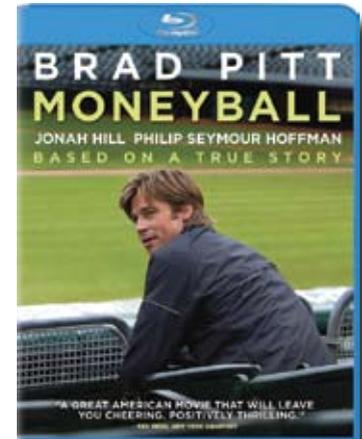


Predicting Job Performance: The Moneyball Factor

J. Scott Armstrong

PREVIEW: Choosing the right person for a given position is a highly complex task, yet experts believe that their experience allows them to do this well. Michael Lewis's 2003 book *Moneyball* and the recent film based on the book provide a counterpoint, showing that the statistical procedures used by Billy Beane, general manager of professional baseball's Oakland Athletics, are more effective in predicting job performance than are experts' judgments.

In this article, Scott Armstrong traces the emergence of the argument in favor of statistical procedures to writings in the 1950s by Paul Meehl and shows how Meehl's principles, carried forward by Billy Beane, can be applied to improve business performance today.



THE PROBLEM

When forecasting in complex, uncertain situations, expertise is of little value. Decades of empirical research have led to the *Seer-Sucker Theory*: “No matter how much evidence exists that seers do not exist, suckers will pay for the existence of seers” (Armstrong, 1980). Subsequent evidence, such as Philip Tetlock's (2005) book *Expert Political Judgment*, provides further support for the conclusion that experts, while good at diagnosis, are poor at prognosis.

Naturally, most experts believe they are immune from the Seer-Sucker Theory. They believe their instincts and experience are superior and their situations unique. The baseball scouts in *Moneyball* were convinced that Billy Beane's player analysis was wrong and that their opinions, based on years of experience, were right. According to Lewis, many people associated with baseball, not just the scouts, thought Beane's reliance on statistics was nuts.

Beane used two key procedures to select and retain his ballplayers: one, he developed models to predict performance; two, he ensured that these models were used properly. Paul Meehl (1954) had recommended such procedures for personnel selection and retention, and the research since then, summarized by Grove and colleagues (2000), supports Meehl's original findings. These have

been widely cited as well as taught in universities for many years.

DEVELOP PREDICTION MODELS

Billy Beane's staff analyzed the data on player performance to create models used to predict which players would be most successful and which ones were underpriced. Beane understood that opinions should be used only when shown to be valid inputs to a model: in short, don't revise the models' recommendations based on opinions. Meehl's example to explain this principle goes something like this:

You're in the supermarket checkout lane, piling up your purchases. You don't say, “This looks like \$178.50 worth to me”; you do the calculations. And once the calculations are done, you don't say, “In my opinion, the groceries were a bit cheaper, so let's reduce it by \$8.00.” You stick with the calculated total.

Regression

Moneyball makes developing a model sound complex. Nevertheless, many people should be able to master it. The tool you need is on your computer—the regression program in a spreadsheet. Regression analysis was developed in the 1800s. The big advance today is that it's now much cheaper and easier to use – perhaps too easy, according to some experts. I've found that you are likely to improve your personnel decisions if you use

Key Points

- Billy Beane used two key procedures to select and retain his ballplayers: one, he developed models to predict performance; two, he ensured that these models were used properly.
- But the baseball scouts in *Moneyball* were convinced that Billy Beane's player analysis was wrong and that their opinions, based on years of experience, were right. Most experts believe their instincts and experience are superior and their situation unique.
- Paul Meehl recommended that the person responsible for screening should not meet the candidates. Many find this preposterous; nevertheless, it leads to better decisions. This was a key aspect of Billy Beane's procedures. He did not want to meet his players nor watch them play.
- Paul Meehl died in 2003 after a long and illustrious career. Thanks to the creativity and tenacity of Billy Beane and the storytelling skills of Michael Lewis, the benefits of models became obvious to others. Today, many baseball, basketball, soccer, hockey, and football teams use prediction models.

prior knowledge to develop a model, keep the model simple, and ignore diagnostic measures such as statistical significance and R-square (Armstrong, 2012).

