices across sectors?
- Unobserved traits may be correlated with observed traits

$$Y_{in} = \underset{\sim}{w}_{no} \underbrace{\underset{\sim}{x}_{io}}_{\text{observed}} + \{\underset{\sim}{w}_{nu} \underbrace{\underset{\sim}{x}_{iu}}_{\text{unobserved}} + \varepsilon_{in}\}, \tag{1}$$

$$i = 1, \ldots, I, \quad n = 1, \ldots, N.$$

- Allow for unobserved skills.
- Skills are assumed constant over time for the individual.
- Suppose that persons stay in one sector and we have T time periods of panel data on those persons.
- Stack these into a vector of length T.
- Let κ_u be the number of unobserved components.
- Let κ_o be the number of observed components.

In matrix form we may write these equations for person i as

$$\underset{\sim}{Y}_i = \underset{\sim}{w}_o \underset{\sim}{x}_{io} + \{\underset{\sim}{w}_u \underset{\sim}{x}_{iu} + \underset{\sim}{\varepsilon}_i\}, \quad \text{for each sector } n \tag{2}$$

(Drop the n subscript for each sector.)

THE UNIVERSITY OF CHICAGO

Following Madansky (1964), Chamberlain (1977) and Pudney (1982), assume $T \geq 2\kappa_u + 1$ and partition (2) into three subsystems:

- We can write a system down for each $n = 1, \ldots, N$.
- Assume for simplicity $\underset{\sim}{x}_{io}$ and $\underset{\sim}{x}_{iu}$ are **time invariant**.

$$\begin{array}{ll} \underset{\sim}{w}_o \, (T \times J_0) & J_0 \text{ is the number of observed variables} \\ \underset{\sim}{w}_u \, (T \times J_1) & J_u \text{ is the number of unobserved variables} \\ \underset{\sim}{x}_{io} \, (J_0 \times 1) & \underset{\sim}{x}_{iu} \text{ is } J_u \times 1 \end{array}$$

- The time invariance of $\underset{\sim}{x}_{iu}$ is essential (at least for a subset).
- Time invariance of $\underset{\sim}{x}_{io}$ is easily relaxed (notationally burdensome).

(i) A basis subsystem of κ_u equations from (2)

$$\underset{\sim}{Y}_{(1)} = \underset{\sim}{w}_{o(1)}\underset{\sim}{x}_{io} + \{\underset{\sim}{w}_{u(1)}\underset{\sim}{x}_{iu} + \underset{\sim}{\varepsilon}_{(1)}\}, \quad n = 1, \ldots, N \tag{3a}$$

$\underset{\sim}{w}_{u(1)}$ is $\kappa_u \times \kappa_u$

(ii) A second subsystem of equations all of which are distinct from the equations used in (i)

$$\underset{\sim}{Y}_{(2)} = \underset{\sim}{w}_{o(2)}\underset{\sim}{x}_{io} + \{\underset{\sim}{w}_{u(2)}\underset{\sim}{x}_{iu} + \underset{\sim}{\varepsilon}_{(2)}\}, \quad n = 1, \ldots, N \tag{3b}$$

(iii) The rest of the equations (at least κ_u in number)

$$\underset{\sim}{Y}_{(3)} = \underset{\sim}{w}_{o(3)}\underset{\sim}{x}_{io} + \{\underset{\sim}{w}_{u(3)}\underset{\sim}{x}_{iu} + \underset{\sim}{\varepsilon}_{(3)}\}. \tag{3c}$$

Assuming that $\underset{\sim}{w}_{u(1)}$ is of full rank, the first system of equations may be solved for $\underset{\sim}{x}_{iu}$, i.e.,

$$\underset{\sim}{x}_{iu} = \underset{\sim}{w}_{u(1)}^{-1}[\underset{\sim}{Y}_{(1)} - \underset{\sim}{w}_{o(1)}\underset{\sim}{x}_{io} - \underset{\sim}{\varepsilon}_{(1)}]. \tag{4}$$

