IABForschungsbericht

Ergebnisse aus der Projektarbeit des Instituts für Arbeitsmarkt- und Berufsforschung

Nr. 3/2007

The effects of collective bargaining on firm performance:

New evidence based on stochastic production frontiers and multiply imputed German establishment data

Uwe Jensen & Susanne Rässler

Bundesagentur für Arbeit

The effects of collective bargaining on firm performance:

New evidence based on stochastic production frontiers and multiply imputed German establishment data

Uwe Jensen, Susanne Rässler (IAB)

Mit der Publikation von Forschungsberichten will das IAB der Fachöffentlichkeit Einblick in seine laufenden Arbeiten geben. Die Berichte sollen aber auch den Forscherinnen und Forschern einen unkomplizierten und raschen Zugang zum Markt verschaffen. Vor allem längere Zwischen- aber auch Endberichte aus der empirischen Projektarbeit bilden die Basis der Reihe, die den bisherigen „IAB-Werkstattbericht“ ablöst.

The effects of collective bargaining on firm performance: new evidence based on stochastic production frontiers and multiply imputed German establishment data

© UWE JENSEN & SUSANNE RÄSSLER*

Uwe Jensen (corresponding author), Institut für Statistik und Ökonometrie,
Christian-Albrechts-Universität, Olshausenstr. 40, 24118 Kiel, Germany,
Tel.: +49-431-880-1627, e-mail: Jensen@stat-econ.uni-kiel.de
Susanne Rässler, Institut für Arbeitsmarkt- und Berufsforschung (IAB),
Weddigenstr. 20-22, 90478 Nürnberg, Germany,
Tel.: +49-911-179-3084, email: Susanne.Raessler@iab.de

*Acknowledgment: The authors like to thank Donald B. Rubin for many helpful comments and discussions about proper multiple imputation procedures. We are also grateful to Hermann Gartner for his help in preparing and multiply imputing the data and to Claus Schnabel for many clarifying comments. All remaining errors are our own.

Abstract

This paper makes three contributions to the literature on the effects of collective bargaining on the performance of German establishments. We include the analysis of firms' efficiency and we model productivity and efficiency simultaneously. Confronted with 25 % observations with missing values, we check the missing data mechanisms and find effects of firm size and collective bargaining on it, among others. After proper multiple imputation of the missing values – thus avoiding obvious nonresponse bias –, the results on the collective bargaining effects on productivity and efficiency change significantly. Finally, we suggest to multiply impute implausible zero values in the capital proxy as well.

Keywords: Collective bargaining, efficiency and productivity, establishment data, missing data, multiple imputation

JEL codes: C15, C24, C81, D24, J50

Contents

1 Introduction

The effects of collective bargaining on wages and firm performance have received a great deal of interest. It is generally accepted among economists that collective bargaining has a positive influence on the wages that are negotiated. On the other hand, the impact on firm performance, e.g., productivity, is not yet resolved. In the highly related literature on collective bargaining, unions or works councils (see Addison et al. (2004) for a recent survey), some authors stress the positive influence of unions on productivity due to workers' higher motivation and satisfaction leading to higher effort, lower turnover costs and more investment in firm-specific human capital. Other authors emphasize the reduced flexibility and power of managers leading to lower productivity. The total effect of collective bargaining is an open empirical question (see e.g. Cahuc and Zylberberg, 2004, p. 424).

Many papers have been written on the effects of collective bargaining on productivity whereas its effect on firms' efficiency is mostly ignored. In our opinion it is highly important to do this with a proper (simultaneous) stochastic frontier model for firms' productivity and efficiency. Until the beginning of this decade, the so-called '2-step approach' has been employed quite often in the frontier literature: the inefficiency estimates from the first step (estimation of productivity with the frontier) were used to find some inefficiency determinants in a second step. But Wang and Schmidt (2002) have shown that this procedure can lead to severely biased results. It is preferable to model productivity and efficiency simultaneously in one step, following, e.g., Reifschneider and Stevenson (1991).

In particular, we are not aware of any study with German data suitably analyzing this aspect of firm performance. International differences in judicial systems, culture, etc. impede the application of empirical results on the effects of collective bargaining from one country to the other. Addison et al. (2003), analyzing the effects of German works councils, are an exception. But they apply the 2-step approach criticized above. Thus, the first contribution of this paper is to analyze the effect of collective bargaining on productivity and efficiency with the suitable tool, i.e., stochastic production frontiers in the 1-step approach, and German establishment data.

When analyzing the data set, we were confronted with missing values, a typical situation in empirical research. A closer look to the data revealed 4 % to 15 % of missing values particularly in the most important variables: output, capital and labor. The typical reaction – e.g., in Schank (2005), Addison et al. (2003) or Jensen (2001, p. 158ff) – to this problem is to include only those observations with no missing values on all relevant variables, i.e., to ignore all the records that have missing values. But ignoring them usually would reduce the complete data records available considerably. Whereas information from 18447 observations from the panel waves of 2002 and 2003 is available in principle, only 13969 observations of them can be used when inference is based only on the complete cases. Furthermore, ignoring missing values is based on very strong assumptions about the missing data mechanism. The question arises whether this item-nonresponse occurs randomly, i.e., whether the remaining data are still representative for the population of interest. If not, the resulting test statistics are no longer valid and the resulting estimates, e.g., on the influence of collective bargaining on productivity and efficiency, will, at least, be less efficient but probably also be biased. Although 13969 observations seem to be ‘a lot’ they are ‘not enough’ if some observations are systematically missing – just to refute a standard argument for ignoring missing data.

Biases can be expected to occur particularly in the establishment’s inefficiency estimates of the stochastic production frontier. Because frontier estimates depend on the extreme efficient establishments in the sample and because the inefficiency estimates are derived from the estimation residuals, the latter are extremely sensitive to any kind of misspecification in the model – see, e.g., Jensen (2005). Therefore, the second contribution of this paper is to demonstrate the dangers of ignoring missing data or the gains of properly imputing them when analyzing effects (of collective bargaining) on productivity and efficiency. Kölling and Rässler (2004) also applied stochastic frontiers to multiply imputed data and found interesting differences but they used the model by Battese and Coelli (1995) and they did not focus on the effects of collective bargaining. The latter also goes for the purely explorative paper by Jensen and Rässler (2006). We finally believe that most studies with German data found no productivity effects of German unions – see the survey in Schnabel (1991) – because they ignored systematically missing data.

The article is structured as follows. In the next section, the stochastic production frontier model is introduced. The following section presents the data. Section 4 analyzes the response behavior and the missing data mechanisms and gives a short introduction to multiple imputation. In the fifth section, the estimation results for the different approaches are given and compared. Finally, section 6 summarizes the paper.

2 Stochastic production frontiers

This section summarizes the theory on stochastic production frontiers necessary in the following.

In microeconomic theory, economic production functions provide maximum possible output for given inputs of, say, n firms in the sample. In reality, inefficient input use may lead to lower outputs for many firms. Therefore, frontier functions (lying on top of the data cloud) have been developed for estimating potential output and inefficiency.

After the seminal work of Aigner and Chu (1968), Aigner et al. (1977) and Meeusen and van den Broeck (1977) introduced the stochastic production frontier

$$Y_i = \exp(\beta_0) \cdot \prod_{j=1}^{k} X_{ij}^{\beta_j} \cdot \exp(v_i) \cdot TE_i, \quad i = 1, \ldots, n, \tag{1}$$

or in logs

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + e_i, \quad e_i = v_i - u_i, \quad u_i \geq 0. \tag{2}$$

The functional form in (1) is Cobb-Douglas. In order to avoid the well-known hard restrictions of this function, we have chosen the rather general translog production function in the estimation.

In (2), y_i is the output (in logs), x_{ij} are k inputs (all in logs) of firm no. i, and β_j are unknown parameters. Then, with $TE_i = 1$ or $u_i = 0$,

$$Y_i^* = \exp(\beta_0) \cdot \prod_{j=1}^{k} X_{ij}^{\beta_j} \cdot \exp(v_i) \qquad \text{or} \qquad y_i^* = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + v_i \tag{3}$$

is maximum possible output for given inputs. The output ratio

$$0 \leq TE_i = \exp(-u_i) = \frac{Y_i}{Y_i^*} \leq 1 \tag{4}$$

is interpreted as technical inefficiency of firm no. i. Finally, the composed error term e_i consists of the one-sided inefficiency term u_i and the symmetric part v_i representing statistical noise. x_{ij}, v_i and u_i are assumed to be independent with the distributional assumptions

$$v_i \sim N(0, \sigma_v^2) \quad \text{and} \quad u_i \sim N_+(\mu, \sigma_u^2) \tag{5}$$

where $N_+(\cdot, \cdot)$ stands for a normal distribution truncated at $u = 0$ (see Stevenson, 1980).

The choice of the truncated normal distribution in (5) contains the half-normal distribution as (testable) special case. Other tractable alternatives for the inefficiency distribution are the exponential and the gamma distribution. Ritter and Simar (1997) have shown the bad performance of the gamma distribution. Jensen (2005) has presented some risks of working with exponentially distributed inefficiency. Therefore, we prefer the truncated normal distribution.

The log-likelihood function is $l(\beta, \sigma, \lambda, \mu) =$

$$-n\left[\ln(\sigma) + const + \ln\left(\Phi\left(\frac{-\mu}{\sigma\lambda}\right)\right)\right] - \sum_{i=1}^{n}\left[\frac{1}{2}\left(\frac{e_i}{\sigma}\right)^2 - \ln\left(\Phi\left(\frac{-\mu}{\sigma\lambda} - \frac{-e_i\lambda}{\sigma}\right)\right)\right] \tag{6}$$

with

$$\lambda = \frac{\sigma_u}{\sigma_v} \quad \text{and} \quad \sigma^2 = \sigma_v^2 + \sigma_u^2 \tag{7}$$

and the standard normal distribution function $\Phi(\cdot)$. Iterative maximization leads to consistent and asymptotically efficient maximum likelihood (ML) estimators $\hat{\beta}_j$, $\hat{\sigma}$, $\hat{\lambda}$ and $\hat{\mu}$.

