



CEP Discussion Paper No 969

February 2010

**New Approaches to Measuring
Management and Firm Organization**

Nick Bloom and John Van Reenen

Abstract

We detail the methodology that we have been using to quantify managerial and organizational practices across firms and countries in recent years. This has been used in many pieces of research at the Centre for Economic Performance. We discuss the pros and cons of such survey tools, describing how our methods lie between the traditional surveys used by economists and the case studies more common in other parts of social science.

JEL Classifications: L2, M2, O32, O33

Keywords: surveys, data, organization, management

This paper was produced as part of the Centre's Productivity and Innovation Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

Acknowledgements

We thank our discussant Ron Jarmin, our long time partners Stephen Dorgan and John Dowdy, our research partners Raffaella Sadun, Christos Genakos and Rebecca Homkes and our 100 management interviewers over the last 10 years. The authors gratefully acknowledge Stiftung Stahlwerk Georgsmarienhütte, the ESRC, Anglo German Foundation, Aim and the Kauffman Foundation for funding this research. Organizational support was provided by the Centre for Economic Performance.

Nick Bloom is an Associate at the Centre for Economic Performance, London School of Economics. He is also Assistant Professor, Department of Economics, Stanford University. John Van Reenen is Director of the Centre for Economic Performance and Professor of Economics, LSE.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means without the prior permission in writing of the publisher nor be issued to the public or circulated in any form other than that in which it is published.

Requests for permission to reproduce any article or part of the Working Paper should be sent to the editor at the above address.

© N. Bloom and J. Van Reenen, submitted 2010

The last three decades have witnessed an explosion of theoretical work on the organization of firms (Robert Gibbons and John Roberts, 2009). In parallel, there has been a massive increase in access to micro data which has revealed huge dispersions in productivity. For example, within narrow industries like cement, oak flooring, and block-ice the total factor productivity of plants at the 90th percentile is about twice that of those at the 10th percentile. (Lucia Foster, John Haltiwanger and Chad Syverson, 2008).

Unfortunately, analyzing to what extent this heterogeneity in productivity is due to management and organizational practices, unmeasured inputs, or other technologies has been held back by a lack of data. National statistical agencies do not usually collect data on the internal organization of companies and nor do firms report this in their accounts. Recently, however, social scientists have been starting to fill this gap by working closely with small numbers of individual firms (e.g. the “Insider Econometrics” approach described in Kathryn Shaw (forthcoming)) or covering wide cross sections of firms (e.g. Nicholas Bloom, Raffaella Sadun and John Van Reenen (2009)). In this paper we describe some of the tools of this research, particularly Bloom and Van Reenen (2007) - henceforth BVR – for measuring management and organizational practices.¹

I. MINIMIZING SURVEY BIAS

A key challenge in surveys is to obtain unbiased responses to questions. Here we outline a series of steps we have found useful, first concentrating on reducing the bias of the respondent (the manager) and then on reducing the bias of the interviewer (typically an MBA student).

I.A Reducing Respondent Bias

Choosing appropriate respondents: It is important to survey somebody junior enough to know day-to-day practices but senior enough to take an overall view of the organization. For example, we targeted plant managers in manufacturing, service line managers in hospitals, principals in schools, and district managers in retail. In manufacturing we phoned firms and requested to speak to plant managers, and if no one fitted that definition we asked for “the person in charge of production at the factory”. It can also be useful to obtain responses from employees in different levels of the firm’s hierarchy to see if there is some systematic difference in response.

Responder Blind Surveys: There is ample evidence in the psychology literature that respondents like to give the answers that they believe the interviewer wants to hear. For example, Norbert Schwartz (1999) asked experimental subjects to discuss newspaper stories about mass murderers. One group was given paper with the letterhead “Institute of *Personality* Research” while the other group was given paper with the letterhead “Institute of *Social* Research”. The former group’s responses concentrated much more on personality and the latter on social environment, highlighting respondents’ desire to provide answers they believe the researcher wants to hear. Thus, in BVR managers were not told in advance they were being scored against a grid of management practices. Instead they were simply told they were being interviewed by a graduate student for a project on “Modern manufacturing practices”.²

