Dresdner Kleinwort

2 August 2006

Global Equity Strategy

Painting by numbers: an ode to quant

What could baseball, wine pricing, medical diagnosis, university admissions, criminal recidivism and I have in common? They are examples of simple quant models consistently outperforming so-called experts. Why should financial markets be any different? So why aren't there more quant funds? Hubristic self belief, self-serving bias and inertia combine to maintain the status quo.

► There is now an overwhelming amount of data to suggest that in many environments, simple quant models significantly outperform human (expert) judgements. For instance, in their study of over 130 different papers, covering decision-making contexts as wide-ranging as occupational choice to the diagnosis of heart attacks, Grove et al located a mere 8 studies that found in favour of human judgements over the quant models.

► All 8 of these studies had one thing in common. The human participants had access to information not available to the quant models. Where quant models and human had the same information set, the models performed much better. Across the full range of papers that Grove et al examined, the average human participant had a 66.5% accuracy rating, whereas the quant models had an average hit ratio of 73.2%.

► Even when the human participants were given access to the quant model's results as an input for them to use if they chose, they still managed to underperform the model. This is an important point. One of the most common responses to quant superiority is that surely this could be a base for qualitative improvements by skilled users. However, the evidence is clear: quant models usually provide a ceiling (from which we detract performance) rather than a floor (on which we can build performance). We tend to overweight our own opinions relative to those of the models.

► The good news is that, in some fields, quant models have become relatively accepted. For instance, over half the states in the US use a quant model when considering parole for convicts. However, in finance a quant approach is far from common. Those that do purse a quant path tend to be rocket scientist uber-geeks. Once in a while a fairly normal 'quant' fund comes to light. Two explicitly behavioural based funds stand out in my mind – LSV and Fuller & Thaler. Both have admirable track records in terms of outperformance. Whilst this is far from conclusive proof of the superiority of quant, it is a step in the right direction.

► So why don't we see more quant funds in the market? The first reason is overconfidence. We all think we can add something to a quant model. However, the quant model has the advantage of a known error rate, whilst our own error rate remains unknown. Secondly, self-serving bias kicks in, after all what a mess our industry would look if 18 out of every 20 of us were replaced by computers. Thirdly, inertia plays a part. It is hard to imagine a large fund management firm turning around and scrapping most of the process they have used for the last 20 years. Finally, quant is often a much harder sell, terms like 'black box' get bandied around, and consultants may question why they are employing you at all, if 'all' you do is turn up and crank the handle of the model. It is for reasons like these that quant investing will remain a fringe activity, no matter how successful it may be.

James Montier
+44 (0)207 475 6821
james.montier@dresdnerkleinwort.com

Global Investment Strategy
Research Analysts
Global Asset Allocation
Albert Edwards
+44 (0)20 7475 2429
albert.edwards@dresdnerkleinwort.com

Global Equity Strategy
James Montier
+44 (0)20 7475 6821
james.montier@dresdnerkleinwort.com

Online research:
www.dresdnerkleinwort.com/research

Bloomberg:
DRKW <GO>

A member of **Allianz**

Painting by numbers: an ode to quant

Don't worry dear reader, there will be no gratuitous use of poetry in this missive, despite the title - I promise. However, pause for a moment and consider what baseball, wine, medical diagnosis, university admissions, criminal recidivism and I might have in common.

The answer is that they all represent realms where simple statistical models have outperformed so-called experts. Long-time readers may recall that a few years ago I designed a tactical asset allocation tool based on a combination of valuation and momentum. At first this model worked just fine, generating signals in line with my own bearish disposition. However, after a few months, the model started to output bullish signals. I chose to override the model, assuming that I knew much better than it did (despite the fact that I had both designed it and back-tested it to prove it worked). Of course, much to my chagrin and the amusement of many readers, I spent about 18 months being thrashed in performance terms by my own model. This is only anecdotal (and economist George Stigler once opined "The plural of anecdote is data"), but it sets the scene for the studies to which I now turn.

Neurosis or psychosis?

The first study I want to discuss is a classic in the field. It centres on the diagnosis of whether someone is neurotic or psychotic. A patient suffering psychosis has lost touch with the external world; whereas someone suffering neurosis is in touch with the external world but suffering from internal emotional distress, which may be immobilising. The treatments for the two conditions are very different, so the diagnosis is not one to be taken lightly.