How can the inefficiency terms be estimated? Since, in a stochastic frontier model, the estimation residuals only estimate the composed error e and not u, the inefficiencies must be estimated indirectly with the help of the minimum mean-squared error predictor

$$E[u_i|e_i] = \frac{\sigma\lambda}{1+\lambda^2}\left(\frac{\phi\left(\frac{e_i\lambda}{\sigma}\right)}{\Phi\left(-\frac{e_i\lambda}{\sigma}\right)} - \frac{e_i\lambda}{\sigma}\right) \tag{8}$$

with the standard normal density function $\phi(\cdot)$.

This basic approach might be too restrictive. Independence of x_{ij} and u_i may be a hard assumption as heteroscedasticity might occur. Despite the iid assumption (5), u might still contain some structure. Until the beginning of this decade, the so-called '2-step approach' has been employed quite often: the inefficiency estimates from (8) – the first step – were used to find some inefficiency determinants in a second step. But Wang and Schmidt (2002) have shown that this procedure – claiming that u_i are iid in the first step and finding structure in them in the second step – can lead to severely biased results.

Therefore, it is advisable to use a '1-step approach', e.g., the procedure already presented by Reifschneider and Stevenson (1991). They allow the inefficiency terms u_i to depend on some explanatory variables z_{ij} (interpreted as sources of inefficiency) which may be partly identical with variables x_{ij}:

$$u_i = \delta_0 + \sum_{j=1}^{l} \delta_j z_{ij} + w_i = d_i + w_i, \quad i = 1, \ldots, n \tag{9}$$

δ_j are unknown parameters. The distributional assumptions are

$$v_i \sim N(0, \sigma_v^2), \quad u_i \sim N_+(d_i, \sigma_u^2) \quad \text{and} \quad w_i \sim \text{trunc}_{-d_i} N(0, \sigma_w^2) \tag{10}$$

where $\text{trunc}_{-d_i} N(\cdot,\cdot)$ stands for a normal distribution truncated at $w = -d_i$. The ML estimators $\hat{\beta}_j$, $\hat{\delta}_j$, $\hat{\sigma}$ and $\hat{\lambda}$ can now be derived simultaneously using iterative ML techniques. See the given references for the likelihood function of the full model. Coelli et al. (1998), Greene (1997) or Jensen (2001) present more details on frontiers.

3 Data

3.1 General description

Our data are taken from two waves (2002 and 2003) of the Establishment Panel of the Institute for Employment Research of the Federal Labor Service (Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit, IAB). The basis for the panel is the employment statistics register of the Federal Labor Service covering all dependent employment in the private and public sector and accounting for almost 85% of total employment in Germany. The survey unit of the register is the establishment or local production unit, rather than the legal and commercial entity of the company.

The IAB Establishment Panel draws a stratified random sample of units from the register, the selection probabilities depend on the number of employees in the respective stratum. The strata comprise some 20 industries and 10 establishment size intervals covering all sectors and employment levels. The overall and size-specific response rates including firms that are interviewed for the first time exceed 60 percent, and, for repeatedly interviewed establishments, more than 80 percent. As usual, we do not have exact information about the reasons for unit-nonresponse in the data. It is commonly assumed that next to the general attitude to take part in a survey there are two main reasons for nonresponse. First, there are questions that are too difficult to understand or the information wanted is not easily available and, second, there are questions that concern sensitive information. In both cases, the interviewee is not willing to participate in the panel. A study for earlier waves of the panel comes to the result that only a few items influence the willingness of firms to participate significantly (see Hartmann and Kohaut, 2000).

The first wave of the establishment panel in 1993 contained data on 4,265 establishments. Since 1993 the panel has been augmented regularly to reflect establishment mortality, other exits, and newly-founded units. In 1996 a panel was initiated for eastern Germany with an initial sample of 4,313 establishments. Currently, the overall number of establishments in the sample approximates 15,000 with the addition of eastern Germany and other regional samples. The panel is designed to meet the needs of the Federal Employment Agency, so that its focus is again on employment-related matters - although its scope is wider than the parent register. Much of the information in the panel concerns worker characteristics and qualifications as well as levels of and changes in establishment employment. There is also information on the training and further training of employees, working time, and overtime. Additionally, information on certain establishment policies, business developments, and investment is similarly collected on an annual basis. Other information is collected biennially or triennially. Up to our knowledge this Establishment Panel is the best and most comprehensive panel survey of firms, companies, and organizations in Germany; for more details see http://betriebspanel.iab.de/.

3.2 Definition of central variables

This subsection documents the measurement of the central variables, i.e., output, capital, labor, and collective bargaining.

Output is measured as value added (see the appendix on variable construction for exact definitions). We excluded all establishments from the sample that do not use turnover as output measure. This affects non-profit organisations, public offices, banks and insurances. In the imputed data-sets, 3 distinct outliers in the output variable had to be eliminated because – particularly with a frontier function – they would significantly bias the estimates.

A reasonable measure for labor input should take into account skill and productivity differences between employees, among others. For labor, the data set provides two possible approximations: full-time equivalents (total number of employees minus 0.5 times total number of part-time employees) or earnings. The first choice would implicitly assume, e.g., that all employees are equally skilled and productive whereas the second choice implicitly assumes that earnings are a good proxy for skills and productivity, among others. We decided for the latter because that assumption seems to be more reasonable as well as seems to lead to better results, particularly in the inefficiency sub-model.

The capital variable is notorious for the difficulties any approximation to the latent value of the capital stock causes in the estimation. With time series data, the capital variable approximated by the perpetual inventory method often shows low variation and non-stationarity. In this paper, with cross-sectional data covering two years, we decided to proxy capital by the replacement investment in the current year. Of course, this choice implicitly assumes that capital is replaced uniformly and sufficiently, among others. An alternative would be to approximate capital by the average replacement investment of, say, 4 years – like e.g. Schank (2005, p. 701) – or 2 years like Addison et al. (2003, p. 14). But, due to the well-known problem of panel attrition, this approximation would lead to the exclusion of many firms: Of those establishments participating in the panel in 2003, only 75 % participated in 2002, only 61 % in 2001 and only 50 % in 2000. So, proceeding like Schank would eliminate 50 % of the observations, proceeding like Addison et al. still 25 %! Moreover, according to the information presented in the following subsection, it is obvious that these observations are not eliminated randomly and even more nonresponse bias may be introduced then. Therefore, we decided to proxy capital by the replacement investment in the current year, despite some possible discontinuities.

In the following subsection, we will show that replacement investment is one of the variables suffering from many missing values. This problem will be soothed by multiple imputation. But another problem is that a large part (7888 of 18447) of the values on investment in the sample are reported as being zero. Addison et al. (2003, p. 14) notice this problem, too. Their reaction is to exclude all establishments reporting zero replacement investment in both years t and $t-1$ from the sample in year t. This technique reduces their sample again by 17 %!

Of course, it can occur that small establishments do not transact any replacement investments during a whole year. But it does not make any sense to assume that the latent value of the capital stock is zero. And eliminating those observations would once again systematically change the sample and possibly lead to another nonresponse bias (see the following subsection). Additionally, there is some unproven suspicion that many of these firms are simply not able or not willing to provide exact non-zero investment numbers. Therefore, a third contribution of our paper is the suggestion to multiply impute these zeroes as well. We are aware of the strong assumptions our suggestion is based on. But the alternatives, i.e.

- taking the average replacement investment of n years (see above)
- assuming the latent value of the capital stock to be zero (see above)
- eliminating the observations with capital zeroes (see the following subsection)

are not more attractive. That is why, in section 5, we will compare the estimation results of three alternative procedures. Notice finally that the imputations are all done in one step. We do not perform a two-step imputation and, therefore, we can still use the usual pooling formulae to get the multiple imputation estimates.

The dummy variable representing the existence of collective agreements for the respective establishment is 1 if an industry pay agreement is effective for the establishment or if the establishment aligns its wages to an industry pay agreement.

After these fundamental decisions, the covariates of labor and capital in the production function and the inefficiency determinants in sub-model (9) had to be selected from the variables available in the IAB Establishment Panel and suggested by diverse economic theories. It is well-known that forward and backward variable selection procedures can lead to very different results when the regressors are correlated. Therefore, we conducted a very detailed data analysis including a factor analysis to examine the correlation structure of the regressors. Then, in a large-scale model selection procedure combining several forward and backward runs (using both the imputed data and only the observed data), the final sets of variables for the production function and the sub-model were fixed. Every variable had several opportunities to enter the production function and the sub-model. A variable is included in all regressions if it was significant in at least one of the 11 regressions (5 + 5 auxiliary regressions with imputed data and one with only the observed data). The appendix on variable construction shows the exact definitions of all variables and the tables show the use of the variables.

4 Item-nonresponse and imputation

4.1 Item-nonresponse

After the application of the restrictions described in the previous section, 9462 establishments remain in the sample. Since not all establishments participated in both years, there are 18447 data records for 2002 and 2003. Multiple imputation, described in the remainder of this section, is able to preserve this sample size. On the other hand, only 13969 observations can be used when inference is based only on the complete cases without any item-nonresponse on the variables used in this study.

Item-nonresponse in the data is mainly found in few variables, particularly those used to construct output, labor and capital. Table 1 gives the variables in the questionnaire with the highest item-nonresponse rates. All the other variables used in our study are distinctly below the rates shown there.