Open rather than Closed Questions: To facilitate blind interviews and to avoid biasing respondents by providing response options we used “open questions”. These are questions with no fixed set of responses, such as “Tell me how you monitor your production process?”. In comparison a closed question is one that admits only a limited set of responses, such as “Do you

monitor your production process daily?” [Yes/No]. Using open questions allows the interviewer to ask a set of questions that feels like a conversation without any strong direction. For example, to score firms on promotion systems the interviewer would start by asking “Tell me about your promotion system?”, followed by “How do you identify and develop top performers?”, “How are decisions made about promotions?” and “Can you describe the most recent promotion round”. The collected responses to these questions would be scored against a grid ranging from 1 for “People are promoted primarily on the basis of tenure (years of service)” to “Top performers are actively identified, developed and promoted”. In contrast the question “Do you actively identify, develop and promote your top performers” [Yes/No] is more leading in that it implies this is a standard practice, so that many firms may (falsely) respond positively.

Absolute rather than subjective scales: Many survey forms offer subjective scales for responses, commonly known as “Likert” scales. For example, a question like “How good is your firm’s performance tracking?” with response choices “Extremely good”, “Good”, “Average”, “Poor” and “Very poor” is subjective because “Poor” means different things to different people. A manager who previously worked in Toyota may view daily production monitoring as “Poor” (Toyota has real time monitoring), while a manager having previously worked in an Indian Textile firm may view daily monitoring as “Extremely good” (Indian textile firms often have no formalized monitoring). Since these responses are not even comparable across respondents, they are certainly not comparable across firms. Using absolute responses avoids this problem – for example asking the question “How frequently do you track performance?” with responses “Yearly”, “Quarterly”, “Monthly”, “Weekly”, “Daily”, “Repeatedly within each day”.

Asking for examples: We have found examples are particularly helpful for topics which are sensitive within firms and where practice often differs from theory. For example, most

organizations in theory have a process for getting rid of underperforming employees, but in practice this rarely happens in many organizations (e.g. the public sector). So we found it was essential that after the initial question “If you had a poor performer what would you do?” to follow up with “Could you give me a recent example?”.

Controlling for respondent characteristics: Different interviewees may respond in different ways to the same question. To address this we collect detailed information on responders (e.g. their position in the hierarchy, tenure in the firm and tenure in their current post). This information can be useful by including these as noise controls in regression analysis. For example, in BVR we found that senior managers had higher management scores, possibly because management practices are better at senior levels of the firm or possibly because senior managers are more positive about their firms. Either way, controlling for these variables in the regressions analysis usually helps reduce measurement error.

I.B Reducing Interviewer Bias

Interviewer Blind Survey: Biases may be due to the interviewer having preconceptions over the firm they are questioning. For example, an MBA student interviewing a Toyota plant is likely to be ex ante prejudiced in giving the firm a high score on their management practices. We tried to mitigate this by choosing medium sized firms (100 to 5000 employees) which the interviewers were unlikely to have previously heard of. We also did not share any financial information on the firms in advance: interviewers were only provided with the firm’s name, telephone number, and industry before the interview. In particular, they were not provided with any prior performance data and did not research their firms on the internet before calling them. This meant the interview had to begin with the question “Could I start by asking you a bit about what you do in your

firm”. If interviewers were challenged over their lack of prior knowledge of the firm they explained “We deliberately do not research firms in advance to make sure we have no preconceptions before running interviews”, which managers seemed happy with.

Calibrate early, calibrate often: When moving away from a single script with closed questions to a more complex script with open questions the concern arises that interviewers will be scoring answers in subtly different ways. To mitigate this problem it is important to have intensive training prior to the survey to explain the scoring grid. For example, we ran Lean manufacturing, target setting and performance management training sessions in the initial training week for BVR. We also ran a series of calibration exercises to ensure consistent scoring. This involved a lead researcher running mock interviews which all the trainee interviewers scored individually, and then discussed together as a group to align scoring. Throughout the survey process we continued to run these mock-interviews to ensure calibration was maintained.

