REach FOR GOLD

CANADIAN REINSURANCE CONFERENCE / CONGRÈS CANADIEN DE RÉASSURANCE

THE 59th ANNUAL CANADIAN REINSURANCE CONFERENCE

Jean-Yves Rioux

Big Data and Analytics' dramatic impacts in the Life Insurance Industry

Agenda

- Drivers and barriers
- Data sources
- The process
- Life and health insurance applications
- Modeling/analyzing longevity

Famous quotes

"Not everything that can be counted counts, and not everything that counts can be counted." – Albert Einstein

"If you can't explain it simply, you don't understand it well enough."
– Albert Einstein

"Those who do not remember the past are condemned to repeat it."
– George Santayana

"The price of light is less than the cost of darkness." – Arthur C. Nielsen

"War is ninety percent information." – Napoleon Bonaparte

"Facts do not cease to exist because they are ignored." – Aldous Huxley

The environment

The interest in predictive modeling has increased substantially in recent years.

Drivers	Barriers
Decrease in cost of storage	Quality of data
Increase in computer processing speed/memory	Rudimentary or basic IT infrastructure
Proliferation of data from third parties	Inadequate size of or expertise from analytics staff
Increase interest in learning more with regard to own data	Lack of executive advocacy

Internal data sources

There is increased interest in understanding, restructuring and using own data.

Representative data categories

Policyholder info
- Age
- Gender
- Smoking status
- Amount inforce
- Province of residence
- Occupational code

Reinsurance info
- Premium Rates
- Allowances
- Coverage

Past/current claims
- Date incurred
- Date reported
- Amount
- Cause

Product design
- Surrender rights/charges
- Vesting
- Market value adj.
- Renewal rights
- Rollover/reset rights
- Conversion rights

Experience results
- Incidence
- Lapse
- Termination
- Mortality
- Persistency

Asset info
- Coupon rates
- Notional amounts
- Market values

External data sources

Companies who are succeeding in advanced analytic analysis are doing so by their commitment to exploring new data.

Data providers

Acxiom
Agriculture and Agri-Food Canada
AM Best
AWCBC
Bank of Canada
Bloomberg
BDC
Canada Hospital Directory
CMHC
Canada Revenue Agency
Canadian Cancer Society
CIHI
CMA
Cap Index
Citizenship and Immigration Canada
Dun & Bradstreet
HRSDC
Environics
Equifax
Experian
GHDx
Industry Canada
Ipsos
LifeGuide
Natural Resources Canada
Office of the Superintendent of Bankruptcy Canada
OECD
PAHO
Public Health Agency of Canada
Public Safety Canada
Statistics Canada
Undata
UIS
World Bank
World Values Survey

Representative data categories

Economic
Real Estate
Equities
Commodities
Interest Rates
Foreign Exchange
Inflation
Economic/Bus. trends

Demographic
Age
Gender
Ethnicity
Income
Immigration Data

Financial
Credit Score
Gross/Total Debt Service Ratio
Credit Ratings

National indices
Wage Data
Wealth/Net Worth
Unemployment Stats
Aggregate CRA Data

DB, LB and health
Death
Diabetes
Cancer
Cardiovascular
Disability
Injury
Depression/Mental

Medical & drugs
Disability Data
Hospital Directory
Nursing Home Data
Hospital Visit Statistics
Prescription Drug Usage
Physician Data

Competitive data
Premium Rates
Crediting Rates
Guaranteed Rates
Product Features

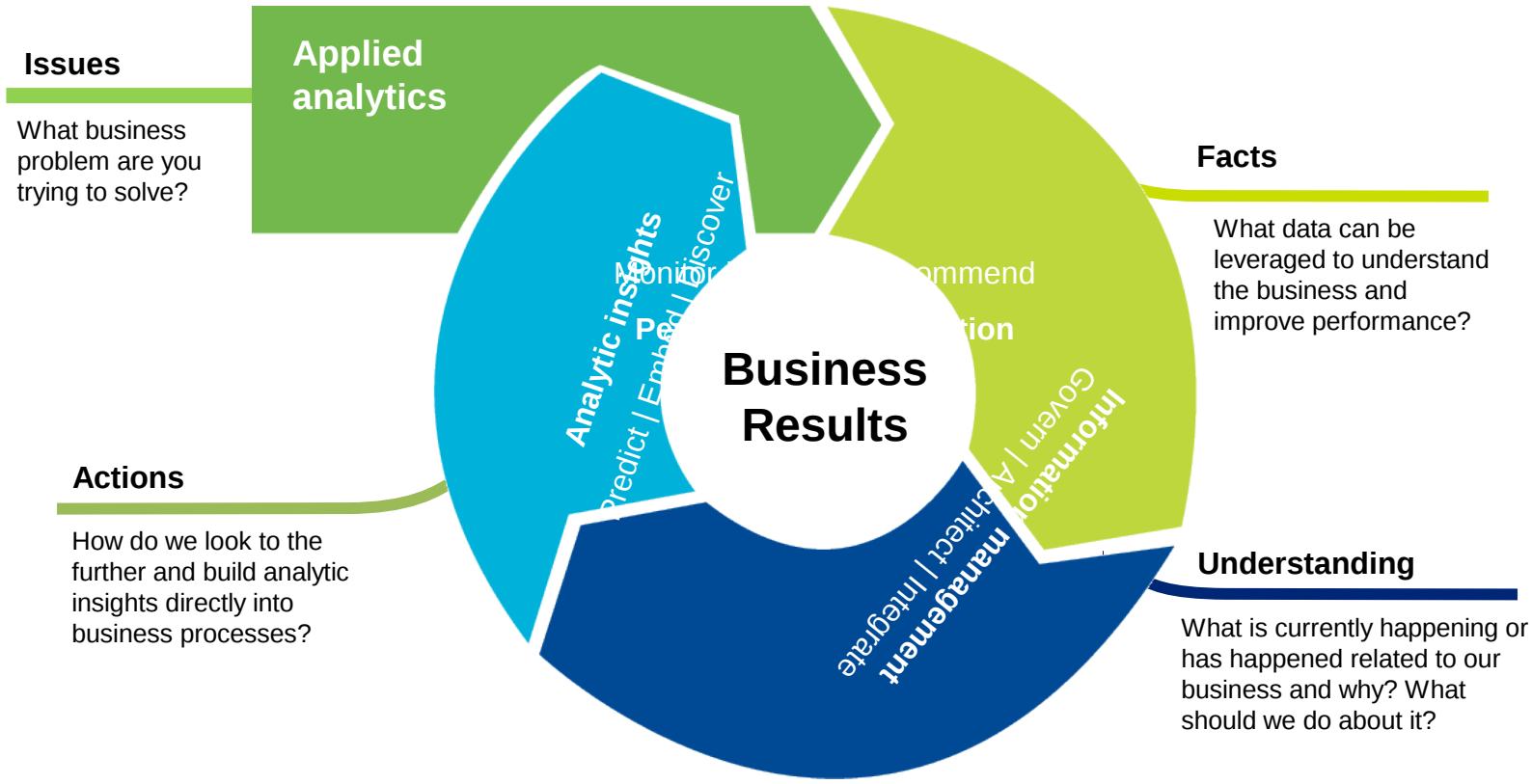
Geographic
Crime Statistics
Climate Data
Geographic Mapping
Population Concentration

