**­Commentaires de Mario Wüthrich**

1. The present version of this article seems too long, and the manuscript would benefit from a more concise style. In the present format, the printed version of the article should not be longer than roughly 20 pages, and additional material can be moved to an online supplementary. E.g., lengthy tables with numbers are often not studied by readers, therefore, it is sufficient to provide them in the supplementary, unless they contribute to a deeper understanding of the article.

We acknowledge that the paper may be lengthy. To address this, we significantly trimmed down Section 3 by removing parts that were judged non-essential for the paper's core message. Additionally, we decided to remove Algorithm 2, which appeared somewhat redundant. Additionally, have relocated lengthy Table 5 to the project’s GitHub page of the project, where all the necessary code to replicate the results is also available.

1. Please remove the table of contents.

Yes.

1. Personally, I find the argument for not using the exponential output function, but the softplus function, not very convincing. Firstly, pre-processing inputs, e.g., MinMaxScaler, should take care of stability issues, and you may additionally use normalization layers. Secondly, I would like see whether your network provides unbiased estimates and predictions, and verifying the auto-calibration property would even be better.

All our CANN models include batch-normalization layers between fully connected layers, including one immediately after the input layer to normalize the inputs. We acknowledge that this might not have been clear in the manuscript, so we have added relevant information (page 18) to address this. Even with normalized inputs and batch-normalization layers, we encountered issues with the exponential function as the output activation, leading to excessively high predictions for some contracts. We therefore opted for the SoftPlus function, which is more stable numerically and resolved the issue.

Following your suggestion, we verified the auto-calibration property in Table 9. It turns out that the CANN models are not too bad, overshooting by only about 3-4% on a holdout sample. This issue can be addressed using, for instance, techniques found in your paper “The balance property in neural network modelling”. We have added relevant details and references about this topic at the end of subsection 5.1 (page 20).

**Reviewer #1**

The authors discuss possible use of CANN for longitudinal claim count data with telematics features. As I understand, the main contributions of the manuscript to the actuarial literature are two-fold

1. (Presumably) the first attempt to understand impacts of telematics features in a longitudinal data, and
2. More sophisticated usage of telematics features to detect possible non-linearlity via CANN.

While the manuscript is generally well-written, I have some concerns on the novelty of the proposed framework and the specification of the empirical analysis. Therefore, I would like to suggest the authors to consider the following comments to emphasize their contributions to the actuarial community.

1. The most concern that I have is, whether the proposed framework has enough novelty to be considered for publication in ASTIN Bulletin. As the authors very well know, each components of the proposed framework such as MVNB distribution, CANN, and use of telematics features are already extensively considered in the actuarial literature so there is little room for technical advancements for these topics (at least under the structure of the manuscript). For example, I am not sure Algorithms 1,2, and 3 are necessary as these are basically implementations of fitting a neural network with pre-specified architecture (for example, Figures 2 and 3) and loss function (negative log-likelihood in this case). In this regard, I wonder if the authors can add more reproducibility of their analysis by providing the R or Keras codes in the appendix that correspond  to Algorithms 1,2, and 3 (like SSRN 3320525) and a 'fake' sample dataset to run the codes.

We believe that the paper’s novelty lies in the integration of a longitudinal model into the framework of the combined actuarial neural network. Implementing a longitudinal model in Torch (or TensorFlow) presents challenges as these frameworks are primarily designed for tasks assuming independence between observations. Consequently, we had to make modifications to the standard training loop, as coded in the “train” method of the “MVNBMLP” R6 class, available in the GitHub repository (<https://github.com/francisduval/article_3_count/blob/master/R/MVNB%20models/MVNBMLP.R>). While we specifically tested the implementation of the MVNB distribution, it is important to note that the approach presented in the manuscript can be readily extended to other longitudinal distributions, such as the beta negative binomial (for further details, see “Turcotte, R. and Boucher, J.-P. (2023). Gamlss for longitudinal multivariate claim count models”). We acknowledge that this generality was not explicitly stated in the manuscript, and we have addressed this by providing additional information in Section 6 (pages 21 and 22).

We acknowledge that Algorithms 1, 2, and 3 are somewhat redundant. We have chosen to remove Algorithm 2, while retaining Algorithms 1 and 3 to emphasize the distinction between the cross-sectional and longitudinal training procedures.

We have made all the code, including the Torch code, available for replicating the analysis in a referenced GitHub repository (<https://github.com/francisduval/article_3_count>). A reference is made at the end of the first paragraph of Section 6 (page 22).

