



EVALUATING TELEMATICS BASED DATA AND DESIGNING UBI RISK & PRICING MODELS

The fast increase of data sources (telematics, mobile, auto-manufacturers), vendors and analytical alternatives poses a great challenge for Usage Based Insurance (UBI) programs. A need arises for methodology that helps compare solutions, assess data and design models. This white paper introduces 2 concepts: Four Layers that define the building blocks, and RQV² approach that defines how to evaluate the different layers and their inter-dependencies.

How To Integrate Apples And Oranges (without making a salad...)

Over the last 2-3 years the number of telematics based data sources has tripled! And so has the number of UBI risk and pricing models. For an insurance carrier to jump on the UBI wagon back in 2012, one had to choose from half a dozen telematics service providers (TSP) using dongles manufactured by about half that number vendors. For the modeling process, one had to decide whether to start a tedious education of its internal actuarial and pricing teams, and / or source a model from a very short list of analytics / scoring providers.

Today one can spend months' on initial meetings and still be confused. Data can come from a very wide variety of dongles, mobile apps, auto manufacturers and a mix of these platforms. There are experienced dongle and TSP vendors, alongside

newcomers with lower cost solutions. There are end-to-end app providers alongside combined app and dongle solutions. Not to mention OEM and tech giants self driving this process into future. When looking at the analytical options, the situation has diversified even more. In addition to past alternatives, one can hire skill that has played around with telematics data and modeling once or twice, and the number of analytics / scoring vendors and partnership opportunities seems to have grown even faster than telematics.

The UBI data world is extremely diversified. How can one compare different data sources? Or integrate them into the same risk model? How can one evaluate and compare driving events that stem from different data sources? What does all this mean for risk modeling? Can a model that was developed using one data source be transformed seamlessly to another (maybe lower cost data source)? How can a model sustain superiority as newer, more accurate data sources become available? Can volume across platforms compensate time-based volume? How can lower quality high volume models stack up against small, rich data, homogenous models?

About the writer



Oren Steinberg is an expert in the field of Usage Based Insurance (UBI). Over the past decade he led half a dozen UBI program deployments with leading insurance companies in the US, as well as PAYD pilots in the UK and Latin America. He was CEO of a scoring company that developed the DrivingDNA[®] solution and is the inventor of several data analysis and telematics related patents. Over the years, he consulted a wide variety of players in the space: equity funds, insurance carriers, service providers, technology vendors, app providers and startups. This White Paper was not sponsored by any company; it represents his opinions and does not constitute advice to make any investment or business decision.

oren.steinberg@gmail.com



The Big Data Questions

Evaluating the existing solutions, selecting the path best fitting each organization and then designing a UBI model can be a daunting task. Product and marketing teams can lead the process, with the help of procurement, however ignoring the Big Data Questions may prove to be shortsighted. Yes, choosing reliable partners and improving efficiencies, are great, but the Big Data Questions remain – What is all this good for? What are short and long term goals for the UBI program? What data to collect? What do with it? When? This White Paper offers a conceptual model for evaluating telematics data and designing UBI models. It does not advocate for any one solution, but rather comes to define the methodology by which different carriers can assess the various options. It can also serve vendors that come to develop a product and look for references. It is meant to make the evaluation and design processes better by taking into account the various aspects that make UBI modeling.

The Four UBI Layers

First we need lay down the foundations. I found it very useful to examine models for UBI in the context of the following Four UBI Layers:

1. Raw Data

Representing the actual measurements as collected by the sensors on whatever device or platform is being used. These can be GPS readings, sub second acceleration measurements, CAN/BUS codes, odometer readings etc. True, that in some devices some of these readings are calculated rather than measured, however at its core, this layer does not deal with calculated values.

