

Panu Kalmi
Derek C. Jones
Antti Kauhanen

Econometric case studies:

overview and evidence from
recent Finnish studies

HELSINGIN KAUPPAKORKEAKOULU
HELSINKI SCHOOL OF ECONOMICS

W-458



Panu Kalmi* – Derek C. Jones** – Antti Kauhanen***

Econometric case studies: overview and evidence from recent Finnish studies

*Helsinki School of Economics,
Department of Economics

**Hamilton College

***Helsinki School of Economics and ETLA

October
2008

HELSINGIN KAUPPAKORKEAKOULU
HELSINKI SCHOOL OF ECONOMICS
PL 1210
FI-00101 HELSINKI
FINLAND

© Panu Kalmi, Derek C. Jones, Antti Kauhanen and
Helsinki School of Economics

ISSN 1235-5674
(Electronic working paper)
ISBN 978-952-488-289-7

Helsinki School of Economics -
HSE Print 2008

Econometric case studies: overview and evidence from recent Finnish studies

Panu Kalmi*, Derek C. Jones** and Antti Kauhanen***

Abstract:

In this report we give an overview of econometric case study method. We review some key econometric case studies and then we present our own case studies that were conducted under the project “Econometric Case Studies on the Impact of Workplace Innovations”. We present some evaluation criteria for econometric case studies as an approach and position our studies in the context of the approach.

* Corresponding author. Helsinki School of Economics. Contact: panu.kalmi@hse.fi,
tel. +358 40 749 6875

** Hamilton College.

*** Helsinki School of Economics and ETLA.

The authors are grateful to support from the Academy of Finland (Project No. 206027) and The Finnish Work Environment Fund (Grant No. 103313). Jones’ work was supported in part by a Foundation for Economic Education Fellowship and NSF SES-0522117 for which he is grateful. This report was prepared as a final report for the Finnish Work Environment Fund.

Introduction

We start this report by giving an overview to the emerging research method in labour economics, namely econometric case studies. The broader purpose of this report is to present the research results of the research project “Econometric Case Studies on the Impact of Workplace Innovations”, funded by the Finnish Work Environment Fund and the Academy of Finland. The funding period for the initial project was January 2004 – May 2008, although many parts of this project are still going on as of September 2008 with new funding decisions.

Econometric case studies employ econometric methods and focus on a single organisation, where the unit of analysis is typically some subunit of the organisation (e.g. plant, production line, team, or individual). Unusually for economics research, it is based on field work and the interaction between the research object and the researcher. It is a quite recent empirical approach that has crucially been influenced by the development of personnel economics (see Lazear (1999) for an overview of personnel economics).

Within personnel economics, several questions were posed to which more traditional methods could not provide answers (e.g., the role of characteristics of team members in influencing team productivity) and thus a need to collect new types of data arose. In addition, econometric case studies enable improved controls and better measures to be used in addressing established questions.

To understand better the distinctive nature of econometric case studies, we continue by providing a detailed comparison of the characteristics of the econometric case study method that distinguish it from other influential empirical methods that also study the relationship between human resource management (HRM) practices and productivity. Econometric case studies are different from traditional case studies and another recently developed approach, namely “informed survey analysis” (Bartel et al., 2004). Then we review nine econometric case studies done by others. For these studies we highlight characteristics of the data sets, discuss the roles played by interviews and surveys and assess the strengths and weaknesses of the method.

After this we present eight case studies that were done under our project. The focus is not so much on the results that we received, but rather on methodological issues: how did these papers address the challenges of the econometric case studies as a method, and how have they succeeded in bringing the approach forward. Next, we present some evaluation criteria for econometric case studies as an approach, noting that construct validity and internal validity appear as significant strengths of the approach, while the approach has often been criticised – perhaps sometimes unfairly – for the lack of external validity.

We conclude by evaluating some plausible future directions for the literature and recommendations for further study.

Comparisons of econometric case studies with other methods

- i) Comparisons with case studies

Sometimes case studies are mistakenly taken to be synonymous with qualitative research methods, even though the main methodological texts within the field are clear in stressing that case studies can – and often do – consist of both qualitative and quantitative analysis (Eisenhardt, 1989; Yin, 1994). Since econometric case studies often utilize qualitative data alongside with the quantitative data, the difference is more one of degree than of kind. In traditional case studies, quantitative data can be used to verify the findings that are gained by qualitative evidence, or even as the main form of evidence. However, traditional case studies rarely, if ever, subject the quantitative data to hypothesis testing using advanced econometric methods. In econometric case studies the balance between quantitative and qualitative analysis is thus different. Qualitative data has a supportive, although often an important, role.

Another difference between econometric case studies and traditional case studies is the role of theory. In traditional case studies, the researcher starts with an initial hypothesis and then builds theory iteratively from the evidence collected from the case (Eisenhardt, 1989). By contrast, in econometric case studies, while there is a feedback from data collection and analysis to theory, testing competing a priori hypotheses is more pronounced relative to theory building.

ii) Comparisons with survey research

In traditional survey research by economists, typically official and / or secondary data are used and the researcher does not influence the data collection process nor interact

with the research object. However, new methods have emerged within the survey tradition that are closely related to econometric case studies. The hallmark study is Ichniowski et al. (1997) who visited 17 US steel plants and collected data from 37 different steel finishing lines. Their data on HRM were collected by standardized interviews. Unlike traditional survey work by economists, this study includes significant amounts of fieldwork and researcher involvement in data collection. More recently, Bartel et al. (2004) have described this research method as “informed survey analysis”. A major motivation for this type of research has been the access to improved measures of productivity.