Substituting (4) into (3b), we reach

$$\underset{\sim}{Y}_{(2)} = \underset{\sim}{x}_{io}\left[\underset{\sim}{w}_{o(2)} - \underset{\sim}{w}_{u(1)}^{-1}\underset{\sim}{w}_{u(2)}\underset{\sim}{w}_{o(1)}\right] \\ + \underset{\sim}{w}_{u(1)}^{-1}\underset{\sim}{w}_{u(2)}\underset{\sim}{Y}_{(1)} + \underbrace{\underset{\sim}{\varepsilon}_{(2)} - \underset{\sim}{w}_{u(1)}^{-1}\underset{\sim}{w}_{u(2)}\underset{\sim}{\varepsilon}_{(1)}}_{\text{unobserved error term}}.$$

- Gets rid of $\underset{\sim}{x}_{iu}$.
- But OLS fails because, by construction, $\underset{\sim}{\varepsilon}_{(1)}$ is correlated with $\underset{\sim}{Y}_{(1)}$.

Internal Instruments

- However, we have an internal instrument
- Use IV to instrument for $Y_{(1)}$. The natural instruments are $Y_{(3)}$. They are valid as long as $\underset{\sim}{w}_{u(3)}$ are nonzero and the rank condition is satisfied.
- Find a lot of evidence against equality of factor prices across sectors.

Simple Example ($\mathbf{J_u = 1}$)

- $X_i^0(1)$: observed variable for i in the first period
- $X_i^u(1)$: unobserved in first period (dimension=1)
- $\varepsilon(j)$: a period j specific shock uncorrelated with $X^u(l), X^0(l)1$, and $\varepsilon(l)$; $l \neq j$.

$$
(*) \quad \begin{aligned}
Y_i(1) &= \beta_1 X_i^0(1) + \lambda_1 X_i^u(1) + \varepsilon_i(1) \\
Y_i(2) &= \beta_2 X_i^0(2) + \lambda_2 X_i^u(1) + \varepsilon_i(2) \\
Y_i(3) &= \beta_3 X_i^0(3) + \lambda_3 X_i^u(1) + \varepsilon_i(3)
\end{aligned}
$$

a. β_j is price of observed skills in period j; X_j is price of unobserved skill
b. Remember: $\varepsilon(j)$ mutually independent, mean zero
c. $X_i^{(0)}(j) \not\perp\!\!\!\perp X_i^{(u)}(l)$; all j, l (omitted variable bias)
d. Assume $X_i^u(1) = X_i^u(2) = X_i^u(3)$
e. λ_j, β_j and $X_i^0(j)$ can change with j

- $\varepsilon(l) \perp\!\!\!\perp \varepsilon(k) \quad \forall l \neq k$
- Steps:
  - Step 1: Use equation for $Y_i(1)$ to solve for $X_i^u(1)$

$$\frac{Y_i(1) - \beta_1 X_i^0(1) - \varepsilon_i(1)}{\lambda_1} = X_i^u(1)$$

  - Assumes $\lambda_1 \neq 0$ (price of unobserved skill in period 1)
  - Step 2: Substitute in the second equation for $Y_i(2)$

$$Y_i(2) = \beta_2 X_i^0(1) + \frac{\lambda_2}{\lambda_1}(Y_i(1) - \beta_1 X_i^0(1) - \varepsilon_i(1))) + \varepsilon_i(2)$$

- Collect terms

$$* \quad Y_i(2) = (\beta_2 - \frac{\lambda_2}{\lambda_1}\beta_1) X_i^0(1) + \frac{\lambda_2}{\lambda_1} Y_i(1) + \varepsilon_i(2) - \frac{\lambda_2}{\lambda_1}\varepsilon_i(1)$$

