A Database of Mincerian Earnings Regressions

Lutz Hendricks
Iowa State University
Department of Economics
May 21, 2004

Abstract
This article provides documentation for the Database of Mincerian Earnings Regressions.

1 Introduction
This document provides documentation for the Database of Mincerian Earnings Regressions.

Additional collections of Mincerian earnings regressions are:

- Psacharopoulos (1994)

2 User Notes

2.1 Computing annual earnings
Earnings regressions are estimated for different frequencies. It is often desirable to convert these into annual earnings estimates (there are obvious caveats).

Assume that annual earnings take the form

\[ Y = \exp \left[ b_0 + b_1 s + b_2 x + b_3 x^2 \right] \]

where \( s \) denotes years of schooling and \( x \) is years of experience.

Studies estimate regressions of the form

\[ \ln \left( \frac{e^Y}{\alpha N} \right) = a_0 + a_1 s + a_2 x + a_3 x^2 \]

where
the time period is $1/N$ years (e.g.: monthly earnings: $N = 12$).

$e$ is the exchange rate that converts local currency earnings into dollar earnings (in local currency units per dollar).

$\alpha$ is a scale factor. For example, if the regression is estimated in terms of thousands of local currency units, then $\alpha = 1,000$.

Therefore,

$$ Y = \alpha N / e \exp \{ a_0 + a_1 s + a_2 x + a_3 x^2 \} $$

$$ = \exp \{ \tilde{a}_0 + a_1 s + a_2 x + a_3 x^2 \} $$

where

$$ \tilde{a}_0 = a_0 + \ln \alpha + \ln N - \ln e $$

It follows that the unit conversions only affect the intercept ($b_0$).

To obtain the PPP adjusted predicted earnings divide by the price level of GDP (e.g., PWT 5.6 variable #3 “P”).

2.2 Problems and caveats

The quadratic functional form tends to fit rather poorly (Murphy and Welch 1990).

Earnings regressions predict (approximately) median earnings, not mean earnings because they compute $\text{mean} (\ln(Y))$ and the distribution of $Y$ is approximately log-normal.

2.2.1 High inflation countries

Latin/South American countries often have high inflation and rapidly depreciating exchange rates. If individuals were surveyed at random over a long period of time, this adds an iid noise term to earnings (later observations have much higher earnings than earlier observations). This makes exchange rates meaningless and it becomes impossible to compute intercepts or mean earnings.

3 Notes on Individual Countries

3.1 Argentina

Psacharopoulos and Ng (1992): For 1980 implied average earnings are only around 10 percent of per capita GDP.

3.2 Australia

Lorenz and Wagner’s (1990): 1981 age-earnings profiles are implausible.

3.3 Austria

Lorenz and Wagner’s (1990): 1987 age-earnings profiles are implausible.

3.4 Bolivia

Exchange rates are fairly constant around November 1989.
3.5 Botswana
Lucas and Stark (1985), table A2: Since average years of schooling are reported only for both sexes combined, I assume they are the same for men and women. The sample year is reported as 1978-79; I assign this to 1979. Earnings may be overstated because of an additional control variable of unknown mean ("1/(1+years in town)").

3.6 Brazil
Psacharopoulos/Arriagada (1989) do not report earnings regressions, but mean earnings by groups (T. B1).

Additional references:
Lam and Schoeni (1993), JPE.

3.7 Canada

3.8 Chile
The exchange rate seems unambiguous. Psacharopoulos and Ng (1992, table 11) implies almost the same value as PWT. Fairly constant exchange rate around November 1989.

3.9 China

3.10 Colombia

3.11 Costa Rica
Exchange rate fairly constant during 1989.
Exchange rate fluctuates over 1981.

3.12 Denmark
Rosholm and Smith (1996): for salaried workers (including unskilled workers would not make a big difference because their earnings structure is similar to that of salaried workers and they constitute only 20 percent of the population). Intercepts do not precisely represent the average worker because of several dummy variables (with small coefficients).
3.13 Dominican Republic
Exchange rate fixed in 1989 at 6.34.

3.14 Ecuador
The data are for November 1989.

3.15 El Salvador
Funkhouser (1992): Table 5.5 report earnings regressions, although not in standard format.

3.16 Greece
No comparable estimates are available for Greece. Lambpropoulos and Psacharopoulos (1992) do not have individuals with less than secondary education and use education dummies, not “years of schooling.”

3.17 Guatemala

3.18 Hong-Kong
Chung (1996) has data to construct earnings profiles, but not a standard regression. Detailed male/female differentials

3.19 India
Psacharopoulos (1989) has additional references.
Rao and Datta’s (1989) are not representative of the population (data are for one firm only and are censored at earnings of $2750 per year).

3.20 Ireland
Callan/Reilly (1993): earnings regressions with a lot of additional variables and no “years of schooling”

3.21 Italy
Lorenz and Wagner’s (1990): implied age-earnings profiles are implausible.

3.22 Jamaica
Estimates for Jamaica from Psacharopoulos and Ng (1992) have implausible slope coefficients. For example, the coefficient on schooling exceeds 30 and hours are not a significant determinant of earnings.
3.23 Korea
Psacharopoulos (1985) contains some references.

3.24 Malaysia
Chapman/Harding: very high average schooling because only ex-students are included in sample.

3.25 Mexico

3.26 Nicaragua

3.27 Norway
Hayfron (1998) for married men. Age is used instead of experience.

3.28 Paraguay

3.29 Peru
Psacharopoulos and Ng (1992, annex 3) report average hours for males of 792 in 1985. This cannot be correct since women work 1564 hours and average hours in the entire sample are 1854. I therefore set male hours to average private sector hours of 1914.

3.30 Philippines

3.31 Portugal
No comparable estimates are available for Portugal after 1977 (Psacharopoulos 1981). Only the estimates for men are used because the earnings equation for women uses firm experience instead of worker experience. Kiker and Santos (1991) include too many control variable to reliably extract returns to education and experience.

3.32 Puerto Rico
Ramos (1992):
Sample pools men and women. I use estimates for the “never migrated” category which is the majority of the sample (table 2.6, column 2). The intercept is adjusted for the coefficient on the “never migrated” dummy.
Hourly wage, males 20-64, in PR: $4.14 / in US: $5.11 (or $6.25 for those who migrated before)
This understates the age/education adjusted differential b/c 0.4 years more schooling in PR; 6 years older in PR
3.33 Singapore
Liu and Wong (1981) use actual instead of potential experience in their earnings regressions (table 3).

3.34 Spain
Lassibille (1998), table 3. The sample consists of private sector workers. From the intercept I subtract the dummies for “married” and the selection bias term (li). The results of Alba-Ramirez and San Segundo (1995) cannot be used because it is not clear what variables are included in the regressions (years of schooling or degrees attained). The regressions also include additional control variables (such as hours worked).

3.35 Thailand
Average years of schooling are imputed based on Chiswick’s (1976) table 1 where I assume that on average workers with “more than primary” education have 12 years of schooling.

3.36 Turkey
Tansel (1994): reports cross-tab of mean earnings by (age/education) classes.

3.37 Uruguay
Psacharopoulos (1994): Correct intercept is $b_0 + 0.551 \ln(hours/week)$. 

T. 2.3, 2.4
T 2.6-2.7: even has earnings regressions for both locations
Ortiz (1986):
no earnings data
educational composition
4 References

4.1 Collections of Mincer Regressions

References


4.2 Econometric Issues and Methods

References


4.3 Data for Individual Countries

References


4.4 Immigrant Earnings

References