Machine learning improves accounting estimates: evidence from insurance payments

Kexing Ding, Baruch Lev, Xuan Peng, Ting Sun, Miklos A. Vasarhelyi

Review of Accounting Studies, 2020.

叶鑫 2021/03/25

Outline

- Introduction
- Research design
- Empirical result
- Additional analyses
- Conclusion

Motivation

- Most financial statement items are based on subjective managerial estimates which are very difficult to audit.
- There is an urgent need to provide an alternative generator of estimates.
- Machine learning has the potential to provide an independent estimates generator(Accounting fraud etc.).

Machine learning may improve the estimate of an account balance, alleviating both intentional and unintentional errors.

Related researches and Innovation

- Perols (2011) and Perols et al. (2017) are among the first in accounting to predict accounting fraud.
- Bao et al. (2020) and Bertomeu et al. (2020), used various accounting variables to improve the detection of ongoing irregularities.
- Barboza et al. (2017) compared several machine learning models with traditional models and found that boosting, bagging, and random forest algorithms provide better prediction performance.
- This research firstly establish the potential of machine learning to independently assess the reliability of estimates underlying financial reports, thereby improving the quality and usefulness of financial information.

Research design

- Research object: Insurance companies' data on estimates and realizations of loss reserves (estimates of future claims related to current policies)

- Research method: Compare four popular machine learning algorithms(linear regression, random forest, gradient boosting machine, and artificial neural network) to predict insurance losses with the actual managerial loss estimates in financial reports.

- Research conclusion: Machine learning are superior to the managerial loss estimates (with a few exceptions).

- Machine learning has the potential to substantially improve auditors' ability to evaluate accounting estimates, thereby enhancing the usefulness of financial information to investors.

Research Data: Sample

- The data were extracted from the SNL FIG website, covering the period 1996 to 2017.
- We focused on five business lines: (1) private passenger auto liability, (2) commercial auto liability, (3) workers' compensation, (4) commercial multi-peril, and (5) homeowner / farm owner.
- For each business line, we only kept observations that had positive total premiums and cumulative paid losses.
- In the final sample, we have a total of 32,939 line-firm-year observations for all five business lines combined.

Table 3 Cumulative payment percentage in the first five years for each business line

Business Line	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5
Private Passenger Auto Liability	40.64%	72.44%	86.76%	94.20%	97.61%	99.06%
Commercial Auto Liability	25.03%	50.74%	70.90%	85.57%	93.88%	97.70%
Workers' Compensation	24.99%	56.11%	72.90%	83.20%	89.09%	93.03%
Commercial Multi-Peril	44.52%	69.22%	80.03%	88.58%	93.85%	97.11%
Homeowner/Farmowner	72.62%	93.50%	96.83%	98.58%	99.48%	99.82%

Research Data

Dependent variable: We measured the ActualLosses for an accident year t as the 10-year cumulative payment of losses incurred in year t. This variable was extracted from the financial report in year t + 9.

Years in which losses were incurred	Incurred net losses and defense and cost containment expenses reported at year end ($000 omitted)										Developme
	1	2	3	4	5	6	7	8	9	10	11
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	One year
1. Prior											
2. 1999	4583	4615	4614	4615	4615	4617	4617	4618	4618	4618	
3. 2000	XXX	4382	4450	4409	4407	4413	4411	4419	4422	4422	
4. 2001	XXX	XXX	4845	4863	5012	5016	4909	4904	4905	4904	−1
5. 2002	XXX	XXX	XXX	7463	7270	7064	7718	7169	7136	7147	11
6. 2003	XXX	XXX	XXX	XXX	18,904	18,091	18,033	17,710	17,465	17,479	14
7. 2004	XXX	XXX	XXX	XXX	XXX	18,201	15,408	15,301	14,754	14,727	−27
8. 2005	XXX	XXX	XXX	XXX	XXX	XXX	24,097	20,611	23,627	24,554	927
9. 2006	XXX	XXX	XXX	XXX	XXX	XXX	XXX	23,828	21,900	21,993	93
10. 2007	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XXX	24,226	24,334	108
11. 2008	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XXX	36,893	XXX

