UBI data management – Granularity and related considerations

March 11, 2015

Casualty Actuarial Society

RPM Seminar, Dallas USA

By : Germain Denoncourt, FCIA, FCAS

Agenda

- Introduction
- Data quantity – data quality
- Devil in the details examples
- UBI program examples from Canada
- Related aspects and a little futurism
- Conclusion

Introduction : CAS website description - overview

- Analytics and modeling are expanding in UBI, and a lot of new info flows in. Actuaries and managers have to prepare data, manage it, - **KNOW IT WELL** - use it, and transform it for good decision making.
- Many questions are to be answered (sometimes before) getting the data; for example : which events to track, ex. : a hard braking event, hard acceleration event, hard curve taking event, use of latitude, longitude, altitude, and so on.
- This session explores the actuarial role and related aspects

Introduction : designing a UBI project … ?

- Though experienced actuaries (at least with long Predictive Modeling/GLMs and preferably with UBI experience) are key people in a UBI project design and implementation, a multi-disciplinary team could be indicated : IT, Marketing, U/W, Actuary(s), … , a top decider, a project manager for key dates

- Learning curve shorter for whole team with experienced actuary in / or leading – officially or not

Data quantity

Compare these projects data quantity …

- 1 - Mid 1990's : a decision tree study in PPA auto: 600 000 earned car-years, ~ 20 risk factors/variables, 3 coverages (Liability, Collision, Comprehensive with several KOL's (theft, glass, animals, all other) ; IT people told me : «You won't take as much as 1 giga for this study ?? Too much space for our IT servers … »

Data quantity

Project 2 data quantity …

- 2 - 2015 UBI pilot project : if target 600 000 earned car-years, ~ 15 UBI risk factors, 4+ coverages (Liability – at least 3 kol's : BI, PD to cars, PD to property), PIP/No Fault, Collision (at least 2 kol's), Comprehensive with several KOL's (theft, glass, animals, vandalism, … all other) ; what IT people will tell you about data quantity ??

Data quantity

Data provider will definitely help … Wikipedia too

Decimal		
Value	Metric	
1000	kB	kilobyte
1000^2	MB	megabyte
1000^3	GB	gigabyte
1000^4	TB	terabyte
1000^5	PB	**petabyte**
1000^6	EB	exabyte
1000^7	ZB	zettabyte
1000^8	YB	yottabyte

Data quantity examples (Source : Wikipedia, February 2015)

- Telecommunications (usage): AT&T transfers about 30 petabytes of data through its networks each day.
- Internet: Google processed about 24 petabytes of data per day in 2009. The BBC's iPlayer is reported to use 7 petabytes of bandwidth each month. Imgur transfers about 4 petabytes of data per month.
- Supercomputers: In January 2012, Cray began construction of the Blue Waters Supercomputer, which will have a capacity of 500 petabytes making it the largest storage array ever if realized.
- Data storage system: In August 2011, IBM was reported to have built the largest storage array ever, with a capacity of 120 petabytes.
- Databases: Teradata Database 12 has capacity of 50 petabytes of compressed data.
- Data mining: In August 2012, Facebook's Hadoop clusters include the largest single HDFS cluster known, with more than 100 PB physical disk space in a single HDFS filesystem. Yahoo stores 2 petabytes of data on behavior.

Data quantity to data itself … and to data quality

Should pilot project / multi-disciplinary team should define which UBI raw data and risk factors wanted asap ? Egg or chicken problem ? Possible UBI data is virtually unlimited ?

- Hard Braking event
- Hard acceleration event
- Speeding
- «Cornering» / «hard curve taking»

Data quantity to data itself … and to data quality

Possible UBI Data is virtually unlimited ? - followed

- Time of day
- Day of week
- Day of year (winter, summer …)
- Type of road / «road segment»
- Other types of «events» : lane change, skidding, backing up, pure speed, entering wrong way, not yielding, Xing under red light …

Data transformation into a GLM file … why, how … ?

Transformation of raw data : not too different than for a PPA Predictive modeling or HO pred. Modelling/GLM project, but …

The devil is in the details ! (both vs quantity needed and quality …)

Devil in the details - examples

Definitions of events / data to collect … and definitions

- Event type 1 : hard braking event – threshold(s)
- Event type 2 : hard acceleration event – threshold(s)
- Event type 3 : hard curve taking event –threshold(s)
- Speeding : how many miles over the limit, if police tolerates X mph over, treshold(s)

Devil in the details – examples

Quality of raw data

- Are speed limits ok for the whole state or province ? What about updates ? How easy to get in Ohio, California, Ontario, Utah … how regulated ?
- Frequency at which raw data is sent to the insurer database – once per second ? Per minute ? (quantity here …)
- Who liable for data quality … IT ? Actuarial ?