The standard test to distinguish the two is the Minnesota Multiphasic Personality Inventory (MMPI). This consists of around 600 statements with which the patient must express either agreement or disagreement. The statements range from "At times I think I am no good at all" to "I like mechanics magazines". Fairly obviously, those feeling depressed are much more likely to agree with the first statement than those in an upbeat mood. More bizarrely, those suffering paranoia are more likely to enjoy mechanics magazines that the rest of us!

In 1968, Lewis Goldberg[1] obtained access to more than 1000 patients' MMPI test responses and final diagnoses as neurotic or psychotic. He developed a simple statistical formula, based on 10 MMPI scores, to predict the final diagnosis. His model was roughly 70% accurate when applied out of sample.

Goldberg then gave MMPI scores to experienced and inexperienced clinical psychologists and asked them to diagnose the patient. As the chart below shows, the simple quant rule significantly outperformed even the best of the psychologists.

---

[1] Goldberg (1968) Simple models or simple processes? Some research on clinical judgements, American Psychologist, 23

Hit rate for diagnosis (%)
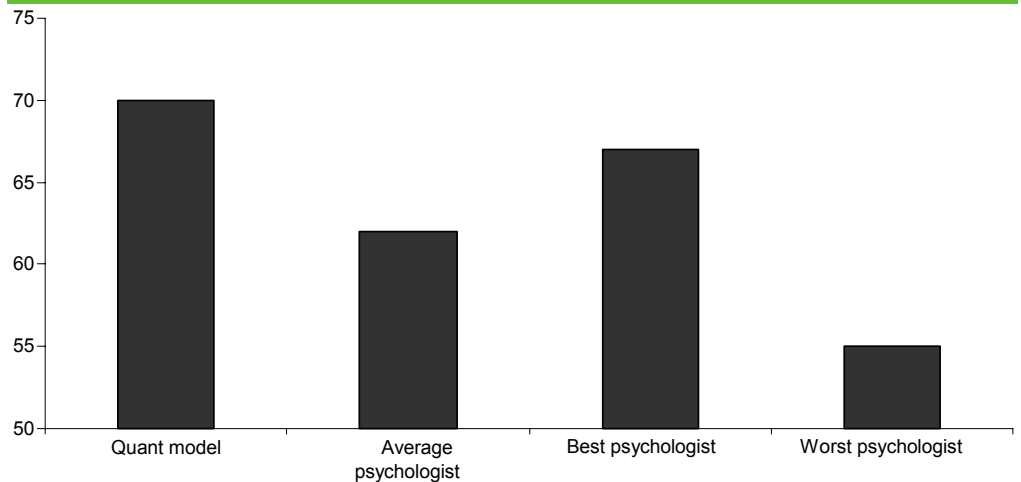
Source: Dresdner Kleinwort Macro research

Even when the results of the rules' predictions were made available to the psychologists, they still underperformed the model. This is a very important point: much as we all like to think we can add something to the quant model output, the truth is that very often quant models represent a ceiling in performance (from which we detract) rather than a floor (to which we can add).

Every so often, and always with the aid of a member of the quant team, I publish a quant note in *Global Equity Strategy*. The last one was based on the little book that beats the market (see *Global Equity Strategy*, 9 March 2006). Whenever we produce such a note, the standard response from fund managers is to ask for a list of stocks that the model would suggest. I can't help but wonder if the findings above apply here as well. Do the fund managers who receive the lists then pick the ones that they like, much like the psychologists above selectively using the Goldberg rule as an input?

Brain damage detection

Similar findings were reported by Leli and Filskov[2], in the realm of assessing intellectual deficit due to brain damage. They studied progressive brain dysfunction and derived a simple rule based on standard tests of intellectual functioning. This model correctly identified 83% of new (out of sample) cases.

However, groups of inexperienced and experienced professionals working from the same data underperformed the model with only 63% and 58% accuracy respectively (that isn't a typo; the inexperienced did better than the experienced!). When given the output from the model the scores improved to 68% and 75% respectively - still both significantly below the accuracy rate of the model. Intriguingly, the improvement appeared to depend upon the extent of the use of the model.

