A Comparison of Small Crowd Selection Methods

Henrik Olsson (olsson@santafe.edu)
Department of Psychology, University of Warwick
Coventry CV4 7AL, UK
Santa Fe Institute, 1399 Hyde Park Road
Santa Fe, New Mexico 87501 USA

Jane Loveday (loveday.jane@gmail.com)
Department of Economics, University of Warwick
Coventry CV4 7AL, UK

Abstract

The literature on the wisdom of crowds argues that in most situations, the aggregated judgments of a large crowd perform well relative to the average individual. There are, however, many real-world cases where crowds perform poorly. A small crowd literature has since developed, finding that better performing small crowds often exist within whole crowds. We compare previously proposed small crowd selection methods based on absolute or relative group performance to a new sequential search method and find that it selects better performing small crowds more consistently for forecasts of real gross domestic product (GDP) growth, inflation (measured by consumer price index, CPI), and unemployment rate made by US and Euro-zone surveys of professional forecasters.

Keywords: Wisdom of crowds; select crowds; US survey of professional forecasters; ECB survey of professional forecasters

Introduction

A large group of people are guessing the number of jellybeans in a jar. To form your own guess would you be better off if you average all previous guesses, make your own, or try and identify an expert to copy? The literature on the wisdom of the crowds (Surowiecki, 2004) suggests the best option will be combining the guesses of the whole group (for example their mean or median). This will outperform at least the average guess and likely any attempt to find an expert, particularly where the signals of expertise are weak, where performance is measured by mean square error (MSE), and outperformance is defined as having a lower MSE (see e.g., Page, 2007; Sunstein 2006).

Psychologists have long studied this phenomenon, finding that crowds consistently outperform the average individual and often the best individual for a range of judgment tasks including estimates of weights (Gordon 1924; 1935) and sizes of piles of buckshot (Bruce, 1935). It has also been observed for forecasting (Armstrong, 2001; Bates & Granger, 1969; Clemen, 1989) where combining forecasts has become a standard econometric methodology (Timmerman, 2006), particularly evident in the implementation of surveys of professional forecasters (SPFs) by central banks around the world, which generate forecasts for key macroeconomic variables by combining the forecasts of large crowds of professional forecasters.

The wisdom of the crowd effect relies on the statistical properties of averaging uncertain estimates and the independent aggregation of these estimates. The level of the crowd's performance depends on the crowd being relatively diverse (Page, 2007), specifically that it is unbiased and that individual members are uncorrelated or even better, negatively correlated (Davis-Stober, Budescu, Dana, & Broomell, 2014; Larrick & Soll, 2006). Both Page (2007) and Brown, Wyatt, Harris, and Yao (2005) show mathematically that a crowd outperforms its average individual member if it is diverse, where diversity measures the extent to which individual's forecasts vary around the whole crowd forecast. The more diverse a crowd, or the more individual's forecasts "bracket" the truth (Larrick & Soll, 2006), the more a crowd will outperform the average individual.

While increasing diversity is important to improve crowd performance, the average individual performance (as measured by squared error) cannot be ignored, creating a tradeoff between expertise and diversity (Brown et al., 2005; Page, 2007). The crowd is only as good as its constituent members, the higher the average individual error (e.g. when individuals are biased and have high variances) the worse performing the whole crowd. Conversely if individual biases bracket the truth, this will reduce the average individual error and therefore increase whole crowd performance.

In reality, experts are not so diverse. Positive correlations of expert judgment have been observed in many fields including sports predictions (Winkler, 1971), forecasts by sales management teams (Ashton, 1986), and assessments of survival probability by physicians in an ICU (Winkler & Poses, 1993), arguably due to experts' access to similar data sources, education and training (Broomell & Budescu, 2009). Where crowds include correlated experts, their combined forecasts perform relatively poorly. Indeed, observations of herding behavior where individuals follow others' behavior in preference to their own private information (Bikhchandani, Hirschleifer, & Welch, 1992) are good examples of crowds performing poorly due to correlated judgments. Bikhchandani et al. (1992) discuss the example of the routine performance of tonsillectomies on children until the 1960s, often unnecessarily and with negative consequences as an example of herding. In this

case, doctors had limited information, thus imitated others, resulting in a crowd of highly correlated experts making poor decisions.