2. Driving Variables

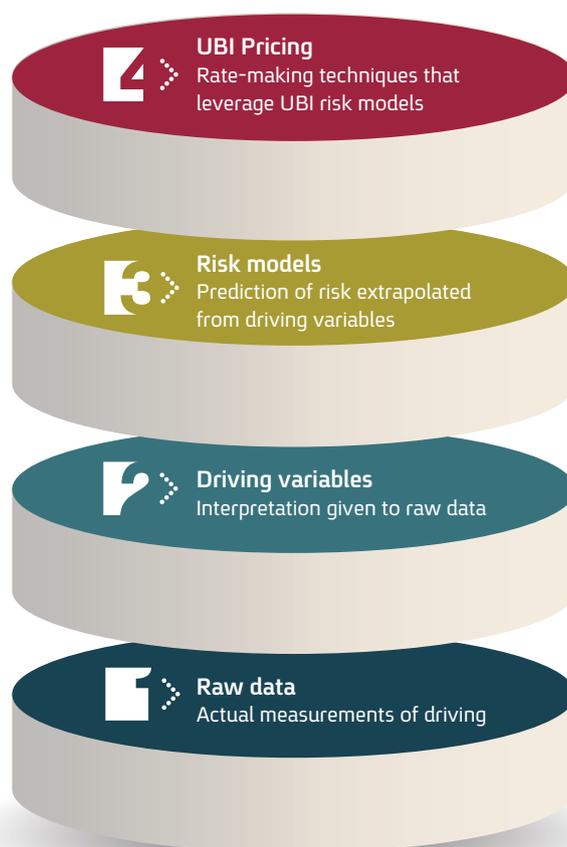
Often called driving events or factors. They represent the first interpretation given to the Raw Data. This interpretation can be done onboard at the moment of raw data collection, for example by setting thresholds that identify braking events in terms of speed drop over time or acceleration value. This interpretation can also be done 'off-line' and off-board by mining through the raw data previously collected, for example calculating a percentage of night time driving mileage in a given month. One of the most valuable aspects of this layer stems from integrating driving variables with contextual data, for example placing a certain braking event in a specific location, road type, weather driver ID, etc.

3. Risk Models

Predictive modeling for extrapolating risk based on the Driving Variables. There are a variety of statistical approaches to modeling, and at the end of the day they all take in variables that could potentially indicate risk on one hand, and correlate them with historic loss data on the other. Some of the challenges of this layer include matching of policy based loss data with VIN based driving data, and selection of applicable loss data – loss cost & frequency of various losses, over different time frames and policies. Because UBI does not stand alone, one needs to decide how to integrate the UBI data with other data and models. This can be done by pooling together all variables (UBI and 'traditional') before risk modeling, which may require deeper changes across the product value chain. Alternatively, many prefer to start with a simple overlay approach.

4. UBI Pricing

Exercising product and marketing considerations, along with price optimization techniques, to determine rates. If one adopts the





simpler overlay approach, it can help communicate the UBI impact on pricing and support behavioral modification. This raises several considerations, such as enrolment discount rates vs. renewal price/discount adjustments, surcharge considerations and more. As price is the main product ‘feature’ that is communicated to the policy holder, UBI pricing will well serve its purpose by addressing issues like privacy, transparency and user empowerment. UBI is often perceived as giving policy holders more control over and visibility into the pricing method.

The most important part of the 4-layer approach is the inter-dependencies that layers have with each other. One cannot assess any one layer without the context of how it serves and impacts the layers around it. For example, when examining an innovative Risk Model, the outcomes depend on the Driving Variable entered into the model as much as on the model itself. Similarly, the ability to extract meaningful and accurate Driving Variables depends heavily on the type and quality of Raw Data collected.

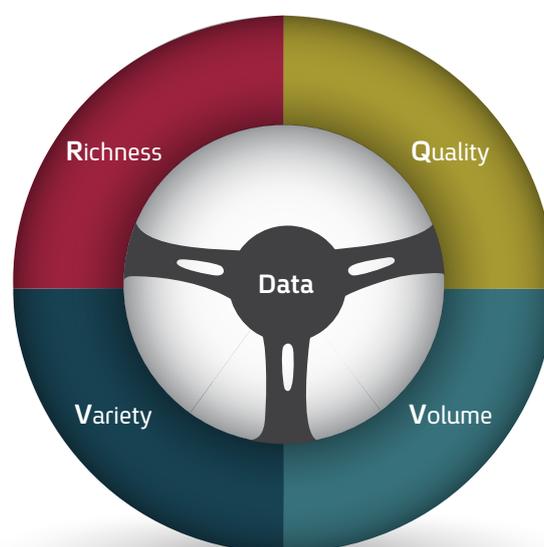
The RQV² Data Approach

The 4 UBI Layers help compartmentalize the various processes. The RQV² Data Approach™ helps ask the right questions, and identify the important tradeoffs. RQV2 stands for DATA: Richness, Quality, Variety and Volume.

Data Richness

In terms of types of data and granularity of data. Data types mean, for example, mileage, braking events, location of events, speed, weather etc. The more types we have the richer the data is. However data richness is not only measured by breadth of data type, but also by granularity of each one. For example, mileage can be measured on 10 second intervals or 1 second.

Data richness also applies to higher layers of the Four UBI Layer approach. For example, a Driving Variable of braking events can be counted using one predetermined threshold or several (despite the granularity of Raw Data). Risk models can be developed in the confines of telematics data or enriched with contextual data (such as road type, weather, etc.). UBI Pricing schemes can be simplified to one ‘black box’ score or communicate (and expose) considerations such as improvements over time.