The key difference between econometric case studies and “informed surveys” is that the former relies on single-firm studies, while the latter compare different organisational units (although some observations may come from the same organisation). Informed surveys usually investigate heterogeneity between organisations, while econometric case studies focus on within-organisation heterogeneity.

Application of the method: key studies

Table 1 presents nine major productivity studies within the field. Five studies analyse piece rates as a compensation method, with four contrasting piece rates and time rates, while the remaining study contrasts the productivity impact of piece rates with a system where workers’ wages are determined by productivity relative to their peers. Two other studies investigate the adoption of teams. One article employs indicators of HRM

environments as derived from employee questionnaires and the remaining paper examines the impact of profit-sharing.

The studies on piece rates have typically found positive and often sizable productivity effects. Lazear (2000) finds that the adoption of piece rates instead of fixed wages increases productivity by 44 % in windshield installing. Bandiera et al. (2005) find even bigger impact when they study a move from relative compensation scheme to piece rates in fruit picking: they report a 59 % productivity increase. Paarsch and Shearer (2000) study the impact of piece rates on tree planting industry and find that a lower bound for the incentive effect of piece rates is 23 %. Shearer (2004) reports the results from an experiment where employees were assigned the compensation system randomly. He finds that productivity under piece rates is at least 22 % higher than under time rates. Unlike these four studies which use data on individuals, Freeman and Kleiner (2005) use data at the plant level. However, they also find that productivity is higher under piece rates.

The general lesson from piece rate studies is that people respond to incentives, but the lesson is not that piece rates should be universally adopted. First, the use of piece rates can be costly. Freeman and Kleiner (2005) report that while productivity fell after the change from piece to time rates, costs fell even more and thus profits improved.

Similarly, Paarsch and Shearer (1999) find that the productivity gains from piece rates are likely to go to employees. Second, econometricians are most likely to collect data when there are precise measures for output and measurement takes place frequently. Piece rates are most likely observed under the same conditions. Thus econometricians are able to

collect data from the impact of piece rates under conditions when the expected effect of piece rates are highest and therefore are likely to overestimate the benefits of piece rates in other settings.

Turning to the team literature, Hamilton et al. (2003) find that worker productivity increases by 16% after the introduction of on-line teams and the change to team compensation in a garment factory. Interestingly, they also note that the first workers who join teams are more able than average workers. This is in marked contrast with earlier theories that predict that teams would suffer from adverse selection problems. They also find that most heterogenous teams are those that are able to improve their collective performance most. While Hamilton et al. (2003) are able to observe individual productivity only before the introduction of teams and thereafter they observe productivity only at the level of teams, Jones and Kato (2006) observe the performance of individuals throughout. They find that the introduction of off-line teams directly improved worker productivity by 3 %.

In a study of the airline industry Knez and Simester (2001) find that the adoption of profit-sharing significantly improves performance, which is measured by the frequency of on-time departures. Compared to most other econometric case studies this study uses data that are at a more aggregate level (the airport) and less frequent (monthly).

The final paper is by Bartel (2004). She discusses the impact of HRM environment in banking. Using annual data, she finds that HRM environment has a statistically and economically significant effect on productivity.

An important difference in research design among studies is whether or not they are event studies. Seven out of the nine studies fall into this category.¹ In five out of seven cases, there is some post-event heterogeneity in the treatment that allows the researcher to control more effectively for other post-event changes that may have affected productivity. Of course, to the extent that this heterogeneity is not random, the researchers have to address the issue of endogeneity in some way. However, sometimes the selection issue itself may be of primary interest. For instance, in Lazear (2000) the least able employees are found to select out of the firm after the adoption of the piece rates, whereas Hamilton et al. (2003) find that the most able employees select into teams first.

Another important thing to notice is that many studies that are conducted at low levels of aggregation or no aggregation (individuals) and have frequent longitudinal observations (in many cases, daily observations) over a relatively long period, this results to a large number of observations. Perhaps surprisingly, econometric case studies focusing on a single organization may yield very large datasets. For instance, the study of Jones and Kato (2006) includes over 77,000 observations, although there is only slightly over 100

¹ Shearer (2004) is a somewhat special case. Other eight event studies are “natural experiments” in the sense that the researchers try to estimate the productivity impact of an exogenous change. In Shearer’s study, the researcher is able to impose the treatment experimentally on the employees.

cross-sectional observations (individuals) that are followed over time. This gives substantial statistical power to these studies.

Embedded designs in econometric case studies

i) The role of interviews

Unlike most research in economics econometric case studies use interviews. One purpose of interviews is to provide auxiliary evidence. From her interviews at bank branches, Bartel (2004) found out how HRM affected productivity and employee motivation. Knez and Simester (2001) also provide evidence derived from interviews on how profit-sharing was perceived both by managers and employees as the main reason for improved efficiency. While most economists would not consider this type of evidence alone as convincing, it is recognized that it provides a useful complement to quantitative evidence.