Selecting small crowds

Observation of poorly performing real-world crowds has led to a range of studies showing the existence of better performing small crowds within the whole crowd in economic forecasting (Budescu & Chen, 2014, Mannes, Soll, & Larrick, 2014), current events (Budescu & Chen, 2014), experimental judgment tasks (Mannes et al., 2014), fantasy soccer prediction games (Goldstein, McAfee, & Suri, 2014), and animal organizational behavior (Kao & Couzin, 2014). Of particular interest to this paper is the work of Mannes et al. (2014) and Budescu and Chen (2014) who both analyze small crowds in the context of surveys of professional forecasters in the US and Euro-zone (EU), respectively. Both find small crowds that can outperform the whole crowd but implement different methods to identify small crowds and analyze data that differs by geography, macroeconomic variable, forecast horizon and forecast type (point vs. probability forecast).

Mannes et al. (2014) select crowds by past performance, finding for an aggregate index of point forecasts of seven variables (the consumer price index, the rate on the 3-month Treasury Bill, the rate on the 10-year Treasury Note, the yield on Moody's AAA corporate bond, nominal GDP, housing starts, and the unemployment rate) drawn from the US Survey of Professional Forecasters published by the Philadelphia Federal Reserve (US SPF) that small crowds outperform the whole crowd.

Budescu and Chen (2014) study probability forecasts (rather than point forecasts) for inflation (measured by consumer price index, CPI) and real gross domestic product (GDP) growth, drawn from the Euro-zone Survey of Professional Forecasters published by the European Central Bank (EU SPF). They develop a "Contribution Weighted Method" (CWM) that select crowds based on each individual's "contribution" to whole crowd performance and find that small crowds outperform whole crowds for CPI but not for real GDP growth. Contribution is measured as the difference between whole crowd performance and crowd performance excluding that individual, with performance measured by a quadratic score out of 100, normalized such that 100 is perfect prediction performance and 0 the worst possible performance. Where the inclusion of an individual improves the performance score, contribution is positive, where it decreases the score, it is negative. The small crowd is then formed from those individuals with positive contributions (roughly 50% of the whole crowd) and weighted either equally (CEWM; this model was called "Contribution" in Budescu & Chen, 2014) or by contribution (CWM).

Budescu and Chen (2014) also analyzed post hoc optimal group sizes for the two methods, where individuals are dropped one by one starting with those with the lowest contribution, to find an "optimal" small crowd with largest possible performance score (CEWM—optimized and CWM—optimized respectively for equally weighted or contribution weighted crowds). In the first case, small crowds weighted by contribution perform best but where individuals are dropped one by one, equally weighted small crowds perform better than those weighted by contribution.

This study builds specifically on these two studies (Mannes et al., 2014; Budescu & Chen, 2014) and more broadly on the small crowd literature by analyzing a range of small crowd selection methods including those already implemented by Mannes et al. (2014) (*Ranked performance*) and Budescu and Chen (2014) (*Contribution*).

We introduce a third *Sequential search* methodology drawing on the machine learning literature (Mendes-Moreira, Soares, Jorge, & Sousa, 2012). Two versions are implemented, *Sequential search—increasing*: where small crowd size increases by one on each step starting with the individual with lowest MSE, and *Sequential search—decreasing*: where crowd size decreases by one on each step, starting with the whole crowd. On each step, individuals are re-analyzed relative to the current small crowd to identify the one that when added to (removed from) the small crowd from the preceding step, gives the greatest reduction in small crowd MSE. The process continues for each subsequent small crowd size until small crowd MSE is minimized.

This has similarities with the *Contribution* methodology (and specifically CEWM—optimized and CWM—optimized) as it selects individuals into the small crowd based on relative rather than absolute performance. It differs, however, in that individuals are re-analyzed relative to each sequential small crowd formed rather than just once relative to the whole crowd as done in the Contribution methodology. This is important if for example the whole crowd contains highly diverse individuals (for example with a range of positive and negative correlations among them), and the makeup of small crowds changes substantially for different sized small crowds. In this case, an individual's "value" to the small crowd (i.e., from error reduction) may change as the makeup and size of the small crowd changes. Sequential search is able to capture this by re-analyzing individuals relative to the small crowd for each new small crowd formed. In contrast, Contribution—optimized cannot as the contribution measure is fixed based on contribution to *whole crowd* performance only and does not change as new small crowds are formed.