Dependent variable

Years in which losses were incurred	Cumulative paid net losses and defense and cost containment expenses reported at year end ($000 omitted)										11 Number claims close with loss pa
	1	2	3	4	5	6	7	8	9	10	
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	
1. Prior	0										
2. 1999	3708	4548	4608	4614	4615	4617	4617	4618	4618	4618	1125
3. 2000	XXX	3486	4293	4393	4397	4403	4404	4404	4422	4422	1683
4. 2001	XXX	XXX	3736	4537	4716	4861	4903	4903	4904	4904	1154
5. 2002	XXX	XXX	XXX	5354	6884	6987	7045	7060	7111	7122	1753
6. 2003	XXX	XXX	XXX	XXX	15,926	17,281	17,141	17,326	17,424	17,438	5985

Research Data

Independent variable:

- Consist of information already known at the year t (no look-ahead bias).
- Operational variables (e.g., claims outstanding, premiums written, or premiums ceded to reinsurers)
- Company characteristics (e.g., total assets or state of operation) for the accident year.
- Exogenous environmental variables (e.g., inflation or GDP growth).

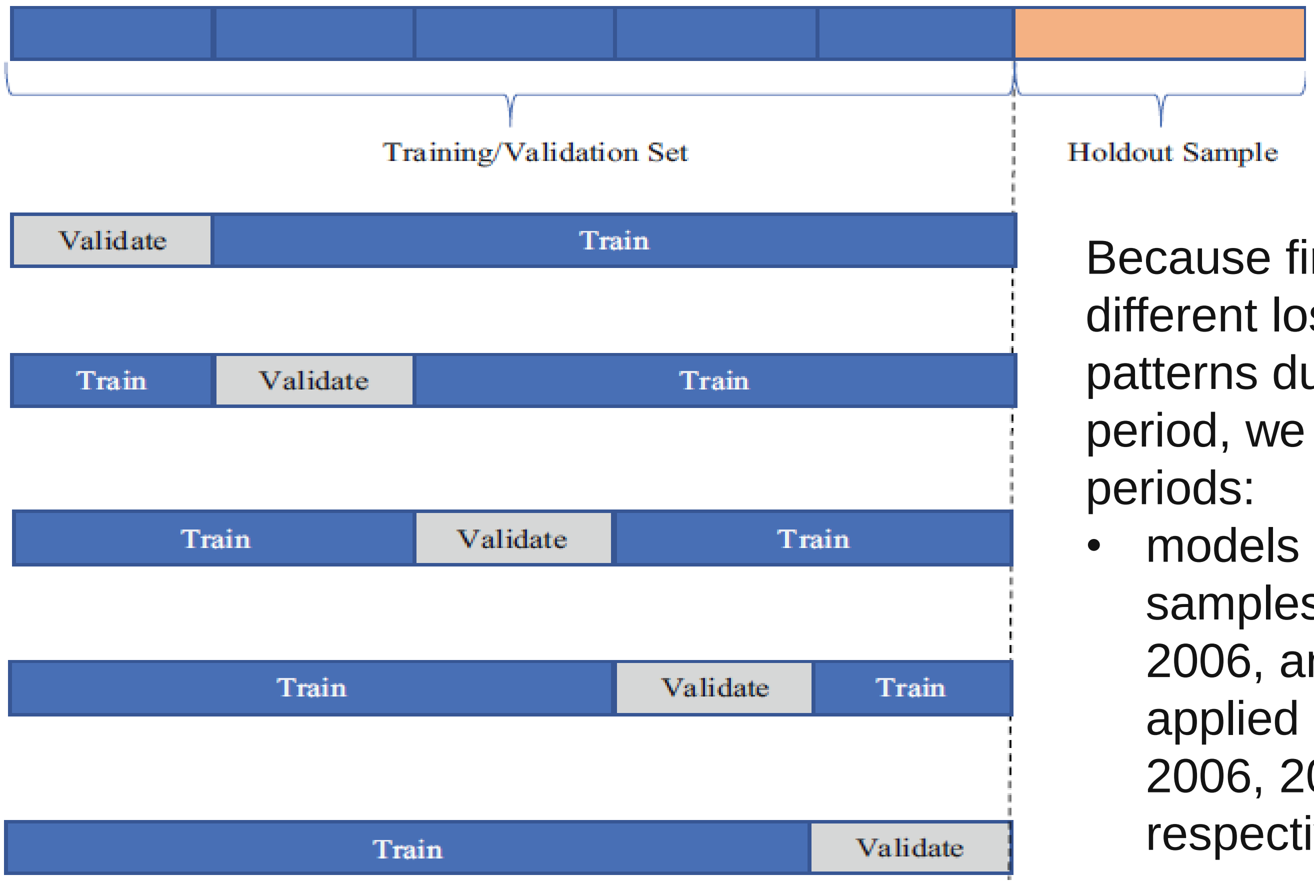
Business Line Operational Variables	
Outstclaim	Cumulative claims outstanding for the current accident year
Reportedclaim	Cumulative reported claims for the current accident year
PaidClaim	Number of loss claims closed with payment
UnpaidClaim	Number of loss claims closed without payment
LinePremiums	Premiums written in the current accident year on the business line
PremiumsCeded	Premiums ceded to reinsurers
LinePayment	Total payments for the current accident year
PaymentCeded	Payments ceded to reinsurers
LineDCC	Defense and cost containment payments direct and assumed
DCC Ceded	Defense and cost containment payments ceded
SSR	Salvage and subrogation received
PaidLoss	Total losses paid for the current accident year