Devil in the details – examples

- Definition of a harsh brake event in Canada
  - Intact Canada = -12km/h per second, Mobiliz -15km/h per second
- Definition of 'drive at night' ?
- Definition of a hard curve taking ? Need centrifugal force measure but crossed with center of gravity of vehicle ? (think motorcycles …)
- Number of harsh brakes per km : same treashold city vs rural ?

Devil in the details - examples

Territories / zones / road segments in a UBI project

- Latitude, longitude, altitude needed as raw data for creation of micro zones ; example : for creating a rate per 100 feet driven per micro zone or road segment ? Rated differently per coverage and KOL and varying per time of day, day of week, season of year ??

Devil in the details - examples

Territories / zones in a UBI project

- If you don’t do it, some other insurer will … sooner than expected maybe
- At least have to get some data soon, a «must do step» : question of live or die ? think about the Google car, automoted cars, «drone cars» …

Devil in the details - examples

So, … what level of raw data is needed to analyze and survive in auto insurance ? (and soon in HO telematics too ?)

Devil in the details – examples

- These aspects have to be considered, too:
  - Data provider – and the right person(s)
  - Cost of data : per Xbyte, per UBI element collected, per detail level
  - Transformation into GLM data file for analysis – quality and quantity – know your data
  - Interactions with traditional actuarial rating variables

UBI examples in PPA from Canada

Pure UBI vs 'Discount UBI' in Canada : Industrial Alliance Mobiliz Program, Intact Canada/Belair Direct Programs, Desjardins' Ajusto program

- Mobiliz : in Quebec province since April 2012 – close to a 'pure' UBI: a fixed $ premium per month per coverage, + a rate per km driven, Δ by behaviour 'factor' applied to both fixed $ and per km $, premium charged **retroactively** after end of month, few traditional rating factors used, on-line quote < 60 seconds

UBI examples in PPA from Canada

Pure UBI vs 'Discount UBI' in Canada : (followed)

Mobiliz :

- Measures various UBI metrics / driving behaviour as harsh brakes per 100 kms driven, harsh accelarations, speeding (need to map the whole province speed limits on each road segment, in real time, and update)
- Insured car's driving Dashboard, compared to 'insured community'
- Can see exactly where and when event occurred

UBI examples in PPA from Canada

Pure UBI vs 'Discount UBI' in Canada (followed)

- Desjardins' Ajusto and Intact/Belair Direct programs (not identical, but similar): use driving behaviour to give a 'discount' at annual renewal of the policy ; all traditional variables still used ; use various metrics to rate as harsh brakes per 100 kms, but not speeding
- Insured car's driving Dashboard, compared to 'insured community'
- Intact – track car 6 months, Desjardins 'forever'

UBI examples in PPA from Canada

Size in Direct Written Premiums, 2013 :

- Intact Canada Group (former ING Canada) : #1 insurer in Canada, all provinces, 7,35$G of DWP
- Desjardins : #2 in Canada : bought State Farm Canadian portion in 2014 ; with 3,9$G of DWP
- Industrial Alliance – (Mobiliz) : < 300 $M DWP in Quebec province only – (small size can succeed !)
- All these 3 insurers have FCAS as presidents

Related aspects and futurism : road types - think granularity needed in the future …

Visually, with photos : compare the insurance risk per ‘road types’ or even ‘road segments’ …

