Analysis on Panel Data

Margrét Vala Gylfadóttir

Increased availability of gender pay gap statistics IV workshop
Tallinn, 31 October 2014

The Context

- In 2008 the "Confederation of employers" and the "Icelandic Confederation of Labour" asked Statistics Iceland to perform a study on wage differentials or the adjusted GPG
  - Based on the dataset from the ISWEL survey
  - Data for the private sector 2000-2007
  - Published early 2010

Opportunities

- Big data set with repeated measurements over time
  - Unbalanced panel data/longitudinal data
    - Repeated measurements over time
      - Data is not present for all individuals at all times
  - Detailed information in the dataset
    - Many variables present
  - Possibility to add variables from other datasets in Statistics Iceland

Obstacles

- Data on education was not comprehensive
  - A lot of work was put into to make existing data compatible
  - Some extra data collection was made
  - Foundation for a database on education that was further extended for the 2011 census
- Data on length of service in enterprise was not very reliable
  - Data from tax declarations and "pay as you earn register" was used to enhance the information

Variables

Individual

year
id
sex
age
non-national citizenship at some point
invalid
living in capital area

Education

education (ISCED97)
student in registers

Family

marital status
number of children undir the age of 1
number of children between 1 and 2
number of children 2-5
number of children 6-9
number of children 10-16

Occupation

member of a union
occupation (3-digit ISCO)
occupational group (1-digit ISCO)
full time
length of service in enterprise
supervisor/management
craft/trade worker
monthly earnings
fixed wage contract

Enterprise/business unit

enterprise/business unit
economic activity
economic acitivity (2-digit NACE)
size
sector
enterprise in captial area

Earnings and working hours

regular earnings
paid total hours

Methodology

- Cross sectional data does not take into account individual effects

  – Not possible to take in account unmeasured variables

- Regression analysis on panel data (like the ISWEL data) gives opportunities to address individual effects

norway grants

Statistics Iceland

Methodology

- Four different models – with the same explanatory variables
  - Cross sectional "ordinary least squares" model on data from 2007 – wage differential of 9.9%
  - "Pooled ordinary least squares" model on data from 2000-2007 – wage differential of 9.2%
    - This model is not suitable as there are individual effects in the data set (Breusch and Pagan Lagrange test)

Methodology

- Generalized least square model (GLS) – with random effects on data 2000-2007 – wage differential of 10.3%
  - Takes into account individual effects
    - This model is not suitable as there exists a correlation between the individual effects and explanatory variables (Hausman test)

norway grants

Statistics Iceland

Methodology

- Generalized least square model (GLS) – with fixed effect should be a proper model
  - Does not give a measurement on variables that are constant over time
    - Sex is constant over time (usually – and if not you are left out of the dataset)
    - thus the wage differential can not be calculated
  - Gives a measure on the effect of different explanatory variables on earnings for men and women

Results

- ## Difference in effect of explanatory variables

Variable		Men		Women	
		Effect	St.dev.	Effect	St.dev.
Upper secondary school					
	RE	4.2	0.21	2.3	0.2
	FE	0.2	0.2	-0.8	0.2
University degree					
	RE	11.3	0.4	9.2	0.3
	FE	4.5	0.4	3.1	0.3
Martial status					
	RE	3.1	0.2	0.7	0.2
	FE	0.6	0.21	0.4	0.19

Results

- Oaxaca-Blinder decomposition
  - Neumark's method, 1988
    - GLS-RE model

Year	Explained	Unexplained	Total
2000 - 2003	21.54	6.0	27.6
2001 - 2004	19.82	6.3	26.1
2002 - 2005	17.07	6.6	23.7
2003 - 2006	14.46	7.2	21.7
2004 - 2007	12.21	7.6	19.8
All years	15.6	7.3	22.9

Ongoing Research

- Following the "Plan of Action on Gender Equality" an agreement was made with Statistics Iceland to repeat this study

- First study in Iceland covering extended part of the labour market

- Data on private and public sector 2008-2013
  - Based on the GLS-RE model
    - With same explanatory variables as last time
    - With fewer explanatory variables
      - The partners in the public sector decided on the following variables: Economic activity, Occupation, Working hours, Education, Age, Length of service
      - This decision is not without criticism

norway grants

Statistics Iceland

Results

- First results indicate that the GPG is decreasing over time
  - Higher GPG in the private sector than in the public sector
  - Some variables have different effect in the private and the public sector

For further research

- Many ways to continue
  - Instrumental variable model (Housman and Taylor) takes into account correlation between the individual effects and explanatory variables
    - Lack of time and resources
    - There is an interest to develop this model
- Study certain groups over time and the effect of different explanatory variables
  - Age groups
  - Occupational groups with mainly one sex
  - Other characteristics – one or more

Conclusions

- The study from 2010 has had influence on the discussion in general
  - Focus on changes over time rather than a number at a given point
  - What is the question?
    - Discrimination in workplaces
      - Unequal pay for same job? Illegal
    - Unequal pay for equal worth jobs
      - What are equal worth jobs?
    - Distribution of wages between men and women in general
      - What is acceptable?

