



UNIVERSITAT DE
BARCELONA



This is a summary [in Spanish] of the paper that was awarded the
SCOR award for the Iberian Peninsula 2016

For more info on the authors and the research visit:

<http://www.ub.edu/riskcenter>

La revolución telemática: ¿cómo se tarifica el seguro por kilómetro?

Jean-Philippe Boucher^a, Steven Côté^b, Montserrat Guillén^c y Ana M. Pérez-Marín^d

^aDépartement de Mathématiques, UQAM, 201, Avenue du Président-Kennedy, Montréal,
Québec (Canada) H2X 3Y7.

E-mail: boucher.jean-philippe@uqam.ca

^bDépartement de Mathématiques, UQAM, 201, Avenue du Président-Kennedy, Montréal,
Québec (Canada) H2X 3Y7.

E-mail: cote.steven@courrier.uqam.ca

^cRiskcenter, Universidad de Barcelona, Av. Diagonal, 690, 08034 Barcelona (Spain).

E-mail: mguillen@ub.edu.

^dRiskcenter, Universidad de Barcelona, Av. Diagonal, 690, 08034 Barcelona (Spain).

E-mail: amperez@ub.edu

Resumen

Desde hace unos años la telemática está revolucionando el mundo del seguro del automóvil a través de un nuevo concepto de seguro llamado *Pay-As-You-Drive (PAYD)*. Es conocido también como Seguro Basado en el Uso (USB, *Usage based insurance*) porque en vez de establecer una prima anual, la prima se fija en base al número de kilómetros realizados por el vehículo y el patrón de conducción del asegurado. De este modo, aquellos que usan más el vehículo pagarán una mayor prima porque tienen una mayor exposición al riesgo de accidente. En realidad se trata de ofrecer un seguro más personalizado en función del uso real que se hace del vehículo. Este tipo de seguros se basan en la información suministrada por un dispositivo GPS, por lo que el asegurado ha de acceder a la instalación del mismo en su vehículo.

A pesar de que este tipo de seguros se comercializa en aseguradoras en todo el mundo, muchas de ellas no disponen de suficiente histórico que permita validar los modelos predictivos utilizados, que en muchos casos se basan en criterios de algún modo subjetivos apoyados en ideas generalmente aceptadas. Por ello, las compañías de seguros que han apostado por este tipo de pólizas se enfrentan a dos grandes retos: manejar grandes volúmenes de información y saber utilizarla correctamente para la fijación de las primas. La principal innovación que ha supuesto la llegada de la telemática al mundo asegurador ha sido la redefinición del concepto de exposición al riesgo. Tradicionalmente en el seguro del automóvil la exposición al riesgo de accidente se medía a través del tiempo que el asegurado estaba cubierto por su póliza, de modo que un asegurado cubierto durante un año tenía el doble de riesgo de accidente que aquel que estaba cubierto sólo seis meses. Ahora se introduce una nueva dimensión: el kilometraje. Algunos trabajos ya han analizado la relación entre el kilometraje acumulado y la siniestralidad y la mayoría de ellos apuntan a que no se cumple esa regla de proporcionalidad que asumimos que existe para el tiempo de exposición (Boucher et al., 2013). Por su parte, Ayuso et al. (2014) modelan el tiempo transcurrido y el número de kilómetros recorridos antes de una primera comunicación de siniestro para jóvenes asegurados que han contratado un seguro

PAYD, utilizando para ello un modelo de regresión de Weibull. Los autores observan diferencias significativas en función del género, de manera que los hombres tienen un patrón de conducción más arriesgado que las mujeres. Además, también observan diferencias entre los conductores noveles y aquellos más experimentados, concluyendo que los jóvenes conductores forman un grupo heterogéneo de riesgo. Los mismos autores (Ayuso et al., 2016a) concluyen que las diferencias observadas en función del género son en buena medida atribuibles a la intensidad de uso del vehículo, dado que mientras el género tiene un efecto significativo a la hora de explicar el tiempo transcurrido hasta el primer accidente, deja de tenerlo cuando la distancia media diaria recorrida se introduce en el modelo (Ayuso et al., 2016b). Más recientemente, Ayuso et al. (2016c) proponen un método para valorar la influencia en la frecuencia esperada de siniestros de las variables relativas al uso del vehículo, que se formula como una corrección del modelo de tarificación clásico y que permite constatar una vez más la importancia de estas variables. Por último, Ayuso et al. (2016d) proponen incluir la distancia recorrida durante el año como variable offset en un modelo de regresión Poisson cero inflado para predecir el número de siniestros en este tipo de pólizas, demostrando que los falsos ceros están relacionados con la distancia recorrida.