Judgmental Bootstrapping

Will models work in your situation? In sports, there is a massive amount of data on performance; in your business, this is likely not the case, and a common problem. Interestingly, a method proposed in the early 1900s for predicting the size of U.S. corn crops is relevant. It involved regressing the corn expert's *predictions* against the information he used. The model's predictions were more accurate than the expert's. Research in the 1970s (in particular, Dawes, 1971) found this method useful for problems in management, such as personnel selection. Called

judgmental bootstrapping, it has been shown to be more accurate than expert predictions because it applies the rules more consistently than the expert does. For example, see Philip Hans Franses' (2009) case study on judgmental bootstrapping in *Foresight*. Additionally, once a simple model is developed, it provides predictions at a much cheaper cost than using experts (Armstrong 2001).

Index Methods

Two colleagues and I are working on a method that is simpler, more effective, and less likely to be misused than regression. It also allows the use of all important variables. This method, developed by Benjamin Franklin, predates bootstrapping – and regression analysis, for that matter. We call it the *index method*. Say you have to select one of a number of candidates. First, list all of the variables that are known to be important, assigning a point for each variable on which the candidate does well (in some cases, it may help to weight the variables). Add the points for each candidate, and then select the candidate with the highest score. We recently used index models to predict the winner of U.S. presidential elections (Armstrong & Graefe, 2011).

Fortunately, an enormous amount of research has been published on what variables help to predict job performance. Schmidt and Hunter's (1998) meta-analysis summarized evidence from 85 years of research. Interestingly, many personnel consultants do not use this literature: when ranking the importance of variables, personnel consultants' rankings were unrelated to the evidence-based rankings (Ahlburg, 1992). Partly this is due to the fact that results from these studies are counterintuitive. For example, general mental ability (GMA) is the most effective predictor variable, while the number of years of education is a poor predictor of job performance.

DO NOT OVERRIDE THE MODEL

When considering job candidates, people are often influenced by biases, some of them

subconscious. Height, weight, gender, and physical appearance are often considered, even when irrelevant to job performance. Thus, the applications should be stripped of all irrelevant information.

Paul Meehl recommended that the person responsible for screening should not meet the candidates. Many find this preposterous; nevertheless, it leads to better decisions. This was a key aspect of Billy Beane's procedures. He did not want to meet his players nor watch them play.

Increasingly, organizations are following these procedures. Goldin and Rouse (2000), in their study of symphony orchestra auditions, found that when the screening committee does not see the applicant (musicians play behind a partition), women were much more likely to pass this stage of the recruitment process.

IMPLEMENTATION: WHY MONEYBALL IS IMPORTANT

Armed with this knowledge about personnel forecasting, I went looking for ways to apply it. My first opportunity came in the early 1970s on an airline flight from Denver to Philadelphia. Some fit young men were on the plane. Wondering who they were, I turned to the person sitting next to me to see if he knew. It happened that he did. He was Ed Snider, owner of hockey's Philadelphia Flyers, and these were his players. Here, I thought, was my big chance. I would persuade Snider to employ me to select hockey players by using predictive models. Sportswriters would learn about me. Other teams would flock to my door; fame and wealth would follow. After a suitable interval, I asked, "Tell me, Ed, how do you select your players?" Snider told me that his managers had recently been using a prediction model to make the decisions. He mentioned that the NFL's Dallas Cowboys used such a model and they were known for making good draft picks. Originally, Snider was the only one in the Flyer organization who thought it would work. His managers resisted, but after a two-year experiment they agreed that the players the model selected

performed better than the ones the managers selected.

Life gave me a second chance at fame. In 1979, while visiting my friend Paul Westhead, I suggested using models to make personnel decisions for the Los Angeles Lakers. Paul, in his first year coaching the NBA team and known for his creativity, loved the idea. I proposed a modest funding figure – petty cash for the Lakers – but team owner Jerry Buss turned it down. As it happened, the Lakers won the championship that season without me.

Others tried to promote acceptance of these procedures. Robin Dawes, who was involved with the successful application for selecting PhD candidates at the University of Oregon (Dawes, 1971), reported eight years later that few schools had adopted the process. However, some sports teams had incorporated it, apparently with great success.