Behaviors & lifestyle
Physical Activity Level
Hobbies
Lifestyle Clusters
Social Values

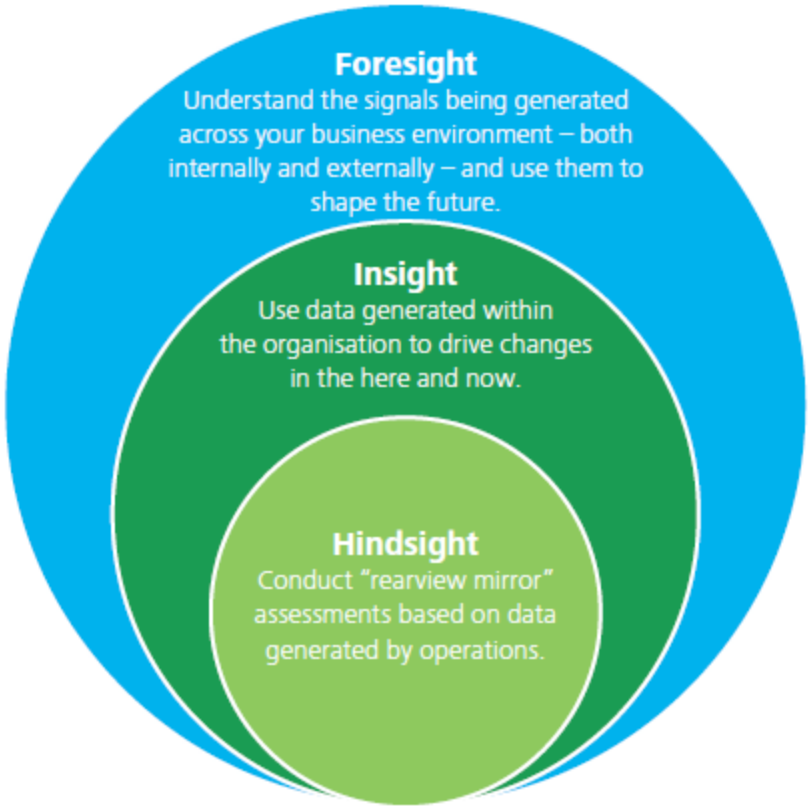
Purchase behaviors
Purchase Propensities
Spend by Category
Purchase Triggers

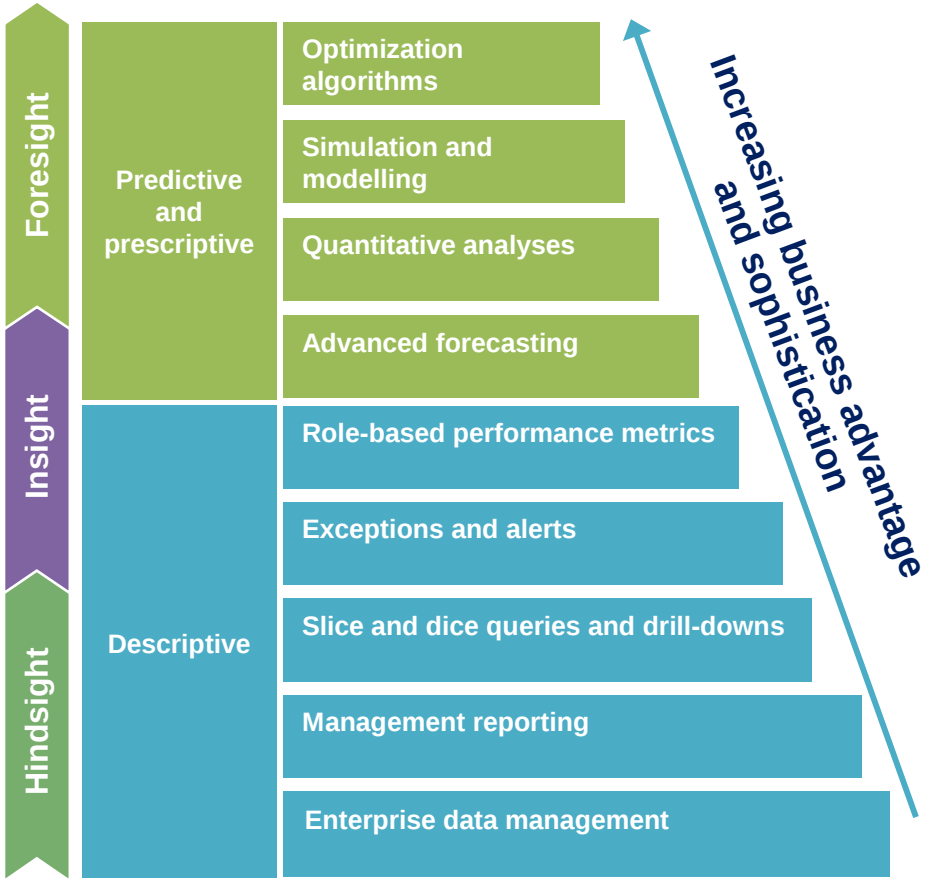
A continuous process…

The process of delivering business analytics results is one of continuous improvement.



… that needs to evolve





Applications – Life and Health

Sales & marketing

- Identify target groups
- Identify characteristics correlated with purchase decision
- Understand purchase behaviors and recommend the right product
- Recruit agents whose characteristics are similar to successful agents
- Monitor existing agents

Underwriting

- Identify best risks and prioritize acceptation efforts
- Identify applicants for whom additional underwriting is needed
- Support simplified underwriting

Claims

- Predict claim frequency and severity
- Prioritize resources
- Identify likely fraudulent/rescinded claims

In-force management

- Identify and retain policyholders likely to surrender
- Offer additional products to current customers
- Profile customers

Pricing/reserving

- Improve pricing accuracy
- Project impact of deviations from pricing parameters
- Reserve more accurately

Experience analysis

- Identify experience drivers
- Handle low credibility data by enhancing the data
- Create own mortality/lapse tables

Application: Modeling Longevity

- **Questions**
  - What is the empirical average death rate (DR)?
  - What is the empirical implied mortality improvement (MI)?
  - What would modeling predict?
  - How does the modeled MI compares to CIA promulgated scale?
  - Are all variables important for forecasting?