La principal aportación de este trabajo es considerar simultáneamente el efecto del kilometraje y el tiempo de exposición sobre la probabilidad de accidente, abordando esta doble dimensión del problema. Para ello proponemos utilizar los modelos aditivos generalizados (en adelante GAM) propuestos por primera vez por Hastie y Tibshirani (1986), que son una extensión de los modelos lineales generalizados (GLM). Los GAM introducen una mayor flexibilidad a la hora de establecer cuál es la relación funcional entre las variables explicativas y la variable de interés. En los GLM la función de enlace relaciona el valor esperado de la variable de respuesta con el predictor lineal. En los GAM esta función de enlace relaciona ese valor esperado con determinadas funciones de las variables explicativas, que no han de ser lineales necesariamente, y que se estiman bajo un enfoque no paramétrico. De este modo, los propios datos determinan la elección de la forma funcional y por lo tanto de la relación que en realidad existe entre las variables. Esto permite explorar si esa regla de proporcionalidad que asumimos que existe entre el tiempo de exposición y la probabilidad de siniestro se cumple o no, y hacer el mismo análisis para el caso del kilometraje. Este enfoque es el que hemos denominado en nuestro trabajo *modelización con splines cúbicos independientes*.

Para considerar simultáneamente el efecto del tiempo de exposición y el kilometraje sobre la probabilidad de accidente, introducimos en el modelo GAM la interacción entre tiempo y kilometraje a través del producto tensorial de funciones básicas de suavizado (*tensor product smoothing base*). Este segundo enfoque es el que hemos denominado en nuestro trabajo *modelización con producto tensorial*. De esta manera estudiamos el efecto de estas dos variables (consideradas simultáneamente) sobre la siniestralidad.

Los modelos propuestos son completamente útiles y pertinentes para analizar el problema que nos ocupa, por una doble razón. En primer lugar, **porque son totalmente flexibles a la hora de establecer la forma funcional de los predictores**, lo cual permite comprobar si existe o no proporcionalidad entre el tiempo y la siniestralidad (y por otra parte, entre el kilometraje y la siniestralidad) y en caso de no haberla, conocer cómo es esa relación. Y en segundo lugar, **porque permiten introducir la interacción entre tiempo y kilometraje** con esa misma flexibilidad en el modelo, y analizar así el problema en su doble dimensión.

La metodología propuesta se pone en práctica con una base de datos reales de asegurados que han suscrito una póliza PAYD. Los datos han sido facilitados por una aseguradora española y todas las pólizas están referidas al año 2011. Los resultados permiten obtener conclusiones muy relevantes.

En primer lugar, la *modelización con splines cúbicos independientes* pone de relieve la existencia de una relación no proporcional entre el número de siniestros y los kilómetros recorridos. En concreto, se observa que el kilometraje hace aumentar de forma muy pronunciada el riesgo de siniestro hasta llegar a los primeros 10.000 kilómetros recorridos. Superado este umbral, se sigue observando este mismo efecto, pero con un incremento mucho más lento, llegando a estabilizarse en torno a los 20.000 kilómetros. Para kilometrajes muy elevados, se pone de manifiesto incluso un decrecimiento en el riesgo de siniestro, pero la escasez de contratos con tal elevado kilometraje impide acabar de confirmar este extremo. En cualquier caso, lo que sí podemos confirmar es la ausencia de proporcionalidad y el hecho de que el riesgo de accidente parezca estabilizarse después de un cierto kilometraje, lo que sugiere que las personas que conducen muchos kilómetros tienden a desarrollar mejores habilidades de conducción.

Por otra parte, en relación con el tiempo de exposición al riesgo, los resultados del modelo están en contradicción con la práctica tradicional, es decir, asumir que el número de reclamaciones es proporcional a la duración el contrato. En concreto, el tiempo tiene un efecto lineal sólo durante los primeros seis meses de exposición. A partir de ese momento, y hasta la finalización del año de cobertura, si la póliza continúa en vigor el tiempo de exposición incrementa en alguna medida la siniestralidad, pero de forma menos pronunciada, y termina por no tener casi ningún impacto notable después de los diez meses de exposición.

La *modelización con producto tensorial* (que admite una interacción entre el kilometraje y el tiempo) ofrece también resultados relevantes. Por una parte, resulta ser mejor a nivel estadístico que el *modelo con splines cúbicos independientes*. La diferencia entre los resultados de los dos modelos es debida a la mayor flexibilidad del *modelo con producto tensorial* en el que existe una forma añadida de dependencia entre tiempo y kilometraje. Sin embargo, la mejora en las predicciones es muy pequeña y la interacción añadida a la modelización hace que la desviación estándar del ajuste sea mucho más alta. Por otro lado, los resultados obtenidos con ambos modelos son en general bastante similares, por lo que llegamos a las mismas conclusiones comentadas anteriormente sobre el efecto del tiempo y el kilometraje en la siniestralidad.