1. I think the telematics data pre-processing schemes between the log-linear model and CANN model are somewhat 'unfair'. Compared to the telematics features used in the log-linear model as in Table 4, the global telematics vector described in page 19 contains richer information due to more sophisticated data pre-processing. In this regard, to assess the effects of CANN model in a fair way, I think it may be better to use the global telematics vector described in page 19 for the log-linear model as well (or add additional benchmark that uses the global telematics vector as it is, on top of the current log-linear telematics model.)

While the global telematics vector contains richer information compared to the handcrafted telematics features in Table 4, the log-linear models may not have the capacity to effectively leverage such detailed low-level data. This is why we computed handcrafted telematics features—to ensure that the log-linear models and CANN models are on an equal footing. We have, however, taken your advice and added additional log-linear model benchmarks using the global telematics vector (page 19).

1. As the authors mentioned, one of the major concerns of CANN model could be lack of interpretability. In this regard, I wonder if the authors could consider a CANN-Boost model as another benchmark - which fixes beta as the plug-in estimate from the GLM but only updates theta over time via neural network. In doing so, one can preserve the interpretability of the original relativity as well as considering additional interaction or non-linearity via CANN.

Following your suggestion, we have included CANN models with fixed beta parameters in our model lineup. These models perform quite similarly to the standard CANN models (those with trainable beta parameters). Our recommendation is to use the CANN models with fixed beta parameters since they offer improved interpretability and slightly faster training. The fixed beta parameters used are the coefficients obtained with the MVNB log-linear model, which considers dependency between contracts. We have included some additional details in the manuscript to cover this topic (pages 19).

1. I think the improvement in prediction performance with CANN model would come with computational costs. So can you report the computation times both for the benchmarks and the proposed model?

We have added some text in subsection 5.1 (page 20) to address this topic. For the Poisson specification, the log-linear model had a training time of about 20 seconds, while the CANN model required about 24 minutes of computations.

1. I think the representation of the global telematics vector in pages 18 and 19 is somewhat similar to that of So et al. (2021, <https://www.mdpi.com/2227-9091/9/4/58>). Also, Jeong (2023, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4281943>) considered dimension reduction of telematics features - where the original representation is from So et al. (2021) - via embedding and PCA. In this regard, I think the authors could consider to add these paper as additional references.

We have added these two references and thank you for the suggestion (page 2).

1. Page 23: I think h\_22 and h\_2 shall be $h\_22$ and $h\_2$ here to maintain consistency.

We are not sure to understand this comment.

1. The authors use theta in two places in different meanings; one as the NN parameter in the CANN architecture and the other as the random effect in (3.16). So please change one of these to another letter to avoid confusion.

Yes, we have replace the random effect theta with the Greek letter psi.

1. It was interesting to see that despite the presence of telematics features, serial association among the contract of the same vehicle was not negligible. Ideally, if the (currently given) telematics features are enough to explain the unobserved heterogeneity of the primary driver of the vehicle, then use of random effect would not be necessary and one only needs to analyze the dataset with cross-sectional model - which is not the case in the analysis results of the manuscript and the longitudinal model still outperforms the cross-sectional model. I think the authors may a bit more of discussion on this (impacts of telematics features in relation to longitudinally observed data) as it could be one of the main contributions of the paper.

This is indeed a good point. The superior performance of longitudinal models over cross-sectional ones, even when telematics data are used, implies that traditional and telematics data alone may not capture the full extent of the dependence between observations due to unobserved variables. We have addressed this by incorporating further discussion in subsection 5.1 (page 20) and section 6 (page 22).

1. Some words in the references are not properly capitalized, for example, poisson, r&d, and cann.

Yes, we have fixed that.

**Reviewer #2**

The article presents some cross-sectional and longitudinal claim count models for car insurance based on the Combined Actuarial Neural Network (CANN) framework. While this article is generally well-organized and there is no serious flaw in the mathematical framework itself, I am not quite sure whether the manuscript has sufficient contributions to the actuarial literature to be published in ASTIN Bulletin: The Journal of the IAA.

1. In the conclusion part, the authors say they develop three novel claim count modes based on CANN, i.e., Poisson, NB, MVNB. Actually the first two models have already been used in the actuarial literature, see e.g.

Yes, we have changed that (page 21).

1. Please make a rough comparison of the networks with different hidden layers and neurons to show the impact of different choices on the model performance.

Unfortunately, our computational resources are limited, which prevented us from exploring a wide range of hyperparameter combinations. We made the deliberate choice to adopt a specific architecture with three hidden layers having 128, 64, and 32 hidden units, respectively, which is reasonable given the dimension of the input vector. Our primary focus has been on fine-tuning other hyperparameters, namely the learning rate, dropout rate, and the factor in the learning rate scheduler. We acknowledge that one may obtain better results using other hyperparameters combinations. However, the goal of the paper is not to have the best improvement possible, but rather to show that there is an improvement over the baseline model.

