Journal Of Investment Management, Vol. 21, No. 1, (2023), pp. 120–121

www.joim.com

BOOK REVIEW

Mark Kritzman, Senior Editor

PREDICTION REVISITED: THE IMPORTANCE OF OBSERVATION, 1ST EDITION

by Mark P. Kritzman, David Turkington and Megan Czasonis

(*Reviewed by Kenneth Winston*)

Prediction Revisited is a delightfully nonlinear romp through the basics of linear statistical inference. Its stated goal is to reframe the subject with a relevance-weighted approach. But there's a second goal: to emulate the singer Mark Anthony. Mark Anthony, of course, is the envy of crossover aspirants everywhere, moving effortlessly back and forth between Spanish hits and English hits. Czasonis, Kritzman and Turkington (CKT hereafter) also seek a crossover hit, moving back and forth between mathematical exposition and conceptual English-based exposition.

In the first chapter, CKT lay out their basic premise: some past observations are more relevant than others for predicting the future. EWMA (exponentially weighted moving average) is a tepid version of this idea: it presumes that more recent experience is more relevant for predicting upcoming behavior than long-ago experience. But CKT point out that time is not the only feature that might be used to judge relevance: financial and economic data provide rich context for past states of the world that might be more effective in predicting the future than, well, yesterday.

Part of the pleasure of reading this book is the way CKT are able to re-examine familiar quantitative methods in ways that gain new insight. For example, what new thing can possibly be said about the mean calculation $\bar{x} = (x_1 + \cdots + x_N)/N$? CKT reframe the calculation as $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} info_i\, obj_i$, where the informativeness $info_i$ of each observation is the same (one), and the object obj_i being evaluated is simply the data x_i.

This simple example serves as the gateway drug to the more addictive stuff, as CKT use the $info_i\, obj_i$ framework over and over again: for variance, for (Pearson) correlation, for linear regression, for asymmetry of features to outcomes, for reliability (R-squared)... The world looks different when seen through the $info_i\, obj_i$ prism: it naturally makes the case that some past data points are better than others for the task of prediction.

Each chapter starts with a section that explains the subject conceptually. These are full of fun facts: who knew that Francis Galton constructed a quincunx to demonstrate the normal distribution? Then concepts like mean and variance, co-occurrence, relevance, fit, asymmetry, and reliability are worked through in language. After that, the subject is re-explored in equations. Finally, the chapter's results are applied to economic data, including the prediction of GDP.

Two recurring themes in the book provide support for the relevance framework. One is the use of pairwise calculation rather than versus-mean calculation. For example, variance is usually defined as the average squared difference of a set of data points to their mean. But it can also be defined as the average squared difference between every unique disjoint pair of points. This allows the relevance of one data point to another to be taken into account.

The second recurring theme is the use of Mahalanobis distance. The informativeness of some set of features is defined as the Mahalanobis distance of that feature set to the mean feature set, and the relevance of one feature set to another is defined as a Mahalanobis cross-term between the two demeaned feature sets. CKT show that these definitions naturally tie the $info_i\, obj_i$ framework into ordinary least squares (OLS) regression, but also allow an easy transition to partial regressions based on more relevant observations. One small carp: their approach is quite similar to weighted least squares (WLS) or generalized least squares (GLS) regression, but I did not see this mentioned or explored.

Like all beautiful ideas, a pure relevance framework is too delicate to survive in the wild. Empirical financial distributions are emphatically leptokurtic; the use of the normal distribution to model them ranges in practice from misleading to disastrous. The idea that the most unusual observations are the most relevant for prediction could be an error maximizer, like the naïve use of optimization to form Markowitz efficient portfolios. However just as subsequent work added techniques like resampling to deal with the unpleasantness of reality in Markowitz portfolio selection, so too should CKT's relevance approach to prediction serve as the base for practical methods that can mud-wrestle dirty data.

As far as the crossover appeal, there's a lot of sugar to help the medicine go down, like quincunxes and an amusing final chapter with biographies of relevant (pun intended) mathematicians. Overall CKT have provided goodies for a wide range of readers, from innumerate traditional MBAs to jaded quants. No matter where you are on that spectrum, you'll enjoy this book.