En la última parte del trabajo, se lleva a cabo una comparación entre los resultados obtenidos con los dos modelos GAM ajustados y un modelo GLM clásico (ampliamente utilizado en la práctica). Los tres modelos se ponen a prueba sobre una muestra de validación, compuesta por observaciones no utilizadas en la estimación de los distintos modelos. También se introducen otras variables explicativas en el predictor lineal, a parte del tiempo de exposición y el kilometraje (en concreto, la edad del asegurado, antigüedad

del vehículo, sexo y tipo de aparcamiento), y se concluye que sólo la edad del conductor resulta tener efecto significativo. En base a nuestros resultados, desarrollamos un prototipo de tarificación para un seguro PAYD utilizando los tres modelos. Este sistema de tarificación permite determinar la prima correspondiente a un valor de tiempo de exposición y kilometraje concretos, fijados de forma simultánea. Como resultado obtenemos que las primas producidas por los modelos GAM son similares entre sí y diferentes de las obtenidas con el modelo GLM clásico, y que estas diferencias se dan en general en cualquier grupo de edad considerado.

La aplicación práctica de la metodología y los resultados obtenidos en este trabajo es directamente trasladable al mercado. Proporcionamos un **estudio avanzado y en profundidad sobre la determinación de una estructura de tarifa para un producto tan innovador como el seguro PAYD.** Además, también se aporta una reflexión en profundidad sobre el papel que juega no solo el kilometraje, sino también el tiempo de exposición en el riesgo de accidente, demostrando que esa suposición de proporcionalidad de la relación entre tiempo y siniestralidad (que es automáticamente aplicada en la práctica), está lejos de ser cierta. Por todo ello, consideramos que **este trabajo es de gran interés no sólo para aquéllas entidades que desean introducir sistemas de pago por uso, sino también para los profesionales del sector asegurador en general.**

La metodología presentada en este trabajo tiene aplicaciones más allá de las mostradas aquí. En concreto, debido a su mayor flexibilidad puede ser utilizada en la tarificación del seguro del automóvil clásico para explorar más la relación entre la siniestralidad y los factores de riesgo. En el ámbito de los seguros PAYD futuras investigaciones podrían valorar la aportación de los modelos GAM a la hora de considerar otros datos de telemetría en la conducción, como el número de frenadas y aceleraciones bruscas u otros elementos respecto al consumo del vehículo, sin necesidad de introducir factores como la hora del día y el lugar donde se lleva a cabo la conducción, que suelen ser datos sensibles y los asegurados pueden considerar que atañen a su privacidad. Por otra parte, la dependencia entre los diferentes tipos de accidentes también se podría estudiar con esta metodología, así como realizar un análisis temporal a fin de agregar los contratos del mismo asegurado. En tales situaciones, los modelos aditivos generalizados para la ubicación, escala y forma, denominados *GAMLSS* (propuestos por Rigby y Stasinopoulos, 2005), deberían ser explorados.

Referencias

- Ayuso, M., Guillén, M. et Pérez-Marín, A. M. (2014). Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. *Accident Analysis & Prevention*, 73, 125-131.
- Ayuso, M., Guillén, M. and Pérez-Marín, A. M. (2016a) "Using GPS data to analyse the distance travelled to the first accident at fault in pay-as-you-drive insurance" *Transportation Research Part C* 68 (2016) 160–167.

- Ayuso, M.; Guillén, M.; Pérez-Marín, A. M. (2016b) “Telematics and gender discrimination: some usage-based evidence on whether men's risk of accident differs from women's”, *Risks* 4, 10; doi: 10.3390/risks4020010.
- Ayuso, M., Guillén, M. and Nielsen, J. P. (2016c) “Improving automobile insurance ratemaking thanks to telematics: mileage and other driving behaviour factors”, *submitted*.
- Ayuso, M., Guillén, M. and Nielsen, J. P. (2016d) “Distance travelled as a risk factor when predicting motor insurance claims”, *submitted*.
- Boucher, J.-P., Pérez-Marín, A. M. et Santolino, M. (2013). Pay-as-you-drive insurance : the effect of the kilometers on the risk of accident. *Anales del Instituto de Actuarios Españoles*, 3a Época, 19, 135-154.
- Hastie, T. y Tibshirani, R. (1986). Generalized additive models. *Statistical science*, 297-310.
- Rigby, R. A. y Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507-554.