A Testing for Moral Hazard
Survey Versus Claims Data

D Rowell
Dr HS Nghiem
Prof. LB Connelly

The University of Queensland

Financial support was provided by
Australian Centre for Economic Research on Health and
Brian Gray Scholarship awarded in 2010 by the APRA and RBA
Introduction

- *Ex ante* moral hazard refers to the tendency of insurance to reduce an individual’s motive to prevent loss and therefore increase the probability of a claim.

- The aim is to present an empiric test for moral hazard using an instrumental variables approach.

- To implement this method, we show why survey data is preferable to claims data.
Introduction

1. Empirical challenge

2. What is wrong with claims data?
   • Testing for Asymmetric information in Insurance Markets
     (Chiappori & Salanié 2000)

3. What is right with survey data?
   • Multiple Dimensions of Private Information: Evidence from Long-term Care Insurance Market
     (Finkelstein & McGarry 2006)

4. Present a test for moral hazard using an Instrumental Variable (IV)
Empirical challenge

• Asymmetric information

\[ \text{Claim} = f(\text{Insurance} \mid X) \]

• Two types of
  i. Moral hazard – *hidden action*

\[ \text{Claim} = \alpha_0 + \alpha_1 \text{Insurance} + \alpha_2 X + \varepsilon_i \]

  ii. Adverse Selection – *hidden information*

\[ \text{Insurance} = \beta_0 + \beta_1 \text{Claim} + \beta_2 X + \eta_i \]
Anatomy of Endogeneity

Generic data

- Correlation
- Causation
- Omitted variable bias

Insurance setting

- Asymmetric information
- Moral hazard
- e.g., Adverse selection
Chiappori & Salanié (2000)

- Test for asymmetric information using French claims data

- \( \text{Insurance} = \alpha_0 + \alpha_1 X + \epsilon_i \)

- \( \text{Claim} = \beta_0 + \beta_1 X + \eta_i \)

- \( X = 55 \) variables insurer’s info. set

  - Gender (1)
  - Type of use (3)
  - Profession (7)
  - Make of car (7)
  - Area (4)
  - Age of car (11)
  - Age of driver (8)
  - Region (9)
  - Performance of car (5)

- Null hypothesis in a sub-sample of young drivers

  - \( H_0 : \text{cov}(\epsilon_i, \eta_i) \)

  - \( H_0 : \rho = 0 \)
Chiappori & Salanié (2000)

• Test for moral hazard using a ‘natural experiment’
  
  – Some young drivers ‘inherit’ their parent’s *bonus malus* discount if car ‘is’ jointly owned
  – *IBM* = 1 if young driver ‘inherits’ their parent’s maximum *bonus malus* discount

• Re-estimated bivariate probit model

\[
\begin{align*}
\text{Insurance} &= \alpha_0 + \alpha_1 \mathbf{X} + \alpha_2 \text{IBM} + \varepsilon_i \\
\text{Claim} &= \beta_0 + \beta_1 \mathbf{X} + \beta_2 \text{IBM} + \eta_i
\end{align*}
\]
Chiappori & Salanié (2000)

**Interpretation of $\beta_2$?**

1. Pure familial correlation hypothesis: Low-risk parents produce low-risk children

2. Pure Moral Hazard hypothesis: The risk rating of premiums means that the cost of each successive claim is increasing. Preventative effort is increased with higher premiums

**Three Options**

A. Familial correlation & no moral hazard
   - $IBM$ will be negatively correlated with RTC

B. No familial correlation & No moral hazard
   - $IBM$ will not be correlated with RTC

C. No familial correlation & moral hazard
   - $IBM$ will be positively correlated with RTC

$\beta_2$ is negative and significant, which rejects the moral hazard hypothesis’ for the alternative hypothesis ‘that the parents’ performances are positively correlated with the child’s.
Following Chiappori and Salanié (2000) the empirical literature has avoided analysis with cross sectional data

- Abbring (2003) use longitudinal data to test for moral hazard using survival analysis


- Cohen (2005) test for asymmetric information using conditional correlation but state that a test for moral hazard is beyond the scope of the paper.

- Israel (2007) use longitudinal data to test for moral hazard using a difference in difference approach
### What’s wrong with claims data?

- **Chiappori and Salanié (2000)**
  - Moral hazard and a familial correlation could both exist
  - There are therefore 4 not 3 possibilities!

