How does granularity affect motor insurance claim predictions in a telematics setting?

Insurance Data Science Conference

Juan Sebastian Yanez, Universitat de Barcelona
Montserrat Guillén, Universitat de Barcelona

June 20, 2025

Bayes Business School

Risk center

Motivation

Motivation

Telematics data:

- Telematics information is collected via a device installed in the vehicle or a mobile application.
- Some variables is geographic (e.g. congestion)
- Others are behavioral (e.g. speed)
- Others provide better risk exposure measures (e.g. distance).

Claim counts vs telematics data

Telematics data:

- Can be collected in a micro-scale
- Can take many forms
- Puntual risky events are observed frequently
- Usually collected for a small sample of drivers

Claim counts:

- Rare events
- Can take into account experience rating but may take time
- Usually available from large data sets (government agencies or insurance companies)

Goals and contributions

- Is it advantageous to work with micro-scale telematics data?
- If telematics data are aggregated how much information is lost?
- How does granularity affect different models for claim frequency?

Results

Dataset description

Key elements

- Drivers: 17,405,
- Exposure: 2 years (2017 et 2018),
- Telematics data: mileage by zone, speed limit violations (among other variables),
- At-fault claims: 1,700.

Dataset description

Variable	Tag	Description
$E_{i,j}$	Years	Exposure in years
$T_{i,j}$	Km all	Total distance (all zones)
$X_{i,j}$	Vehicle Power	-
	Gender Male	-
	Driver Age	-
$Z_{i,j}$	Km at Night %	% of km traveled at night
	Km overspeed %	% of km traveled over the speed limit
	Km urban %	% of km traveled at urban areas

Results

		PO	NBI	NBII	ZIP	CANN
MSE	Years	0.05034320	0.05034439	0.05034257	0.05029249	0.05026365
	Trimesters	0.05030701	0.05030743	0.05030421	0.05027107	0.05026191
	Months	0.05030367	0.05030405	0.05029935	0.05029168	0.05030466
RSE	Years	0.99101254	0.99103591	0.99100005	0.99001431	0.98944662
	Trimesters	0.99030005	0.99030842	0.99024499	0.98959258	0.98941236
	Months	0.99023435	0.99024184	0.99014925	0.98999833	0.99025388

Out-of-sample MSE and RSE for models trained on 2017 data and tested on 2018 data

Conclusion

Conclusion

- Granularity affects models differently
- Simple models benefit from a higher degree of granularity
- Complex models do not benefit from a higher granularity

For the future:

- Add random effects for longitudinal data.
- Consider binary models for higher granularity.
- Use more diverse types of telematics data.
- Consider seasonality (e.g., winter)

Thank you for your attention!

- Duval, F., Boucher, J. P., & Pigeon, M. (2024). Telematics combined actuarial neural networks for cross-sectional and longitudinal claim count data. ASTIN Bulletin: The Journal of the IAA, 54(2), 239-262.
- Verbelen, R., Antonio, K., & Claeskens, G. (2018). Unravelling the predictive power of telematics data in car insurance pricing. Journal of the Royal Statistical Society Series C: Applied Statistics, 67(5), 1275-1304.
- Yanez, J. S., Guillén, M., & Nielsen, J. P. (2024). Weekly dynamic motor insurance ratemaking with a telematics signals bonus-malus score. ASTIN Bulletin: The Journal of the IAA, 1-28.