The application of the multiple imputation procedure requires some assumptions on the nonresponse mechanism. Therefore we analyzed the probability for item-nonresponse with a Probit model (see, e.g., Greene, 2003) first, according to

$$P(y_i = 1|x_i) = \Phi(x_i\beta) \tag{11}$$

with the normal distribution function $\Phi(\cdot)$. We did this in two ways: The endogenous binary variable $y =$ MISS is one if item-nonresponse occurs. The endogenous binary variable $y =$ MISS0 is one if item-nonresponse occurs or if the capital variable is zero (see the previous section). We used the same exogenous variables x_j as in the frontier estimation described later (see the appendix on variable construction for exact definitions) except output, capital and labor – of course – because they are affected by item-nonresponse. We added the log of the number of employees as measure of firm-size. All variables with $|t|$ values less than 1.0 were eliminated.

Table 2 provides the results. For better interpretation, we do not show the parameter estimates but the marginal effects

$$\frac{\partial E(y_i)}{\partial x_{ij}} = \frac{\partial \Phi(x_i\beta)}{\partial x_i\beta}\beta_j = \phi(x_i\beta)\beta_j \tag{12}$$

with the standard normal density function $\phi(\cdot)$. For metric exogenous variables, the marginal effects are evaluated at the mean. For exogenous dummy variables d_j, they are calculated by

$$\Phi(x_i\beta|d_i = 1) - \Phi(x_i\beta|d_i = 0) \tag{13}$$

Marginal effects with absolute values less than 0.00005 are not shown.

Table 2 shows that many variables in the data-set have a significant influence on the probability for item-nonresponse. E.g., an establishment with one unit higher log of the number of employees (as proxy for firm-size) shows roughly 2 % lower probability for item-nonresponse (ceteris paribus and on average). And an establishment where an industry pay agreement is effective or which at least aligns its wages to an industry pay agreement has roughly 4 % lower probability for item-nonresponse than an establishment without any industry pay agreement effects. Most absolute marginal effects and absolute t values decrease in the Probit model for MISS0 (item-nonresponse or capital zeroes) but the main results remain valid. Among others, we also see that item-nonresponse rates are lower for establishments with relatively

- much paid overtime (a proxy for high labor utilization)
- high technical condition
- and many high-skilled employees

This means that by ignoring missing data we are systematically reducing the share of establishments in the sample with smaller size, low degree of labor utilization, low technical condition, few high-skilled employees and no industry pay agreement effects. And this means particularly that we are not able to declare (without closer analysis) that ignoring missing data will have no influence on the estimated effects of collective bargaining on the productivity and efficiency of the establishments in the sample. Therefore, one should try a proper imputation technique to fill the missing values and allow the use of all valuable information that is observed.

4.2 Missing data mechanism

First formalized by Rubin (1976), in modern statistical literature (see Little and Rubin 1987, 2002, p. 12) the missing data mechanisms are commonly distinguished according to the probability of response yielding the following three cases:

- The missing data are said to be missing completely at random (MCAR), if the nonresponse process is independent of both unobserved and observed data.
- If, conditional on the observed data, the nonresponse process is independent only of the unobserved data, then the data are missing at random (MAR). This is the case, e.g., if the probability of answering the turnover question varies according to the size of the company, and the size is observed.
- Finally, data are termed not missing at random (NMAR), if the nonresponse process depends on the values of the variables that are actually not observed. This might be the case for turnover reporting, where companies with higher turnover tend to be less likely to report their turnover.

In the context of likelihood-based inference and when the parameters describing the measurement process are functionally independent of the parameter describing the nonresponse process, MCAR and MAR are said to be ignorable; otherwise we call it non-ignorable missingness which is the hardest case to deal with analytically because the missingness mechanism has to be modeled itself.

As mentioned above, the highest amount of missing values occurs in the most important variables for production function estimation: output, capital and labor. The Probit analysis of item-nonresponse in the previous subsection shows that the assumption of

MCAR obviously is violated but, fortunately, we seem to find good predictors of the non-response behavior. Therefore, we assume that the missing values of the variables used in the productivity model are missing at random (MAR).

It is important to realize that basing inference only on the complete cases – still the standard choice in most empirical papers – would implicitly assume that the data are missing completely at random (MCAR) which obviously is not the case. To ensure the MAR-assumption and allow to estimate a sophisticated econometric model with missing data, we decided to use a multiple imputation procedure. Using a single imputation technique such as mean imputation, hot deck, or regression imputation, in general results in confidence intervals and p-values that ignore the uncertainty due to the missing data, because the imputed data were treated as if they were fixed known values. Thus, basing standard complete data inference on singly imputed data will typically lead to standard error estimates that are too small, p-values that are too significant, and confidence intervals that undercover – see, e.g., Rässler et al. (2003). To correct for these effects using singly imputed data, special variance estimation techniques have to be applied. For a recent discussion of the merits and demerits of single and multiple imputation see Groves et al. (2002).

Notice that the ignorability assumption can never be contradicted by the observed data. However, Schafer (2001) provides evidence that even the erroneous assumption of MAR might have only minor impact on estimates and standard errors using a proper multiple imputation strategy. Only when NMAR is a serious concern, it is obviously necessary to jointly model the data and the missingness, although such models are based on other untestable assumptions. Therefore, a multiple imputation procedure seems to be the best alternative at hand in our situation to account for missingness, to exploit all valuable information, and to get statistically valid subsequent results based on standard complete data inference.

4.3 Multiple imputation

Multiple imputation (MI), introduced by Rubin (1978) and discussed in detail in Rubin (1987), is a Monte Carlo technique replacing missing values by $m > 1$ simulated versions, generated according to a probability distribution or, more generally, any density function indicating how likely imputed values are given the observed data. MI therefore is an approach that retains the advantages of imputation while allowing the data analyst to make valid assessments of uncertainty. The concept of multiple imputation reflects uncertainty in the imputation of the missing values through wider confidence intervals and larger p-values than under single imputation. Typically m is small, with $m = 3$ or $m = 5$. Each of the imputed and thus completed data sets is first analyzed by standard methods. Then, the results are combined or pooled to produce estimates and confidence intervals that reflect the missing data uncertainty.

The theoretical motivation for multiple imputation is Bayesian. Let Y_{obs} denote the observed components of any uni- or multivariate variable Y, and Y_{mis} its missing components. Basically, MI requires independent random draws from the posterior predictive distribution

$$f(y_{\text{mis}}|y_{\text{obs}}) = \int f(y_{\text{mis}}, \psi|y_{\text{obs}})\, d\psi = \int f(y_{\text{mis}}|y_{\text{obs}}, \psi)\, f(\psi|y_{\text{obs}})\, d\psi \tag{14}$$

of the missing data Y_{mis} given the observed data Y_{obs} with parameter vector ψ. Since $f(y_{\text{mis}}|y_{\text{obs}})$ itself often is difficult to derive, we may alternatively perform

- random draws of the parameters according to their observed-data posterior distribution $f(\psi|y_{\text{obs}})$ as well as
- random draws of the missing data according to their conditional predictive distribution $f(y_{\text{mis}}|y_{\text{obs}}, \psi)$ given the drawn parameter values.

For many models the conditional predictive distribution $f(y_{\text{mis}}|y_{\text{obs}}, \psi)$ is rather straightforward due to the data model used. On the contrary, the corresponding observed-data posterior

$$f(\psi|y_{\text{obs}}) = L(\psi; y_{\text{obs}}) \frac{f(\psi)}{f(y_{\text{obs}})} \tag{15}$$

(with the likelihood function $L(\psi; y_{\text{obs}}) = f(y_{\text{obs}}|\psi)$) usually is difficult to derive, especially when the data have a multivariate structure and different, non-monotone missing data patterns. The observed-data posteriors often are not standard distributions from which random numbers could easily be generated. Therefore, simpler methods have been developed to enable multiple imputation on the grounds of Markov chain Monte Carlo (MCMC) techniques. They are extensively discussed by Schafer (1997). Such data augmentation procedures (Tanner and Wong, 1987), which include, for example, Gibbs Sampling, yield a stochastic sequence $\left\{y_{\text{mis}}^{(t)}, \psi^{(t)} : t = 1, 2, \ldots\right\}$ whose stationary distribution is $f(y_{\text{mis}}, \psi|y_{\text{obs}})$. More specifically, an iterative sampling scheme is created, as follows. Given a current guess $\psi^{(t)}$ of the parameter, first draw a value $y_{\text{mis}}^{(t+1)}$ of the missing data from the conditional predictive distribution $f(y_{\text{mis}}|y_{\text{obs}}, \psi^{(t)})$. Then, conditioning on $y_{\text{mis}}^{(t+1)}$, draw a new value $\psi^{(t+1)}$ of ψ from its complete-data posterior $f(\psi|y_{\text{obs}}, y_{\text{mis}}^{(t+1)})$. Assuming that t is suitably large, m independent draws from such chains can be used as multiple imputations of Y_{mis} from its posterior predictive distribution $f(y_{\text{mis}}|y_{\text{obs}})$.

Based on these m imputed data sets we calculate m complete data statistics $\hat{\theta}^{(r)}$ and their variance estimates $\hat{\mathrm{V}}(\hat{\theta}^{(r)})$, $r = 1, \ldots, m$. The complete-case estimates are combined according to Rubin's rule such that the MI point estimate $\hat{\theta}_{\text{MI}}$ for parameter θ is the average

$$\hat{\theta}_{\text{MI}} = \frac{1}{m} \sum_{r=1}^{m} \hat{\theta}^{(r)} \tag{16}$$

Its estimated total variance T is calculated according to the analysis of variance principle:

$$\begin{aligned}
\text{'between-imputation variance':}\quad B &= \frac{1}{m-1}\sum_{r=1}^{m}\left(\hat{\theta}^{(r)}-\hat{\theta}_{\mathrm{MI}}\right)^2 \\
\text{'within-imputation variance':}\quad W &= \frac{1}{m}\sum_{r=1}^{m}\hat{\mathrm{V}}\left(\hat{\theta}^{(r)}\right) \\
\text{'total variance':}\quad T &= W+\left(1+\frac{1}{m}\right)B
\end{aligned} \tag{17}$$

For large sample sizes, tests and two-sided interval estimates can be based on the Student's t-distribution

$$\frac{\hat{\theta}_{\mathrm{MI}}-\theta}{\sqrt{T}} \dot{\sim} t(v) \quad \text{with} \quad v=(m-1)\left(1+\frac{W}{(1+m^{-1})B}\right)^2 \tag{18}$$

degrees of freedom. For a comprehensive overview of MI see Schafer (1999a).