<table>
<thead>
<tr>
<th>Nil Moral Hazard</th>
<th>Moral Hazard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nil Correlation</td>
<td>+ve correlation</td>
</tr>
<tr>
<td>+ve correlation</td>
<td>Ambiguous</td>
</tr>
<tr>
<td>-ve correlation</td>
<td></td>
</tr>
<tr>
<td>Nil Familial Correlation</td>
<td>Nil Correlation</td>
</tr>
</tbody>
</table>

- **Chiappori and Salanié (2000)**
  - Moral hazard and a familial correlation could both exist
  - There are therefore 4 not 3 possibilities!
What’s wrong with claims data?

- **A Natural experiment**
  - occurs when some (often unintended) feature of the setup we are studying produces exogenous variation in an otherwise endogenous explanatory variable” (Wooldridge 2002)

- **But the IBM is**
  - Observable to insurer
  - Used by insurers to risk rate some young policyholders
  - Correlated with driving ability

- **Therefore**
  - $IBM$ should be in $X$
  - $X$ should comprise 56 exogenous variables not 55
  - Analysis of claims data are invariably constrained in this way
What’s right with survey data?

Finkelstein and McGarry (2006)

• **Aim**
  - To investigate asymmetric information in a market for long-term care insurance (LTCINS)

• **Data**
  - Health and Retirement Study (HRS) uses survey data!
    - Nursing home placement (NHP)
    - LTCINS
    - Insurer’s information set $X$
    - Contained private information not observable to the insurer
      - Prob. nursing home placement
      - Preference for insurance
What’s right with survey data?

Method

1. Provide evidence of private information by estimating 2 probit models

\[ NHP = \alpha_o + \alpha_1 X + \alpha_2 \text{Prob}_\text{NHP} + \varepsilon_i \]
\[ LTCINS = \beta_o + \beta_1 X + \beta_2 \text{Prob}_\text{NHP} + \eta_i \]

2. Traditional tests for asymmetric information fail to identify this private information

1. Bivariate probit model: \( \rho = 0 \)

\[ NHP = \alpha_o + \alpha_1 X + \epsilon_i \]
\[ LTCINS = \beta_o + \beta_1 X + \eta_i \]

2. Probit model: \( \alpha_2 = 0 \)

\[ NHP = \alpha_o + \alpha_1 X + \alpha_2 LTCINS + \epsilon_i \]
What is right with survey data?

3. Hypothesise that other dimensions of private information e.g. preference for insurance \((PI^*)\)

4. Re-estimate the probit models using seat-belt \((SB)\) use as a proxy for \(PI^*\)

\[
\begin{align*}
NHP &= \alpha_0 + \alpha_1 X + \alpha_2 \text{Prob}_- NHP + \alpha_3 SB + \varepsilon_i \\
LTCINS &= \beta_0 + \beta_1 X + \beta_2 \text{Prob}_- NHP + \beta_3 SB + \eta_i
\end{align*}
\]

“...the existence of multiple forms of private information in an insurance market and demonstrate how these factors can have offsetting effects on the correlation between insurance coverage and risk occurrence, thus invalidating the standard test of asymmetric information.”
A Test for Moral Hazard

• Structural Model

\[ NHP = \beta_0 + X\beta_1 + \beta_2LTCINS + PI^* + RT^* + \mu_i \]

• Empirical test for Moral Hazard
  – Bivariate probit with one instrumental variable and proxy

\[ NHP = \beta_0 + \beta_1X + \beta_2LTCINS_R + \beta_3SB + \epsilon_i \]
\[ LTCINS_R = \alpha_0 + \alpha_1X + \alpha_2LTCINS_S + \beta_3SB + \eta_i \]
Selection of Instrumental Variable

Insurance status of spouse (LTCINS_S) selected as IV for LTCINS_R

1. LTCINS_S correlated with LTCINS_R?
   - Becker (1981) has argued that within couples matching occurs on many criteria.
   - Specifically intra-spousal correlation occurs in
     • Risk aversion (Spivey 2010)
     • Health status (Wilson 2002)
     • Health maintenance activities (Macken, Yates et al. 2000)

2. LTCINS_S is not correlated with error term
   - Once living arrangements are controlled for the respondent’s decision to enter a nursing home will not be correlated with spouse’s insurance status
## Results from Biprobit model