Interviews may also provide important clues as to how to interpret other data. Paarsch and Shearer (1999) provide a good illustration of this point. Contrary to theory, when they regress trees planted on the intensity of the piece rate, they identify a negative relationship. However, interviews revealed that piece rates are adjusted upward when the planting conditions are particularly difficult, thus generating the observed negative relationship (see pp. 651 – 2). Similarly, Paarsch and Shearer (2000) report that the productivity of workers receiving both fixed wages and piece rates was higher than those receiving only piece rates. This finding is contrary to the standard prediction that high-

ability workers sort to piece rates and low ability workers sort into time rates. However, interviews revealed that those working only under piece rates tended to be students hired just for summer employment and having lower ability than permanent workers (p. 77). In explaining their result that profit-sharing improved efficiency, Knez and Simester (2001) use evidence from interviews and fieldwork to show that the autonomous work groups used at Continental Airlines provided a conducive environment for mutual monitoring.

A third function of interviews is to use them to rule out rival explanations. By formulating a plausible rival hypothesis that could have explained the observed findings, interview evidence can be used to refute alternative hypotheses, especially where appropriate quantitative data are unavailable. As documented in Knez and Simester (2001), interviews may also help to identify appropriate control variables.

As Helper (2000) points out, there is no established protocol on how economists should report their qualitative methods. Often the use of interviews is omitted from the “methods” section and there are only occasional sentences where it can be inferred that some interviews were indeed conducted. In our view, when interviews or surveys are used, this should be recognised in the text. An exemplary discussion is in Bartel (2004, 185-6).

ii) Use of surveys

While interviews are conducted in nearly all econometric case studies, the use of surveys is rarer. One exception is Jones and Kato (2006), where the authors report auxiliary evidence from a survey of individual employees. This evidence shows that individual team members put more effort into their work, a finding that was supported by the panel data evidence on enhanced individual productivity. Bandiera et al. (2005) use survey evidence collected on the social networks of employees to determine whether lower productivity that was observed under the relative incentive system was attributable to pure employee altruism, or whether it reflected collusion among friends. Their results strongly supported the latter interpretation.

Finnish Econometric Case Studies

In this section we discuss the case studies that were conducted within the project “Econometric Case Studies on the Impact of Workplace Innovations”. At the time of writing (September 2008), these studies are at different stages of completion. One of them has been published in an international academic journal, some are in review process, and others are in preparation to be submitted. For each study, we briefly outline its context, methodological features and key results. After that, we evaluate them on the basis of the preceding methodological discussion and literature review.

Five of the studies are based on retail trade, on two different chains. Two studies are based on the data of Finnish co-operative banks, and one study on a food producing company. We start with the retail trade studies.

1: HRM and Productivity in Retail Trade (Jones, Kalmi, Kauhanen 2007a)

In this paper we use data from a Finnish retailer having 47 establishments all over Finland. We use data from three years (2001-2003). Internal financial records are combined with HRM data that comes from a survey organized by an independent consultant.

In this study we used interview evidence to understand the context in which the stores were operating and the HRM might contribute to productivity. This is reflected in the discussions of the job context.

Dependent variable is the level of productivity, measured as value added relative to inputs. We use fixed effect analysis including store and manager dummies, and do the estimation in the Cobb Douglas framework. The analysis is conducted at annual level.

The results indicate that participation in substantive decisions and receiving relevant information are related to productivity in statistically significant ways. One problem is that the HRM variables are heavily correlated with each other, so we have to take multicollinearity into account in our estimations. We do this by means of principal component analysis (PCA). The new scale constructed by PCA is also statistically significantly related to productivity. An increase of standard deviation of the first component increases productivity approximately by 3%.

2: ERP introduction and performance in retail trade (Jones, Kalmi, Kauhanen 2008a)

This paper builds on data from the same firm as in our paper # 1 but it uses a longer data set. We are interested in the development of productivity and inventory turnover after the introduction of ERP system. The new ERP system was introduced gradually during 2004. We look at the time profile of the introduction of the new system. It is well known that the introduction of ERP may lead into a performance dip and thereafter a resurgence of performance. Inventory turnover is included because this is one of the main things to be influenced by the ERP system. Second, we broaden the existing literature by examining how the use of ERP affects the subsequent performance development. We also examine the effects of HRM and late adoption. Finally, we look at how employees viewed the adoption of the system and how they perceived it.

In this study we used interviews and also additional survey data collection in order to understand the developments. The survey data was collected in three different stages: twice from department managers and once from employees. The additional survey analysis is used both in descriptive analysis and as an input to the econometric analysis.

The basic findings are that we find that the introduction of ERP was accompanied with a variety of use-related problems and therefore a significant increase in the amount of work employees had to cope with. However these problems gradually decreased over time and this was reflected as lower perceived increase in the amount of work and as increased motivation by the employees. In our financial data, this was indicated as a U-shaped

curve on performance; a deep initial decline and thereafter a resurgence in performance. We find that broader use of ERP has mostly positive performance effects. However, contrary to some hypotheses, we do not find evidence on positive interactions between HRM climate and ERP adoption. Firms that adopt ERP later fare somewhat better. While the quantitative analysis contributes to the measurement of the effect, the underlying causes behind these effects become much clearer when combined with the interview and case study evidence.