---

[2] Leli and Filskov (1981) Clinical-Actuarial detection and description of brain impairment with the W-B Form 1, Journal of Clinical Psychology, 37

Hit rate for diagnosis (%)
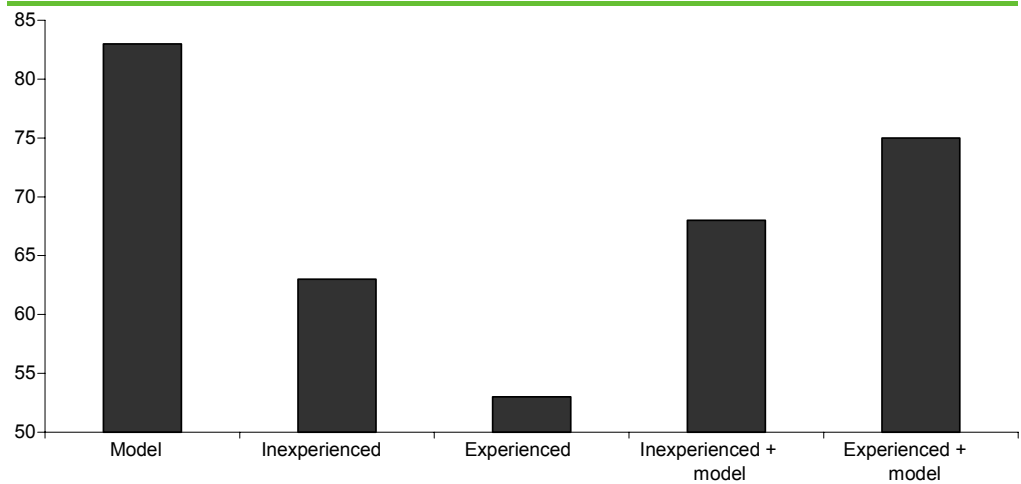
Source: Dresdner Kleinwort Macro research

University admissions

Dawes[3] gives a great example of the impotence of interviews (further bolstering our arguments as to the pointlessness of meeting company managements – *The seven sins of fund management*, November 2005). In 1979, the Texas legislature required the University of Texas to increase its intake of medical students from 150 to 200.

The prior 150 had been selected by first examining the academic credentials of approximately 2200 students, and then selecting the highest 800. These 800 were called for an interview by the admissions committee and one other faculty member. At the conclusion of the interview, each member of the committee ranked the interviewee on a scale of 0 (unacceptable) to 7 (excellent). These rankings were then averaged to give each applicant a score.

The 150 applicants who ended up going to Texas were all in the top 350, as ranked by the interview procedure. When the school was told to add another 50 students, all that were available were those ranked between 700 and 800. 86% of this sample had failed to get into any medical school at all.

No one within the academic staff was told which students had come from the first selection and which had come from the second. Robert DeVaul and colleagues[4] decided to track the performance of the two groups at various stages – i.e. the end of the second year, the end of the clinical rotation (fourth year) and after their first year of residency.

The results they obtained showed no difference between the two groups at any point in time; they were exactly equal at all stages. For instance, 82% of each group were granted the M.D. degree, and the proportion granted honours was constant etc. The obvious conclusion: the interview served absolutely no useful function at all.

Criminal recidivism

Between October 1977 and May 1987, 1035 convicts became eligible for parole in Pennsylvania. They were interviewed by a parole specialist who assigned them a score on a five point scale based on the prognosis for supervision, risk of future crime, etc. 743 of these cases were then put before a parole board. 85% of those appearing before the board were granted parole, the decisions (bar one) following the recommendation of the parole specialist.

---

[3] Dawes (1989) House of Cards: Psychology and Psychotherapy built on myth, Free Press
[4] DeVaul, Jervey, Chappell, Carver, Short, and O'Keefe (1957) Medical school performance of initially rejected students, Journal of American Medical Association, 257

25% of the parolees were recommitted to prison, absconded, or arrested for another crime within the year. The parole board predicted none of these. Carroll et al[5] compared the accuracy of prediction from the parole board's ranking, with that of a prediction based on a three factor model driven by the type of offence, the number of past convictions and the number of violations of prison rules. The parole boards ranking was correlated 6% with recidivism. The three factor model had a correlation of 22%.

Bordeaux wine

So far we have tackled some pretty heavy areas of social importance. Now for something lighter. In 1995, a classic quant model was revealed to the world: a pricing equation for Bordeaux wine!

Ashenfelter et al[6] computed a simple equation based on just four factors; the age of the vintage, the average temperature over the growing season (April-September), rain in September and August, and the rain during the months preceding the vintage (October-March). This model could explain 83% of the variation of the prices of Bordeaux wines.