<table>
<thead>
<tr>
<th></th>
<th>NHP(1995-2000)</th>
<th></th>
<th>Coefficient</th>
<th>p-value</th>
<th>Coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondent's long-term care insurance (1995)</td>
<td></td>
<td>n.a.</td>
<td>-0.176</td>
<td>0.43</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Spouse's long-term care insurance (1995)</td>
<td>n.a.</td>
<td>n.a.</td>
<td>1.942</td>
<td>&lt;0.01</td>
<td>0.145</td>
<td>0.08</td>
</tr>
<tr>
<td>Wears seatbelt</td>
<td>0.009</td>
<td>0.89</td>
<td>1.942</td>
<td>&lt;0.01</td>
<td>0.145</td>
<td>0.08</td>
</tr>
<tr>
<td>Female</td>
<td>-2.745</td>
<td>&lt;0.01</td>
<td>0.360</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coupled household</td>
<td>-1.435</td>
<td>0.01</td>
<td>-0.367</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spouse's age (=0 if unmarried)</td>
<td>0.017</td>
<td>0.01</td>
<td>-0.003</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd quartile -Income</td>
<td>0.045</td>
<td>0.55</td>
<td>0.191</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd quartile -Income</td>
<td>0.032</td>
<td>0.71</td>
<td>0.312</td>
<td>&lt;0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th quartile -Income</td>
<td>0.028</td>
<td>0.79</td>
<td>0.519</td>
<td>&lt;0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd quartile -Assets</td>
<td>0.013</td>
<td>0.86</td>
<td>0.253</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd quartile -Assets</td>
<td>0.036</td>
<td>0.67</td>
<td>0.363</td>
<td>&lt;0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th quartile -Assets</td>
<td>0.055</td>
<td>0.56</td>
<td>0.310</td>
<td>&lt;0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADL limitation: walking</td>
<td>-0.205</td>
<td>0.21</td>
<td>-0.132</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADL limitation: dressing</td>
<td>-0.142</td>
<td>0.33</td>
<td>-0.347</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADL limitation: bathing</td>
<td>0.144</td>
<td>0.36</td>
<td>0.111</td>
<td>0.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADL limitation: eating</td>
<td>-0.138</td>
<td>0.45</td>
<td>0.169</td>
<td>0.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADL limitation: toileting</td>
<td>-0.148</td>
<td>0.36</td>
<td>-0.078</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IADL limitation: medication</td>
<td>0.091</td>
<td>0.60</td>
<td>0.151</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IADL limitation: shopping</td>
<td>0.155</td>
<td>0.14</td>
<td>0.027</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use wheelchair</td>
<td>-0.094</td>
<td>0.58</td>
<td>0.264</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use walker</td>
<td>0.265</td>
<td>0.02</td>
<td>-0.180</td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use crutch</td>
<td>-0.016</td>
<td>0.97</td>
<td>-4.514</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use oxygen</td>
<td>0.012</td>
<td>0.97</td>
<td>-5.768</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use cane</td>
<td>0.103</td>
<td>0.17</td>
<td>0.024</td>
<td>0.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI &gt; 30</td>
<td>0.108</td>
<td>0.14</td>
<td>-0.179</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI &lt; 20</td>
<td>0.037</td>
<td>0.68</td>
<td>0.051</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incontinent</td>
<td>-0.054</td>
<td>0.39</td>
<td>0.067</td>
<td>0.34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results…continued

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>p-value</td>
</tr>
<tr>
<td>Takes medication</td>
<td>0.067</td>
</tr>
<tr>
<td>Smokes</td>
<td>0.000</td>
</tr>
<tr>
<td>Depressed</td>
<td>0.135</td>
</tr>
<tr>
<td>ETOH</td>
<td>0.044</td>
</tr>
<tr>
<td>Cognitively impaired</td>
<td>-0.356</td>
</tr>
<tr>
<td>Diabetic</td>
<td>0.022</td>
</tr>
<tr>
<td>Insulin dependent diabetic</td>
<td>0.102</td>
</tr>
<tr>
<td>Diabetic with renal Dx.</td>
<td>0.124</td>
</tr>
<tr>
<td>CVA</td>
<td>0.222</td>
</tr>
<tr>
<td>Heart Dx.</td>
<td>0.101</td>
</tr>
<tr>
<td>Heart Rx.</td>
<td>-0.196</td>
</tr>
<tr>
<td>AMI</td>
<td>0.063</td>
</tr>
<tr>
<td>CCF</td>
<td>0.242</td>
</tr>
<tr>
<td>Hypertensive</td>
<td>-0.013</td>
</tr>
<tr>
<td>Hip replacement</td>
<td>0.206</td>
</tr>
<tr>
<td>Lung Dx.</td>
<td>-0.066</td>
</tr>
<tr>
<td>Cancer</td>
<td>0.012</td>
</tr>
<tr>
<td>Psychiatric Dx.</td>
<td>0.149</td>
</tr>
<tr>
<td>Arthritis</td>
<td>-0.071</td>
</tr>
<tr>
<td>Any days in nursing home (pre 1995)</td>
<td>0.469</td>
</tr>
<tr>
<td>Receives home help</td>
<td>0.129</td>
</tr>
<tr>
<td>Injury from fall</td>
<td>0.101</td>
</tr>
<tr>
<td># ADL's by age</td>
<td>0.003</td>
</tr>
<tr>
<td># IADL's by age</td>
<td>-0.001</td>
</tr>
<tr>
<td>Cognitive impairment by age</td>
<td>0.015</td>
</tr>
<tr>
<td>Sex by age</td>
<td>0.036</td>
</tr>
<tr>
<td>Sex by cognitive impairment</td>
<td>3.399</td>
</tr>
<tr>
<td>Sex by # ADL's</td>
<td>1.201</td>
</tr>
<tr>
<td>Sex by # IADL's</td>
<td>-0.330</td>
</tr>
<tr>
<td>Sex by cognitive impairment by age</td>
<td>-0.054</td>
</tr>
<tr>
<td>Sex by # ADL's by age</td>
<td>-0.017</td>
</tr>
<tr>
<td>Sex by # IADL's by age</td>
<td>0.005</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.701</td>
</tr>
</tbody>
</table>
### Results from Biprobit model