3: The impact of new sales incentive scheme on incentives in retail trade (Jones, Kalmi, Kauhanen 2008b)

This paper has a focus on an actual HRM event and its impact. We study the effects of an introduction of a sales incentive scheme using a before-after analysis. The sales incentive scheme is such that sales are measured over a three-month period. Bonus is paid if sales budget is exceeded. The bonus pay-outs depend on realized sales according to a step-wise (non-linear) scheme. The paper addresses the questions i) whether the sales bonus scheme increases productivity (sales and gross proceeds relative to use of labor and capital); ii) whether employees adjust effort based on realized performance.

The data we use is based on the 53 retail stores for 45 months in 2004 – 2007. The incentive plan is introduced roughly in the middle of the period. The results indicate that productivity increases significantly after the introduction of the performance based pay. The estimated increase in productivity is roughly 7.5%. Our results also indicate that the

realized performance in the end of the period depends on the accumulated performance during the earlier months. This is consistent with the interpretation that if the performance has been lagging behind the target, employees may not consider worthwhile exerting extra effort, since reaching the target may be unlikely.

4: HRM and Productivity in Retail Trade: a Replication Study (Kalmi, Jones and Kauhanen 2006)

A novel feature of this paper is that it is a replication of the study # 1. The research designs in the two cases are very similar and this study is intended to check whether the results are similar when the study is replicated. In this particular study, we have 35 establishments and 48 months. As in the other case, the HRM data comes from a separate survey. Financial data is very similar to the earlier case. We use Cobb-Douglas production functions and store level fixed effects.

The results we find are broadly similar and therefore increase the external validity of the previous research results. The impact is however somewhat smaller, a standard deviation in the HRM scale increases productivity by 1%. However even this effect is statistically significant. We argue that the difference in productivity impact might be because stores are much more capital intensive in the latter case.

5: Productivity dispersion in retail trade (Kauhanen and Roponen 2008)

This paper extends the analysis of paper # 4. Instead of looking what affects average productivity of the establishments, it focuses on the dispersion of productivity between establishments and its development over time. We attempt to explain the productivity with HRM environment and employee skills. The data is monthly, with 33 stores and 48 months (2002 – 2005).

The findings indicate that productivity is highly dispersed between stores. This dispersion decreases over time, but the productivity rankings are very consistent. HRM and employee skills appear to explain only a small proportion of these findings.

6: The impact of training on efficiency in cooperative banks (Jones, Kalmi and Kauhanen 2008c)

In this paper we combine financial data on cooperative banks with HRM records and data on training. Training is measured by financial expenses used in training and time used for training. We also have data that is related to the generality and specificity of training. This is important because much literature has stressed that employers should have stronger incentives to provide specific rather than general training.

We have data for all cooperative banks that belong to the OP – Pohjola group, though we have to exclude the smallest banks from the analysis. In this analysis we have done specific interviews with employees and management. Main evidence comes from

econometric analysis. One of the advantage of the data is that we are able to take the costs of training more fully into account than in the previous analysis.

Our findings show that general training, but not specific training, is related to higher wages at the level of the bank. However, there is not much evidence that higher levels of training at the bank level are related to higher efficiency, measured as cost efficiency and profits. This does not have to indicate that training is ineffective. Instead, it may mean that banks are also training optimally and increases in training do not significantly raise efficiency (net of costs). It may also mean that the gains from training go disproportionately to workers in the context of co-operatives, where sharing of residual revenues is limited.

7: Determinants of membership in cooperative banks (Kalmi, Jones, Jussila 2007)

This paper differs from others by having a focus on customers, instead of employees. We discuss the determination of membership in cooperatives. This paper uses basically the same data as paper #6. We analyze the determination of cooperative membership on the basis of theory that membership motivations are partially determined by individual monetary motivations and partly by concern for others. We find that average transaction intensity of banks is significantly related to membership rates, supporting the idea that individual pecuniary motivations are related to the decision to become a member. We delve deeper than previous literature into the question of the sources of non-pecuniary motivations. We identify that these would be desire to participate in decision-making,

ensuring the availability of regional services, and utilizing collective reserves. Our results generally support the idea that there are both types of motivations explaining membership in the cooperative.

#8: Teams, performance related pay and productivity in a food production plant (Jones, Kalmi and Kauhanen 2007b)

In this paper we study the effects of various actual HRM events on productivity. This study is done at a more disaggregate level than other studies and it focuses on one plant of a meat producer. This plant has four different lines. We obtain for these lines long and detailed weekly data on productivity for seven years. The productivity measure is adjusted for used inputs (labor, capital) and change in equipment.

There happened several changes in the company that were related to the introduction of teams and subsequent changes in performance pay. The theory of complementarity would lead us to expect that one type of HRM practice is not very efficient without the other.

The long time dimension of the data (364 weekly observations) makes it possible to study the different HRM regimes by the means of time series analysis. We use advanced time series techniques, especially the Bai-Perron method which assigns the break dates endogenously. The results are consistent with the hypothesis that complementary packages of work organization and reward structure are needed before any observable changes in productivity occur. There is also some heterogeneity in the results, as the

positive impact does not show up in one of the lines. Interview evidence is used to explain this puzzle.

Putting the Finnish case studies in a context

Type of data

Almost all of our studies use panel data. The advantage of panel data is that it makes a stronger base to make causal inferences. One reason for this is that panel data makes possible the use of fixed effects methodology that allows controlling for unobserved time-invariant heterogeneity between observations. Second advantage is that it allows before-after comparisons in event studies, where one can observe changes induced by distinct events.