Common Location with Cross Group Interviewing: To compare different subjects groups – for example different countries – it is useful to base the interview team in one location and rotate interviewers across groups. Rotation across groups means that interviewer fixed effects can be removed when making comparisons. For example, if Ron and Pierre are interviewing the US and France respectively using field based surveys (i.e. visiting the firms in person) then the difference between the scores could either be due to real differences in management practices, or due to differences in the interview approach of Ron and Pierre. Instead, if they are both running the survey from the UK by telephone, undertaking regular calibration, and switch countries throughout the survey (because Ron speaks French and Pierre speaks English), then differences between countries should be more informative of management differences. This is a substantial

advantage of telephone based surveys, in that it enables interviewers to rotate between organizations from different locations.

Interviewer quality: Open questions with absolute scoring grids are demanding on the ability of interviewers. Thus, the human capital of the interviewers is important – for example, interviewers need to be able to rapidly understand a range of modern management practices in training, and ideally have some prior business experience (as well as having language skills!). This is especially important if the target respondents are senior managers, such as plant managers, who can give short shrift to people who they think do not know what they are talking about. So, we usually hire international MBA students from good schools as interviewers.

Incentives and Monitoring: Personnel Economics emphasizes the importance of the right kind of incentive pay contracts. In early survey waves we paid interviewers flat-rate salaries, using personal encouragement to persuade them to make calls. However, since scheduling and running interviews is hard and repetitive, we found flat rate salaries led to only moderate levels of productivity. So in later survey waves we moved to piece rate pay for interviewers, with a supervisor for each group of four interviewers. The supervisors did not run interviewers, but instead they silently listened in to the interviews to ensure quality control. Supervisors were paid a flat-rate salary and were usually part of the research team or a trusted PhD student. While this change led to a 20% drop in interview manpower (one of out every five people became a supervisor) we discovered a doubling in productivity.

I.C Reducing Interview Bias

The context of the interview itself can cause a bias – for example, Daniel Kahnemann et al (2004) report how happiness varies with the time of day. So it is helpful to record the time of the

day (for the interviewee and the respondent), the day of the week, and the interview duration as factors and control for these in regression analysis. For example, in BVR we found that responders were more likely to give higher responses in the morning and towards the end of the week. We also realized that interviewers could evaluate the quality of an interview – for example some respondents were very forthcoming with information and knowledgeable about their firm, while others were guarded and less informed. So we asked interviewers to score the perceived reliability of the interview score. This turned out to be valuable, because in interviews with higher reliability scores the management practices scores were more strongly correlated with firm performance, suggesting self-assessed interview quality can proxy for interview quality.

I.D Getting people to Respond

The surveys we have conducted are voluntary and we often are asked “why do people respond to your survey?”. Of course, one could ask this of almost any voluntary survey (e.g. the Current Population Survey), but it is more likely to be an issue because researcher’s surveys are not directly Government supported. Our experience is once you get respondents talking on the phone they usually like to talk about themselves and their jobs. The hardest part is getting an interview scheduled in the first place. Several strategies can help. First, firms are inundated by marketing research so switchboards refuse to connect calls from people wanting to conduct “surveys” or mentioning “research” (because of the link to market research). We found it was best to confidently ask to be connected to the production manager, and if questioned state we were “doing a *piece of work* on manufacturing management”. Our interviewers tried to avoid ever using the word “survey” or “research”.

Second, try to obtain some endorsement for your study from an official body – like the Central Bank - to distinguish what you are doing from marketing companies. Third, persistence is a virtue. Responders will frequently procrastinate and miss the scheduled time for interviews. It is necessary to persistently follow up, which requires running the survey over several months. Also phoning outside regular business hours is helpful – in the US in particular many managers will only pick up their phones to an unknown caller ID outside the working day.

Fourth, avoid asking for information that can be obtained from other sources. Financial information can be sensitive, and is anyway often publicly available in company accounts. Having the respondent obtain this data wastes valuable interview time. Finally, open ended questions like “Tell me about your promotion system” are more engaging than closed ended factual questions like “How many people were promoted in your firm last year”. Open ended questions feel more like a conversation, while closed ended questions feel like a dry data extraction. In BSV we minimized the time on closed ended questions, and located these at the end of the interview.

In BVR we obtained a 54% completed interview response rate, which is extremely high by the standards of large-scale surveys which are not Government mandated. Interestingly, we also found that response rates were uncorrelated with observables like firm performance, suggesting more individualistic reasons for non-response.