The Data

- Canadian Standard Ordinary Mortality 2005-2012, Canadian Institute of Actuaries [Insured population experience]
- Human Mortality Database 1921-2012, Statistics Canada [General population experience] (as a secondary data set)

Data (CIA CSO) – Empirical rates

Death rates

DR	’05	’06	’07	’08	’09	’10	’11	Avg
20-25	0.0004	0.0003	0.0003	0.0004	0.0003	0.0002	0.0001	0.000
40-45	0.0013	0.0016	0.0018	0.0020	0.0023	0.0032	0.0008	0.002
60-65	0.0088	0.0064	0.0063	0.0067	0.0053	0.0091	0.0057	0.007
80-85	0.0547	0.0498	0.0546	0.0497	0.0410	0.0401	0.0493	0.048

Mortality improvement rates

MI	’06	’07	’08	’09	’10	’11	Avg
20-25	29.2%	3.4%	-48.5%	11.1%	39.5%	29.1%	10.6%
40-45	-20.1%	-12.6%	-13.6%	-13.7%	-41.2%	74.3%	-4.5%
60-65	26.8%	2.7%	-6.5%	20.1%	-69.9%	37.2%	1.7%
80-85	8.9%	-9.5%	8.9%	17.5%	2.1%	-22.9%	1.7%

The Approach

Item	Specifics
Target variable	Death rate ($ claims/ $ exposure)
Predictive variables	Year, Sex, Smoker, Type of Underwriting, Insured amount, Duration, Attained Age
Predictive models	• Generalized Linear Model (GLM) • Lee-Carter (L-C)
Approach	• Fitted model to 2005-2011 data • Tested the model fit • Tested the predictive powers on 2012 • Derived and compared the MI rates

The GLM Model (Probit version)

- $$\phi^{-1}(DR_i) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_M x_{M,i} + \varepsilon_i$$

Where

- DR_i is the death rate for occurrence i
- $\beta_0 \ldots \beta_M$ are the regression coefficients indicating the relative effect of a particular explanatory variable on the outcome
- $x_{1,i} \ldots x_{M,i}$ are the explanatory variables
- ε_i is the error term

GLM Model – Predictive variables

- Most important predicting variables have lower p-values
- Eliminated of product type and Preferred classification due to high p-values
- Odd ratios indicate the level of change per unit increase in the variable

Year	Odds ratio
t	0.99

Sex	Odds ratio
Male	1.00
Female	0.84

Duration	Odds ratio
D	1.00

Amount insured	Odds ratio
0-10K	1.00
10K-50K	0.95
50K-100K	0.91
100K-250K	0.87
250K-500K	0.88
500K-1M	0.78
1M+	0.83

Attained Age	Odds ratio
x	1.04

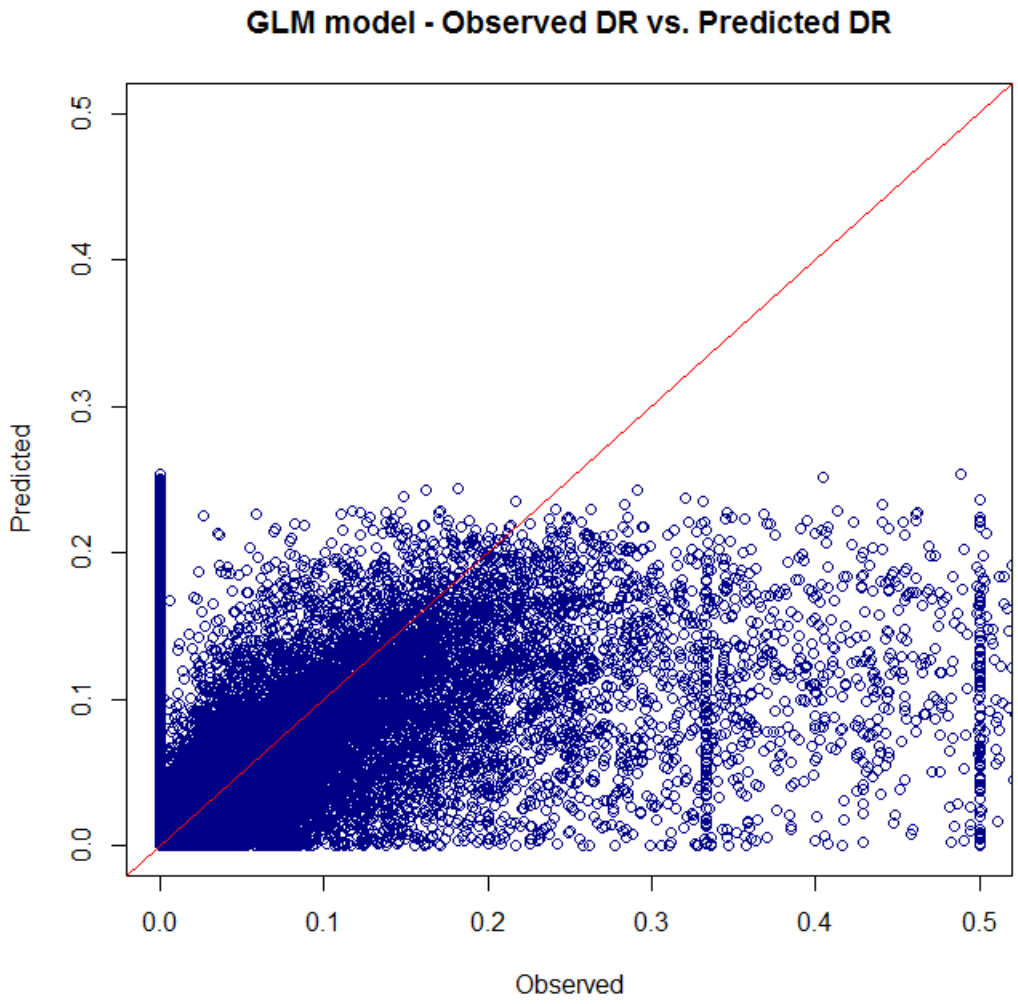
Smoker	Odds ratio
Smoker	1.00
Non-Smoker	0.79
Smoker status unknown	0.93

Type of Underwriting	Odds ratio
Medical	1.00
Non-medical	1.00
Paramedical	0.96
Unknown	0.97

GLM Model – Fit

- **Model is a good fit**
  - Coefficients of determination:
    - $R^2 = 22\%$
    - $Adjusted\ R^2 = 44\%$
  - Goodness of fit test (Chi Square) = 5506
  - P-value from Chi Square test <0.001
  - Modeling error and projection range