Multiple imputation is in general applicable when the complete-data estimates are asymptotically normal or t distributed; e.g., see Rubin and Schenker (1986), Rubin (1987), Barnard and Rubin (1999), or Little and Rubin (2002). Notice that the usual maximum-likelihood estimates and their asymptotic variances derived from the inverted Fisher information matrix typically satisfy these assumptions. In this paper we use ML estimation for the analyst's model.

For the creation of the multiple imputations we have used the stand alone software NORM provided for free by Schafer (1999b). A very detailed description of this data augmentation algorithm is given by Schafer (1997).

4.4 Data preparation

When using NORM, the data are assumed to follow a multivariate normal distribution. Clearly, our survey data are not normally distributed: some are bounded between zero and one, others are skewed and some have large proportions of zeros; the latter are called semi-continuous variables (see Schafer and Olsen, 1999). A way to handle non-normality of the data is by applying suitable transformations (e.g. logit, log or Box-Cox) to the variables which is done in our application. Moreover, if non-normal variables (such as discrete or binary ones) are completely observed, then it is quite plausible to still use the multivariate normal model because incomplete variables are modeled as conditional normal given a linear function of the complete variables – see, e.g., Schafer (1997). The variables and their transformations used in our models are listed in the appendix.

Theoretically, we should transform the data to achieve multivariate normality. Practically, such transformations are not yet available: the usual transformations are performed on a univariate scale. Investigations show that such deviations from normality (for the variables to be imputed) should not harm the imputation process too much – see Schafer (1997) or Gelman et al. (1998). A growing body of evidence supports the claim to use a normal model to create multiple imputations even when the observed data are somewhat non-normal. The focus of the transformations is rather to achieve a range for continuous variables to be imputed that theoretically have support on the whole real line than to achieve normality itself. Even for populations that are skewed or heavy-tailed, the actual coverage of multiple imputation interval estimates is reported to be very close to the nominal coverage. The multiple imputation framework has been shown to be quite robust against moderate departures from the data model – see Schafer (1997).

With NORM 2.03, the imputations are created very easily. After a burn-in period of 2000 iterations, every further 200 iterations the imputed data sets are stored. Finally, $m = 5$ multiply imputed data sets are used for our analysis. Investigations of time-series and autocorrelation plots did not suggest any convergence problems. Notice that in the imputer's and analyst's model the same set of input data, i.e., variables and observations, is used to avoid problems of misspecification – see Meng (1995) or Schafer (2001).

5 Results

5.1 Approaches

The stochastic production frontier (2) with inefficiency sub-model (9) has been estimated with the IAB German establishment data described in section 3. The production function follows the translog form in capital and labor and includes further variables given in the appendix where the variables of the inefficiency sub-model are given as well. As described in the previous section, 11 regressions have been run for 3 approaches:

- Approach NONR: One regression with only the observed data.
- Approach MIC0: $m = 5$ auxiliary regressions with the full data set where all missing values have been filled by multiple imputation (see section 4) but where the zeroes in the capital variable are maintained.
- Approach MIMI: $m = 5$ auxiliary regressions with the full data set where all missing values and the zeroes in the capital variable have been filled by multiple imputation.

Estimation has been performed with LIMDEP 8.0. Tables 3 and 3a provide the results. In the following, 'significance' means 'significance on the 5 % level', 'weak significance' means 'significance on the 10 % level'.

5.2 The effects of collective bargaining

The positive influence of collective bargaining on wages and its negative impact on firm profits are generally accepted among economists. But the effects on productivity and efficiency are unclear. In the highly related literature on collective bargaining, unions (see Freeman and Medoff, 1984) and works councils, some authors stress the positive influence of unions on productivity due to workers' higher motivation and satisfaction leading to higher effort, lower turnover costs and more investment in firm-specific human capital. Other authors emphasize the reduced flexibility and power of managers leading to lower productivity. The total effect of collective bargaining on firms' productivity is an open empirical question (see e.g. Cahuc and Zylberberg, 2004, p. 424). Studies with German data mostly seem to have not found any effects (see the survey in Schnabel, 1991).

Indeed, excluding missing observations, the parameter measuring the effect of collective bargaining on productivity in table 3 is insignificantly positive. But these results rely on the obviously violated assumption of the data being MCAR. On the other hand, assuming the less restrictive MAR assumption and with multiply imputed data, the effect is significantly negative. With the untransformed frontier (1), the interpretation is straightforward. E.g., in the approach with imputed missing values and imputed capital zeroes, the potential output of an establishment where an industry pay agreement is effective or which at least aligns its wages to an industry pay agreement is only

$$\exp(-0.0473 \cdot 1) \cdot 100 = 95.8\% \tag{19}$$

of the potential output of an establishment without collective bargaining, ceteris paribus and on average.

This parameter estimate should be interpreted very cautiously, of course. There might be no causal relation between collective bargaining and potential output because unionized workers can have unobserved characteristics different from those of non-unionized ones leading to selection bias (Cahuc and Zylberberg, 2004, p. 421 and 424, and Gürtzgen, 2006). This heterogeneity problem is certainly aggravated by putting 'all' German establishments into one sample assuming that they have the same technology. Nevertheless, properly imputing missing data significantly changes the collective bargaining parameter.

The effect on firms' efficiency is still largely unexplored. In their theoretical analysis, Addison et al. (2003, p. 3) describe possible positive effects due to lower labor turnover, lower training costs, and more investment in firm-specific human capital. Particularly, we do not know of any attempt to analyze this effect with stochastic production frontiers with the 1-step approach and German establishment data. And international differences in judicial systems, culture, etc. impede the application of empirical results on the effects of collective bargaining from one country to the other.

Addison et al. (2003, p. 3) and Schank et al. (2004) – analyzing the effects of works councils – are exceptions, but they used a fixed-effects panel frontier model, not the Reifschneider/Stevenson model used in this paper. Hence, apart from providing fixed individual effects, they did not include inefficiency determinants in their estimation models. In their approach, they did not find any significant impact of works councils on firms' efficiency. This non-result might be caused by their method: They tried to find significant differences by analyzing the confidence intervals of median plants of subgroups. But Horrace and Schmidt (1996) and Jensen (2000) have shown that efficiency differences across individuals mostly are insignificant when the analysis is based on individual confidence intervals. Jensen (2000) also provided an alternative by a simple Monte-Carlo procedure, but, of course, this is still a 2-step approach already been criticized.

Table 3a provides the effects of collective bargaining on firms' inefficiency. Excluding missing observations, the parameter is (distinctly) insignificantly positive whereas, with multiply imputed data, it is significantly (weakly in one case) negative. This means that – see (4) – an establishment where an industry pay agreement is effective or which at least aligns its wages to an industry pay agreement is more efficient than an establishment without collective bargaining, ceteris paribus and on average. With (9), one could interpret the parameter estimates. We do not dare to do this because table 3a shows that all parameter estimates increase distinctly when based on the multiply imputed data (compared to the data set with excluded missing observations).

Inefficiency estimates of stochastic production frontiers are well-known to be very sensitive. Frontier estimates depend on the extreme efficient establishments in the sample, and the inefficiency estimates are derived from the estimation residuals of these frontiers – see e.g. Jensen (2005). This problem is even increased in the Reifschneider/Stevenson model (9) when many variables appear simultaneously in the productivity model and in the inefficiency sub-model, a very demanding approach (regarding the degrees of freedom) used in this paper.

Therefore, in table 4, we checked the sensitivity of the inefficiency estimates in the 5 auxiliary regressions with imputed missing values and imputed capital zeroes (u01 - u05), in the 5 auxiliary regressions with imputed missing values (u1 - u5) and in the regression with only non-missing values (um). We see that the correlations of these most sensitive parts of any frontier estimations are very high (see Jensen, 2005) and the performance, e.g. in the subgroups, is quite reasonable. Furthermore, the large parameter estimates are quite stable across the $m = 5$ auxiliary regressions. This means that, although the size of the parameter estimates in the inefficiency sub-model is not trustworthy, the inefficiency estimates based on them are well-behaved. Hence, we believe that properly imputing missing data significantly changes the collective bargaining parameter in the inefficiency sub-model, too.

What is the reason for the significant effect of multiple imputation on the collective bargaining parameter in both parts of the model? Recently, many papers have been published on the productivity effects of German works councils, a highly related problem. Addison et al. (2004) survey the literature and report, among others, that some recent studies like Addison et al. (2003) find considerable sensitivity of the collective bargaining parameter estimates to sector, region, and especially to establishment size. Unfortunately, due to the very demanding approach (see above) used in this paper, we were not able to conduct a sensible subgroup analysis in this paper. But the results of subsection 4.1 are helpful. There, it has been shown that by ignoring missing data we are systematically reducing the share of establishments in the sample with smaller size, low degree of labor utilization, low technical condition, few high-skilled employees, and no industry pay agreement effects. When there are establishment size effects in the impact of collective bargaining on firm performance, systematically reducing the share of establishments in the sample with smaller size by ignoring missing data has a significant influence on the estimated effects of collective bargaining on the productivity and efficiency of the establishments in the sample.

As a consequence, it seems to be very dangerous to proceed like most empirical economists did and do when confronted with missing data (i.e. ignore all the records with missing values).

5.3 Further controversial results

We begin by comparing the remaining results on the production frontier in table 3. Here, all 3 approaches perform rather similar – with one important exception. In the MIC0 approach, one labor parameter is insignificant, even with changing signs in the auxiliary regressions. This certainly is a severe drawback of this approach.