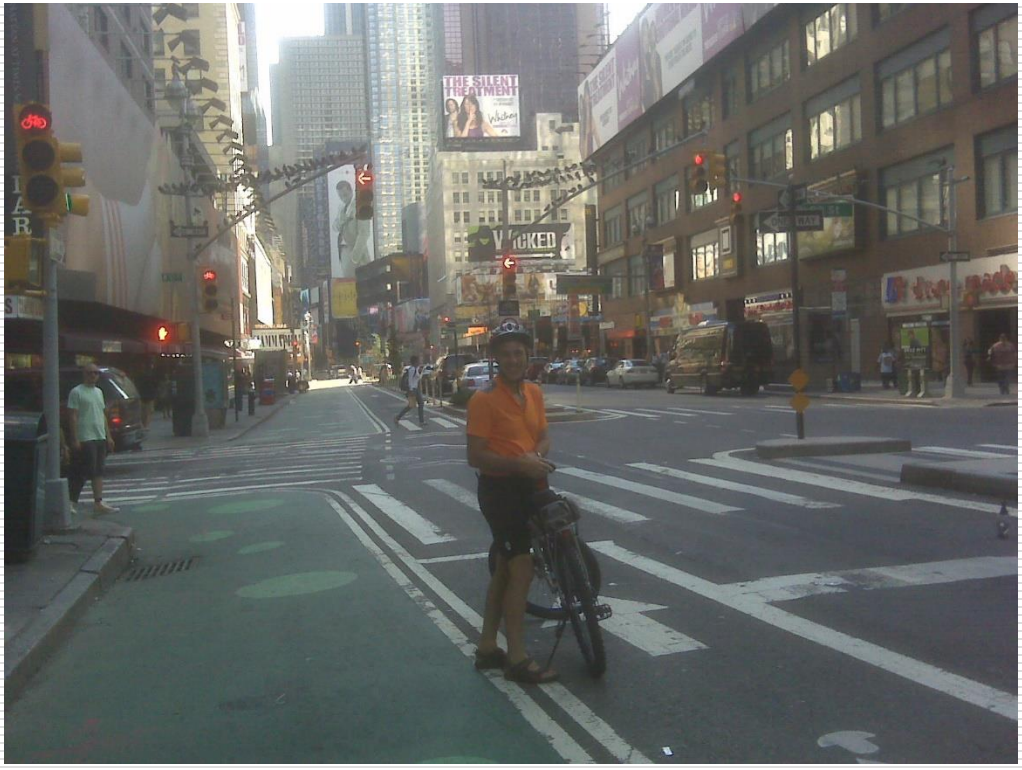
(disregard my bike on several of them)

Type of road/road segments

Quebec province 'Very' Rural road

Type of road/road segments

Real urban – Manhattan Broadway !



Type of road/road segments

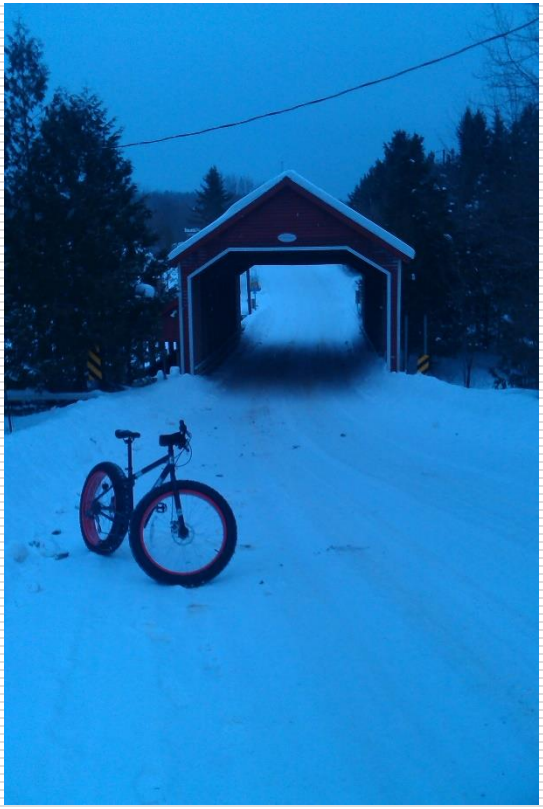
Frontier road : Vermont / Canada border (see next slide)