Conclusions

- – Different effect of explanatory variables on the sexes
  - Why is that?
  - Are there changes over time?
- – Explanatory variables
  - Should every variable known to influence earnings be used?
  - Is there a set of objective/factual variables that is acceptable to have influence on earnings?
    - – If yes – who should decide?

Conclusions

– Should this fact be a part of the adjusted GPG?

  • GLS-FE model

Variable	Men		Women	
	Effect	St.dev.	Effect	St.dev.
Number of children under the age of 1	0.82	0.20	-1.64	0.21
Number of children between 1 and 2	1.14	0.21	-0.60	0.23
Number of children 2-5	0.88	0.15	-1.67	0.17
Number of children 6-9	0.70	0.15	-1.07	0.17
Number of children 10-16	0.42	0.13	-0.60	0.15

Conclusions
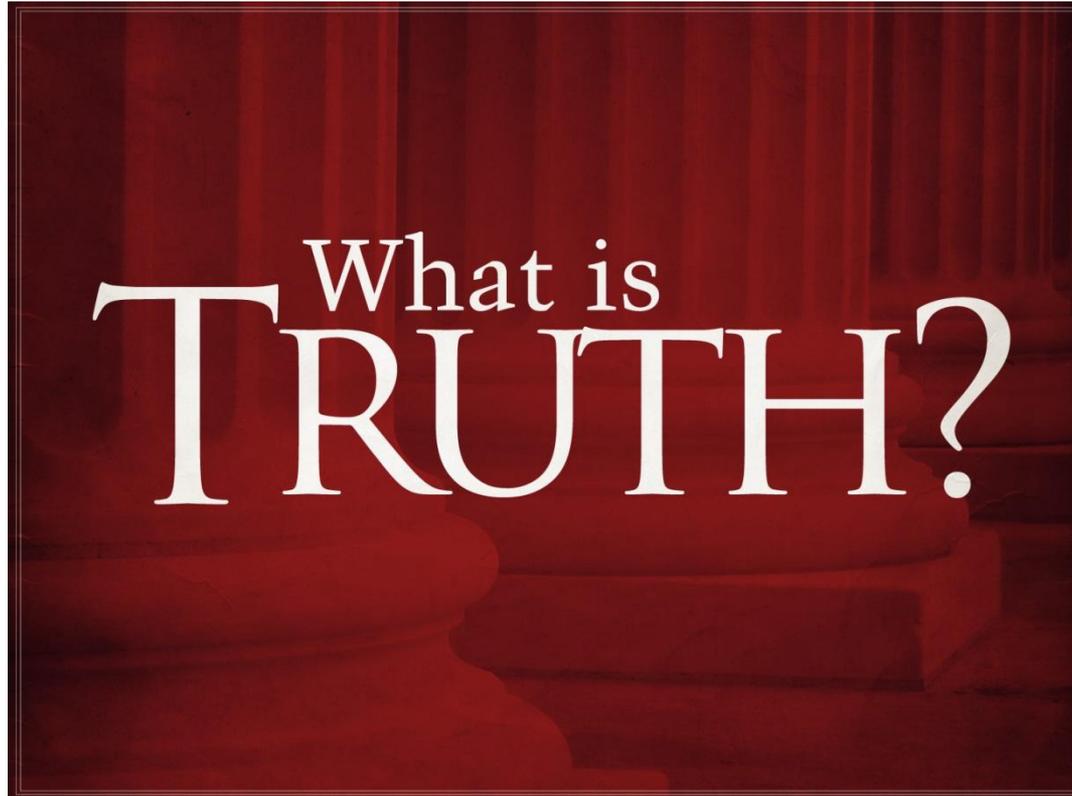
– There is no true number
  - Data availability, methodology and other factors influence the results

- Is the solution to stop asking like my four year old?
  - Is this a man or a woman?
- And start asking what are the individual characteristics and interests
  - Giving every individual a possibility to develop it potentials