Apart from that, it strikes that higher export activity leads to higher productivity only when missing observations remain missing whereas, after multiple imputation, the export parameter becomes insignificantly or weakly significantly negative. See the next paragraphs for the relation between export activity and efficiency.

More striking differences between the approaches are found in the results on the inefficiency sub-model in table 3a. This is in line with our expectations in the first section because this part of the model is very sensitive to any kind of misspecification, e.g., due to wrong assumptions about the missing data mechanism. With multiply imputed data,

- labor has a weakly significantly positive effect on u, i.e., a weakly significantly negative effect on efficiency – see (4) – whereas, excluding missing observations, higher wage costs significantly increase efficiency. The negative effect of wages on efficiency seems more sensible and can be explained by noticing that labor is proxied by total gross wages and that these total gross wages directly stand for labor costs, too. So, standard arguments from labor economics, namely shirking theory (Lazear, 1981), can help to explain the result: Larger firms with many employees have problems with monitoring the work effort of their employees. A well-known solution are higher relative wages and the threat of being discharged, a powerful disciplinary threat. This can increase productivity but, of course, this might be inefficient, i.e., too costly.

- higher exports significantly coincide with higher efficiency whereas the relation is weakly significantly negative with excluded missing observations.
- firms receiving relatively more wage subsidies are significantly less efficient. Employees receiving wage subsidies might not work efficiently. This effect is only weakly significant with excluded missing observations.
- firms supporting relatively more on-the-job-training cases are less efficient. This can make sense because the returns to the firm costs of on-the-job-training might not be sufficient. This effect is insignificant with excluded missing observations, where firms supporting the use of PCs for on-the-job-training cases are significantly less efficient.
- mean technical efficiency – see (4) – is distinctly higher (55 %) than with excluded missing observations (48 %).

Since we are working with real data and not with simulated data, we don't know anything about the true parameter values. Hence, we are not able to say which results come closer to the truth. Nevertheless, apart from the results on the effects of collective bargaining, we see particularly in the inefficiency sub-model that working with multiply imputed data reveals some interesting and plausible results which are not available when ignoring missing observations. Moreover, with multiply imputed data we account for the more realistic and less restrictive assumption of the data being missing at random. Summarizing the performance of the two multiple imputation approaches, the MIC0 approach suffers from the serious drawback of counterintuitively producing an insignificant labor parameter in the production function. To conclude, we have a small but distinct preference for the results obtained with multiple imputation where the capital zeroes are imputed as well.

5.4 The unanimous results

In this subsection, a larger part of the unanimous and significant results are interpreted. We start with the results on the production function.

- Except the labor parameter in the MIC0 approach (see the previous subsection), the capital and labor parameters show the expected signs. Particularly, the t values for the capital and labor parameter estimates show that the Translog function is a more sensible choice than the restrictive Cobb-Douglas function.

- OUTPROGP/OUTPROGN: If turnover is expected to increase (decrease), it seems to be rather low (high). Thus, an expected increase (decrease) goes in line with lower (higher) productivity.

- DEVELOP: If the technical condition of a firm is up to date, the productivity is higher.

- NEWWORK: Firms with relatively many new hires (having little firm-specific human capital) are less productive.

- SKSEARCH: Firms searching relatively many skilled employees as of now are more productive.

- FLUCT: Stronger production fluctuations lead to lower productivity.

- EAST: Enterprises which are by majority in East German property are less productive, a well-known result.

- TRAIND/TRAINPC: Firms supporting on-the-job-training (with or without PCs) are more productive.

- PROP1: Firms offering many jobs for whom experience is important are less productive.

Finally, two stable significant results on the inefficiency sub-model are:

- SKILL: Firms with relatively many skilled employees are producing more efficiently.

- PROP4: Firms offering many jobs for whom creativity is important might be exposed to relatively many production risks leading to lower efficiency.

6 Conclusions

This paper has made three contributions to the literature. First, we have analyzed the effects of collective bargaining on the productivity and efficiency of German establishments with the suitable tool, i.e., stochastic production frontiers in the 1-step approach (modeling productivity and efficiency simultaneously). The effect of collective bargaining on efficiency is mostly ignored. And when efficiency analysis is included (in some papers on the effects of German works councils), the so-called '2-step approach' has been employed where the inefficiency estimates from the first step (estimation of productivity with the frontier) were used to find some inefficiency determinants in a second step.

The second contribution has been to demonstrate the dangers of ignoring missing data or the gains of properly imputing them when analyzing effects (of collective bargaining) on productivity and efficiency. It has been shown that the standard choice in most empirical economic papers (ignoring missing values) is based on very strong assumptions about the missing data mechanism. These assumptions are obviously not fulfilled in our case, as will be the case for many other survey data sets, too. By ignoring missing data we would have systematically reduced the share of establishments in the sample with smaller size and no industry pay agreement effects, among others. The results on the effects of collective bargaining on productivity and efficiency changed after multiple imputation of the missing values, i.e., after avoiding nonresponse bias, in a plausible way. Whereas the collective bargaining effect on productivity and efficiency was insignificant with ignored missing data, it turned out to be significantly negative for productivity and significantly (weakly in one case) positive for efficiency. Moreover, several additional parameter estimates, particularly in the very sensitive inefficiency sub-model, changed significantly and plausibly after multiple imputation.

A third contribution has been the discussion of another important problem in the estimation of production functions that deserves more attention than it receives: zeroes in the capital variable proxied by replacement investment. Our suggestion is to multiply impute these zeroes as well. Of course, our suggestion is based on strong assumptions but the alternatives are not better:

- The standard choice in most empirical papers (ignoring the observations with capital zeroes) is based on very strong assumptions about the missing data mechanism. These assumptions seem not to be fulfilled, thus the standard choice will lead to even more biased results.
- Taking the average replacement investment of n years also leads to an additional non-random elimination of observations and, therefore, to an additional nonresponse bias.
- Using the capital variable with its reported zeroes, i.e. assuming the latent value of the capital stock to be zero is implausible and turned out to be problematic due to an insignificant estimate of the labor parameter.

We conclude that the literature on the effects of collective bargaining – and works councils, too – on firm performance would gain distinctly from adopting recent developments in the frontier literature as well as from properly checking their implicit assumptions about the missing data mechanisms, thus, working with (multiply) imputed data.

References

Addison, J.T., T. Schank, C. Schnabel and J. Wagner. (2003). "German works councils in the production process." *IZA Discussion Paper* 812.

Addison, J.T., C. Schnabel and J. Wagner. (2004). "The course of research into the economic consequences of German works councils." *British Journal of Industrial Relations* 42/2, 255 - 281.

Aigner, D.J. and S.-F. Chu. (1968). "On estimating the industry production function." *American Economic Review* 58, 826 - 839.

Aigner, D.J., C.A.K. Lovell and P. Schmidt. (1977). "Formulation and estimation of stochastic frontier production function models." *Journal of Econometrics* 6, 21 - 37.

Barnard, J. and D.B. Rubin. (1999). "Small-sample degrees of freedom with multiple imputation." *Biometrika* 86, 948 - 955.

Battese, G.E. and T.J. Coelli. (1995). "A model for technical inefficiency effects in a stochastic frontier production function for panel data." *Empirical Economics* 20, 325 - 332.

Cahuc, P. and A. Zylberberg. (2004). *Labor Economics*. Cambridge: MIT Press.

Coelli, T., D.S.P. Rao and G.E. Battese. (1998). *An introduction to efficiency and productivity analysis*. Boston: Kluwer.

Freeman, R.B. and J.L. Medoff. (1984). *What do unions do?* New York: Basic Books.

Gelman, A., G. King and C. Liu. (1998). "Not asked and not answered – multiple imputation for multiple surveys" (with discussion). *Journal of the American Statistical Association* 93, 846 - 869.

Greene, W.H. (1997). "Frontier production functions." In Pesaran, M.H. and P. Schmidt, *Handbook of applied econometrics*. Oxford: Blackwell, 81 - 166.

Greene, W.H. (2003). *Econometric analysis*, 5th ed. Upper Saddle River: Prentice Hall.

Groves, R.M., D.A. Dillman, J.L. Eltinge and R.J.A. Little. (2002). *Survey nonresponse*. New York: Wiley.

Gürtzgen, N. (2006). "The effect of firm- and industry-level contracts on wages – evidence from longitudinal linked employer-employee data." *ZEW Discussion Paper* 06-082.

Hartmann, J. and S. Kohaut. (2000). "Analysen zu Ausfällen (unit-nonresponse) im IAB-Betriebspanel." *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 39, 609 - 618.

Horrace, W.C. and P. Schmidt. (1996). "Confidence statements for efficiency estimates from stochastic frontier models." *Journal of Productivity Analysis* 7, 257 - 282.

Jensen, U. (2000). "Is it efficient to analyze efficiency rankings?" *Empirical Economics* 25/2, 189-208.

Jensen, U. (2001). *Robuste Frontierfunktionen, methodologische Anmerkungen und Ausbildungsadäquanzmessung*. Frankfurt: Peter Lang.

Jensen, U. (2005). "Misspecification preferred: The sensitivity of inefficiency rankings." *Journal of Productivity Analysis* 23/2, 223 - 244.

Jensen, U. and S. Rässler. (2006). "Stochastic production frontiers with multiply imputed German establishment data." *Zeitschrift für ArbeitsmarktForschung* 39/2, 277-295.

Kölling, A. and S. Rässler. (2004). "Editing and multiply imputing German establishment panel data to estimate stochastic production frontier models." *Zeitschrift für Arbeitsmarktforschung* 37/3, 306-318.

Lazear, E. (1981). "Agency, earnings profiles, productivity and hours restrictions." *American Economic Review* 71, 606 - 620.

Little, R.J.A. and D.B. Rubin. (1987, 2002). *Statistical analysis with missing data.* New York: Wiley.