Type of road/road segments

Frontier road … Vermont and Quebec/Canada border

Type of road/road segments

Rural winter road …

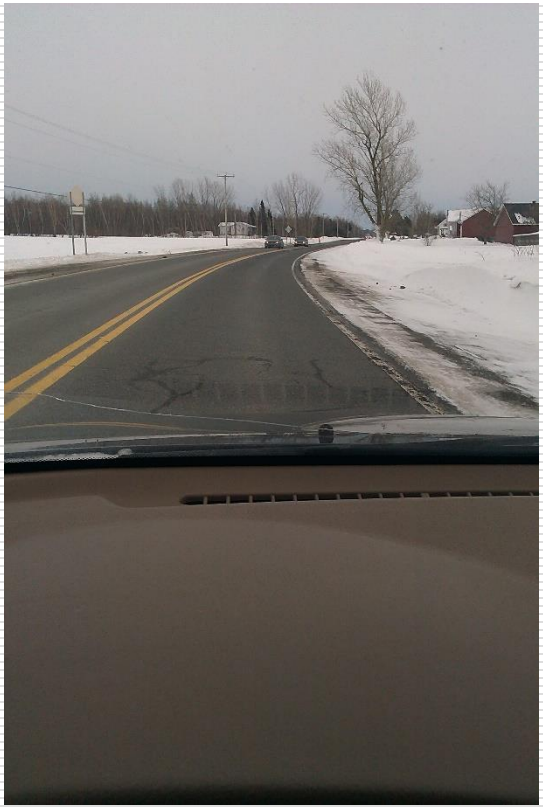


Type of road/road segments

Rural winter roads …

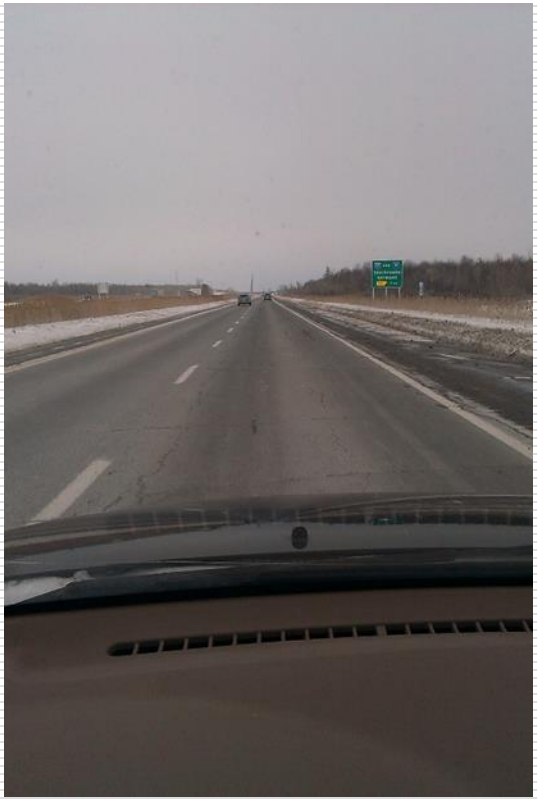
Type of road/road segments

Two-way secondary road with high speed traffic both ways



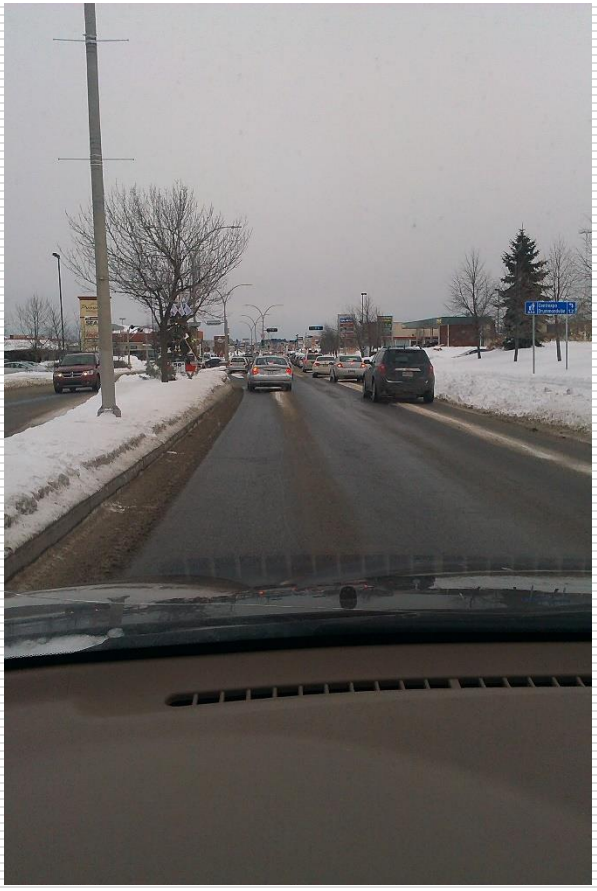
Type of road/road segments

Highways and major roads as 95,40, … in USA



Type of road/road segments

Urban Boulevard (with traffic lights …)



Related aspects : The 'Scoring Method' – Pure UBI vs 'Discount UBI'

So …

Consider the following potential relativity tables, based on driving behaviour relativities per 'UBI events' per coverage and several KOL's (kind of loss), compared with a single amount of 'discount' applied to every coverage of the insured car :

Relativities per type of road, per mile driven, for GLM frequency

Example …

Type of road	Liab. BI	Collision (all kol's)	Comp. Theft	Comp. Animals
Highway (95,40,…)	0.25	0.35	1.00	2.00
Main road – both sides traffic)	0.80	2.00	1.00	3.00
Urban boulevard	3.50	3.25	1.00	1.00
Resi. street	**1.00**	**1.00**	**1.00**	**1.00**
Secondary road	0.65	0.75	1.00	5.00
Very rural road	0.45	0.60	1.00	10.00
etc				