1. In the application part, please indicate whether the exposure for each contract is one year. Otherwise the exposure term should be considered in the model construction.

We have chosen to treat the exposure as a standard covariate, and additional clarification on this point can be found in the text added on page 15.

1. As mentioned in the conclusion, I wonder if the authors could include some more longitudinal models and conduct a more comprehensive comparison among them.

While we specifically tested the longitudinal model implementation for the MVNB distribution, it is important to note that the approach presented in the manuscript can be readily extended to other longitudinal distributions. Additional information has been added in the conclusion (page 22) to highlight this. In fact, one could leverage Algorithm 2, the parameter estimation procedure for the MVNB CANN model, or the code that can be found on the project’s GitHub page (<https://github.com/francisduval/article_3_count>), to implement other longitudinal models, like the beta-binomial distribution (see “Turcotte, R. and Boucher, J.-P. (2023). Gamlss for longitudinal multivariate claim count models”).

1. Some features should be pre-processed (e.g. MinMaxScaler) before entering the neural network.

All our CANN models include batch-normalization layers between fully connected layers, including one immediately after the input layer to normalize the inputs. We acknowledge that this might not have been clear in the manuscript, so we have added relevant information (page 18) to address this.

1. The authors use the softplus function instead of the exponential function in the output layer. The exponential function is often adopted in insurance because of its multiplicative property. I wonder if the authors have compared these two options based on real data.

Even with normalized inputs and batch-normalization layers, we encountered issues with the exponential function as the output activation, leading to excessively high predictions for some contracts. We therefore opted for the SoftPlus function, which is more stable numerically and resolved the issue.

1. The title of Section 5 is not appropriate: Analyzes → Analysis: Yes
2. Page 4 line 40: is only given → only gives: Yes
3. Page 5 line 22: a dependence → dependence : Yes
4. Page 5: The title of Table 1 has two full stops. Same mistakes for the following tables and figures: Yes
5. Page 7: generalize of → the generalization of: Yes
6. Page 8: Equation (3.4): yit! → ln(yit!) : Yes
7. Page 12: There is no wˆφ in the Poisson CANN model in Algorithm 1: Yes
8. Page 12: The formula for computing the empirical risk is incorrect: Yes
9. Page 20 line 15: four vectors → five vectors: Yes
10. Page 22 line 19: significative → significant : Yes

**Reviewer #3**

The paper is very well written and educational. The explanations are very clear and comprehensive, with the drawback that they are sometimes (too?) long for a scientific journal such as Astin Bulletin. However, in all subjectivity, I think that such a detailed and complete presentation is in fact an excellent thing to bring more practitioners back to journals such as Astin Bulletin. So, I would be in favour of keeping the presentation as it is, even if this is debatable as it is a rather subjective point of view. The paper presents claim count models based on the Combined Actuarial Neural Network (CANN) approach. The novelty of the paper lies in the application of the MVNB distribution by incorporating it into the neural network framework, specifically the CANN architecture, for modeling longitudinal claim count data.

I have three main comments:

1. It is undoubtedly regrettable to reject so easily the approach that involves drawing upon domain knowledge to craft telematics features from raw data, as is done on page 3. Of course this approach has its drawbacks, which are well described on page 3, but it can also have its advantages, such as its simplicity and ease of interpretation, or the processing of a posteriori variables.

We genuinely did not intend to reject the “domain knowledge” approach. As you rightly point out, it carries its own benefits (which likely explains why many insurers still use it in their telematics programs). We included additional information (see page 2) to emphasize these benefits and underscore that our data-driven approach is not meant to replace the domain knowledge approach but rather to serve as an alternative or complement.

1. Section 4.2 is very brief compared to the rest of the paper. However, choices that may seem important are made in this section (for example, see lines 34-36 on page 20). The choices made in section 4.2 should be better justified.

We have improved subsections 4.2 and 4.3 by providing justifications for our choices, including the number of hidden layers/units, the activation function, batch normalization, dropout, and the learning rate. You can find these justifications on page 18.

1. Similarly, in section 4.3 (although the rest of the paper is intended to be pedagogical and accessible to a wide audience, due to the long and comprehensive explanations of the methods used), the a priori values chosen (lines 8-11 on page 21) for the hyperparameters lstart, factor and p should be better justified.

Please refer to the previous comment. We have added justifications on page 18.

The results are convincing, and it is undeniable that the approach here allows dependen- cies between contracts of the same vehicle to be taken into account, which is a nice property of the approach. Moreover, the approach implemented is at the cutting edge of machine learning techniques used in insurance for supervised learning. Even if the paper does not shed any new methodological light on the actuarial community, in that it consists of estimating MVNB models using specific neural networks (CANN), the application is very interesting for the actuarial community, so that I would recommend the paper for publication in Astin Bulletin.