Panel datasets are formed quite naturally for two of our cases. The retail firms in our dataset each have a fairly large number of stores that have a considerable degree of independence, and each of them also keep their own accounts. This makes it possible to use the stores as the units of observation. Similarly, in cooperative banks, the individual banks (which are 240 in total) are natural units of observation.

The food-producing plant is different in the sense that there are only four production lines, which is not a sufficient number of cross-sections for panel data analysis. In this case we are lucky that there is a very large longitudinal dimension in the analysis that

makes it possible to do a detailed time series analysis. Conducting a time-series analysis within an econometric case study is a novel approach to the literature.

Event vs. non-event studies

Quite many of the classic papers in the field of econometric case studies, including the classic paper by Lazear (2000), are event studies in the sense that they incorporate data before and after a certain event has taken place. Three of our papers (#2,3,7) are event studies in the sense that they analyze the effect of a discrete event (the introduction of the ERP system and the introduction of the incentive pay in the retail firm, the introduction of team organization and incentive pay in the food producer), while the other studies are based on panel data where the explanatory variables are continuous variables that vary on annual basis.² This requires that attention will be paid to endogeneity issues.³

Another issue arising in event studies is whether there is cross-sectional variation after the event in the sense that the event occurs in some units and not in others. This helps to isolate the true effect of the event from other changes that may take place simultaneously.

² An exception here is the study on the cooperative memberships, that uses cross-sectional methods due to the fact that some of the explanatory variables change very little from year to year.

³ An independent variable is said to be endogenous if the dependent variable is having a causal effect on the independent variable. For instance, the HRM variables used in the retail trade analysis would be endogenous if the financial performance of the stores would have an effect on the HRM variables. This would mean that the coefficient of the HRM variable would be biased.

It may often be quite difficult to find examples of cases where there is such cross-sectional variation. The HRM policies of the firm are often such that the changes in the workplace are introduced simultaneously in all units. This is often a requirement of the workplace cooperation laws and collective bargaining agreements.

For these reasons, we do not observe much cross-sectional variation in our event studies. A partial exception is the study on the ERP adoption, where the introduction of the system takes place over a period of time, though this variation is rather limited. The lack of cross-sectional variation means that we have to exclude alternative explanations by other means. For instance, in the paper focusing on the productivity effects in the food producer we include a general index of productivity within the food producing industry in our time series regression.

Evaluating econometric case studies: different types of validity

In the social sciences methodology, four different types of validity are distinguished: construct validity, internal validity, external validity (generalizability) and reliability (replicability) (e.g. Yin, 1994, 33). Of course the importance of these types of validity will vary across individual studies, and there will also be trade-offs depending on empirical approaches. For instance, there appears to be a trade-off between construct validity and external validity that also others have stressed (e.g. Ichniowski and Shaw, 2003). In this section, we explain the types of validity in econometric case studies, and

how we aim to use the strengths of the approach to our advantage, and how we respond to possible criticism.

i) Construct validity

By construct validity we mean that the measures really represent what they purport to measure. Problems of construct validity have been a major impetus in the recent shift in labour economics towards the “insider econometrics” approach (Ichniowski and Shaw, 2003). In the analysis of the links between HRM and productivity, measurement problems plague both measures of the dependent variable as well as the HRM independent variables and other independent variables. These problems are acute in traditional studies, especially those that use firm-level data. Thus many studies examining the relationship between HRM and productivity have had to use value added and, in many cases, even sales as the dependent variable. While this is necessary when the sample consists of heterogeneous units, the use of sales or value added is not ideal since the effects of output prices means that such measures often poorly capture underlying productivity. Moreover, several econometric case studies have been able to use efficiency indices as their dependent variables.

Similar problems arise when measuring labour or capital inputs. Labour might sometimes be measured as wages or (somewhat preferably) as the number of workers. By contrast, in econometric case studies, labour is usually measured as the number of hours. Also often labour quality can be controlled for and the measures of capital may also be better

in econometric case studies than in traditional studies. For instance, efficiency norms already control for relevant characteristics of labour and capital, including characteristics that may otherwise remain unmeasured.

The HRM characteristics are also poorly measured in many survey studies (see also the discussion in Edwards et al., 2002). First, the concepts often invoked in the studies concerning HRM may be somewhat vague – for instance, the meaning of the concept of “teams” may differ between respondents. In an econometric case study, one can be at least more specific on what type of teamwork is discussed. Second, the respondents in large-scale surveys are often persons relatively high in the organisation, who may either not have the appropriate knowledge, or they may give an idealised description of the reality. These problems are also minimized in econometric case studies that rely on multiple and well-informed respondents.

We argue that our studies are also rather strong in construct validity. The stores in the retail trade case are very similar to each other, as are the individual cooperative banks. This increases the reliability of productivity comparisons. The explanatory variables are also measured in a consistent way – for instance the HRM data that comes from the HRM surveys in the retail trade, or the training data in the banking case that is much more detailed than the data that is usually used in the training research. The bulk of the training is organized internally by the central unit of the bank, so our training measure is also very homogenous.

A good example of construct validity is our measure of productivity in the food producing case. The management had constructed an efficiency measure that took into account labor and capital inputs. Similar measures have been used also in the previous econometric case studies, including the studies of Hamilton et al. (2003) and Jones and Kato (2006). This reduces the need to use control variables for labor and capital in the equations.

ii) Internal validity

Internal validity refers to the ability to make causal inferences based on the available data. In practice this means that the researcher tries to rule out alternative explanations for the observed phenomenon. In survey analysis, practically the only way to do this is through the inclusion of various control variables and fixed effects.