We use the US and EU surveys of professional forecasters, drawn from professional economic forecasters in the public and private sectors, as our real-world data. The nature of these crowds is well suited to our analysis as forecasters are recognized experts in their fields, the data for forecasts and actual values is easily accessible and the surveys are referenced in both the small crowd and economic forecasting literature. A particular advantage of analyzing the SPFs is the literature demonstrating their outperformance of traditional macroeconomic forecasting methods. For example, Ang, Bekaert, and Wei (2007) and

showed that the US SPF outperformed traditional forecasting methods for predictions of CPI. Others have compared the simple average of forecasts to other combination schemes. Genre, Kenny, Meyler, and Timmermann (2013) showed that the simple average can outperform many other combination methods for predictions of real GDP growth and the unemployment rate in the EU SPF, although it could not outperform them for CPI. Genre et al. concluded, however, that there is no combination method that consistently outperforms all the other methods across variables and time horizons.

In this study we focus on real GDP growth, CPI, and unemployment, as simple averages from expert forecasters have been shown to outperform traditional forecasting methods and other combination schemes for these variables. In addition, these variables are common across both the US and the EU SPF.

Method

Data for the EU and US SPF are extracted from the publicly available databases of the Philadelphia Federal Reserve (http://www.philadelphiafed.org) and the European Central Bank (http://www.ecb.europa.eu) for the variables real GDP growth, unemployment rate (civilian unemployment, aged > 16 years), and inflation (HCIP in Europe, CPI in the US). Data are taken for the time period 1999 – 2013 and split into two time periods for analysis, 1999 – 2008 and 2009 – 2013. This split segregates the period of high volatility following the most recent financial crisis from earlier time-periods as done by Genre et al. (2013). Each variable is analyzed for forecasts made quarterly for 6 months in the future (2Q) (US SPF only), the current year (1Y) and the next calendar year (2Y). These forecast horizons align with those used by Mannes et al. (2014) who evaluated 5 forecast horizons up to a year of the US SPF and Budescu and Chen (2014) who evaluated 1 year forecasts of the EU SPF.

Table 1: Whole crowd sizes.

	1999-2008			2009-2013		
Variable	2Q	1Y	2Y	2Q	1Y	2Y
US						
Unemployment	26	26	25	40	37	41
CPI	23	23	20	41	36	40
Real GDP growth	26	26	26	42	38	43
EU						
Unemployment	-	48	47	-	54	56
CPI	-	50	49	-	55	58
Real GDP growth	-	50	49	-	55	58

To generate a clean panel, we adopt the methodology of Genre et al. (2013) and cleanse the raw survey data to account for individuals entering and exiting the panel over time as well as missing an occasional forecast. Only individuals that are not absent from the survey for four or more consecutive periods are included. Remaining missing data are estimated using a basic panel regression model, assuming individuals' deviations from the mean forecast are consistent with their previous period's performance.

The resulting whole crowd sizes are shown in Table 1 giving crowds of between 20 and 43 for each trial in the US data and between 47 and 58 in the EU data. Forecasts are compared to final adjusted actual data drawn from Eurostat (http://epp.eurostat.ec.europa.eu), the US Department of Labor Statistics (http://data.bls.gov), and the US Bureau of Economic Analysis (http://www.bea.gov).

Procedure

We compared the performance of selected crowds derived from ten selection methods, four based on absolute performance (Ranked performance) and six based on relative performance (two Sequential search and four Contribution selection methods). Inspired by the post-hoc optimal group size analyses in Budescu and Chen (2014), two of the Contribution selection methods were "optimized" on the training set, dropping the lowest contributing individuals one by one (see Table 2 for descriptions). In addition, we compared the whole crowd performance against the average individual and a randomly chosen individual.