- $X_i^u(2) = X_i^u(1)$ eliminated; $\therefore$ omitted variable eliminated
- From first equation: $Y_i(1) \not\perp\!\!\!\perp \varepsilon_i$ (out of the frying pan and into the fire)
- Step 3: $Y_i(3)$ is an instrument for $Y_i(1)$ in equation $(*)$
- Why? (Depends on $X_i^u(1)$ as does $Y_i(1)$)
- $\varepsilon_i(3) \perp\!\!\!\perp (\varepsilon_i(2) - \lambda_2\varepsilon_i(1))$
- Conclusion: $\therefore$ we get $(\beta_2 - \frac{\lambda_2}{\lambda_1}\beta_1)$ and $\frac{\lambda_2}{\lambda_1}$

- Switching the roles of 1, 2, and 3, we can get $\frac{\lambda_j}{\lambda_k}; j \neq k$
- All assumed to be non-zero
- Notice we need one normalization to separate λ_j from X_i^u (both unobserved)
- Set $\lambda_1 = 1$, $\therefore$ we know λ_2, λ_3
- This normalization is essential: we do not directly observe $X_i^u(i), X_i^u(2)$ or $X_i^u(3)$ or the λ.
- They enter the wage equation as $\left[\lambda_1 X_i^u(1)\right], \left[\lambda_2 X_i^u(2)\right], \left[\lambda_3 X_i^u(3)\right]$.

$$\left.\begin{aligned}\beta_3 - \lambda_3\beta_1 &= \phi_{31}\\ \beta_3 - \lambda_3\beta_2 &= \phi_{32}\\ \beta_1 - \lambda_1\beta_2 &= \phi_{12}\\ \beta_1 - \lambda_1\beta_3 &= \phi_{13}\\ \beta_2 - \lambda_2\beta_1 &= \phi_{21}\\ \beta_2 - \lambda_2\beta_3 &= \phi_{23}\end{aligned}\right\} \phi_{l,k} \text{ all known}^1$$

- [1]But not necessarily the individual parameters on the left hand side (except λ_j)
- From previous analysis, the ϕ_{ij} all known as are λ_j
- 3 equations; 3 unknowns
- $\therefore \beta_1, \beta_2, \beta_3$ known (rank condition requires "sufficient" variation in prices of skills)
- Everything identified (prices of observed and unobserved skills) up to normalization.

TABLE I

(*Basis described in the appendix*)

(1) Sector	(2) System MSE	(3) Test	(4) F(DFN, DFD) =	(5) Prob > F	(6) Number of observations in each year
Durable vs. Nondurable	3·208210	1	(117, 1143) = 1·1448	0·1491	153
		2	(90, 1143) = 0·9213	0·6840	
		3	(27, 1143) = 1·7777	0·0087	
Manufacturing vs. Service	3·447400	1	(117, 3411) = 1·6754	0·0001	405
		2	(90, 3411) = 0·7336	0·9717	
		3	(27, 3411) = 3·0062	0·0001	
Blue vs. White Collar	2·600956	1	(156, 6648) = 2·4197	0·0006	580
		2	(120, 6648) = 1·2943	0·0176	
		3	(36, 6648) = 3·0714	0·0001	
North vs. South	2·299067	1	(156, 7056) = 1·9586	0·0001	614
		2	(120, 7056) = 1·4981	0·0007	
		3	(36, 7056) = 3·0844	0·0008	
Manufacturing vs. Non-mfg	4·746601	1	(117, 5787) = 1·4411	0·0015	669
		2	(90, 5787) = 1·1062	0·2323	
		3	(27, 5787) = 3·0978	0·0001	

Notes.

1. Test 1 tests equality of the coefficients of (12) in both sectors.
   Test 2 tests equality of the coefficients associated with observed characteristics in (12).
   Test 3 tests equality of the coefficients associated with the unobserved characteristics in (12) ($w_{u(1)}^{-1}$, $w_{u(2)}$).

Notes.