Meng, X.L. (1995). "Multiple imputation inferences with uncongenial source of input" (with discussion). *Statistical Science* 10, 538 - 573.

Meeusen, W. and J. van den Broeck. (1977). "Efficiency estimation from Cobb-Douglas production functions with composed error." *International Economic Review* 18, 435 - 444.

Rässler, S., D.B. Rubin and N. Schenker. (2003). "Imputation." In Bryman, A., M. Lewis-Beck and T.F. Liao (eds.), *Encyclopedia of social science research methods.* Thousand Oaks: Sage.

Reifschneider, D. and R.E. Stevenson. (1991). "Systematic departures from the frontier: a framework for the analysis of firm inefficiency." *International Economic Review* 32, 715 - 723.

Ritter, C. and L. Simar. (1997). "Pitfalls of normal-gamma stochastic frontier models." *Journal of Productivity Analysis* 8, 167 - 182.

Rubin, D.B. (1976). "Inference and missing data." *Biometrika* 63, 581 - 592.

Rubin, D.B. (1978). "Multiple imputation in sample surveys - a phenomenological Bayesian approach to nonresponse." *Proceedings of the Survey Research Methods Sections of the American Statistical Association*, 20 - 40.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Rubin, D.B. and N. Schenker. (1986). "Multiple imputation for interval estimation from simple random samples with ignorable nonresponse." *Journal of the American Statistical Association* 81, 366 - 374.

Schafer, J.L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Schafer, J.L. (1999a). "Multiple imputation – a primer." *Statistical Methods in Medical Research* 8, 3 - 15.

Schafer, J.L. (1999b). "Multiple imputation under a normal model," version 2, Software for Windows 95/98/NT. http://www.stat.psu.edu/ jls/misoftwa.html.

Schafer, J.L. (2001). "Multiple imputation in multivariate problems when the imputation and the analysis models differ." In Bethlehem, J. and S. van Buuren (eds.), *Missing*

values proceedings of a symposium on incomplete data. Utrecht, 1 - 21.
Schafer, J.L. and M.K. Olsen. (1999). "Modeling and imputation of semi-continuous survey variables." Technical report 00-39, Pennsylvania State University.
Schank, T. (2005). "Are overtime plants more efficient than standard-time plants? A stochastic production frontier analysis using the IAB establishment panel." *Empirical Economics* 30, 693 - 710.
Schank, T., C. Schnabel and J. Wagner. (2004). "Works councils - sand or grease in the operation of German firms?" *Applied Economics Letters* 11/3, 159 - 161.
Schnabel, C. (1991). "Trade unions and productivity: the German evidence." *British Journal of Industrial Relations* 19, 15 - 24.
Stevenson, R.E. (1980). "Likelihood functions for generalized stochastic frontier estimation." *Journal of Econometrics* 13, 57 - 66.
Tanner, M.A. and W.H. Wong. (1987). "The calculation of posterior distributions by data augmentation (with discussion)." *Journal of the American Statistical Association* 82, 528-550.
Wang, H.-J. and P. Schmidt. (2002). "One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels." *Journal of Productivity Analysis* 18, 129 - 144.

Appendix: Data preparation, variable construction

Variables in the questionnaire (to be transformed)

SALE	turnover in EUR
INPUT	input of materials, goods and services in % of turnover
INVEST	investment in EUR
ADDINV	investment to enlarge capital in % of investment
EMP	total number of employees
NOVERTIM	total number of employees with paid overtime in previous year
EXPORT	export in EUR
NSKILL	total number of highly skilled employees
NTEMP	total number of temporary employees
NONEWHIR	dummy: NONEWHIR = 1 if no new hires in first half-year
WOULD	dummy: WOULD = 1 if employer wanted to hire new employees
NNEWHIR	total number of new hires in first half-year
QUIT	total number of quits in first half-year
NTERMIN	total number of terminations by employees in first half-year
NSEARC	total number of employees searched as of now
NSKSEARC	total number of skilled employees searched as of now
NSUBSIDL	total number of employees supported by wage subsidies in previous year
NSHORT	total number of short-time workers in first half-year
NTRAINP	total number of employees in on-the-job-training in first half-year
NTRAINC	total number of on-the-job-training cases in first half-year

Variables in the regressions

MISS	dummy: MISS = 1 if item-nonresponse occurs
MISS0	dummy: MISS0 = 1 if item-nonresponse occurs or if C is zero
Y	output: SALE * (1 - INPUT/100)
C	capital: INVEST * (1 - ADDINV/100), C = 1 if no investment
L	labor: total gross monthly wages in June
YEAR	dummy: YEAR = 1 if observation in 2003
OVERTIM	NOVERTIM/EMP
OUTPROGP	dummy: OUTPROGP = 1 if turnover is expected to increase
OUTPROGN	dummy: OUTPROGN = 1 if turnover is expected to decrease
EXP	EXPORT/SALE
DEVELOP	ordinal: Rating of technical condition of enterprise (0 = completely out-of-date, 4 = up to date)
COLLECT	dummy: COLLECT = 1 for collective agreements
SKILL	NSKILL/EMP
TEMP	NTEMP/EMP
NOLABSUP	dummy: NOLABSUP = NONEWHIR * WOULD
NEWWORK	NNEWHIR/EMP
TERMIN	NTERMIN/QUIT
SEARCH	NSEARC/EMP
SKSEARCH	NSKSEARC/EMP
SUBSIDYL	NSUBSIDL/EMP
FLUCT	dummy: FLUCT = 1 for stronger production fluctuations in previous year
EAST	dummy: EAST = 1 if enterprise by majority in East German ownership
PUBLIC	dummy: PUBLIC = 1 if enterprise by majority in public ownership
SHORTTIM	NSHORT/EMP
TRAIND	dummy: TRAIND = 1 if employer has supported on-the-job-training in first half-year
TRAINPER	NTRAINP/EMP
TRAINCAS	NTRAINC/EMP
TRAINPC	dummy: TRAINPC = 1 if employer supports use of PCs for on-the-job-training
TYPE1	dummy: TYPE1 = 1 for independent enterprise without any establishments elsewhere
TYPE2	dummy: TYPE2 = 1 for head office of an enterprise with establishments elsewhere
TYPE3	dummy: TYPE3 = 1 for branch establishment of a larger enterprise
TYPE4	dummy: TYPE4 = 1 for intermediate authority of a larger enterprise
PROP1	dummy: PROP1 = 1 if experience is important for most jobs in the firm
PROP2	dummy: PROP2 = 1 if physical endurance is important for most jobs in the firm
PROP4	dummy: PROP4 = 1 if creativity is important for most jobs in the firm
PROP5	dummy: PROP5 = 1 if discipline is important for most jobs in the firm
PROP6	dummy: PROP6 = 1 if flexibility is important for most jobs in the firm
PROP8	dummy: PROP8 = 1 if superior workmanship is important for most jobs in the firm
PROP9	dummy: PROP9 = 1 if theoretical knowledge is important for most jobs in the firm
PROP11	dummy: PROP11 = 1 if loyalty is important for most jobs in the firm
PROP12	dummy: PROP12 = 1 if willingness to learn is important for most jobs in the firm

Data transformation for MI procedure

Y	Box-Cox
C	log, dummy*
L	Box-Cox
OVERTIM	logit
EXP	log, dummy*
DEVELOP	no transformation
SKILL	logit
NEWWORK	Box-Cox
TERMIN	logit
SKSEARCH	Box-Cox
SUBSIDYL	Box-Cox
SHORTTIM	Box-Cox
TRAINPER	Box-Cox
TRAINCAS	Box-Cox

1. Variables marked with an asterisk are treated as semi-continuous, i.e., a major part of the observations are at the minimum or the maximum of values. Therefore, we defined dummy variables that indicate whether an observation is at the respective minimum or maximum. The transformation procedure is performed only for the continuous part of the variable (see subsection 4.4).

2. All variables not mentioned in this list are dummies which remain untransformed (see subsection 4.4).

Tables

Table 1: Variables with the highest nonresponse (in %)

Variable	2002	2003
Turnover	13.69	15.05
Input of materials, goods and services	11.99	12.67
Total gross monthly wages in June	11.07	12.78
Investment to enlarge capital	8.38	6.92
Investment	4.19	4.51

Table 2: Estimates of Probit Model for item-nonresponse

Endogenous var. →	MISS		MISS0	
Exogenous var. ↓	Marg. effect	t value	Marg. effect	t value
Const.	0.0876	2.97	0.1366	3.22
$\ln(EMP)$	-0.0195	-5.44	-0.0182	-3.09
YEAR	0.0211	2.99	0.0085	2.54
OVERTIM	-0.0003	-2.15	-0.0002	-2.78
OUTPROGP	-0.0419	-7.83	-0.0127	-4.96
OUTPROGN	-0.0214	-3.82	-0.0023	-1.32
DEVELOP	-0.0011	-3.08	-0.0076	-3.10
COLLECT	-0.0390	-4.48	-0.0067	-2.27
SKILL	-0.1526	-6.41	-0.0179	-2.64
TEMP			0.0134	1.37
NEWWORK	0.0359	1.35		
TERMIN	-0.0012	-1.59	-0.0004	-1.14
SEARCH	0.0554	1.52		
SUBSIDYL	-0.0663	-2.28		
FLUCT	-0.0347	-4.00	-0.0105	-2.58
EAST	-0.0847	-6.28	0.0054	1.82
PUBLIC			-0.0176	-1.65
TRAIND			-0.0257	-2.93
TRAINPER	0.0001	3.28		
TRAINPC			-0.0107	-2.51
Property dummies	yes		yes	
Type dummies	yes		yes	
Industry dummies	yes		yes	
	18447 observations		18447 observations	