Research Method: Machine Learning

Linear regression	
Learning rate	0.1, 0.01, 0.001, 0.0001
Number of iterations	10, 50, 100, 500, 1000
Random forest	
Number of trees	50, 100
Maximum depth of the tree	20, 30, 50
Minimum leaf size	1, 5, 10, 50
Gradient boosting machine	
Number of trees	50, 100
Maximum depth of the tree	5, 10, 20, 30, 50
Minimum leaf size	1, 5, 10, 50
Artificial neural networks	
Activation function	Rectifier, Tanh, Max out, Rectifier with Dropout, Tanh with Dropout, Max out with Dropout
Number of hidden layers	2,3, 4
Number of nodes	The first layer had a number of nodes equal to the number of independent variables, in each additional layer the number of nodes decreased by approximately 50%
Number of epochs	10, 50, 100
Learning rate	Examined at 100 possible learning rates, scaled from 0.0001 to 0.000001.

Research Method: Machine Learning

For each algorithm, we developed machine learning models using the fivefold cross-validation method and used the holdout set to evaluate the practical usefulness of the models.



Because firms may experience different loss claim and payment patterns during the financial crisis period, we construct three test periods:

- models developed from training samples of 1996–2005, 1996–2006, and 1996–2007 were applied to the holdout sets in 2006, 2007, and 2008, respectively.

Research Method: Control Variable

We conducted two sets of tests for each line of insurance (private passenger auto liability, commercial auto liability, etc.).

- The first set of tests did not include managers' loss estimates as an input attribute.
- In the second set of tests, we added managers' initial estimates to the machine learning models.
- This design enables us to evaluate the performance of machine learning techniques on a standalone basis as well as when managers' inputs are incorporated into the algorithms.

- **Evaluation metrics:** We used two metrics to compare estimates with actuals: the mean absolute error (MAE) and the root mean square error (RMSE).

$$MAE = \frac{1}{n} \sum_{j=1}^{n} \left| TrueValue_j - ModelEstimate_j \right|.$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left(TrueValue_j - ModelEstimate_j \right)^2}$$

Empirical result

- The random forest algorithm produced good predictions for four lines and the linear regression model performed well for the fifth homeowner/farm owner line.
- Machine learning models generate more accurate loss predictions than managers in most circumstances.
- In general, models incorporating manager's estimate have higher predictive accuracy.