Bordeaux 1952-1980, summer temperature and harvest rain related to price
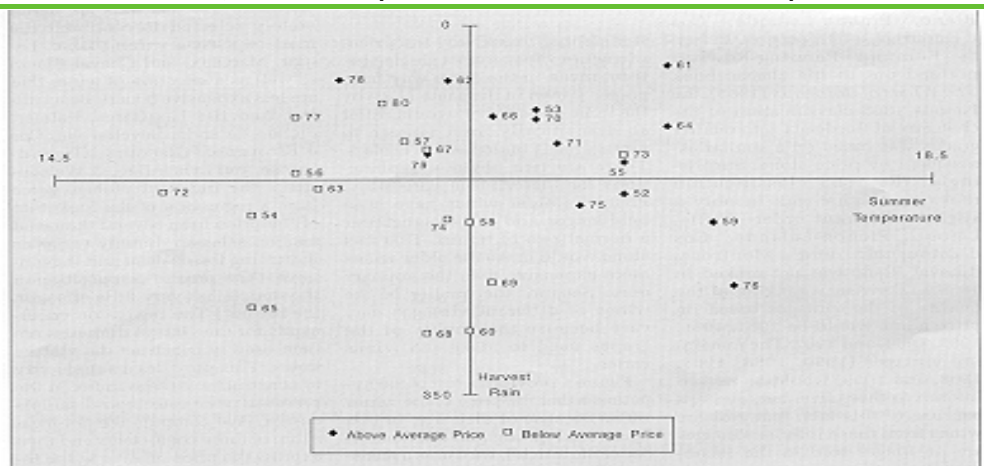
Figure 2. Bordeaux: 1952–1980, summer temperature and harvest rain related to price.

Source: Ashenfelter et al

Ashenfelter et al also uncovered that young wines are usually overpriced relative to what one would expect based on the weather and the price of old wines. As the wine matures, prices converge to the predictions of the equation. This implies that "bad" vintages are overpriced when they are young, and "good" vintages may be underpriced.

The table below shows the basic pattern. It shows the price of a portfolio of wines from each vintage relative to the (simple average) price of the portfolio of wines from the 1961, 62, 64 and 66 vintages. The second column gives the value of the benchmark portfolio in GBP. The entries for each of the vintages in the remaining columns are simply the ratios of the prices of the wines in each vintage to the benchmark portfolio. The predicted price from the equation is also shown. Incidentally, this data is from a different sample than the original estimation of the equation, so it amounts to an out of sample test[7].

[5] Carroll, Winer, Coates, Galegher and Alibrio (1988) Evaluation, Diagnosis, and Prediction in Parole Decision Making, Law and Society Review, 17

[6] Ashenfelter, Ashmore and LaLonde (1995) Bordeaux Wine Vintage Quality and the Weather, Chance

[7] For those with an interest, Ashenfelter publishes a news letter called liquid asset.

Price per case of a portfolio of Bordeaux relative to benchmark portfolio

Year of sale	Bench-mark portfolio (£)	Vintage											
		1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
1971	54	1.66	0.79	0.41	0.76		0.79						
1972	97	1.58	0.76	0.26	0.70	0.27	0.96	0.77		0.75			
1973	119	1.62	0.71	0.28	0.74	0.24	0.93	0.62	0.28	0.70	0.83		
1974	85	1.31	0.77	0.39	0.84		1.08	0.78	0.30	0.70	0.88		0.30
1975	76	1.65	0.77	0.29	0.78	0.35	0.80	0.57	0.31	0.41	0.84	0.61	0.44
1976	109	1.67	0.83	0.30	0.65	0.29	0.85	0.51	0.23	0.36	0.69	0.54	
1977	165	1.67	0.83	0.26	0.63	0.26	0.87	0.50	0.23	0.36	0.70	0.51	0.32
1978	215	1.67	0.76	0.26	0.65	0.18	0.91	0.45	0.25	0.31	0.70	0.53	0.25
1979	274	1.61	0.73	0.20	0.66	0.23	1.00	0.49	0.24	0.29	0.71	0.50	0.23
1981	296	1.75	0.62	0.22	0.70	0.04	0.93	0.47	0.25	0.29	0.82	0.52	0.22
1982	420	1.80	0.71	0.15	0.60	0.18	0.89	0.39	0.17	0.24	0.77	0.55	0.19
1983	586	1.77	0.53	0.10	0.59	0.18	1.11	0.36	0.18	0.21	0.91	0.48	0.20
1985	952	2.19	0.53	0.12	0.50	0.21	0.78	0.30	0.11	0.14	0.68	0.46	0.13
1986	888	2.10	0.56	0.25	0.54	0.17	0.80	0.30	0.15	0.19	0.65	0.46	0.14
1987	901	2.11	0.56		0.53		0.80	0.32	0.19	0.20	0.64	0.49	0.18
1988	854	2.01	0.56	0.21	0.61	0.14	0.82	0.34	0.23	0.20	0.67	0.58	0.17
1989	1048	2.09	0.61	0.28	0.53	0.19	0.77	0.27	0.24	0.18	0.66	0.43	0.15
Predicted price		**1.74**	**0.72**	**0.29**	**0.76**	**0.16**	**0.78**	**0.49**	**0.21**	**0.29**	**0.60**	**0.53**	**0.14**