Source: own calculations, based on IAB data

Table 3: Estimates of stochastic production frontier

	Imputed missing values, imputed capital zeroes		Imputed missing values, with capital zeroes		Non-missing values	
Variable	Coefficient	t value	Coefficient	t value	Coeff.	t value
Const.	7.8103	30.52	9.3327	28.72	8.6089	73.44
$\ln(C)$	0.1541	3.19	0.0253	3.44	0.0227	3.78
$\ln(L)$	0.1144	2.70	0.0125	0.46	0.1721	7.38
$(\ln(C))^2$	0.0115	3.25	0.0070	13.37	0.0064	16.54
$(\ln(L))^2$	0.0486	17.70	0.0399	27.79	0.0317	25.64
$\ln(C) \cdot \ln(L)$	-0.0309	-4.08	-0.0088	-11.15	-0.0079	-12.63
YEAR	-0.0386	-1.34	-0.0026	-0.19	0.0136	1.21
OVERTIM	-0.0453	-1.44	-0.0382	-1.23	0.0292	1.34
OUTPROGP	-0.0508	-2.54	-0.0517	-2.69	-0.0565	-3.55
OUTPROGN	0.0678	4.52	0.0701	4.59	0.0852	6.73
EXP	-0.0750	-1.34	-0.0999	-1.79	0.0781	5.39
DEVELOP	0.0708	4.02	0.0640	4.65	0.0567	5.13
COLLECT	-0.0473	-2.02	-0.0606	-2.36	0.0126	0.64
NEWWORK	-0.4025	-6.60	-0.4282	-7.63	-0.5289	-11.64
SKSEARCH	0.2057	2.31	0.1956	2.21	0.1725	3.42
FLUCT	-0.0436	-2.11	-0.0471	-2.27	-0.0411	-2.56
TYPE2	0.1652	4.77	0.1646	5.02	0.0783	3.14
TYPE3	0.3810	12.01	0.3922	13.21	0.3203	14.44
TYPE4	0.4094	5.80	0.4197	5.91	0.3707	7.35
EAST	-0.1681	-7.00	-0.1695	-7.11	-0.1657	-8.35
TRAIND	0.0662	3.08	0.0551	2.30	0.0682	3.61
TRAINPER	-0.0081	-1.79	-0.0117	-2.60	-0.0059	-1.20
TRAINPC	0.0796	3.63	0.0770	3.54	0.0766	4.38
PROP1	-0.0587	-2.69	-0.0614	-2.89	-0.0557	-3.41
PROP2	-0.0396	-2.08	-0.0317	-1.69	-0.0703	-5.40
PROP5	0.0412	1.96	0.0441	2.09	0.0458	3.05
PROP6	0.0362	1.62	0.0373	1.74	0.0238	1.38
PROP8	-0.0651	-2.95	-0.0651	-2.91	-0.0548	-2.83
PROP9	-0.0700	-3.85	-0.0726	-3.98	-0.0529	-3.65
PROP11	0.0470	2.69	0.0381	2.10	0.0511	3.81
PROP12	0.0416	2.29	0.0385	2.12	0.0444	3.29
Industry dummies	yes		yes		yes	
	18447 observations		18447 observations		13969 observations	

Source: own calculations, based on IAB data

Table 3a: Estimates of inefficiency submodel

	Imputed missing values, imputed capital zeroes		Imputed missing values, with capital zeroes		Non-missing values	
Variable	Coefficient	t value	Coefficient	t value	Coeff.	t value
Const.	-32.816	-2.15	-29.564	-2.30	-0.1646	-0.41
$\ln(L)$	0.809	1.61	0.793	1.77	-0.0874	-2.90
EXP	-32.826	-2.74	-31.184	-2.99	0.0633	1.71
DEVELOP	1.039	1.30	0.708	1.03	0.1195	1.82
COLLECT	-3.100	-1.85	-3.407	-2.14	0.0148	0.12
SKILL	-5.615	-2.11	-4.750	-2.28	-0.3442	-2.62
NOLABSUP	4.066	1.53	3.745	1.59	0.0907	0.48
TERMIN	-4.482	-1.74	-4.599	-1.83	-0.2006	-1.64
SUBSIDYL	6.064	2.46	5.599	2.71	0.2723	1.84
FLUCT	-2.667	-1.61	-2.313	-1.64	-0.1489	-1.55
TYPE1	-7.468	-2.58	-6.533	-2.76	-0.5958	-5.15
EAST	-2.135	-1.28	-2.174	-1.40	-0.2946	-2.48
SHORTTIM	-4.841	-1.17	-5.279	-1.40	-0.0541	-0.22
TRAIND	-1.671	-1.11	-1.698	-1.26	0.1614	1.37
TRAINCAS	0.207	2.72	0.202	3.07	0.0121	1.46
TRAINPC	1.721	1.22	1.797	1.27	0.2222	2.08
PROP1	-1.766	-1.25	-1.945	-1.48	-0.1842	-1.84
PROP4	4.423	2.22	4.042	2.46	0.4545	5.38
PROP6	2.876	1.62	2.422	1.63	0.1528	1.44
PROP8	-4.468	-1.38	-4.003	-1.55	-0.2531	-2.21
Industry dummies	yes		yes		yes	
λ	6.428	2.88	6.024	3.21	2.6818	26.78

Technical inefficiency estimates

Variable	Mean		Mean		Mean	
u_i	0.5924		0.5908		0.7433	
	18447 observations		18447 observations		13969 observations	

Source: own calculations, based on IAB data

Table 4: Correlation of inefficiency estimates

	u01	u02	u03	u04	u05	u1	u2	u3	u4	u5	um
u01	1.000	0.999	0.999	0.999	0.999	0.984	0.985	0.987	0.980	0.983	0.960
u02	0.999	1.000	0.999	0.999	0.999	0.984	0.986	0.987	0.981	0.984	0.962
u03	0.999	0.999	1.000	0.998	0.999	0.984	0.985	0.988	0.980	0.983	0.959
u04	0.998	0.999	0.998	1.000	0.998	0.984	0.985	0.986	0.982	0.983	0.965
u05	0.999	0.999	0.999	0.998	1.000	0.984	0.985	0.987	0.980	0.984	0.962
u1	0.984	0.984	0.984	0.984	0.984	1.000	0.985	0.982	0.984	0.986	0.946
u2	0.985	0.986	0.985	0.985	0.985	0.985	1.000	0.982	0.984	0.985	0.948
u3	0.987	0.987	0.988	0.986	0.987	0.982	0.982	1.000	0.978	0.981	0.948
u4	0.980	0.981	0.980	0.982	0.980	0.984	0.984	0.978	1.000	0.987	0.949
u5	0.983	0.984	0.983	0.983	0.984	0.986	0.985	0.981	0.987	1.000	0.949
um	0.960	0.962	0.959	0.965	0.962	0.946	0.948	0.948	0.949	0.949	1.000

Source: own calculations, based on IAB data

u01 - u05 from 5 auxiliary regressions with imputed missing values and imputed capital zeroes

u1 - u5 from 5 auxiliary regressions with imputed missing values

um from regression with non-missing values

In dieser Reihe sind zuletzt erschienen

Nr.	Autor(en)	Titel	Datum
1/2004	Sabine Hagemann, Werner Sörgel, Eberhard Wiedemann	Vermittlungsgutscheine nach § 421g SGB III - Zwischenergebnisse aus der Begleitforschung zur Vermittlung	9/2004
2/2004	Lutz Bellmann, Vera Dahms, Jürgen Wahse	IAB-Betriebspanel Ost - Ergebnisse der achten Welle 2003 – Teil I: Entwicklung und Struktur der Betriebe und Beschäftigten, Auszubildende	9/2004
3/2004	Lutz Bellmann, Vera Dahms, Jürgen Wahse	IAB-Betriebspanel Ost - Ergebnisse der achten Welle 2003 – Teil II: Personalpolitik, Betriebliche Flexibilität, Weiterbildung	9/2004
4/2004	Lutz Bellmann, Vera Dahms, Jürgen Wahse	IAB-Betriebspanel Ost - Ergebnisse der achten Welle 2003 – Teil III: Wirtschaftliche Lage der Betriebe, Öffentliche Förderung	9/2004
5/2004	Eugen Spitznagel, Susanne Wanger	Mehr Beschäftigung durch längere Arbeitszeiten? Ein Beitrag zu der Diskussion um eine generelle Erhöhung der Arbeitszeit	10/2004
6/2004	IAB-Autorengemeinschaft	Forschung zum SGB II des IAB: Die neuen Forschungsaufgaben im Überblick	12/2004
1/2005	Anja Heinze, Friedhelm Pfeiffer, Alexander Spermann, Henrik Winterhager, Amelie Wuppermann	Vermittlungsgutscheine - Zwischenergebnisse der Begleitforschung 2004 Teil I: Datenstruktur und deskriptive Analysen	3/2005
2/2005	Sabine Dann, Günther Klee, Martin Rosemann	Vermittlungsgutscheine - Zwischenergebnisse der Begleitforschung 2004 Teil II: Typisierung der Arbeitsagenturen	2/2005
3/2005	Anja Heinze, Friedhelm Pfeiffer, Alexander Spermann, Henrik Winterhager	Vermittlungsgutscheine - Zwischenergebnisse der Begleitforschung 2004 Teil III: Mikroökonometrische Wirkungsanalyse	3/2005

