2010 Russell Ackoff Doctoral Student Fellowship Proposal
Forecasting: Expectations, Probabilities, and Observable Characteristics

David Rothschild, PhD Candidate, Business and Public Policy Department
Advisor: Justin Wolfers

In a far-reaching 2004 paper in *Econometrica*, "Measuring Expectations", Charles Manski demonstrates that playing a simple economic game, a subject with one of three different expectations and one of two different utility functions will make the same move in four out of six possible scenarios. The common practice, to assume subjects have perfect expectations and differ in their revealed behavior because of variation in their fully realized budget constraints and utility functions, misses an important source of heterogeneity in decision making under uncertainty. Is that woman carrying an umbrella because she thinks there is a high probability of rain or because she values having an umbrella with the slightest chance of rain? Did that man move all of his money into a savings account because he thinks the stock market is going to crash or because he is extremely risk adverse? While there is consensus in the literature on the importance of expectations in decision making, there is little research on expectations separated from utility functions. First, was their revealed preference determined with reasonably accurate expectations of the world and how does that accuracy shift by observable characteristics? Second, how would their actions differ with perfect expectations and how do I help induce those expectations? Third, how can I exploit their preferences or signals in forecasting the probability of rain or the trajectory of stock market?

For this project, I build on the most recent methods of surveying expectations, many designed by students of Manski, to create a web-based graphical interface that gathers expectations and studies their development. The expectations, including both the value of outcomes and a full distribution of subjective probabilities of the value of the outcomes, are collected with a set of observable characteristics about the respondents that may influence their responses. The web-based interface is series of questions; in the first question the respondent states his expected value for the upcoming episode and in the second question the respondent creates a distribution of subjective probabilities. The design of the second question, and all subsequent follow-up questions, builds from the answers to the previous questions. The graphical nature of the interface guides the respondent to efficiently understand and answer the questions without any possibility of being inconsistent (a big issue in polling). Further, it takes measures to avoid any anchoring or suggestive examples for the respondent (a big issue in prediction markets). Partiality questions follow in two groups. First, within each episode, the two main questions are followed by one or two questions examining the respondents' personal partiality related to the episode. Second, after all of the episode related questions are finished, the respondent is asked a series of questions on observable characteristics. In the calibration phase, besides exploring a range of methodological questions, I study a host of questions about anchoring, information, and the presentation of choices, and how they affect both the expectation and its certainty.

The data from the fully calibrated interface helps explain the systematic discrepancies or biases in expectations by several observable characteristics of the respondent a few examples:

geographical region, education, age, partiality (as it pertains to the outcome in question) and certainty (possibly correlated with risk profile). First, by comparing the value of the data as I randomize different variables I can explore existing expectations, by observable characteristics, and their interaction with revealed preference. Second, understanding these biases is essential to creating methods of inducing people to make efficient predictions; more efficient predictions lead to more efficient decisions, which is beneficial to the consumer and society.

Two recent papers of mine address the confusion in determining the root of differences in forecasts and demonstrate how further research on expectations will prove valuable. Rothschild and Wolfers (work-in-progress) uses aggregate and individual-level responses to two different questions, asked to the same respondents at the same time, to model a new set of responses that are either not publicly available or are answers to questions that are never asked. Using these modeled responses, the new data points create more accurate forecasts than the available data. Rothschild (2009, POQ) demonstrates the conditions under which the accuracy and informational advantage of a debiased prediction market-based forecast is statistically significant relative to a debiased poll-based forecast. It is informative in answering the question of why forecasts diverge; prediction markets are relatively strong when public information is low. But there are numerous differences between polls and prediction markets beyond the differing information pools made available by scientifically weighted samples (polls) and self-selected samples (prediction markets). There are different questions, aggregations techniques, and reporting incentives and biases. Thus, with multiple variables having differences, it is hard to identify the cause of the distinctions. The two papers suggest a need for two new data sources. The first data source is the individual-level expectation data noted above. The second data source is the massive amount of forecasts (computer models, experts, prediction markets), and the multitude of individual-level signals that can be exploited as forecasts, available, but fleeting, on the internet; recent studies have explored signals as diverse as EBay pricing, Google searches, Facebook supporters, etc., as forecasts or inputs into forecasts. These two sources of data can be combined to model and create more accurate and efficient forecasts.

This project utilizes an interactive web-based interface to improve upon ineffective methods of collecting and utilizing individual-level expectation and preference data in a wide range of fields including: economics, finance, politics, marketing, sports, and miscellaneous examples. The results are more accurate, provide more useful information and are more efficient in understanding consumer behavior under uncertainty, making their behavior more efficient and utilizing individual-level data for forecasts. The project expands the current literature on consumer expectation, especially as it is separated from utility functions, and on forecasts, moving well beyond the dialog in the current debate between computer models and prediction markets. The deeper understanding of expectations helps explain heterogeneity in revealed preferences. Better forecasts help researchers connect shocks with changes in the underlying values and investors make more efficient use of their time and money.

Proposal is written by David Rothschild

PhD Candidate
Department of Business and Public Policy (Applied Economics PhD Program)
Cell: (973) 818-8997
Email: RothscDM@Wharton.UPenn.edu
Web: http://Assets.Wharton.UPenn.edu/~rothscdm

Signed by Advisor

Justin Wolfers