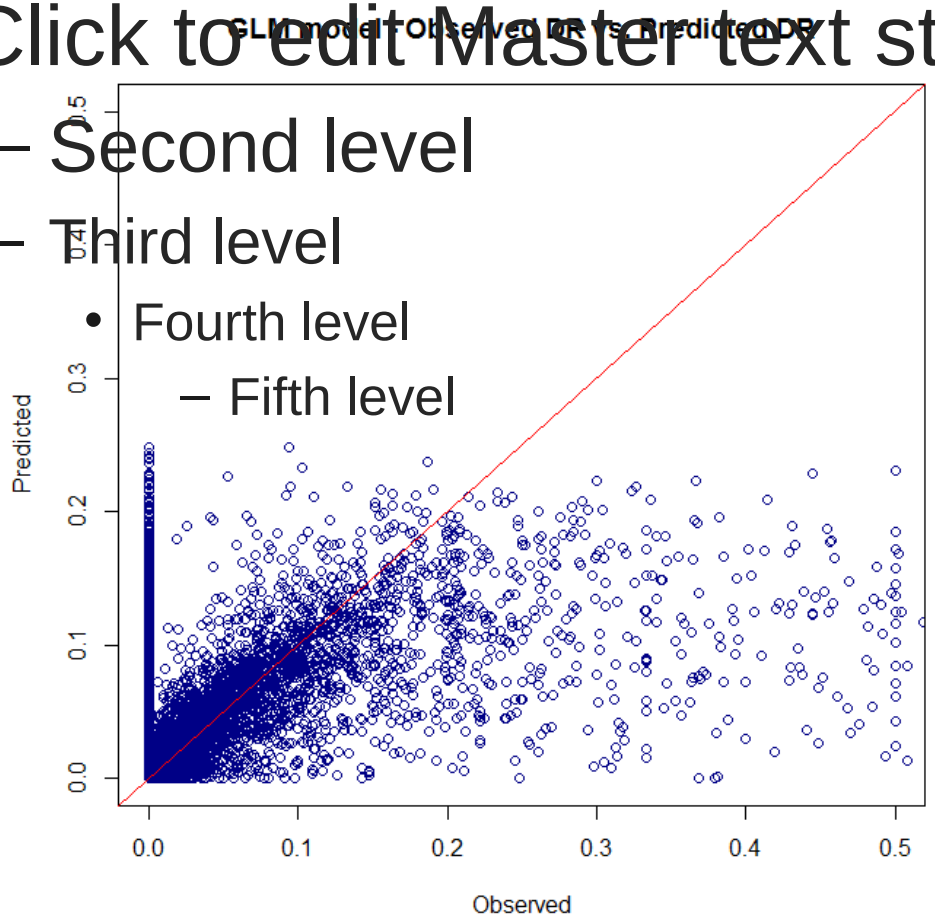
GLM Model – Fit (cont'd)



GLM Model – Power of Predictability

$Coefficient\ of\ determination\ (R^2)$ for 2012 predicted DR is 22%

- Click to edit Master text styles
  - Second level
    - Third level
      - Fourth level
        - Fifth level



GLM Model – Results

Death rates

DR	’05	’06	’07	’08	’09	’10	’11	Avg
20-25	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002
40-45	0.00049	0.00049	0.00048	0.00048	0.00047	0.00047	0.00047	0.00048
60-65	0.00738	0.00729	0.00725	0.00716	0.00707	0.00701	0.00692	0.00715
80-85	0.05679	0.05642	0.05551	0.05517	0.05486	0.05438	0.05402	0.05531

Mortality improvement rates

MI	’06	’07	’08	’09	’10	’11	Avg
20-25	1.5%	1.1%	1.0%	2.9%	0.2%	1.2%	1.3%
40-45	0.9%	0.5%	1.5%	0.9%	0.9%	0.9%	0.9%
60-65	1.1%	0.6%	1.3%	1.2%	0.8%	1.3%	1.0%
80-85	0.7%	1.6%	0.6%	0.6%	0.9%	0.7%	0.8%

The Lee-Carter Model

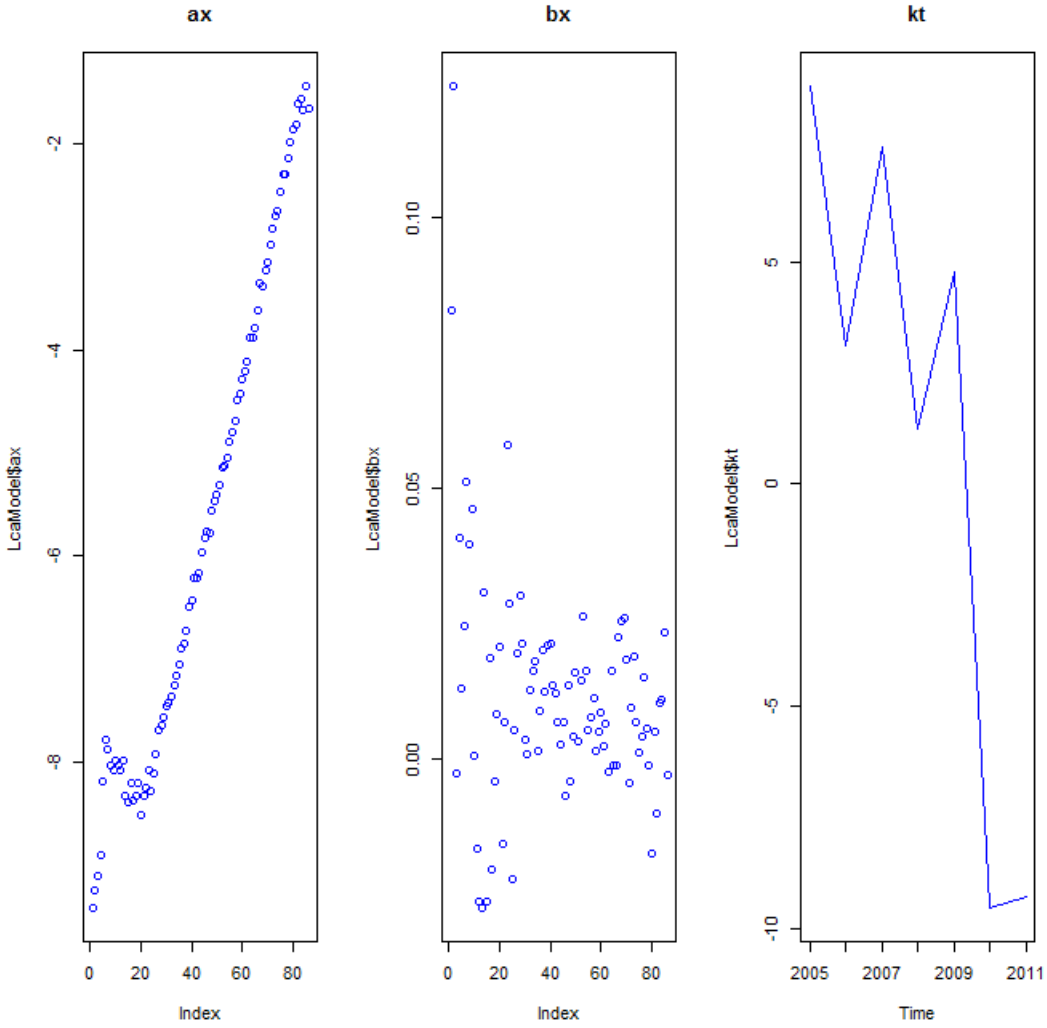
- $$log(DR_{xt}) = a_x + b_x k_t + \varepsilon_{xt}$$

Where

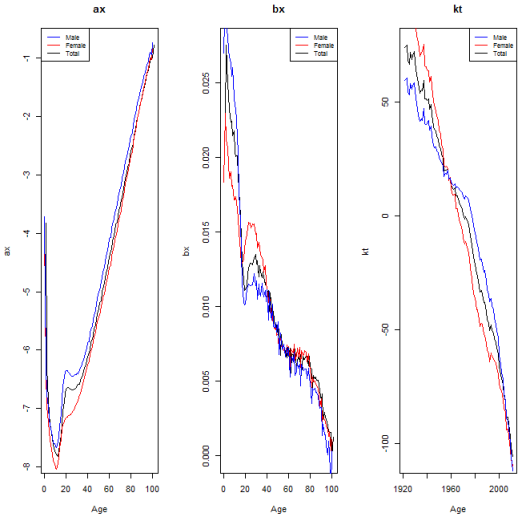
- DR_{xt} is the death rate for age x at time t
- a_x and b_x are the regression coefficients relating to the age
- k_t is the regression coefficient relating to the time
- ε_{xt} is the error term