Paul Meehl died in 2003 after a long and illustrious career. Thanks to the creativity and tenacity of Billy Beane and the storytelling skills of Michael Lewis, the benefits of models became obvious to others after *Moneyball* was published that same year. It was a game changer. Today, many baseball, basketball, soccer, hockey, and football teams use prediction models. In the first part of the NBA's 2009-10 season, the 15 teams with at least one full-time statistician on their staff won 59% of their games, while the 15 teams with no statisticians won only 41% (Biderman, 2010).

IMPLICATIONS FOR BUSINESS FIRMS

Sports teams have a big incentive to use predictive models for personnel selection. Teams that don't use them will fall behind, and when they do, it will be obvious. The benefits of better personnel selection are equally large for other businesses (Schmidt & Hunter 1998), but results are harder to see in the short term. However, firms ultimately may benefit even more than sports teams from prediction models. In sports, the advantage begins to disappear when other

teams adopt similar models. By contrast, every large organization would gain long-term benefits by selecting employees through these methodologies. Applicants benefit too, since they end up in jobs for which they are better suited.

Innovators, beware. *Moneyball* describes the difficulties Billy Beane faced, given the reverence accorded to the status quo. My suggestion is to start off by asking whether managers are willing to consider these techniques for personnel selection. In the 1980s, my friend Professor Morris Hamburg was chairing a faculty committee responsible for student admissions. I explained how they could make better admissions decisions, overcome claims of bias, and save time and money by developing a judgmental bootstrapping model. I suggested that he ask the committee to describe what evidence would convince them that they should adopt such a model, and then I would move ahead to assess the evidence (at no cost to the committee). After a few weeks, Morris called me and said that the committee had considered the proposal and that there was no evidence that would convince them to change. End of story.

On the other hand, when you adopt predictive models for personnel selection, you gain an advantage over competitors that refuse to do so. As Winston Churchill said, "Men occasionally stumble across the truth, but most of them pick themselves up and hurry off as if nothing had happened." Billy Beane stumbled, and then showed that persistence can produce huge benefits.



Scott Armstrong is Editor of the *Principles of Forecasting: A Handbook for Researchers and Practitioners*. In addition he recently published *Persuasive Advertising: Evidence-based Principles*. As he found in forecasting, evidence-based principles provide opportunities to design more persuasive advertisements as well as to forecast which ads will be most effective.

REFERENCES

- Ahlburg, D.A. (1992). Predicting the job performance of managers: What do the experts know? *International Journal of Forecasting*, 7, 467-472.
- Armstrong, J.S. (2012). Illusions in regression analysis, *International Journal of Forecasting* (forthcoming).
- Armstrong, J.S. (2001). Judgmental bootstrapping: Inferring experts' rules for forecasting, in J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Norwell, MA: Kluwer Academic Publishers.
- Armstrong, J.S. (1980). The seer-sucker theory: The value of experts in forecasting, *Technology Review*, 83 (June/July), 18-24.
- Armstrong, J.S. & Graefe, A. (2011). Predicting elections from biographical information about candidates: A test of the index method, *Journal of Business Research*, 64, 699-706.
- Biderman, D. (2010). *Wall Street Journal*, March 12.
- Dawes, R.M. (1971). A case study of graduate admissions: Application of three principles of human decision making, *American Psychologist*, 26, 180-188.
- Franses, P.H. (2009). Can managers' judgmental forecasts be made scientifically?, *Foresight*, Issue 15 (Fall), 32-36.
- Goldin, C. & Rouse, C. (2000). Orchestrating impartiality: The effect of "blind" auditions on female musicians, *American Economic Review*, 90 (4), 715-741.
- Grove, W.M. et al. (2000). Clinical versus mechanical prediction: A meta-analysis, *Psychological Assessment*, 12, 19-30.
- Meehl, P.E. (1954). *Clinical vs. Statistical Prediction*, Minneapolis: University of Minnesota Press.
- Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings, *Psychological Bulletin*, 124, 262-274.
- Tetlock, P. (2005). *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press.
- Acknowledgements: Fred Collopy, Kesten C. Green, and Philippe Jacquart provided useful suggestions on this paper. Editing was provided by Jennifer Armstrong, Nathan Fleetwood, and Kevin Stawicki and the *Foresight* editorial staff.