My previous blog post (“Beware! Estimates Dominate Financial Reports,” October 13, 2018) alerted you the hidden secret of accounting: Most balance sheet and income statement items are based on managerial subjective estimates and projections (like depreciation, asset write-offs, bad debts, expected gain on pension assets, etc.). The financial data that many investors are using literally “float on a sea of estimates,” many of which are no better than sheer guesses, while others are manipulated. Worse yet (if this is possible), the number and impact of estimates in financial reports is fast increasing, “thanks” to numerous Financial Accounting Standards Board (FASB) regulations requiring enhanced use of estimates (fair values, asset write-offs, etc.). This is clearly indicated by the figure below (from The End of Accounting) showing a five-fold increase in the number of estimates underlying S&P 500 companies.

I am not optimistic about reversing the trend of the ever-rising estimates, but I found a way of substantially improving their accuracy and reliability, thereby enhancing the usefulness of reported earnings and asset values to investors.
Enter machine learning (ML). ML is particularly useful in improving predictions, given a large number of observations, which is often the case with accounting data. Considering accounting estimates, there are two major causes of estimation errors affecting earnings:

1. **Economic uncertainty**: Unexpected changes in customers’ demand, input prices, regulations, economic conditions, etc., will result in estimation errors.
2. **Managers’ manipulation**: It’s relatively easy to affect reported results, like earnings, by small tweaks of the assumptions underlying the estimates (e.g., percent of credit sales that will default).

ML eliminates error no. 2 (manipulations—machines may be dumb, but they don’t lie) and decreases error no. 1 (uncertainty) by systematically taking into account more underlying factors (like the state of the economy) than managers do. A win-win situation for ML.

To prove this I teamed up with a great group of ML experts at Rutgers University (Miklos Vasarhelyi, Kexing Ding, Xuan Peng, and Ting Sun). In the first research on the use of ML in accounting, we focus on property & casualty insurance companies. The reason: a major cost item of these companies is the “future loss estimate.” Explanation: For insurance policies written (earned) in a given year, the major cost item are the payments made on insurance claims (car accidents, home damage, etc.). But that’s not the end of the story. People can, and often do claim damages years later, like long-term effects of concussions. So, in addition to claims paid during the insured year, there will be claims to be paid in the future years (“insurance long tails”). Thus, a proper measurement of insurance companies’ earnings requires accounting for both claims paid during the year, and an estimate of future claims—“future loss estimates.” These are generally large estimates having a material effect on insurance companies’ reported earnings. We focus on these estimates in our ML research.

In essence, we conducted a horse-race between managers’ annual loss estimates and the ML predictions of future claims, both compared to the actual claims paid over the subsequent 10 years. Humans, with all their frailties (manipulation) vs. machines, like former world chess champion Gary Kasparov playing against a computer.

I will spare you all the involved technicalities of generating ML predictions (we will discuss them fully in a forthcoming paper), and get straight to the results which are summarized in the following table.

<table>
<thead>
<tr>
<th>Insurance Companies' Loss Reserve Estimate</th>
<th>Comparing managers' with ML estimates of loss reserve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private Passenger Auto Liability</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>ML Model</td>
</tr>
<tr>
<td>All insurers</td>
<td>Random Forest</td>
</tr>
<tr>
<td>Public insurance firms (with compensation variables)</td>
<td>Random Forest</td>
</tr>
<tr>
<td>Public insurance firms (w/o compensation variables)</td>
<td>Random Forest</td>
</tr>
</tbody>
</table>
The table shows the percent improvement of the ML future loss estimates over managers’ estimates (which are, by the way, audited by public accountants). We use two, widely-used, error (estimate vs. actual) measures: root mean square error (RMSE) and mean absolute error (MAE). The first row of the table reflects all U.S. insurers (including subsidiaries and related companies) in our database, and the second row portrays only publicly traded insurance companies. The improvement achieved by ML predictions over managers’ estimates is astounding: 22-37%. Kasparov lost (May 11, 1997).

By using ML, you can cut managers’ errors by more than a third! And this is just our initial run of ML predictions. Improvements are coming. (The third row in the table shows the impact of one of our predictors—top management compensation. Without this key variable, the average error improvement decreases by 5-10%. Second vs. third rows).

Summarizing, this is the first demonstration of the use of Machine Learning in improving the quality and reliability of financial information. In addition to insurance companies’ loss estimates, ML can be applied to a wide range of managerial accounting estimates, like the bad debt reserve, future payments on warranties, pension estimates, etc., etc. Very promising indeed, leading me to the final figure:

Perhaps,
Machines Aren’t
That Dumb
After All

Baruch Lev is the Philip Bardes Professor of Accounting and Finance at the Stern School of Business, NYU. This article first appeared at the Lev End Of Accounting Blog and is shared here with his permission. Professor Lev is not affiliated with Knowledge Leaders Capital and his opinions are his own.

© Knowledge Leaders Capital