1. Test 1 tests equality of the coefficients of (12) in both sectors.
   Test 2 tests equality of the coefficients associated with observed characteristics in (12).
   Test 3 tests equality of the coefficients associated with the unobserved characteristics in (12) $(w_{u(1)}^{-1}, w_{u(2)})$.
2. Durable: Metal Industries, Machinery including Electrical, Motor Vehicles and other Transportation Equipment, other durables.
   Non Durable: Food, Tobacco, Textile, Paper, Chemical and other Non Durables.
   Manufacturing: All Durable and Non Durable plus "manufacturing unknown".
   Services: Retail Trade, Wholesale Trade, Finance, Insurance, Real Estate, Repair Service, Business Service, Personal Service, Amusement, Recreation and Related Services, Printing, Publishing and Allied Services, Medical and Dental Services, Educational Services, Professional and Related Services.
   North: Conn., Del., Ill., Ind., Maine, Mass., Mich., Minn., N.H., N.J., N.Y., Ohio, Penn., R.I., W. Va., Wis., Vermont.
   South: Alab., Ark., Fla., Geo., Ky., La., Miss., N.C., S.C., Tenn., Tex., Va., Ok.
   White Collar: Professional, Technical and Kindred; Managers, Officials and Proprietors; Self Employed Businessmen; Clerical and Sales Work.
   Blue Collar: Craftsmen, Foremen and Kindred Workers; Operatives and Kindred Workers; Labourers and Service Workers, Farm Labourers.

TABLE II

4 *Factor models*

(1) Sector	(2) System MSE	(3) Test	(4) F(DFN, DFD) =	(5) Prob > F	(6) Number of observations in each year
Durable vs. Nondurable	1·480446	1	(144, 1089) = 1·2902	0·0166	153
		2	(108, 1089) = 1·1722	1·1197	
		3	(36, 1089) = 1·3644	0·0756	
Manufacturing vs. Service	1·271277	1	(144, 3357) = 2·6513	0·0001	405
		2	(108, 3357) = 1·2957	0·0231	
		3	(36, 3357) = 6·6334	0·0001	
Blue vs. White Collar	3·830300	1	(192, 6576) = 1·7228	0·0001	580
		2	(144, 6576) = 1·3400	0·0045	
		3	(48, 6576) = 1·8698	0·0003	
North vs. South	2·456318	1	(192, 6984) = 1·9893	0·0001	614
		2	(144, 6984) = 0·8240	0·9381	
		3	(48, 1836) = 2·3018	0·0001	
Manufacturing vs. Non-mfg.	1·617166	1	(180, 1836) = 1·7121	0·0001	669
		2	(132, 1836) = 1·4107	0·0020	
		3	(48, 1836) = 2·0701	0·0001	

TABLE III

5 *Factor models*

(1) Sector	(2) System MSE	(3) Test	(4) F(DFN, DFD) =	(5) Prob > F	(6) Number of observations in each year
Blue vs. White Collar	1·573852	1	(228, 6912) = 2·0534	0·0001	580
		2	(168, 6912) = 1·6639	0·0001	
		3	(60, 6912) = 3·8733	0·0001	
North vs. South	1·418750	1	(228, 6504) = 3·8840	0·0001	614
		2	(168, 6504) = 2·2027	0·0001	
		3	(60, 6504) = 10·0017	0·0001	

APPENDIX

For the 3 factor models we adopt the following basis:

Years for wages ($Y_{(2)}$)	Basis years
1968, 1969, 1970	1971, 1972, 1973
1971, 1972, 1973	1968, 1969, 1970
1974, 1975, 1976	1971, 1972, 1973
1977, 1978, 1979	1974, 1975, 1976

For the 4 factor models we adopt the following choice of basis:

Years for wages ($Y_{(2)}$)	Basis years
1968, 1969, 1970, 1971	1972, 1973, 1974, 1975
1972, 1973, 1974, 1975	1968, 1969, 1970, 1971
1976, 1977, 1978, 1979	1972, 1973, 1974, 1975

For the 5 factor models we adopt the following choice of basis:

Years for wages ($Y_{(2)}$)	Basis years
1968, 1969, 1970, 1971, 1972	1973, 1974, 1975, 1976, 1977
1973, 1974, 1975, 1976, 1977	1968, 1969, 1970, 1971, 1972
1978, 1979	1968, 1969, 1970, 1971, 1972