Business line	Training/Validation Sample	Obs	Managers' estimates		Machine learning without manager estimates				Machine learning with manager estimates			
			MAE	RMSE	MAE	RMSE	Accuracy edge (MAE)	Accuracy edge (RMSE)	MAE	RMSE	Accuracy edge (MAE)	Accuracy edge (RMSE)
Private Passenger Auto Liability					**Random forest**				**Random forest**			
	1996–2005	5949	9461	37,494	8213	34,687	13%	7%	7758	36,071	18%	4%
	1996–2006	6298	9793	38,266	7848	34,547	20%	10%	7220	30,305	26%	21%
	1996–2007	6602	9575	37,940	7869	35,047	18%	8%	6902	30,220	28%	20%
Homeowner/Farmowner					**Linear regression**				**Linear regression**			
	1996–2005	6121	3905	16,789	5674	22,069	−45%	−31%	4402	16,359	−13%	3%
	1996–2006	6544	3878	16,611	5687	21,070	−47%	−27%	4203	16,201	−8%	2%
	1996–2007	6946	3962	16,826	5548	21,269	−40%	−26%	4321	16,674	−9%	1%

Empirical result: Holdout tests

- Overall, the findings are consistent with the cross-validation results, indicating that machine learning models have superior predictive accuracy.
- After we added manager estimate to the model, its performance improved.

Panel A Holdout test results

Business line	Holdout Sample	Obs	Managers' estimates		Machine learning without manager estimates				Machine learning with manager estimates			
			MAE	RMSE	MAE	RMSE	Accuracy edge		MAE	RMSE	Accuracy edge	
							(MAE)	(RMSE)			(MAE)	(RMSE)
Private Passenger Auto Liability					Random forest				Random forest			
	2006	670	9337	36,120	8434	35,271	10%	2%	8083	33,037	13%	9%
	2007	659	9435	36,221	8390	37,857	11%	−5%	7670	33,500	19%	8%
	2008	637	10,616	50,851	8507	41,440	20%	19%	8664	39,732	18%	22%
Commercial Auto Liability					Random forest				Random forest			
	2006	620	3852	17,287	3679	14,527	4%	16%	3475	14,468	10%	16%
	2007	609	3288	13,481	3056	10,413	7%	23%	2912	10,228	11%	24%
	2008	592	4219	19.361	3216	9638	24%	50%	3268	11.353	23%	41%
Homeowner/Farmowner					Linear regression				Linear regression			
	2006	697	2964	12,231	5219	14,457	−76%	−18%	3413	11,225	−15%	8%
	2007	692	3525	14,042	5202	15,297	−48%	−9%	4064	13,748	−15%	2%
	2008	678	5565	24,434	7968	23,580	−43%	3%	5628	20,881	−1%	15%

Additional analyses: Estimation errors

- We defined managers' estimation error as the reported loss estimate minus the actual loss, scaled by total assets.
- We focused on the signed estimation errors, instead of absolute errors, as in the previous section.
- We used the holdout prediction results generated by random forest models for the years 2006, 2007, and 2008 to calculate model estimation errors.

	ManagerError		ModelError (Machine learning without manager estimates)					ModelError (Machine learning with manager estimates)			
	Mean	Median	Mean	Median	Mean Diff.	*p* value		Mean	Median	Mean Diff.	p value
Panel A: Comparing Manager Estimation Errors and Model Estimation Errors for Line Losses											
Unsigned Error	0.0120	0.0035	0.0110	0.0033	0.0010	0.02	**	0.0106	0.0032	0.0014	0.00
Signed Error	0.0039	0.0015	0.0014	0.0006	0.0025	0.00	***	0.0015	0.0007	0.0010	0.00
Panel B: Comparing Aggregated Manager Estimation Errors and Model Estimation Errors											
Unsigned Error	0.0286	0.0139	0.0257	0.0120	0.0029	0.01	***	0.0249	0.0118	0.0038	0.00
Signed Error	0.0102	0.0088	0.0038	0.0031	0.0071	0.00	***	0.0040	0.0041	0.0063	0.00

Managers' signed estimation errors were larger than model errors on average, suggesting that managers tended to overstate insurance losses during our sample period.