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>p-value</td>
<td>Coeff.</td>
<td>p-value</td>
</tr>
<tr>
<td>LTCINS_R (1995)</td>
<td>-0.176</td>
<td>0.43</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>LTCINS_S (1995)</td>
<td>n.a.</td>
<td>n.a.</td>
<td>1.942</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Wears seatbelt</td>
<td>0.009</td>
<td>0.89</td>
<td>0.145</td>
<td>0.08</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.701</td>
<td>&lt;0.01</td>
<td>-1.924</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>
The Instrumental Variable

1. \( \text{LTCINS}_S \notin X \)

2. Correlated with endogenous variable \( \text{LTCINS}_R \)
   i. Coefficient for \( \text{LTCINS}_S = 1.942 \) (\( p \)-value < 0.01)
   ii. Test for weak instruments \( R^2 \) increased from 7.9% to 22.8%

\[
\text{LTCINS}_R = \alpha_0 + \alpha_1 X + \alpha_2 \text{LTCINS}_S + \beta_3 \text{SB} + \eta_i
\]

3. IV should be uncorrelated with \( \varepsilon_i \)
   - This assertion is usually taken on faith since \( RT^* \) is a component of \( \varepsilon_i \) which is unobservable to the econometrician
   - But the HRS contains \( \text{Prob}_NHP \) which is used as a proxy for \( RT^* \)

\[
\text{NHP} = \beta_0 + \beta_1 X + \beta_2 \text{LTCINS}_R + \beta_3 \text{SB} + \beta_4 \text{Prob}_NHP + \varepsilon_i
\]

\[
\text{LTCINS}_R = \alpha_0 + \alpha_1 X + \alpha_2 \text{LTCINS}_S + \alpha_3 \text{SB} + \alpha_4 \text{Prob}_NHP + \eta_i
\]
### Quality of the Instrumental Variable?

-0.176 ($p$-value = 0.43)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondent's LTC insurance (1995)</td>
<td>-0.209</td>
<td>0.37</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Spouse's LTC insurance (1995)</td>
<td>n.a.</td>
<td>n.a.</td>
<td>1.937</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Wears seatbelt</td>
<td>-0.030</td>
<td>0.69</td>
<td>0.181</td>
<td>0.04</td>
</tr>
<tr>
<td>Predicted probability of NHP</td>
<td>0.268</td>
<td>0.01</td>
<td>0.487</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.734</td>
<td>&lt; 0.01</td>
<td>-2.008</td>
<td>&lt; 0.01</td>
</tr>
</tbody>
</table>
Conclusions

1. **Empirical result**: No *ex ante* moral hazard in insurance market for long-term care
   
   i. Plausible?
   
   ii. Moral hazard in this market may be *ex post*

2. **Method**: Suggest testing for moral hazard with an IV

   1. Claims data are the wrong data because they can not discriminate between the different dimensions of private information
   
   2. The test for moral hazard used by Chiappori & Salanié (2000) was compromised because of the use of claims data
   
   3. The analysis of survey data as outlined by Finkelstein and McGarry (2006) offers a new and promising line of investigation for econometricians who are looking for moral hazard.