In Ranked performance the four methods analyzed were "All" (training period is all available past periods), "1q" (training period is immediately prior quarter), "4q" (training period is immediately prior 4 quarters) and "8q" (training period is immediately prior 8 quarters). We did not evaluate the performance of all small crowds generated post hoc, instead we selected the small crowd that minimized MSE in the training period.

Initial analysis of Ranked performance methods showed 1q to give the greatest performance improvement and we also found 1q to be more accurate for all Sequential search and Contribution methods. Therefore the main results are only reported for 1q training periods.

Due to the requirement to use the previous 8 quarters in the performance – 8q selection method, the first forecast period used in each of 1999 – 2008 and 2009 – 2013 is the ninth quarterly time period, respectively 2001 Q1 and 2011 Q1. Forecast performance is measured for each available forecast period and final point estimates of forecast performance are the average of all forecast periods.

For each of 30 trials (a trial being a combination of variable, time-period, forecast horizon, and geography) selected and whole crowd forecasts are aggregated using the mean and performance in comparison to the actual level of the variables was measured using both MSE and MAD. Analysis of results for MSE and MAD error measurement showed the same relative performance of selection methodologies for both measures, therefore only MSE are reported in the results section.

Table 2: Selection methods.

Selection method	Method description
Ranked performance	Individuals ranked based on MSE for period 1 (training period). Forecasts formed for small crowd sizes 2 – 9 in ranking order and the crowd size with minimum MSE in period 1 selected. These individuals form the small crowd from which period 2 forecasts are taken.
Sequential search	
Increasing	Starting with individual with lowest MSE, individuals added one by one choosing the one delivering the greatest reduction in period 1 MSE until MSE is minimized. MSE is recalculated for each crowd size. Small crowd size is then fixed and period 2 forecast is taken from these individuals.
Decreasing	As for the above method but beginning with the whole crowd and removing individuals one by one choosing the individual delivering the greatest reduction in period 1 MSE until MSE is minimized. Small crowd size is then fixed and period 2 forecast formed from these individuals.
Contribution	
CEWM	Contribution Equal Weighted Model. Individual's contribution to whole crowd MSE computed in period 1 as defined by Budescu and Chen (2014). Small crowd forecast in period 2 is an equal weighting of all individuals with positive contribution.
CWM	Contribution Weighted Model. As above but small crowd forecast weighted according to individual contributions.
CEWM—optimized	Starting with the small crowd from CEWM, individuals removed one by one starting with the individual with the lowest contribution until period 1 crowd MSE is minimized. These individuals then form the small crowd from which the period 2 forecast is drawn.
CWM—optimized	As for the above but applied to the small crowd from CWM with the weights re-weighted for each new small crowd size.

Results

Results are reported in terms of performance ratios of small crowd MSE to whole crowd MSE ("Avg. ratio" in Table 3). This enables averaging across different variables, geographies, time-scales, and forecast horizons, to generate an overall picture of the best performing selection methodology. The consistency of selection methodologies is also reported, where consistency is the percentage of trials on which that selection method identifies a small crowd that outperforms the whole crowd ("% trials <1" in Table 3). Finally, the number of wins is also reported where a win is an occurrence of a particular selection methodology delivering the best performance for a trial. Where selection methodologies deliver the same performance, each methodology is allocated a win ("Trials won" in Table 3).

Performance of the Whole Crowd

The whole crowd outperforms the average individual for all trials (average MSE ratio = 0.79), but performs the same as a randomly selected individual (average MSE ratio = 0.99, whole crowd outperforms on 16 out of 30 trials).

Selection of Small Crowds

Table 3 shows that Sequential search—decreasing is the best performing selection methodology followed by CEWM—optimized, and Ranked performance 4q. In terms of average ratios of small crowd MSE to whole crowd MSE, Sequential search—decreasing and CEWM—optimized deliver performance improvements of more than 20% over the whole crowd and outperform all Ranked performance selection methods. Both also find an outperforming small crowd more consistently than all other selection methodologies (90% of the time), implying they not only perform better but also more consistently. Ranked performance 4q comes close with 87%. CEWM does not perform as well in terms of average ratios, but is consistent due to low standard deviation. In line with the average performance, Sequential search—decreasing wins the most trials. The remaining wins are relatively evenly spread across methodologies.