Additional analyses: Estimation errors

What causes the advantage of machine learning models over managers? Various incentives may motivate managers to report biased estimates intentionally.

- **Tax shield**: because determining the taxable income involves loss estimates, over-reserving is more beneficial if more income is classified as a reserve.
- **The income smoothing incentive**.
- **Small Profit**: firms with small positive earnings are likely to have boosted reported income by understating loss reserves.
- **IRIS ratio violation**: financially weak firms tend to under-reserve to appear adequate in capital and avoid regulatory scrutiny.
- **Risk-based capital ratio:** equals one if the ratio is smaller than two and 0 otherwise.

$$\begin{aligned} ManagerError = {} & \alpha_0 + \alpha_1 TaxShield + \alpha_2 Smooth + \alpha_3 SmallProfit \\ & + \alpha_4 Insolvency + \alpha_5 Violation + \alpha_6 Liab + \alpha_7 Crisis + \alpha_8 Size \\ & + \alpha_9 SmallLoss + \alpha_{10} Profit + \alpha_{11} Loss + \alpha_{12} Linesize \\ & + \alpha_{13} Reinsurance + \alpha_{14} Public + \alpha_{15} Mutual + \alpha_{16} Group \\ & + LineFixedEffects + \epsilon. \end{aligned}$$

Additional analyses: Estimation errors

	Pred. Sign	ManagerError			ModelError (Machine learning without manager estimates)				ModelError (Machine learning with manager estimates)				
Panel A: The association between estimation errors and managerial incentives													
		Coeff.	Std Err.		Coeff.	Std Err.	Diff.		Coeff.	Std Err.		Diff.	
TaxShield	+	0.055	0.0222	***	−0.048	0.0214	0.102	***	−0.022	0.0215		0.076	***
Smooth	−	0.013	0.0211		0.088	0.0509	−0.076		0.021	0.0208		−0.008	
SmallProfit	−	−0.005	0.0028	**	−0.002	0.0025	−0.002		−0.004	0.0028	*	0.000	
Insolvency	−	0.003	0.0031		0.002	0.0027	0.001		0.005	0.0037		−0.002	
Violation	−	−0.003	0.0011	***	0.001	0.0013	−0.003	***	−0.001	0.0010		−0.002	***
Liab	−	−0.001	0.0067		−0.005	0.0102	0.004		−0.003	0.0083		0.001	
Crisis	−	−0.002	0.0009	**	−0.001	0.0016	−0.001		−0.001	0.0010		−0.001	

- Managerial incentives, including tax reduction, income smoothing, and financial strength concerns, Crisis, Violation affect insurance firms' reserve levels.
- Machine model estimates are less affected by managers' incentives, as none of the incentive variables were statistically significant.
- The coefficient of Small Profit became marginally significant in the models including manager estimates, suggesting that incorporating managers' estimates might also bring their biases into the models.
- Overall, the results indicate that the influence of managerial incentives is hardly present in the model estimation, which explains, in part, the model's superior performance.

Conclusion

1. Our results, based on a large set of insurance companies' loss (future claim payments) estimates, revisions, and realizations, indicate that, with one exception (homeowner/ farm owner insurance), loss estimates generated by machine learning are more accurate than managers' actual estimates underlying financial reports.
2. Accounting estimates generated by machine learning are potentially superior to managerial estimates because they may use the archival (training) data more consistently and systematically than managers.
3. On the other hand, managers may include in their estimates (forecasts) forward-looking information (e.g., on expected inflation or the state of the economy) that machines obviously ignore. Accordingly, we assess the superiority of machines over humans in generating accounting estimates.