Data Quality

In terms of how reliably does the data depict the behavior or context it comes to represent. High quality data means that every driving or contextual event that we set out to capture is manifested properly in our data. Data quality can be measured in two aspects:

- a. Showing false or inaccurate data. For example, showing driving data from the wrong vehicle (e.g. not identifying properly which vehicle the data comes from); or showing a braking event that is a result of anything but the actual driving (e.g. pothole or bad sensor reading).
- b. Missing data or data gaps. For example, not detecting part of a trip or not showing relevant weather at the time and location of an event. Because of auto-complete algorithms, missing data often becomes inaccurate data.

In higher layers Quality can be measured using a similar approach. For example, assuming the Raw Data captured a driving event, is there a Driving Variable that shows it? Or do we see Driving Variables of events that never occurred in real life. To an extent, a high quality Driving Variables layer can compensate for some shortcomings of a lower quality Raw Data, but if a driving event was never captured to begin with, there is no super-hero algorithm that can identify it and show a Driving Variable.

Data Variety

In terms of variety of data sources. From hard wired telematics prevalent more in commercial lines and Europe, through self installed dongles, smartphones, combination of smartphone with ‘simpler’ dongle, OEM installed solution, combination of OEM and smartphone, or OEM and aftermarket device (ecall button, visor...). Data variety assumes similar data types (mileage, accelerations...) but from different technology platforms.



In higher layers, Variety can account for using multiple sources for contextual data (e.g. multiple weather providers). In the Risk Modeling layer, Data Variety means integrating Driving Variables that stem from telematics and contextual sources, together with traditional risk factors (such as age, gender, etc.). Similarly in UBI Pricing, Variety measures the different product types or segments. For example, one rate for all UBI customers vs. a large variety of UBI options based on various considerations (e.g. young driver UBI, low mileage UBI, accident forgiveness UBI etc.).

Data Volume

In terms of the amount of data we have to play with. The custom units are vehicle months or years. Total time or mileage can also be used for measuring, however when we get to the Risk Modeling layer, it boils down to the number of vehicles over policy terms. Data volume does not state where the data is being kept (if any), and it can relate to any of the Four UBI Layers. For example, one can capture huge data volumes on-board but communicate only aggregated variables for Driving Variables layer for analysis; or one can collect, transfer and analyze large volumes of data for the training set of Risk Modeling, but keep a much smaller and efficient balance for the UBI Pricing layer.

It's Good To Be Young Rich n' Healthy

Just like we all want to be young rich n' healthy, it's better to have rich, high quality data from a variety of technology platforms, over large number of vehicle years. But why?

In light of the still early stage of UBI modeling, one cannot know what type or resolution of data is most valuable. So in a world without constraints, it is safe to assume that the richer the data (type and granularity) the better our chances of finding the optimal set over time. Similarly, we want to have only high quality data. Because of the ever-changing technology landscape and customer preferences, we don't want to be locked out of any platform. Why take chances if tech giants take over in-vehicle OS faster than OEM and carriers come to terms, or if some killer-app captures imagination of consumers giving the lead to phone based solutions. We prefer our modeling approach to be truly platform agnostic, so we can migrate and integrate easily across dongles, OEMs smartphones and any combination thereof.

And finally, if we could handle it (and afford it...) we want to have all the data volume possible. Just in case an actuarial Einstein will come along in a few years and we can leverage our existing data base to re-slice the data in a new way.

Identify The Important Tradeoffs

Just as we can't all be young, rich n' healthy, we can't have rich, high quality data from a variety of technology platforms, over large number of vehicle years. So the important (and interesting) thing is to manage the tradeoffs. These can be done within each of the 4 UBI Layers or across layers. It helps to first know what tradeoffs are solved by new technology or innovative approaches, so we can make only the hard choices that are really necessary.

For examples, advanced compression algorithms can help overcome tradeoff of data richness vs. cost; pattern recognition algorithms can help increase quality of smartphone based solutions; combination of tech platforms can be designed to compensate for each platform's individual shortcomings; data pooling solutions can help increase and expedite volumes; contextual analysis can help optimize volume-richness tradeoffs; claims focused approaches help overcome volume-cost tradeoffs; business partnerships can help combine best of breed solutions and save time & effort; and more.



Summary

Assessing telematics data and designing UBI models can be daunting. As the industry grows in terms of technology platforms (dongle, mobile, OEM) and vendors (telematics and analytics) the complexity grows exponentially. The Four UBI Layers help identify the building blocks that need to be designed or assessed: Raw Data collection, Driving Variables extraction, Risk Models creation and evaluation and UBI Ratemaking. The methodology of RQV² helps identify the important questions and trade-offs by looking at data: Richness, Quality, Variety and Volume. Implementing the RQV² approach across all Four UBI Layers and mainly to identify inter-layer trade-offs can go a long way to assess telematics data and design UBI models.