4/2005	Reinhard Hujer, Christopher Zeiss	Vermittlungsgutscheine - Zwischenergebnisse der Begleitforschung 2004 Teil IV: Makroökonomische Wirkungsanalyse	2/2005
5/2005	Friedhelm Pfeiffer, Henrik Winterhager	Vermittlungsgutscheine - Zwischenergebnisse der Begleitforschung 2004 Teil V: Kosten-Nutzen-Analyse	2/2005
6/2005	Sabine Hagemann, Werner Sörgel	Vermittlungsgutscheine - Zwischenergebnisse der Begleitforschung 2004 Teil VIa: Implementations- und Strukturanalysen - Private Arbeitsvermittler	7/2005
7/2005	Sabine Hagemann, Werner Sörgel	Vermittlungsgutscheine - Zwischenergebnisse der Begleitforschung 2004 Teil Vb: Implementations- und Strukturanalysen - Tabellenanhang	7/2005
8/2005	Reinhard Hujer, Günther Klee, Alexander Spermann, Werner Sörgel	Vermittlungsgutscheine - Zwischenergebnisse der Begleitforschung 2004 Teil VII: Zusammenfassung der Projektergebnisse	7/2005
9/2005	Regina Konle-Seidl	Lessons learned – Internationale Evaluierungsergebnisse zu Wirkungen aktiver und aktivierender Arbeitsmarktpolitik	2/2005
10/2005	Ch. Brinkmann, J. Passenberger, H. Rudolph, E. Spitznagel, G. Stephan, U. Thomsen, H. Roß	SGB II – Neue Herausforderungen an Statistik und Forschung	2/2005
11/2005	Corinna Kleinert, Hans Dietrich	Aus- und Weiterbildungen im Pflegebereich - Eine Analyse des Eingliederungsprozesses in Erwerbstätigkeit	3/2005
12/2005	Axel Deeke	Kurzarbeit als Instrument betrieblicher Flexiblität - Ergebnisse aus dem IAB-Betriebspanel 2003	3/2005
13/2005	Oliver Falck	Das Scheitern junger Betriebe Ein Überlebensdauermodell auf Basis des IAB-Betriebspanels	3/2005
14/2005	Helmut Rudolph, Kerstin Blos	Schätzung der Auswirkungen des Hartz-IV-Gesetzes auf Arbeitslosenhilfe-Bezieher	4/2005

15/2005	Johann Fuchs, Brigitte Weber	Neuschätzung der Stillen Reserve und des Erwerbspersonenpotenzials für Westdeutschland (inkl. Berlin-West)	5/2005
16/2005	Johann Fuchs, Doris Söhnlein	Vorausschätzung der Erwerbsbevölkerung bis 2050	5/2005
17/2005	Michael Feil, Gerd Zika	Politikberatung mit dem Simulationsmodell PACE-L – Möglichkeiten und Grenzen am Beispiel einer Senkung der Sozialabgaben	5/2005
18/2005	Johann Fuchs, Brigitte Weber	Neuschätzung der Stillen Reserve und des Erwerbspersonenpotenzials für Ostdeutschland (einschl. Berlin-Ost)	6/2005
19/2005	Stefan Schiel, Ralph Cramer, Reiner Gilberg, Doris Hess, Helmut Schröder	Das arbeitsmarktpolitische Programm FAIR - Zwischenergebnisse der Begleitforschung 2004 -	7/2005
20/2005	Lutz Bellmann, Vera Dahms, Jürgen Wahse	IAB-Betriebspanel Ost – Ergebnisse der neunten Welle 2004 – Teil I: Entwicklung und Struktur der Betriebe und Beschäftigten, Auszubildende	7/2005
21/2005	Lutz Bellmann, Vera Dahms, Jürgen Wahse	IAB-Betriebspanel Ost – Ergebnisse der neunten Welle 2004 – Teil II: Personalpolitik, Betriebliche Flexibilität, betriebliche Arbeitszeiten, ältere Arbeitnehmer	7/2005
22/2005	Lutz Bellmann, Vera Dahms, Jürgen Wahse	IAB-Betriebspanel Ost – Ergebnisse der neunten Welle 2004 – Teil III: Innovationen im Betrieb, wirtschaftliche Lage der Betriebe	7/2005
23/2005	Aderonke Osikominu	Eine Analyse der Teilnehmerselektion in die berufliche Weiterbildung auf Basis der Integrierten Erwerbsbiografien (IEB)	9/2005
24/2005	Uwe Blien, Franziska Hirschenauer	Vergleichstypen 2005: Neufassung der Regionaltypisierung für Vergleiche zwischen Agenturbezirke	9/2005
25/2005	Johann Fuchs, Katrin Dörfler	Projektion des Erwerbspersonenpotenzials bis 2050 – Annahmen und Grundlagen	9/2005
26/2005	Axel Deeke	Das ESF-BA-Programm im Kontext der arbeitsmarktpolitischen Neuausrichtung der Bundesagentur für Arbeit – Zur Umsetzung des Programms von 2000 bis Anfang 2005	10/2005

1/2006	Lena Koller, Ulrike Kress, Kerstin Windhövel	Blinde Kuh war gestern – heute ist FIS Das Forschungs-Informations-System – ein neuer Weg wissenschaftlicher Politikberatung	1/2006
2/2006	Susanne Wanger	Erwerbstätigkeit, Arbeitszeit und Arbeitsvolumen nach Geschlecht und Altersgruppe – Ergebnisse der IAB-Arbeitszeitrechnung nach Geschlecht und Alter für die Jahre 1991-2004	1/2006
3/2006	Sarah Heinemann, Hermann Gartner, Eva Jozwiak	Arbeitsförderung für Langzeitarbeitslose - Erste Befunde zu Eingliederungsleistungen des SGB III im Rechtskreis SGB II	2/2006
4/2006	Jan Binder, Barbara Schwengler	Neuer Gebietszuschnitt der Arbeitsmarktregionen im Raum Berlin und Brandenburg – Kritische Überprüfung der bisher gültigen Arbeitsmarktregionen und Vorschläge für einen Neuzuschnitt	2/2006
5/2006	Ch. Brinkmann, M. Caliendo, R. Hujer, St. L. Thomsen	Zielgruppenspezifische Evaluation von Arbeitsbeschaffungsmaßnahmen – Gewinner und Verlierer	2/2006
6/2006	Ch. Gaggermeier	Indikatoren-Modelle zur Kurzfristprognose der Beschäftigung in Deutschland	4/2006
7/2006	St. Schiel, R. Gilberg, H. Schröder	Evaluation des arbeitsmarktpolitischen Programms FAIR - 3. Zwischenbericht	4/2006
8/2006	K. Blos	Die Bedeutung der Ausgaben und Einnahmen der Sozialversicherungssysteme für die Regionen in Deutschland	3/2006
9/2006	A. Haas, Th. Rothe	Regionale Arbeitsmarktströme - Analysemöglichkeiten auf Basis eines Mehrkontenmodells	4/2006
10/2006	J. Wolff, K. Hohmeyer	Förderung von arbeitslosen Personen im Rechtskreis des SGB II durch Arbeitsgelegenheiten: Bislang wenig zielgruppenorientiert	6/2006
11/2006	L. Bellmann, H. Bielenski, F. Bilger, V. Dahms, G. Fischer, M. Frei, J. Wahse	Personalbewegungen und Fachkräfterekrutierung – Ergebnisse des IAB-Betriebspanels 2005	6/2006

12/2006	Th. Rhein, M. Stamm	Niedriglohnbeschäftigung in Deutschland: Deskriptive Befunde zur Entwicklung seit 1980 und Verteilung auf Berufe und Wirtschaftszweige	7/2006
13/2006	B. Rudolph, C. Klement	Arbeitsmarktpartizipation von Frauen im Transformationsprozess - Sozio-ökonomische Realität in den EU-Beitrittsländern Polen, Tschechien und Ungarn	7/2006
14/2006	Th. Rothe	Die Arbeitskräftegesamtrechnung für Ost- und Westdeutschland – Konzeption und ausgewählte Ergebnisse	7/2006
15/2006	R. Konle-Seidl, Kristina Lang	Von der Reduzierung zur Mobilisierung des Arbeitskräftepotenzials	8/2006
16/2006	Johanna Dornette, Marita Jacob	Zielgruppenerreichung und Teilnehmerstruktur des Jugendsofortprogramms JUMP	8/2006
17/2006	Andreas Damelang, Anette Haas	Arbeitsmarkteinstieg nach dualer Berufsausbildung – Migranten und Deutsche im Vergleich	8/2006
18/2006	Susanne Rässler	Der Einsatz von Missing Data Techniken in der Arbeitsmarktforschung des IAB	10/2006
19/2006	Kerstin Blos	Haushalte im Umfeld des SGB II	11/2006
20/2006	György Barabas, Roland Döhrn	Konjunktur und Arbeitsmarkt: Simulationen und Projektionen mit der IAB-Version des RWI-Konjunkturmodells	12/2006
21/2006	Axel Deeke	Berufsbezogene Sprachförderung für Arbeitslose mit Migrationshintergrund: Erste Ergebnisse aus der Begleitforschung zum ESF-BA-Programm	12/2006
22/2006	K. Hohmeyer, Ch. Schöll, J. Wolff	Arbeitsgelegenheiten in der Entgeltvariante: Viele Zielgruppen werden noch vernachlässigt	12/2006
23/2006	S. Noll, A. Novorozhkin, J. Wolff	Förderung mit dem Einstiegsgeld nach § 29 SGB II – Erste Befunde zur Implementation und Deskription	12/2006
1/2007	F. Rüb, D. Werner	Typisierung von SGB II-Trägern	1/2007

2/2007	A. Kettner, M. Rebien	Soziale Arbeitsgelegenheiten * Einsatz und Wirkungsweise aus betrieblicher und arbeitsmarktpolitischer Perspektive)	1/2007

Impressum

IABForschungsbericht
Nr. 3 / 2007

Editorial address
Institut für Arbeitsmarkt- und Berufsforschung
der Bundesagentur für Arbeit
Weddigenstr. 20-22
D-90478 Nürnberg

Editorial staff
Regina Stoll, Jutta Palm-Nowak

Technical completion
Jutta Sebald

Download of this Forschungsbericht:
http://doku.iab.de/forschungsbericht/2007/fb0307.pdf

Website
http://www.iab.de

For further inquires contact the author:
Susanne Rässler, Tel. 0911/179-3084,
oder E-Mail: susanne.raessler@iab.de