Source: Ashenfelter et al

Purchasing managers

Professor Chris Snijders has been examining the behaviour of models versus purchasing managers[8]. He has examined purchasing managers at 300 different organizations. The results will not be surprising to those reading this note. Snijders concludes "We find that (a) judgments of professional managers are meagre at best, and (b) certainly not better than the judgments by less experienced managers or even amateurs. Furthermore, (c) neither general nor specific human capital of managers has an impact on their performance, and (d) a simple formula outperforms the average (and the above average) manager even when the formula only has half of the information as compared to the manager."

Meta-analysis

Ok enough already, you may cry[9]. I agree. But, to conclude, let me show you that the range of evidence I've presented here is not somehow a biased selection designed to prove my point.

Grove et al[10] consider an impressive 136 studies of simple quant models versus human judgements. The range of studies covered areas as diverse as criminal recidivism to occupational choice, diagnosis of heart attacks to academic performance. Across these studies 64 clearly favoured the model, 64 showed approximately the same result between the model and human judgement, and a mere 8 studies found in favour of human judgements. All of these eight shared one trait in common; the humans had more information than the quant models. If the quant models had the same information it is highly likely they would have outperformed.

---

[8] http://www.tue-tm.org/snijders/home/
[9] For those wondering about the mention baseball at the beginning this wasn't a red herring…Michael Lewis' book Moneyball is a great example of the triumph of statistics over judgement.
[10] Grove, Zald, Lebow, Snitz, Nelson (2000) Clinical Versus Mechanical Prediction: A meta-analysis, Psychological Assessment, 12

The chart below shows the aggregate average 'hit' rate across the 136 studies that Grove et al examined. The average person in the study (remember they were all specialists in their respective fields) got 66.5% of the cases they were presented with correct. However, the quant models did significantly better with an average hit ratio of 73.2%.

Hit rate across studies (%)
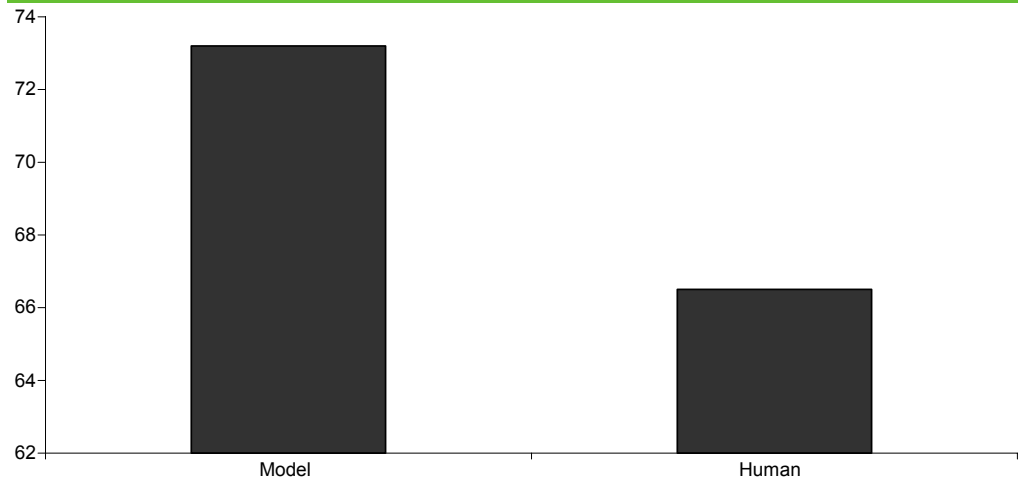

Source: Dresdner Kleinwort Macro research

As Paul Meehl (one of the founding fathers of the importance of quant models versus human judgements) wrote:

There is no controversy in social science which shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one… predicting everything from the outcomes of football games to the diagnosis of liver disease and when you can hardly come up with a half a dozen studies showing even a weak tendency in favour of the clinician, it is time to draw a practical conclusion.