There are several reasons why internal validity in econometric case studies compares favourably to large-scale surveys. One is the higher construct validity, as discussed above. Better measures of productivity and HRM enable the researchers to make firmer inferences. Also interviews can be used to improve the internal validity. The discussions between the researchers and case informants often guide the researcher to choose the appropriate control variables. The interviews also provide valuable clues on how to interpret the results, and may help in identifying exceptional observations.

For instance, in the case of food producing company, we would not have had good explanations on the heterogeneity of the HRM effects without interviews that pointed out differences in the organization, differences in the composition of workforce and wage structure, that are helpful in understanding this heterogeneity. Another example how interaction with company management increases internal validity of the research is the study on ERP effects. Interviews with company management and store employees were instrumental in drafting the questionnaires. When survey data had been initially analyzed, the preliminary results were presented to the management representatives, they were discussed and new interpretations were suggested. The discussions in these sessions also helped in planning subsequent questionnaires.

iii) External validity, replicability and reliability

A criticism that is often voiced of case studies is that their external validity (generalizability) is poor. In our view, this point is often overstated. There are two types of generalizations: statistical generalizations and analytical generalizations (Yin, 1994, 36). Statistical generalization takes place when results from a representative sample are generalized to the relevant population. This can be done if the initial sample was chosen so that it is representative of the population and no significant response bias affects the realized outcomes. Analytical generalization means that the findings of the study can plausibly be extended beyond the immediate application. This requires that the findings are consistent with a theoretical hypothesis. Analytic generalizations are typically routinely reported in the “conclusions” section of the paper, and they rely on the

judgement of the researcher. The subsequent research far more typically challenges the analytical generalizations made in the study rather than the statistical generalizations.

One cannot generalize on the basis of econometric case studies on any population of firms, since one observation cannot form a representative sample. Similarly, even if the case study consists of hundreds or thousands of individual observations (e.g. individuals), the effects cannot still typically be generalized. A sample of a thousand workers within one firm is not a representative sample of workers, when one considers generalizations outside the case firm. However, this limitation is not devastating. Most survey studies have also limited statistical generalizability. For instance, Ichniowski et al. have data on 36 steel finishing lines located in the United States. In total, there were 60 lines of that type during the time the study was made (Ichniowski et al., 1997; 292). Thus, their sample can be taken to be representative of US integrated steel finishing. However, their results cannot be generalised statistically outside the steel industry, to Chinese steel mills, or even to other US steel mills using a different technology. Cross-industry studies such as Black and Lynch (2004) have obviously broader statistical generalizability, but even they have limitations. The findings from US studies cannot be statistically generalized to other industrial relations environments, such as Germany. Of course, the predictions can be analytically generalized to Germany and this generalization can be tested by a replication study. Ichniowski and Shaw (2003) note that by extending the domain of statistical generalizability the study usually loses in some other dimension, such as productivity measurement (construct validity).

A consequence of the impossibility of doing statistical generalisations based on case studies is that it is sometimes advised to choose a “representative” case. This is wrong. Apart from the difficulties of determining what is a “representative” case, there is no reason to expect that the average observation would somehow produce the average effect.⁴

Moreover, the crucial strength of insider econometric approaches, such as econometric case studies and informed surveys, is that they provide a reliable basis for analytical generalizations. This follows from their careful use of theory, high internal validity, and careful reflections on the context. It is always helpful to outline the context where the HRM innovations are taking place, in order to understand to what extent the results can be analytically generalized. While there are no firm guidelines as to how analytical generalisations should be done (and they always require deliberation based on the features of particular cases), as more econometric case studies are conducted one is able to assess whether similar practices produce consistent results in different circumstances or whether results differ. As Helper (2000) puts it, “... the solution to the generalizability problem is to do more field research, not less!”

For individual studies, it is helpful to outline the context where the HRM innovations are taking place to understand to what extent the results can be analytically generalized. In the food industry paper, we discuss in detail the context where the HRM innovations were adopted. These help to understand the conditions where firms may be willing to

⁴ This would require very specific and implausible assumptions on the population variance of the disturbance term.

undertake various HRM innovations. We use similar strategies in discussing the retail trade cases.

Another point related to external validity of econometric case studies is their replicability. External validity can be increased if one set of results can be replicated in a different context but using similar research design. This is very common in sciences that are conducted in laboratory or otherwise controlled settings, such as in medicine. In economics they are rarer (however see Ehrenberg and Bognanno 1990 and Hamermesh 2007 for a survey). Our study # 4 presents a rare example of a study where the research design of another study (#1) is replicated in a different context.

A different sense of replicability or reliability is that another researcher could use the same evidence to replicate the results of the original study. And in economics it is increasingly common to transfer quantitative data sets to other researchers so as to alleviate concerns concerning the reliability of results. However, since the data source (the firm) usually regards the data as confidential, this practice does not normally extend to econometric case studies. Indeed, sometimes confidentiality requirements even prevent the presentation of key summary statistics (Bandiera et al., 2005, 928; Table 1).