II. EVALUATING SURVEY BIAS

The previous section outlined a number of steps to minimize survey bias. But from our experience of presenting organizational survey research it is also helpful to provide validation

results to convince audience that the survey was measuring real organizational differences across firms. Below we outline a number of survey exercises we have found useful for this.

Response bias: After collecting survey data it is important to evaluate response reliability. One way is to compare the observable characteristics of the responders to the non-responders, and for this one needs variables that are in both samples (i.e. not variables collected as part of the survey). For firms this typically includes location, industry, size and accounting information.

Independent Resurvey Tests: It is a good idea to re-survey about 5% of the firms using different interviewers and responders. This can be used to gauge the extent of measurement error at the question level. For example, in BVR we resurveyed 64 firms using a different interviewer to survey a different plant manager in the same firm to obtain two independent surveys from the same firm. We found that the question level scores were correlated at 0.502 and the management average score at 0.734. This indicated survey reliability, and was useful for convincing potential sceptics that the survey was really measuring differences in management practices across firms.

Different Survey Instruments: Another good design is to collect the data using more than one instrument and compare the results across the methods. For example, Grous (2009) implemented the BVR approach for UK aerospace firms asking the same questions to plant managers using a telephone based survey, and then visited the factory and interviewed plant managers, shopfloor workers, CEOs and IT managers. He found a high degree of consistency between the responders' answers, suggesting that the single interview of the plant manager was a low cost method of eliciting this type of information. Bloom et al. (2009) interviewed a set of Indian firms using the BSV methodology and then had a management consulting firm independently evaluate the practices of these firms using factory visits, and again found a high degree of consistency.

Finally, European Bank of Reconstruction and Development (2009) organized a face to face survey using a closed-question paper based instrument on 418 UK, German, Indian and Polish firms that had previously been interviewed using the BVR methodology. Again, these responses were well correlated across the different survey tools.

V CONCLUSIONS

We have described some practical strategies for economists who want to conduct surveys on organizational practices. These involve using double-blind survey techniques, with open ended questions against absolute scoring grids, while also collecting background controls for potential survey bias. We also suggested the resurvey of a 5% sample of the original group using different interviewers and respondents within the same organizations to help to validate the methodology. The exact methodology will rest on the research question at hand, but we believe most of these steps should be adopted by researchers aiming to run high quality organizational surveys.

Methodologically, what we are trying to do is somewhere between the traditional approach of economists generating and using large-sample secondary data and the approach of qualitative social scientists using a case study approach. For the types of questions in a range of research areas like personnel economics, organizational economics, and contract theory we believe this methodology has major advantages over more standard approaches.

References:

Bloom, Nick, Benn Eifert, Aprajit Mahajan, David McKenzie and John Roberts. 2009. “Does management matter: evidence from India”, Stanford mimeo.

Bloom, Nick and John Van Reenen. 2007. “Measuring and Explaining Management Practices Across Firms and Countries”, *Quarterly Journal of Economics*, 122(4), 1341-1408.

Bloom Nicholas, Raffaella Sadun and John Van Reenen. 2009. “The Organization of Firms Across Countries”, NBER working paper w15129.

European Bank of Reconstruction and Development. 2009. *Transition Report*. European Bank of Reconstruction, London.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2009. “Reallocation, Firm Turnover and Efficiency: Selection on Productivity or Profitability”. *American Economic Review*, 98(1): 394-425.

Kahnemann, Daniel, Alan Krueger, David Schkade and Arthur Stone. 2004. A survey method for characterizing daily life experience. *Science*, 306 (December), 1776-1780.

Gibbons, Robert and John Roberts. 2008. *The Handbook of Organizational Economics*. Princeton: Princeton University Press.

Grous, Alexander. 2009. *Managerial Practices, ICT and Productivity in Emerging Business Organisation Forms*. London School of Economics PhD thesis, London.

Schwartz, Norbert. 1999. “Self reports: how the questions shape the answers”, *American Psychologist*, Vol 54, pp93-105.

Shaw, Kathryn. Forthcoming. “Insider Econometrics: A Roadmap with stops along the way”,
Labor Economics

ENDNOTES

¹ Further material can be found on-line at: http://cep.lse.ac.uk/_new/research/productivity/management.asp and http://www.stanford.edu/~nbloom/index_files/Page371.htm .