The good news

The good news is that in some fields quant models have become far more accepted. For instance, the chart below shows the number of states in which the decision to parole has a quant prediction instrument involved.

Number of States using a quant parole input
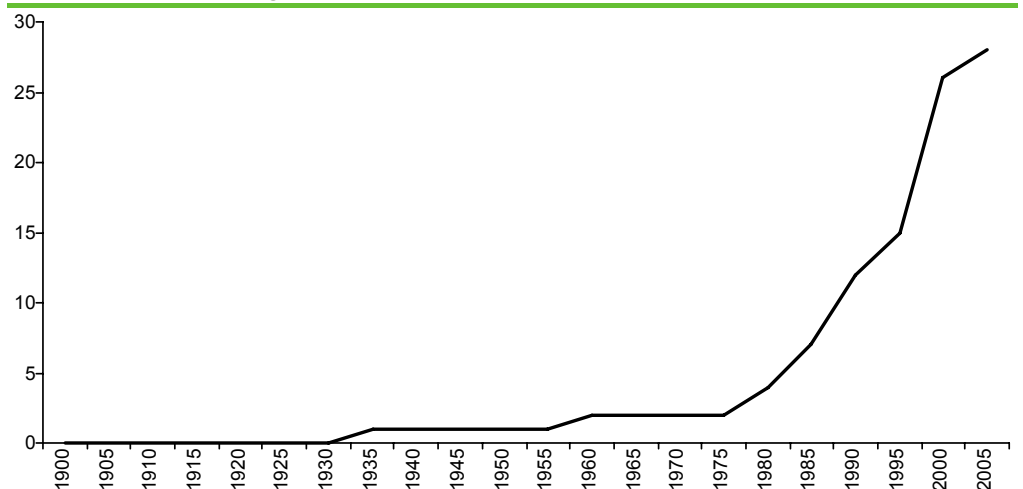
Source: Harcourt (2006)

However, in the field of finance most still shy away from an explicit quant process. A few brave souls have gone down this road. Two explicitly behavioural finance groups stand out as using an explicitly quantitative process – LSV and Fuller & Thaler. The chart below shows the performance of their funds relative to benchmark since inception. With only one

exception all of these funds have delivered pretty significant positive alpha. Now, of course, this doesn't prove that quant investing is superior; I would need a much larger sample to draw any valid conclusions. But it is a nice illustration of the point I suspect is true.

Performance versus benchmark since inception (% p.a.)
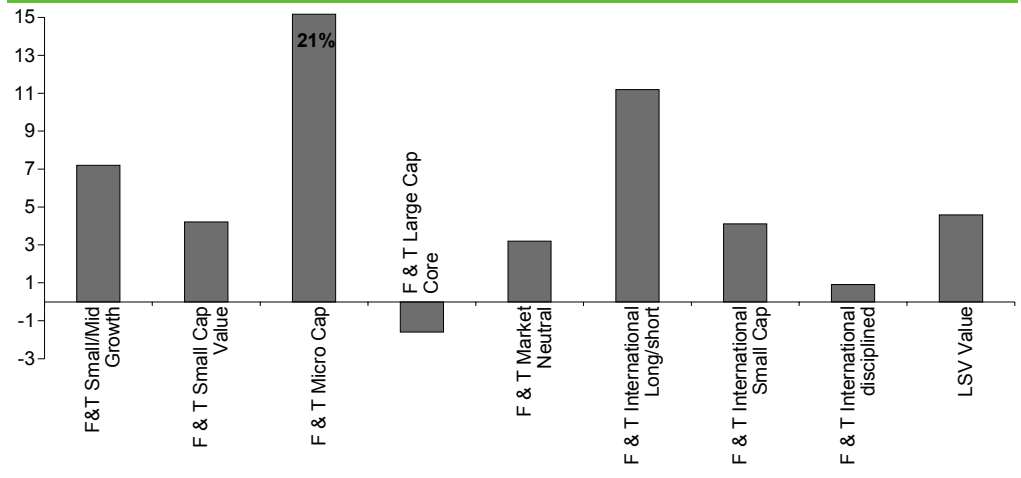
Source: Dresdner Kleinwort Macro research

So why not quant?