Table 3: Average small to whole crowd comparisons.

Selection method	Avg. ratio	*St. dev.*	% trials <1	Trials won
Ranked performance				
All	1.01	*0.34*	73	3
1q	0.84	*0.30*	83	2
4q	0.84	*0.27*	87	3
8q	0.91	*0.32*	77	3
Sequential search				
Increasing	0.86	*0.33*	80	4
Decreasing	0.74	*0.24*	90	7
Contribution				
CEWM	0.89	*0.18*	90	0
CWM	0.96	*0.62*	87	3
CEWM—optimized	0.78	*0.32*	90	3
CWM—optimized	1.22	*1.35*	80	2

Table 4 shows that despite the overall crowd size varying by trial (mean = 40.3), consistent patterns in average small crowd size can be observed. Ranked performance and Sequential search—increasing tend to result in small crowd sizes with small standard deviations, while Contribution methods and Sequential search—decreasing give higher small crowd sizes with larger standard deviations. As found by Budescu and Chen (2014), CEWM and CWM find small crowds roughly half the size of the whole crowd.

Table 4: Mean small crowd size and standard deviation.

Selection methodology	Mean size	*St. dev.*
Ranked performance		
All	2.8	*1.5*
1q	2.8	*2.3*
4q	2.6	*1.0*
8q	2.2	*0.7*
Sequential search		
Increasing	1.5	*0.3*
Decreasing	10.3	*5.9*
Contribution		
CEWM	20.3	*6.0*
CWM	20.3	*6.0*
CEWM—optimized	8.1	*4.3*
CWM—optimized	7.9	*3.4*

We confirm the findings of Budescu and Chen (2014) for their time-period (1999-2011) that when using cumulative history, CWM outperforms CEWM for EU CPI (ratios of 0.73 and 0.88 respectively). When individuals are dropped one by one to minimize MSE in the forecast period, this pattern reverses and CEWM delivers a greater performance than CWM (ratios of 0.39 and 0.52 respectively).

Using recent performance, however, has some drawbacks, especially for CWM compared to CEWM. We find that when looking across geographies, variables, timescales and forecast horizons, and particularly when the history is constrained to just the prior one quarter, CEWM delivers a greater performance improvement than CWM. This is due to the inconsistency and high standard deviation found for CWM, with large performance improvements in some cases but not in others. It may be that limiting the history leads to unreliable estimates of contribution magnitude and an equally weighted aggregation is to be preferred in terms of stability of forecasts.

Discussion

The best selection method is Sequential search—decreasing, followed by CEWM—optimized. The Ranked performance method with a 4q training period is a close third in terms of consistently finding a small crowd that outperforms the whole crowd. The Ranked performance method, however, seems to suffer from selecting too small crowds (2 to 3 individuals). Indeed, Mannes et al.'s Figure 6 shows that crowds of sizes between approximately 5 and 13 maximize performance. Consequently, Mannes et al. argue for a "take the top five" rule of thumb when selecting small crowds. In future research it would be interesting to investigate how well the top five rule generalizes to other data sets and compare it with other selection methods such as Sequential search and Contribution.

The iterative nature of Sequential search—decreasing and its re-analysis of individuals relative to each new small crowd size could explain its outperformance of CEWM—optimized. While CEWM—optimized analyzes individuals relative to the whole crowd and then drops individuals one by one starting with those with the lowest contribution, Sequential search—decreasing, starts at the same point but then re-analyzes individuals relative to the new small crowd at each step. This allows it to capture any changes in the makeup of the small crowd and the value of individuals to that particular small crowd. Preliminary simulations, not presented here, suggested that it is where there is more diversity across individuals that Sequential search—decreasing has an advantage over CEWM—optimized. Where diversity is lower, crowd makeup will change little as the small crowd size changes, and in this case the two methods should give similar results.