A more serious variant of this problem is that the company, who has veto powers over the use of the data, wishes to terminate the research, perhaps because of management changes, because the results are putting the company into a negative light, or the results are not consistent with the strategic guidelines under which the company is operating. We

observed constraints of the first kind, i.e. the data was not proprietary and cannot be made available to third parties. Fortunately, we did not experience much problems of the first kind.⁵

Conclusions

In this report we present a review and a methodological discussion of the econometric case study method. Our discussion clarifies what econometric case studies are and how they differ from other related approaches. We outline key features of the main studies within the field and we also evaluate the strengths and weaknesses of econometric case studies. We argue that econometric case studies have higher construct validity and internal validity than competing approaches, but that there continue to be doubts concerning its external validity. We discuss how these concerns can be alleviated by doing more econometric case studies, including replication studies. We review our own eight econometric case studies, conducted within the project “Econometric Case Studies on the Impact of Workplace Innovations”, and position them within the broader approach.

Our final task is to briefly consider the potential of the econometric case study method for future studies. We note that there are still many potential advantages of the method that do not appear to have been used to best advantage to date. For example with some limited

⁵ We entered into agreement with one manufacturing firm on research that did not go forward because of the passive attitude of the company. The costs in terms of foregone effort remained limited.

exceptions the potential for realizing major gains exists by combining different types of data within the case. For example by combining customized survey data within a case with objective performance data one can begin to better understand the specific channels through which HRM policies may enhance enterprise performance. Another possible extension of the method that promises large payoffs is to use multiple units such as employee outcomes and firm outcomes. One obvious route is to try to assemble and combine panels of data for employee outcomes with panels of objective performance data. Finally, as the researchers develop trust with the case, the possibility arises of persuading the case to permit experiments to be undertaken. This would largely solve the endogeneity problems that are pervasive in non-experimental work. Major published studies along these lines are Shearer (2004) and Bandiera et al. (2007).

REFERENCES

Bandiera, O., Barankay, I., and Rasul, I. (2005), 'Social Preferences and Responses to Incentives: Evidence from Personnel Data', *Quarterly Journal of Economics*, **120** (3), 917-962.

Bandiera, O., Barankay, I., and Rasul, I. (2007), 'Incentives for Managers and Inequality Among Workers: Evidence from a Firm-Level Experiment', *Quarterly Journal of Economics*, **122** (2), 729 – 73

Bartel, A.P. (2004), 'Human Resource Management and Organizational Performance: Evidence from Retail Banking', *Industrial and Labor Relations Review*, **57** (2), 181-203.

Bartel, A., Ichniowski, C. and Shaw, C. (2004), 'Using 'Insider Econometrics' to Study Productivity', *American Economic Review*, **94** (2), 217-223.

Black, S., and Lynch, L. (2004), 'What's Driving the New Economy: The Benefits of Workplace Innovation', *Economic Journal*, **117** (February), F97-F116.

Edwards, P., Geary, J., Sisson, K. (2002), 'New Forms of Organization in the Workplace: Transformative, Exploitative, or Limited and Controlled?', in G. Murray, J. Belanger, A. Giles and P.-A. Lapointe, *Work and Employment Relations in the High Performance Workplace*, London: Continuum, 72 – 119.

Ehrenberg, R. G. and Bognanno, M.L. (1990), 'The Incentive Effects of Tournaments Revisited: Evidence from the European PGA Tour', *Industrial and Labor Relations Review*, **43**, 74-S-88-S.

Eisenhardt, K. M. (1989), 'Building Theories from Case Study Research', *Academy of Management Review*, **14** (4), 532-550.

Freeman, R. and Kleiner, M. (2005), 'The last American Shoe Manufacturer: Decreasing Productivity and Increasing Profits in the Shift from Piece rates to Continuous Flow production', *Industrial Relations*, **44** (2), 307-330.

Hamermesh, D.S. (2007): 'Viewpoint: Replication in Economics', *Canadian Journal of Economics*, **40** (3): 715-33.

Hamilton, B. H., Nickerson, J. A., and Owan, H. (2003), 'Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation', *Journal of Political Economy*, **111** (3), 465-497.

Helper, S. (2000), 'Economists and Field Research: "You Can Observe a lot Just by Watching"', *American Economic Review*, **90** (2), 228-232.

Ichniowski, C. and Shaw, K. (2003), 'Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices', *Journal of Economic Perspectives*, **17** (1), 155-80.

Ichniowski, C., Shaw, K., and Prenzushi, G. (1997), 'The Effects of Human Resource Management on Productivity', *American Economic Review*, **87** (June), 291-313.

Jones, D.C., Kalmi, P., and Kauhanen, A. (2006): 'Human Resource Management Policies and Productivity: New Evidence from an Econometric Case Study', **22**(4): 526-38.

Jones, D.C., Kalmi, P, and Kauhanen, A. (2007a), 'How Does Employee Involvement Stack Up? The Effects of Human Resource Management Policies on Performance in a Retail Firm', Manuscript, Helsinki School of Economics.

Jones, D.C., Kalmi, P, and Kauhanen, A. (2007b), 'Teams, Performance-Related Pay, Profit-Sharing and Productive Efficiency: Evidence from a Food-Processing Plant', Manuscript, Helsinki School of Economics.

Jones, D.C., Kalmi, P., and Kauhanen, A. (2008a): 'Enterprise Information Systems, Human Resource Management and Performance', Manuscript, Helsinki School of Economics.