The most likely answer is overconfidence. We all think that we know better than simple models. My own confession at the start of this note is a prime example of such hubris. The key to the quant model's performance is that it has a known error rate, whereas our error rates are unknown.

The most common response to these findings is to argue that surely a fund manager should be able to use quant as an input, but still have the flexibility to override the model when appropriate. However, as mentioned above, the evidence suggests that quant models tend to act as a ceiling rather than a floor for our behaviour.

Additionally there is plenty of evidence to suggest that we tend to overweight our own opinions and experiences against statistical evidence. For instance, Yaniv and Kleinberger[11] have a clever experiment based on general knowledge questions such as: In which year were the Dead Sea scrolls discovered?

Participants are asked to give a point estimate and a 95% confidence interval. Having done this they are then presented with an advisor's suggested answer, and asked for their final best estimate and rate of estimates. The chart below shows the average mean absolute error in years for the original answer and the final answer. The final answer is more accurate than the initial guess.

The most logical way of combining your view with that of the advisor is to give equal weight to each answer. However, participants were not doing this (they would have been even more accurate if they had done so). Instead they were putting a 71% weight on their own answer. In over half the trials the weight on their own view was actually 90-100%! This represents egocentric discounting - the weighing of one's own opinions as much more important than another's view.

---

[11] Yaniv and Kleinberger (2000) Advice taking in decision making: Egocentric discounting and reputation formation, Organizational Behavior and Human Decision Processes. 83

Mean absolute error (in years)
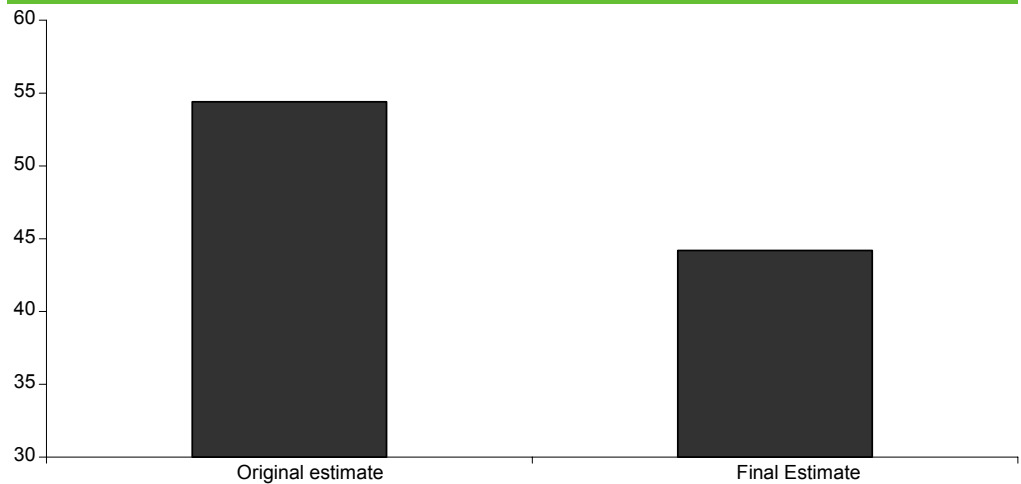
Source: Dresdner Kleinwort Macro research

Similarly, Simonsohn et al[12] showed that in a series of experiments direct experience is frequently much more heavily weighted than general experience, even if the information is equally relevant and objective. They note, "If people use their direct experience to assess the likelihood of events, they are likely to overweight the importance of unlikely events that have occurred to them, and to underestimate the importance of these that have not". In fact, in one of their experiments, Simonsohn et al found that personal experience was weighted twice as heavily as vicarious experience! This is an uncannily close estimate to that obtained by Yaniv and Kleinberger in an entirely different setting.

Grove and Meehl[13] suggest many possible reasons for ignoring the evidence presented in this note; two in particular stand out as relevant to the discussion here. Firstly, the fear of technological unemployment. This is obviously an example of a self serving bias. If, say, 18 out of every 20 analysts and fund managers could be replaced by a computer, the results are unlikely to be welcomed by the industry at large.