In a similar vein, Sequential search—decreasing outperforms Sequential search—increasing by starting with a broader analysis of individuals. By starting with one individual and adding individuals to the crowd one by one, the increasing method only analyzes each individual relative to the first crowd member and is therefore much more sensitive to the choice of starting individuals and their diversity relative to the rest of the crowd. By starting with the whole crowd, Sequential search—decreasing reduces this sensitivity and enables a better performing small crowd to be identified. There is still a risk, however, that this procedure finds a local rather than global minimum for MSE and therefore cannot find the best performing small crowd. Machine learning addresses this issue by combining increasing and decreasing methods and adding a "drop-one" step into the increasing method (Mendes-Moreira et al., 2012). This method has not been implemented in this study but provides an area of future study to further improve the performance of small crowds. Another limitation of the Sequential search method is that it does not guarantee diversity in the small crowd. One possibility that has been explored in the machine learning literature is to add another step in the search process and select individuals with low correlations with other individuals (Rooney, Patterson, Anand, & Tsymbal, 2004).

References

Ang, A., Bekaert, G., & Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics, 54*, 1163-1212.

Armstrong, J. S. (2001). *Principles of forecasting: A handbook for researchers and practitioners*. Norwell, MA: Kluwer.

Ashton, R. H. (1986). Combining the judgments of experts: How many and which ones? *Organizational Behavior and Human Decision Processes, 38*, 405-414.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly, 20*, 451-468.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *The Journal of Political Economy, 100,* 992-1026.

Broomell, S. B., & Budescu, D. V. (2009). Why are experts correlated? Decomposing correlations between judges. *Psychometrika, 74*, 531-553.

Brown, G., Wyatt, J., Harris, R., & Yao, X. (2005). Diversity creation methods: A survey and categorization. *Information Fusion, 6*, 5-20.

Budescu, D. V., & Chen, E. (2014). Identifying expertise to extract the wisdom of crowds. *Management Science*. Advance online publication. doi:10.1287/mnsc.2014.1909

Clemen, R. T. (1989) Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*, 559-609.

Davis-Stober, C. P., Budescu, D. V., Dana, J., & Broomell, S. B. (2014). When is a crowd wise? *Decision, 1*, 79-101.

Genre, V., Kenny, G., Meyler, A., & Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting, 29*, 108-121.

Goldstein, D. G., McAfee, R. P., & Suri, S. (2014). The wisdom of smaller, smarter crowds. In *Proceedings of the fifteenth ACM conference on Economics and computation* (pp. 471-488). New York: ACM press.

Gordon, K. (1924). Group judgments in the field of lifted weights. *Journal of Experimental Psychology, 7*, 389-400.

Gordon, K. (1935). Further observations on group judgments of lifted weights. *Journal of Psychology, 1*, 105-115.

Kao, A. B., & Couzin, I. D. (2014). Decision accuracy in complex environments is often maximized by small group sizes. *Proceedings of the Royal Society Biological Sciences, 281*, 20133305. doi: 10.1098/rspb.2013.3305

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science, 52*, 111-127.

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of Personality and Social Psychology, 107*, 276-299.

Mendes-Moreira, J., Soares, C., Jorge, A. M., & De Sousa, J.F. (2012). Ensemble approaches for regression: A survey. *ACM Computing Surveys, 45*, 10-39.

Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.

Rooney, N., Patterson, D., Anand, S., & Tsymbal, A. (2004). Dynamic integration of regression models. In *Lecture Notes in Computer Science: Vol. 3181. International Workshop on Multiple Classifier Systems* (pp. 164–173). Berlin, Germany: Springer.

Timmerman, A. (2006). Forecast combinations. In Elliott, G., Granger, C. W. J., & Timmermann, A (Eds.), *Handbook of Economic Forecasting Volume 1* (pp. 135 – 196). Amsterdam: Elsevier.

Sunstein, C. R. (2006). *Infotopia: How many minds produce knowledge*. New York: Oxford University Press.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. London, UK: Little Brown.

Winkler, R. L. (1971). Probabilistic prediction: Some experimental results. *Journal of the American Statistical Association, 66,* 675-685.

Winkler, R. L., & Poses, R. M. (1993). Evaluating and combining physicians' probabilities of survival in an intensive care unit. *Management Science, 39,* 1526-1543.