LC Model – Calibration

CIA CSO Data

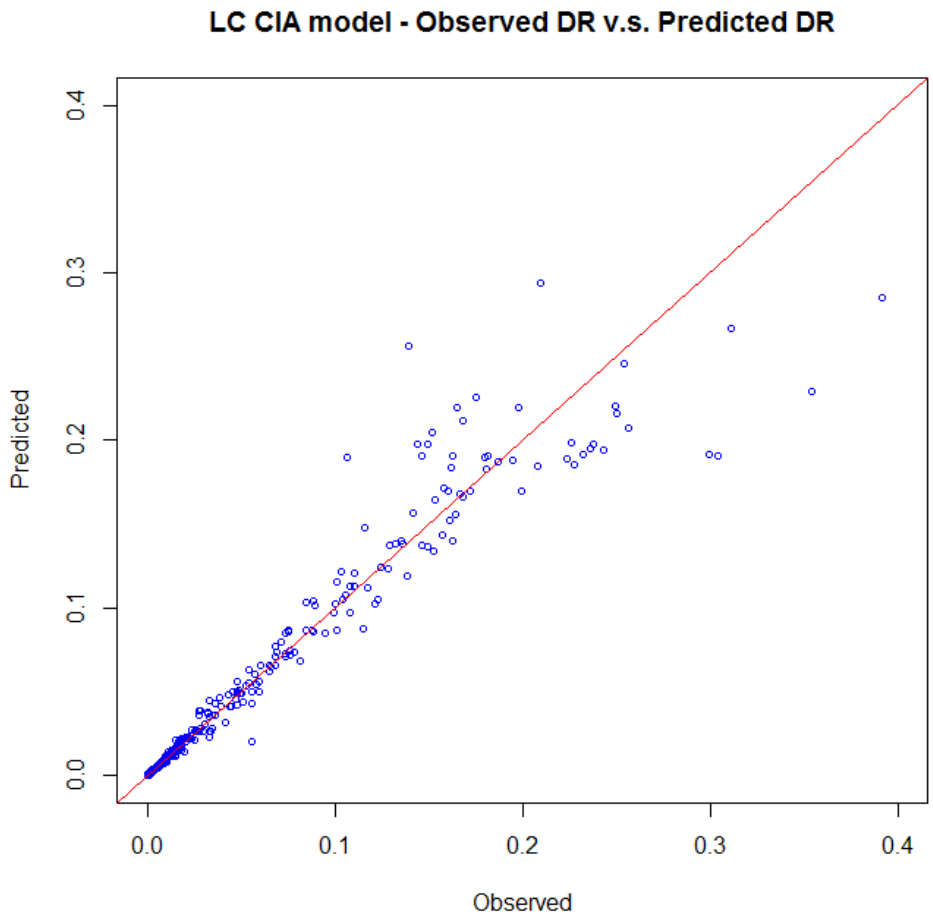


HMD Data

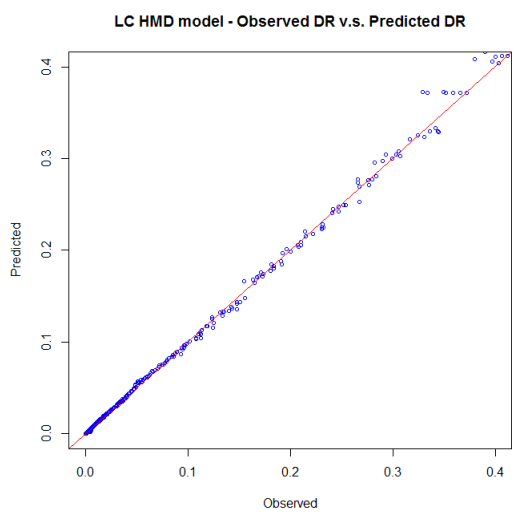


LC Model – Fit

CIA CSO Data



HMD Data

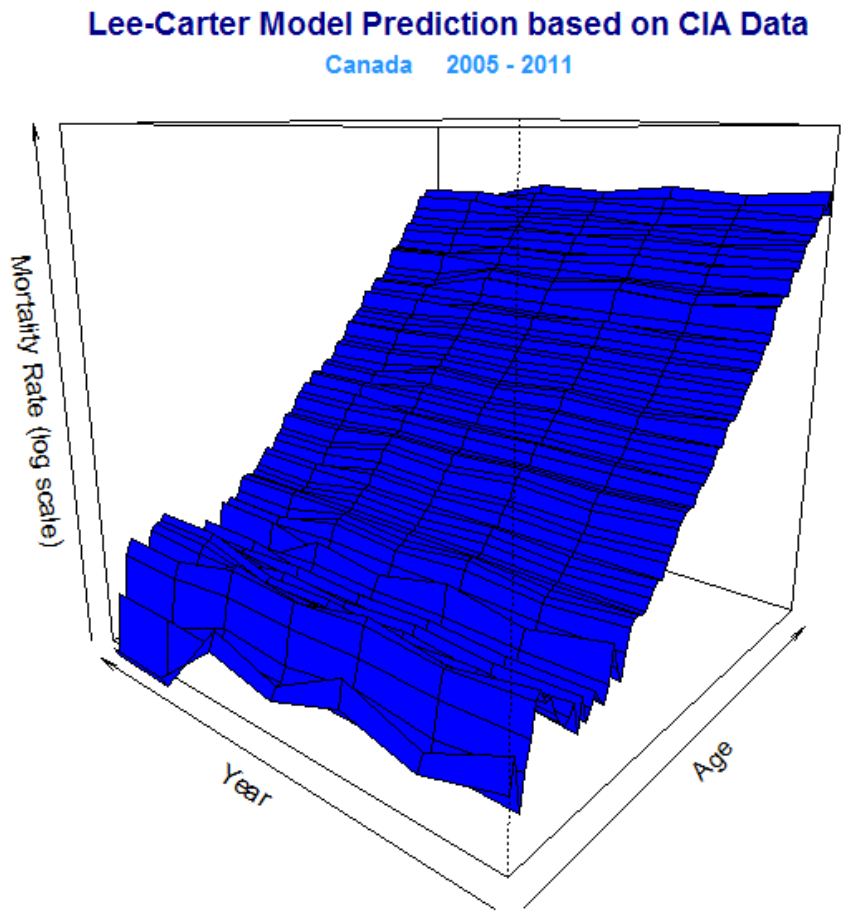


LC Model – Power of Predictability

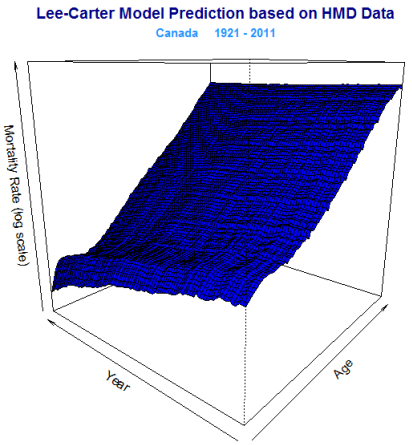
- $Coefficient\ of\ determination\ (R^2)$ for 2012 observed v.s. predicted death rate is 85%