Jones, D.C., Kalmi, P., and Kauhanen, A. (2008b): 'Non-Linear Incentives in Retailing', Manuscript, Helsinki School of Economics.

Jones, D.C., Kalmi, P. and Kauhanen, A. (2008c): 'The Effects of Training on Performance and Incomes: Econometric Evidence from Finnish Co-operative Banks', Manuscript, Helsinki School of Economics.

Jones, D. C. and Kato, T. (2006), 'The Impact of Teams on Output, Quality and Downtime: An Empirical Analysis Using Individual Panel Data', working paper, Hamilton College.

Kalmi, P., Jones, D.C. and Jussila, I. (2007): 'What Determines Membership in Cooperatives? Evidence from Finnish Banks on the Role of Ownership', Manuscript, Helsinki School of Economics.

Kauhanen, A. and Roponen, S. (2008): 'Productivity Dispersion: A Case Study in the Finnish Retail Trade', Manuscript, Helsinki School of Economics.

Knez, M. and Simester, D. (2001), 'Firm-Wide Incentives and Mutual Monitoring at Continental Airlines', *Journal of Labor Economics*, **19** (4), 743-772.

Lazear, E.P. (1999), 'Personnel Economics: Past Lessons and Future Directions', *Journal of Labor Economics*, **17** (2), iv + 199-236.

Lazear, E. P. (2000), 'Performance Pay and Productivity', *American Economic Review*, **90** (5), 1346-1361

Paarsch, H.J. and Shearer, B.S. (1999), 'The Response of Worker Effort to Piece Rates: Evidence from the British Columbia Tree-Planting Industry', *Journal of Human Resources*, **34** (4), 643-667.

Paarsch, H.J. and Shearer, B.S. (2000): 'Piece Rates, Fixed Wages and Incentive Effects: Statistical Evidence from Payroll Records', *International Economic Review*, **41** (1): 59 – 92.

Shearer, B. (2004), 'Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment', *Review of Economic Studies*, **71** (2), 513-534.

Yin, R.K. (1994), *Case Study Research: Design and Methods*, Thousand Oaks, CA: Sage.

Table 1: Selected examples of econometric case studies

Author	Lazear	Paarsch and Shearer	Knez and Simester	Hamilton et al.	Shearer	Bartel	Bandiera et al.	Freeman and Kleiner	Jones and Kato
Year	2000	2000	2001	2003	2004	2004	2005	2005	2006
Research question	Productivity and sorting effects of piece rates	Productivity effects of piece rates	Impact of profit-sharing on on-time departures	Productivity and sorting effects of team compensation	Productivity effects of piece rates	impact of HRM on sales growth	productivity effects of relative vs. absolute incentives	productivity and profits after the change from piece to time rates	performance effects of off-line teams
Type of activity	windshield installing	tree planting	Airport departures	Textile	Tree planting	Banking	Fruit picking	Shoe manufacturing	Light manufacturing
Productivity measure	# of units	# of units	% of on-time departures	Efficiency standard	# of units	Sales growth	Kg / hour	Actual production / planned production	Efficiency standard
Unit of observation	Individual	Individual	Airport	Individual	Individual	Branch	Individual	Plant	Individual
Frequency of observation	Monthly	Daily	Monthly	Weekly	Daily	Yearly	Daily	Monthly	Daily
Event study	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Cross-sectional variation after the event	Yes	N/A	Yes	Yes	Yes	N/A	No	Yes	Yes
Main results	adoption of piece rates improve productivity by 44 %	piece rates increase productivity at least 23 %	profit-sharing improves on-time departures	adoption of teams leads to 14 % productivity increase	piece rates increase productivity at least 22 %	Some elements of the HRM environment improve performance	piece rates increase productivity by 59 %	shift from piece to time rates decrease productivity and wages but increase profits	off-line teams increase productivity by 3 %; effects dissipate over time

Table 2: Summary of Finnish Econometric Case Studies

Studies	CM	CM II	3	Anttila I	Anttila II	OP	OPII	HK
Field	Retail	Retail	Retail	Retail	Retail	Banking	Banking	Food production
Research question	Productivity	Productivity, inventory turnover	Productivity, timing of incentives	Productivity	Productivity dispersion	Productivity	Co-op membership rates	Productivity
Key explanatory variables	HRM	ERP introduction	Sales incentive scheme,	HRM	HRM, employee skills	Training	Transaction intensity, bank size, collective reserves, competition	Teams, performance related pay
Source of data	Internal records, HRM survey, interviews	Internal records, interviews, specific survey	Internal records	Internal records, HRM survey, interviews	Internal records	Internal records, interviews	Internal records	Internal records, interviews
Unit of observation	Store	Store	Store	Store	Store	Bank	Bank	Production line
Frequency of observation	Annual	Monthly	Monthly	Monthly	Monthly	Annual	Annual	Weekly
Main results	Elements of HRM environment increase productivity	ERP introduction is associated with initial drop of performance and then increase; impact depends on how ERP is used	Sales incentive scheme is association with increase in productivity and it influences timing of effort	Elements of HRM environment increase productivity	Productivity dispersion decreases over time but persists	General training is associated with higher incomes but no statistically significant effect on organizational performance	Membership is motivated both by individual-level pecuniary motives and other-regarding motives	Adoption of teams insufficient for performance impact; after PRP, productivity increases in 3 out of 4 lines