Relativities per type of road, per mile driven, for GLM severity

Example …

Type of road	Liab. BI	Collision (all kol's)	Comp. Theft	Comp. Animals
Highway (95,40,…)	2.50	2.75	1.00	4.00
Main road – both sides traffic)	4.00	5.00	1.00	3.00
Urban boulevard	1.50	2.00	1.00	1.00
Resi. street	**1.00**	**1.00**	**1.00**	**1.00**
Secondary road	2.00	4.00	1.00	5.00
Very rural road	3.00	3.00	1.00	10.00
etc				

Relativities per type of road, per mile driven, for GLM pure premium

Example … GLM freq * GLM sev

Type of road	Liab.BI	Collision (all kol's)	Comp. Theft	Comp. Animals
Highway (95,40,…)	0.63	0.96	1.00	8.00
Main road – both sides traffic)	3.20	10.00	1.00	9.00
Urban boulevard	5.25	6.50	1.00	1.00
Resi. street	**1.00**	**1.00**	**1.00**	**1.00**
Secondary road	1.30	3.00	1.00	25.00
Very rural road	1.13	1.80	1.00	100.00
etc				

Relativities per Season

Estimated risk per month of year, numbers (SAAQ – Société de l'Assurance Automobile du Québec public data)*

Season	Death	Serious injury	Minor injury	All 3 categories
Spring	0.78	0.78	0.84	0.83
Summer	1.46	1.30	1.18	1.20
Fall	**1.00**	**1.00**	**1.00**	**1.00**
Winter	0.89	0.96	0.97	0.97

* For one given past year, whole province, not controlled for factor effects other than season itself, whole Quebec province SAAQ government insurer data

Relativities per day of week

Estimated risk per day of week (SAAQ data)*

Day of week	Death	Serious injury	Minor injury	All injuries
Monday	1.18	1.04	0.96	0.98
Tuesday	1.25	0.95	1.01	1.01
Wednesday	**1.00**	**1.00**	**1.00**	**1.00**
Thursday	1.27	1.19	1.14	1.15
Friday	2.17	1.53	1.31	1.35
Saturday	1.78	1.41	1.15	1.19
Sunday	1.68	1.37	1.06	1.11

* For a given year, idem as previous slide, ‘oneway’, not ‘GLM’

Relativities per time of day

Estimated risk per time of day, per km driven **

Hour	Death	Serious injury	Minor injury	All injuries
0h-6ham	20.8	14.5	11.0	11.6
6h-9ham	**1.0**	**1.0**	**1.0**	**1.0**
9h-3hpm	2.7	3.2	3.4	3.3
3h-6h30pm	1.6	1.9	1.8	1.8
6h30-12hpm	3.5	3.2	2.6	2.7

* For a given year, with approximation hypotheses, (SAAQ and AMT data**)

The scoring method – vs Pure UBI

Based on the previous relativity tables :

Which insurer is segmenting the best way ? (Insurer A with one single 'discount' ex. - 20% on each coverage, or Insurer B using all the previous tables ?)

Which 'raw data' details and granularity will you need to cope with future competition, and how manage it ?

Should insurers fear more 'pure UBI' insurers coming to the market, given easier and easier 'Big Data' use ? Any 'Colossus with feet of clay' in Canada, USA, and abroad ?

A little futurism on open data

From a February 2015 recent text in the local Montreal subway newspaper (see annex) : open data from the City of Montreal, will be released in March 2015, namely on :

- Where road accidents and infractions occur
- Where fires occur
- Traffic per road segment
- … and other data elements.

A little futurism on data

Examples of new kind of data tracked : at least a few Canadian insurers are already working on it …

- Driving with sunshine low in front
- Types of ditches along each 'road segment' (influences GLM severity risk for Collision and No-Fault PIP per road segment/mile driven, at least)
- Snowing, wind and snow, raining events …
- Etc … other matches of geographic/weather data with auto insurance claims

Conclusion

(Japanese proverb)

A plan without action, is a dream.

But it could be worse.

Action without a (clear) plan …

It’s a nightmare !

Conclusion

So, a UBI, R&D or other project, without a clear plan about data management … and other aspects …

Could be a nightmare ?

Questions ?

Mr. Germain Denoncourt, FCIA, FCAS
Coefficient Actuarial Consulting
Montreal, Canada

cell. 514 706 7866
gdenoncourt@coefficient-act.com