LC Model – Power of Predictability

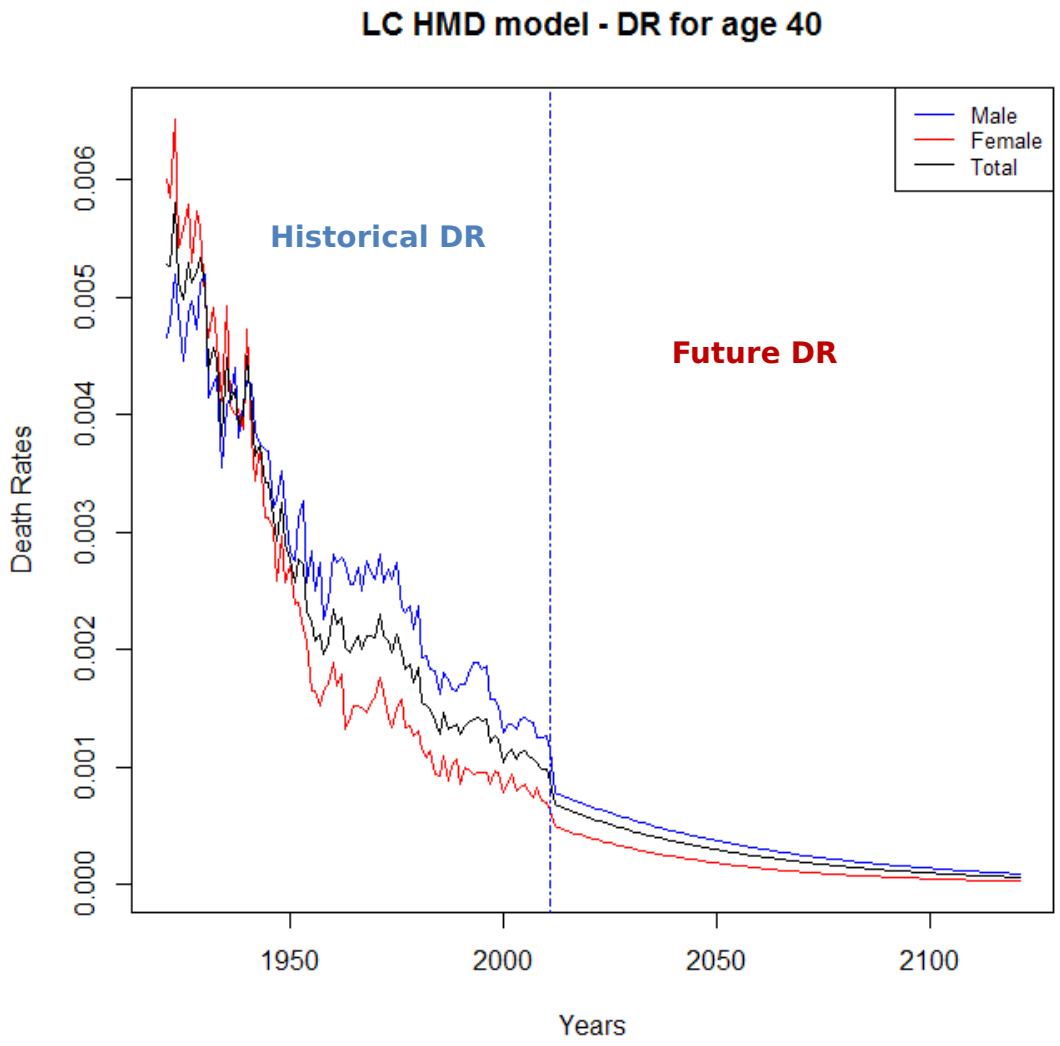
CIA CSO Data



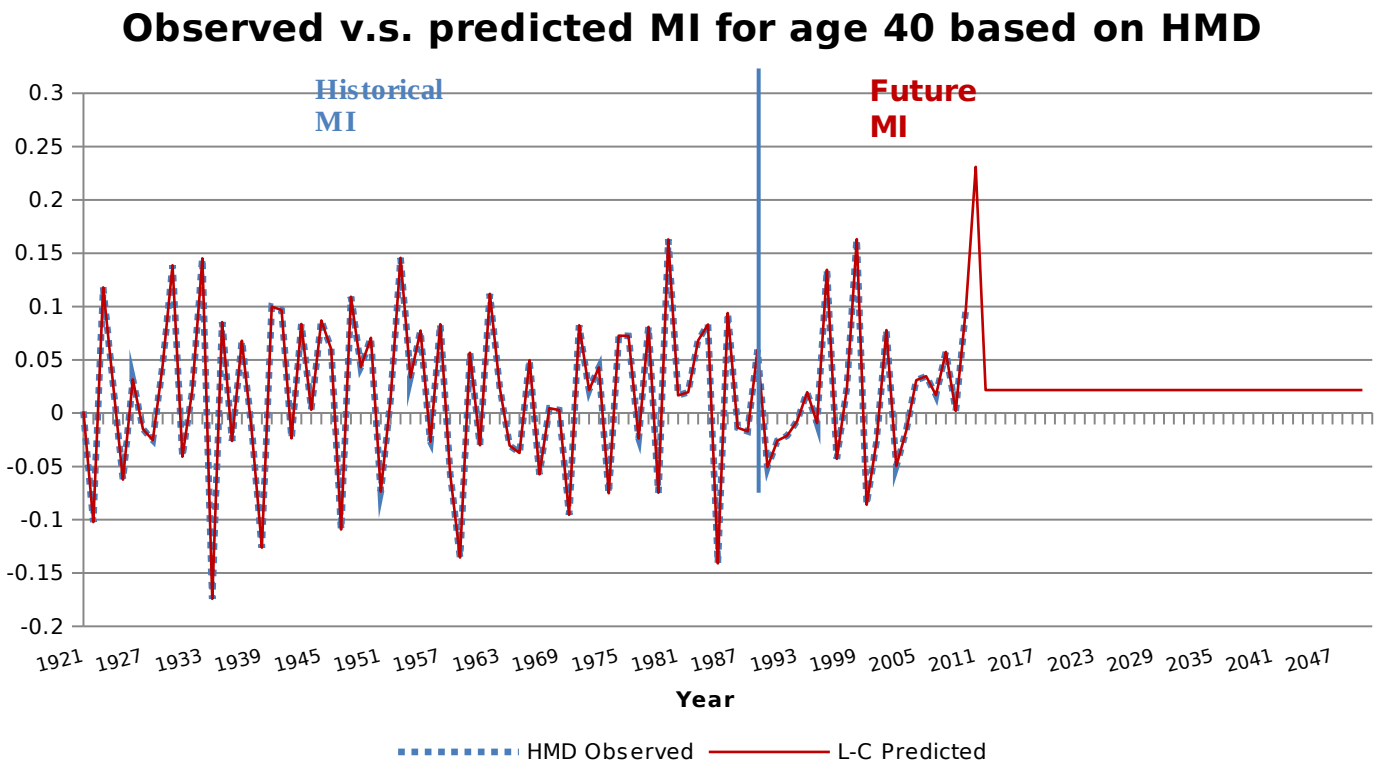
HMD Data



LC Model – DR Illustration



LC Model – MI Illustration



LC Model – Results

Death rates

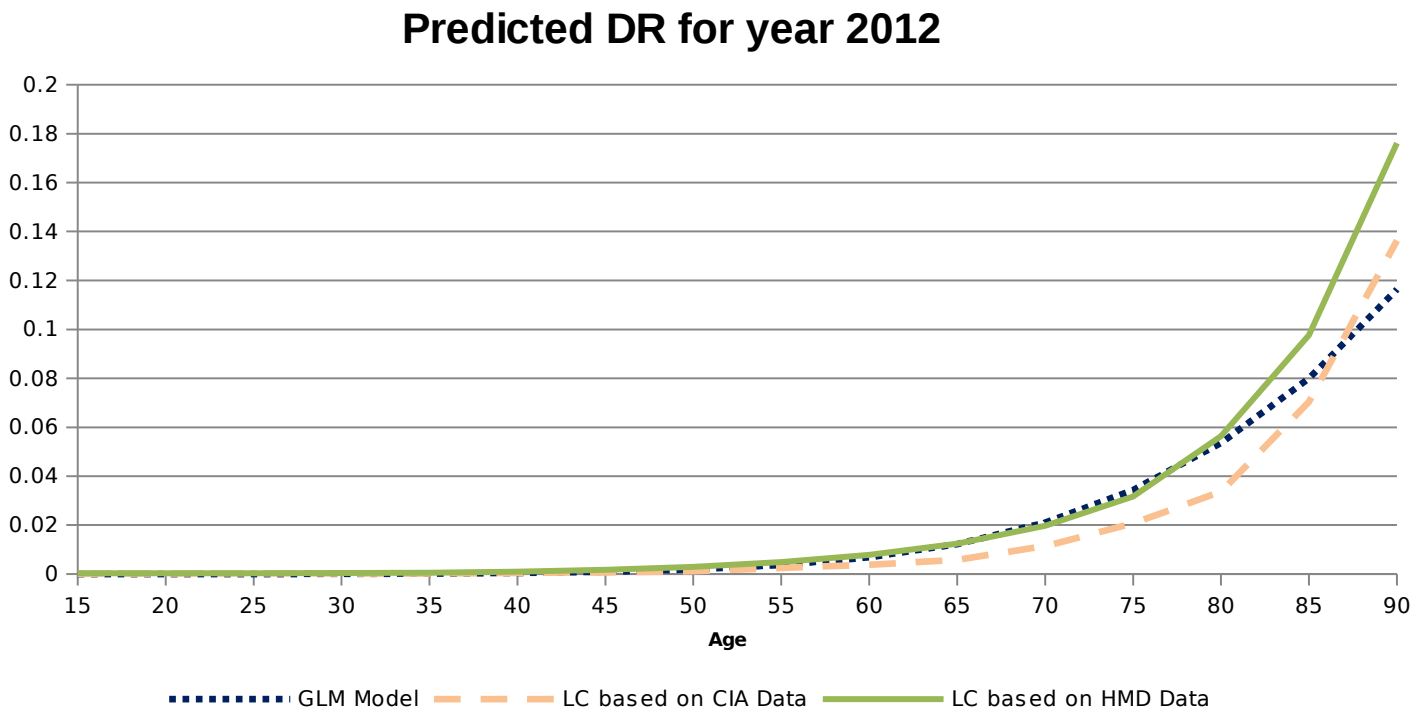
DR	'05	'06	'07	'08	'09	'10	'11	Avg
20-25	0.00043	0.00036	0.00041	0.00034	0.00038	0.00027	0.00027	0.000
40-45	0.00060	0.00055	0.00059	0.00053	0.00056	0.00046	0.00046	0.001
