Appendix E: Robustness Tests.



Figure E1. Scatterplot of Predicted Values of the Dependent Variable versus Residuals. Left-hand plot plots estimates for the untransformed overclaiming data, the right-hand plot for the log-transformed data. N.B Scales are necessarily different.



Figure E2. Scatterplot of Dependent Variable versus estimated distance from regression line using approach of Verardi & McCathie (2012). Observations classified as outliers by Verardi & McCathie (2012) estimator are plotted in light blue, other observations in black. Left-hand plot reports results for untransformed data, right-hand plot for log-transformed data. In the untransformed data clear that the largest overclaims are all classified as outliers, unlike for the transformed data.





References:

Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo. *The Annals of Statistics* 1: 799-821.

Neelon, B., O'Malley, A. J., & Smith, V. A. (2016). Modeling zero‐modified count and semicontinuous data in health services research part 1: background and overview. *Statistics in Medicine, 35*(27), 5070-5093.

Verardi, V., & McCathie, A. (2012). The S-Estimator of Multivariate Location and Scatter in Stata. *The Stata Journal*, *12*(2), 299–307. https://doi.org/10.1177/1536867X1201200208