² This raises ethical issues, but Human Subjects Committees have accepted this approach is appropriate since the deception is: (i) necessary to get accurate responses; (ii) minimized to sensitive questions and is temporary (i.e. managers are informed afterwards), and (iii) presents no risk as the data is kept confidential.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

- | | | |
|-----|--|--|
| 968 | Andrew B. Bernard
J. Bradford Jensen
Stephen J. Redding
Peter K. Schott | Wholesalers and Retailers in U.S. Trade
(Long Version) |
| 967 | Pooyan Amir Ahmadi
Albrecht Ritschl | Depression Econometrics: A FAVAR
Model of Monetary Policy During the Great
Depression |
| 966 | Nicholas Bloom
Raffaella Sadun
John Van Reenen | Does Product Market Competition Lead
Firms to Decentralize? |
| 965 | Ralf Martin | Why is the US so Energy Intensive?
Evidence from Multinationals in the UK |
| 964 | Christian A. L. Hilber
Frédéric Robert-Nicoud | Origins of Land Use Regulations: Theory
and Evidence from US Metro Areas |
| 963 | Maria Bas
Juan Carluccio | Wage Bargaining and the Boundaries of the
Multinational Firm |
| 962 | L. Rachel Ngai
Christopher A. Pissarides | Welfare Policy and the Distribution of
Hours of Work |
| 961 | Caroline Freund
Emanuel Ornelas | Regional Trade Agreements |
| 960 | Francesco Caselli
Guy Michaels | Do Oil Windfalls Improve Living
Standards? Evidence from Brazil |
| 959 | Iga Magda
David Marsden
Simone Moriconi | Collective Agreements, Wages and
Restructuring in Transition |
| 958 | Carlos Daniel Santos | Recovering the Sunk Costs of R&D: the
Moulds Industry Case |
| 957 | Nicholas Oulton
Ana Rincon-Aznar | Rates of Return and Alternative Measures of
Capital Input: 14 Countries and 10
Branches, 1971-2005 |
| 956 | Tim Leunig
Chris Minns
Patrick Wallis | Networks in the Premodern Economy: the
Market for London Apprenticeships, 1600-
1749 |
| 955 | Urban Sila | Can Family-Support Policies Help Explain
Differences in Working Hours Across
Countries? |

- | | | |
|-----|--|--|
| 954 | John T. Addison
Alex Bryson
Paulino Teixeira
André Pahnke
Lutz Bellman | The Extent of Collective Bargaining and Workplace Representation: Transitions between States and their Determinants. A Comparative Analysis of Germany and Great Britain |
| 953 | Alex Bryson
Harald Dale-Olsen
Erling Barth | How Does Innovation Affect Worker Well-being? |
| 952 | Nathan Foley-Fisher
Bernardo Guimaraes | US Real Interest Rates and Default Risk in Emerging Economies |
| 951 | Yann Algan
Christian Dustmann
Albrecht Glitz
Alan Manning | The Economic Situation of First- and Second-Generation Immigrants in France, Germany and the United Kingdom |
| 950 | Jérôme Adda
Francesca Cornaglia | The Effect of Bans and Taxes on Passive Smoking |
| 949 | Nicholas Oulton | How to Measure Living Standards and Productivity |
| 948 | Alex Bryson
Bernd Frick
Rob Simmons | The Returns to Scarce Talent: Footedness and Player Remuneration in European Soccer |
| 947 | Jonathan Wadsworth | Did the National Minimum Wage Affect UK Wages? |
| 946 | David Marsden | The Paradox of Performance Related Pay Systems: ‘Why Do We Keep Adopting Them in the Face of Evidence that they Fail to Motivate?’ |
| 945 | David Marsden
Almudena Cañibano | Participation in Organisations: Economic Approaches |
| 944 | Andreas Georgiadis
Alan Manning | One Nation Under a Groove? Identity and Multiculturalism in Britain |
| 943 | Andreas Georgiadis
Alan Manning | Theory of Values |
| 942 | Kristian Behrens
Giordano Mion
Yasusada Murata
Jens Südekum | Trade, Wages and Productivity |