60-65	0.00436	0.00419	0.00432	0.00414	0.00424	0.00387	0.00388	0.004
80-85	0.04702	0.04262	0.04591	0.04133	0.04380	0.03502	0.03515	0.042

Mortality improvement rates

MI	'06	'07	'08	'09	'10	'11	Avg
20-25	15.5%	-13.5%	16.1%	-10.0%	27.9%	-0.4%	6.0%
40-45	8.3%	-6.8%	8.9%	-5.2%	17.9%	-0.3%	3.8%
60-65	3.9%	-3.0%	4.1%	-2.4%	8.6%	-0.1%	1.8%
80-85	9.4%	-7.7%	10.0%	-6.0%	20.0%	-0.4%	4.2%

Comparison of the results

Comparison GLM and LC models



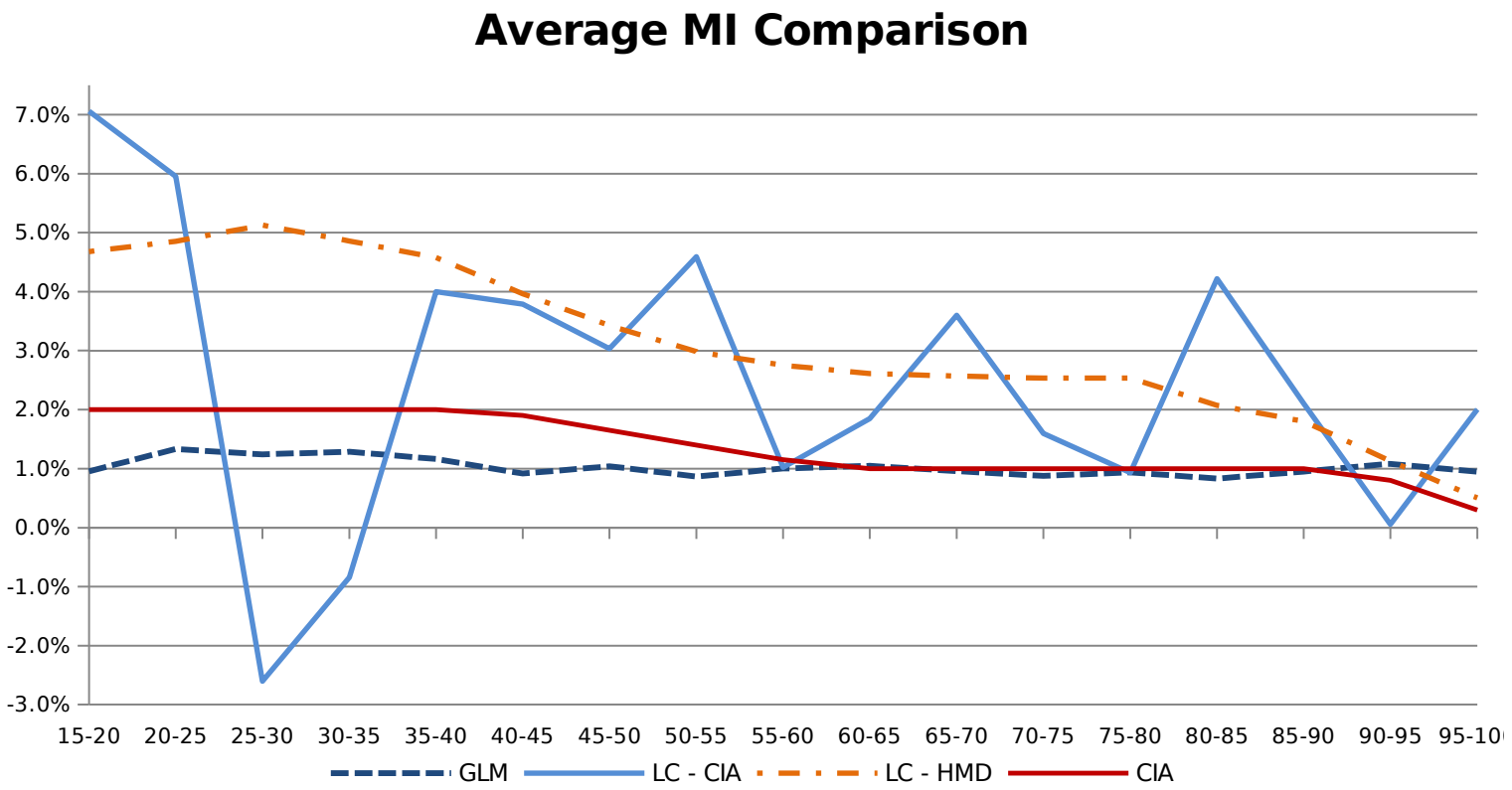
Comparison of the results (cont'd)

