**Measuring short-term and long-term forecast efficacy**

Let be the time series of market-day opening prices for a fixed ticker. Suppose that a congressional member, denoted , buys shares of ticker at time (for dollars per share). Define the indicator function to be 1 if this is the case—i.e., congressional member purchases shares of worth at time . Without knowing the congressional member’s eventual sale price, we want to quantify if the purchase at is an effective forecast for “good” performance of the ticker—that is, if,

Now, the monotonic condition is hard to satisfy given market fluctuations, so a more reasonable statement of “good” performance” is:

Finally, it is useful to put bounds on to confine assessment of forecast efficacy to a discrete time period. For this, we define the true values for a short-term forecast from time as the set: And the true values for a medium-term forecast from time as the set: Thus, a congressional member’s purchase of a fixed ticker at time is a good forecast for positive short-term performance and positive medium-term performance respectively if and only if:

These two statements are equivalent to:

We set the right side of the implications equal to and corresponding to short-term forecast efficacy and medium-term forecast efficacy respectively (note that accounts for the volatility of the stock). Notice that we can compute and for every transaction a congressional member makes from the data we have available. Given the union of transaction data (e.g., purchase industry) and congressional data (committee membership), the machine learning task is to develop model and . If the models perform better than a naïve bayes one, we can infer correlation.