Earnings and Languages in the Family: Second-Generation Australians

(Economic Record, vol. 85, Special Issue, S59-S73)

George Messinis
Centre for Strategic Economic Studies
28 October 2009

Motivation

- First-generation (OB): Persons born O/S
- Second-generation (2G): Native-born with at least one parent born O/S
- Third-generation (3G): Native-born and both parents native-born

- How does 2G perform in the labour market?
- The international evidence suggests
  - Intergenerational transmission of disadvantage in 2G
  - Language is a key component of human capital
- What drives the 2G disadvantage?
  - An official language deficit?
  - Lack of labour market socialisation?
  - Cultural barriers?
  - Discrimination?
- Are languages other than English (LOTE) valuable in Australia?
Methodology

- DATA: HILDA, Wave 7 (cross-section)

- We account for undereducation, overeducation & severe overskilling

- 1st Generation of Australians (1G) effects:
  - English-speaking (ESOB) and Non-English-speaking (NESOB)
  - Skilled and Unskilled
  - LOTE

- 2nd Generation (2G) effects:
  - Ancestry/Cultural effects
  - Linguistic background (i.e., LOTE)
  - Linguistic diversity in the family (i.e., LOTE and English-speaking)

- We use Quantile regressions to minimise selection bias

The Model

- An expanded Mincer equation (Mavromaras et al. 2007)

\[
\ln W_i = \alpha T_i + \beta VET_i + \gamma Y12_i + \sum_{k=1}^{3} \phi_k M_k + X_i \varphi + \eta_i
\]  

- \( \ln W \) = log of weekly Wages (adjusted for time NILF)
- \( T \) = Tertiary is highest education level
- \( VET \) = Cert III, IV or Diploma or Apprenticeship
- \( Y12 \) = completed Year 12
- \( M_k \) = mismatch indicator k
- \( X \) = vector of covariates
- \( \eta \) = random error term
PROBLEM: Non-Random Sample Selection

- NOTE: Pervasive in cross-section data
- CAUSES: unobservable factors that correlate with dependent variable
- RESULT: omitted variables bias → biased (inaccurate) coeff. estimates
- EXAMPLES:
  - Earnings (missing: those not-in-employment or not in training)
  - Health care (missing: those not using medicines)
  - Innovation (missing: firms that do not apply for patents)
  - Energy (missing: countries that do not record energy use)
PROBLEM: Non-Random Sample Selection

SIMULATION: Population → No relation by design

MISSING: workers (firms) that do not have a wage (file for patents)

PROBLEM: non-selection or missing is systematic
PROBLEM: Non-Random Sample Selection

SELECTED: a non-random sample of workers (firms) that participate in employment/training (patents)

![Graph showing the relationship between work experience/training/R&D spending and weekly wage/patent number with a regression equation y = 38x + 416.]

SIMULATION: Observed → Positive relation

![Graph showing the same relationship with a regression equation y = 38x + 416 and labeled Δx and Δy.]
PROBLEM: Non-Random Sample Selection

GROUPS: Is sub-group in RED different to others?

BIASED RESULT: group in RED seems to have a stronger relation.
Accounting for Sample Selection Bias

ESTIMATION TECHNIQUES:
1) Heckman (1979): models jointly (a) employment participation & (b) wages conditional on (a)
2) “Differences in Difference” estimators
3) Copulae functions
4) Quantile regressions

LIMITATIONS:
- (1) assumes bivariate normal distribution & hard to identify (a) & (b)
- (2) assumes group differences do NOT change over time
- (3) hard to identify the right distribution for each equation
- (4) assumes unobservable variable changes little within a quartile

SOLUTION: Quantile Regressions

IDEA: Slice distribution in quartiles & expect omitted variable (eg. ability) to change little within the quantile
## RESULTS: Quantile Regression (Table 3, part A)

### Wages and Cultural Diversity: Full-time Workers, 2007

<table>
<thead>
<tr>
<th>Variables</th>
<th>MALES</th>
<th>FEMALES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education: Tertiary</td>
<td>0.136**</td>
<td>0.281**</td>
</tr>
<tr>
<td>Education: VET</td>
<td>0.453**</td>
<td>0.599**</td>
</tr>
<tr>
<td>Education: Year 12</td>
<td>0.186**</td>
<td>0.292**</td>
</tr>
<tr>
<td>Undereducation</td>
<td>0.176**</td>
<td>0.221**</td>
</tr>
<tr>
<td>Overeducation</td>
<td>0.079**</td>
<td>0.096**</td>
</tr>
<tr>
<td>Severe overskilling</td>
<td>-0.046</td>
<td>-0.103**</td>
</tr>
<tr>
<td>Work experience/100</td>
<td>0.150</td>
<td>0.170</td>
</tr>
<tr>
<td>History: Unemployed or NILF</td>
<td>-0.450**</td>
<td>-0.116**</td>
</tr>
<tr>
<td>Job tenure</td>
<td>0.005**</td>
<td>0.004**</td>
</tr>
<tr>
<td>Union member</td>
<td>0.109**</td>
<td>0.077**</td>
</tr>
<tr>
<td>Married</td>
<td>0.092**</td>
<td>0.118**</td>
</tr>
<tr>
<td>Public sector employment</td>
<td>0.019</td>
<td>-0.022</td>
</tr>
<tr>
<td>Indigenous person</td>
<td>-0.013</td>
<td>-0.034</td>
</tr>
</tbody>
</table>