Comparison of mortality improvement rates GLM, L-C and CIA promulgated

Average MI	Empirical data	GLM	Lee Carter CIA CSO	Lee Carter HMD	CIA
20-25	10.6%	1.3%	6.0%	4.9%	2.0%
30-35	-14.4%	1.3%	-0.8%	4.9%	2.0%
40-45	-4.5%	0.9%	3.8%	4.0%	1.9%
50-55	-7.3%	0.9%	4.6%	3.0%	1.4%
60-65	1.7%	1.0%	1.8%	2.6%	1.0%
70-75	-1.4%	0.9%	1.6%	2.5%	1.0%
80-85	1.7%	0.8%	4.2%	2.1%	1.0%
90-95	-7.0%	1.1%	0.1%	1.1%	0.8%

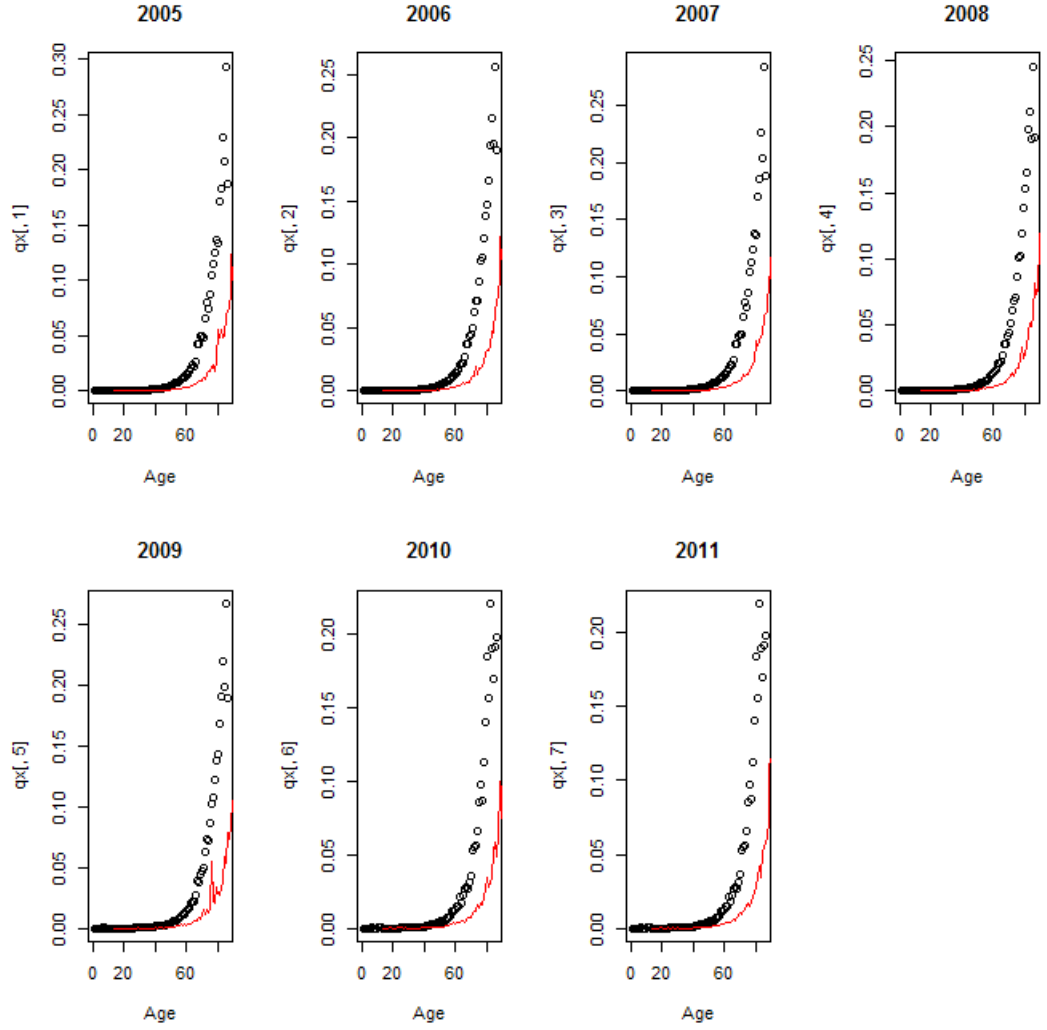
Comparison of the results (cont'd)

Comparison of mortality improvement rates GLM, L-C and CIA promulgated

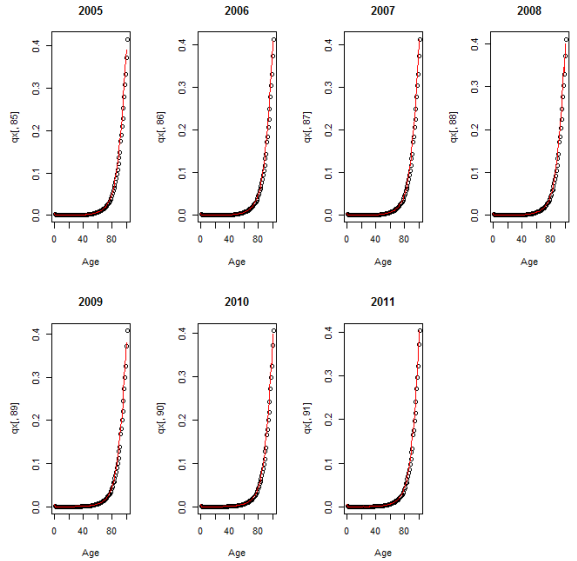


Appendix – LC Model – Fit

CIA CSO Data



HMD Data



Appendix – LC Model – Illustrations