Secondly, the industry has a large dose of inertia contained within it. It is pretty inconceivable for a large fund management house to turn around and say they are scrapping most of the processes they had used for the last 20 years, in order to implement a quant model instead.

Another consideration may be the ease of selling. We find it 'easy' to understand the idea of analysts searching for value, and fund managers rooting out hidden opportunities. However, selling a quant model will be much harder. The term 'black box' will be bandied around in a highly pejorative way. Consultants may question why they are employing you at all, if 'all' you do is turn up and run the model and then walk away again.

It is for reasons like these that quant investing is likely to remain a fringe activity, no matter how successful it may be.

---

[12] Simonsohn, Karlsson, Loewenstein and Ariely (2004) The tree of experience in the forest of knowledge: Overweighting personal over vicarious experience

[13] Grove and Meehl (1996) Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy, Psychology, Public Policy and Law, 2

Notes

Disclosure appendix

Disclosures under US regulations

The relevant research analyst(s), as named on the front cover of this report, certify that (a) all of the views expressed in this research report accurately reflect their personal views about the securities and companies mentioned in this report; and (b) no part of their compensation was, is, or will be directly or indirectly related to the specific recommendation(s) or views expressed by them contained in this report.

Any forecasts or price targets shown for companies and/or securities discussed in this report may not be achieved due to multiple risk factors including without limitation market volatility, sector volatility, corporate actions, the unavailability of complete and accurate information and/or the subsequent transpiration that underlying assumptions made by Dresdner Kleinwort or by other sources relied upon in the report were inapposite.

Recommendation history charts

Past performance is not an indicator of future performance.

Dresdner Kleinwort Research – Recommendation definition
(Except as otherwise noted, expected performance over next 12 months)

Buy:	10% or greater increase in share price	Sell:	10% or more decrease in share price
Add:	5-10% increase in share price	Reduce:	5-10% decrease in share price
Hold:	+5%/-5% variation in share price		

Distribution of Dresdner Kleinwort equity recommendations as of 30 Jun 2006

	All covered companies		Companies where a Dresdner Kleinwort company has provided investment banking services (in the last 12 months)	
Buy/Add	280	56%	42	15%
Hold	146	29%	15	10%
Sell/Reduce	71	14%	7	10%
Total	**497**		**64**	

Source: Dresdner Kleinwort

Additional disclosures under other non-US regulations
The disclosures under US regulations above should be read together with these additional disclosures.

In respect of any compendium report covering six or more listed companies, please refer to the following website for all relevant disclosures: www.dresdnerkleinwort.com/research/disclosures

Disclaimer

This report has been prepared by Dresdner Kleinwort, by the specific legal entity named on the cover or inside cover page.

United Kingdom: This report is a communication made, or approved for communication in the UK, by Dresdner Kleinwort Wasserstein Securities Limited (authorised and regulated by the Financial Services Authority and a Member Firm of the London Stock Exchange). It is directed exclusively to market counterparties and intermediate customers. It is not directed at private customers and any investments or services to which the report may relate are not available to private customers. No persons other than a market counterparty or an intermediate customer should read or rely on any information in this report. Dresdner Kleinwort Wasserstein Securities Limited does not deal for, or advise or otherwise offer any investment services to private customers.

European Economic Area: Where this report has been produced by a legal entity outside of the EEA, the report has been re-issued by Dresdner Kleinwort Wasserstein Securities Limited for distribution into the EEA. Dresdner Kleinwort Wasserstein Research GmbH is authorised and regulated by the Federal Financial Supervisory Authority ('BaFin') by the laws of Germany.

United States: Where this report has been approved for distribution in the US, such distribution is by either: (i) Dresdner Kleinwort Wasserstein Securities LLC (DrKWS LLC); or (ii) other Dresdner Kleinwort companies to US Institutional Investors and Major US Institutional Investors only ; or (iii) if the report relates to non-US exchange traded futures, Dresdner Kleinwort Wasserstein Limited (DrKWL). DrKWS LLC, or in case (iii) DrKWL, accepts responsibility for this report in the US. Any US persons wishing to effect a transaction through Dresdner Kleinwort (a) in any security mentioned in this report may only do so through DrKWS LLC, telephone: (+1 212) 429 2000; or (b) in a non-US exchange traded future may only do so through DrKWL, telephone: (+ 11 44) 20 7623 8000; or (c) in a banking product may only do so through Dresdner Bank AG New York Branch, telephone (+1 212) 969 2700.