## RESULTS: Quantile Regression (Table 3, part B)

<table>
<thead>
<tr>
<th>Variables</th>
<th>MALES</th>
<th>FEMALES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education: Year 12</td>
<td>0.150</td>
<td>0.170</td>
</tr>
<tr>
<td>Job tenure</td>
<td>0.005**</td>
<td>0.004**</td>
</tr>
<tr>
<td>Union member</td>
<td>0.109**</td>
<td>0.077**</td>
</tr>
<tr>
<td>Married</td>
<td>0.092**</td>
<td>0.118**</td>
</tr>
<tr>
<td>Public sector employment</td>
<td>0.019</td>
<td>-0.022</td>
</tr>
<tr>
<td>Indigenous person</td>
<td>-0.013</td>
<td>-0.034</td>
</tr>
</tbody>
</table>
**RESULTS: Quantile Regression (Table 4, part B)**

<table>
<thead>
<tr>
<th>Wages and Linguistic Diversity in Family: Full-time Male Workers, 2007</th>
<th>(1) Q25</th>
<th>(2) Q50</th>
<th>(3) Q75</th>
<th>(4) Q25</th>
<th>(5) Q50</th>
<th>(6) Q75</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESOB: Skilled</td>
<td>-0.023</td>
<td>-0.017</td>
<td>0.015</td>
<td>-0.028</td>
<td>-0.019</td>
<td>0.021</td>
</tr>
<tr>
<td>ESOB: Other</td>
<td>0.134**</td>
<td>0.124**</td>
<td>-0.021</td>
<td>0.140**</td>
<td>0.129**</td>
<td>-0.006</td>
</tr>
<tr>
<td>NESOB: Skilled</td>
<td>0.053</td>
<td>-0.062</td>
<td>-0.105</td>
<td>0.047</td>
<td>-0.055</td>
<td>-0.108</td>
</tr>
<tr>
<td>NESOB: Other</td>
<td>0.059</td>
<td>0.047</td>
<td>-0.081</td>
<td>0.064</td>
<td>0.040</td>
<td>-0.108</td>
</tr>
<tr>
<td>2G: English Speaking (ES) Background</td>
<td>-0.092**</td>
<td>-0.084**</td>
<td>-0.098*</td>
<td>-0.079**</td>
<td>-0.075**</td>
<td>-0.087*</td>
</tr>
<tr>
<td>2G: NES North Europe (native-born)</td>
<td>-0.049</td>
<td>-0.008</td>
<td>-0.029</td>
<td>-0.029</td>
<td>0.010</td>
<td>0.007</td>
</tr>
<tr>
<td>2G: NES East Europe (native-born)</td>
<td>0.193**</td>
<td>0.085</td>
<td>-0.016</td>
<td>0.200**</td>
<td>0.120</td>
<td>0.067</td>
</tr>
<tr>
<td>2G: NES South Europe (native-born)</td>
<td>-0.041</td>
<td>-0.206**</td>
<td>-0.022</td>
<td>-0.030</td>
<td>-0.139**</td>
<td>0.058</td>
</tr>
<tr>
<td>2G: NES Other</td>
<td>0.094</td>
<td>-0.019</td>
<td>-0.082</td>
<td>0.088</td>
<td>-0.007</td>
<td>-0.017</td>
</tr>
<tr>
<td>LOTE: 1G (NESOB only)</td>
<td>-0.109*</td>
<td>-0.083</td>
<td>-0.048</td>
<td>-0.106*</td>
<td>-0.089</td>
<td>-0.039</td>
</tr>
<tr>
<td>LOTE: 2G NES Background (native-born)</td>
<td>0.007</td>
<td>0.042</td>
<td>-0.143</td>
<td>0.248**</td>
<td>-0.069</td>
<td>-0.341**</td>
</tr>
<tr>
<td>2G: Spouse is from ES Background</td>
<td>0.068</td>
<td>0.072*</td>
<td>0.105*</td>
<td>0.324**</td>
<td>0.135</td>
<td>0.239*</td>
</tr>
</tbody>
</table>

*Note: * and ** denote 5% and 5% level of significance respectively.

---

**Summary of results**

- Little evidence of a skill premium for 1G (ESOB or NESOB)
- 2G of NES background is not at a disadvantage
- 2G of ES background seems disadvantaged (puzzling)
  - Low SES background?
- LOTE attracts a wage penalty for both 1G & 2G
- A spouse with an ES background helps ameliorate the LOTE disadvantage for 2G → English language deficit?
- Results cast doubt on the hypotheses of (i) cultural barriers or (ii) discrimination as drivers of 2G disadvantage