Methods in empirical economics - a selective review with applications

Olaf Hübler*

Discussion Paper No.513

ISSN 0949-9962

July 2013

Abstract

This paper presents some selective aspects of standard econometric methods and of new developments in econometrics that are important for applications with microeconomic data. The range includes variance estimators, measurement of outliers, problems of partially identified parameters, nonlinear models, possibilities of instrumental variables, panel methods for models with time-invariant regressors, difference-in-differences estimators, matching procedures, treatment effects in quantile regression analysis and regression discontinuity approaches. These methods are applied to production functions with IAB establishment panel data.

Keywords: Significance, standard errors, outliers, influential observations, partially identified parameters, unobserved heterogeneity, instrumental variables, panel estimators, quantile regressions, causality, treatment effects, DiD estimators, regression discontinuity

JEL classification: C21, C26, D22, J53

*Institut für Empirische Wirtschaftsforschung, Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover, huebler@ewifo.uni-hannover.de

1 Introduction

Contents, questions and methods have changed in empirical economics in the last 50 years. Many methods were developed in the past but the application in empirical economics followed with delay. Some methods are well-known but have experienced only little attention. New approaches focus on characteristics of the data, on modified estimators, on correct specifications, on unobserved heterogeneity, on endogeneity and on causal effects. Actual data sets are not compatible with the assumptions of classical models. Modified methods were presented to inference. Especially nonlinear relationships are in the focus.

Users have to follow specific principles at the selection of appropriate methods. Kennedy (2002) has formulated 10 rules:
rule 1: Thou shalt use common sense and economic theory.
rule 2: Thou shalt ask the right questions.
rule 3: Thou shalt know the context.
rule 4: Thou shalt inspect the data.
rule 5: Thou shalt not worship complexity.
rule 6: Thou shalt look long and hard at thy results.
rule 7: Thou shalt beware the costs of data mining.
rule 8: Thou shalt be willing to compromise.
rule 9: Thou shalt not confuse significance with substance.
rule 10: Thou shalt confess in the presence of sensitivity.

The following considerations are based on five hypotheses:

(1) Significance is an important indicator in empirical economics but the results are sometimes misleading.

(2) Assumptions' violation, outliers and only partially identified parameters are often the reason of wrong standard errors.

(3) OLS estimation is the working horse in empirical economics but especially unobserved heterogeneity and endogeneity require alternative methods.

(4) The estimation of average effects is useful but subgroup analysis and quantile regressions are important supplements.

(5) Causal effects are of great interest but the determination is based on disparate approaches with varying results.

2 Econometric methods

2.1 Significance and standard errors in regression models

The working horse in empirical economics is the classical linear model

$$y_i = x_i'\beta + u_i \qquad i = 1, ..., n.$$

with some specific assumptions which are often not fulfilled using real data sets. The coefficient vector β is estimated by ordinary least squares (OLS)

$$\hat{\beta} = (X'X)^{-1}X'y$$

and the covariance matrix by

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}.$$

The influence of a regressor, e.g. x_k on the regressand y is called significant if $|t| = |\hat{\beta}_k/\sqrt{\hat{V}(\hat{\beta}_k)}| > t_{0.975}$. Ziliak/McCloskey(2008) and Krämer(2011) have criticized this procedure. Three types of mistakes can lead to a misleading interpretation:

(1) There does not exist any effect but due to technical inefficiencies a significant effect is reported.

(2) The effect is small but due to the precision of the estimates a significant effect is determined.

(3) There exists a strong effect but due to the variability of the estimates the effect cannot be detected.

The consequence cannot be to neglect the instrument of significance but what can we do? Some proposals may help:

- Compute robust standard errors.
- Analyze whether variation within clusters is only small in comparison with variation between the clusters.
- Check whether dummies as regressors with low or high probability are responsible for insignificance.
- Test whether outliers induce large standard errors.
- Consider the problem of partially identified parameters.
- Detect whether collinearity is effective.
- Investigate alternative specifications.
- Use sub-samples and compare the results.
- Execute sensitivity analyzes.
- Employ Hamermesh's sniff test in order to detect whether econometric results are in accord with economic plausibility.

2.1.1 White's covariance estimator and modifications

Heteroscedasticity induces inefficient OLS estimates. This problem can be avoided by transformation of the regression model

$$\frac{y_i}{\sigma_i} = \frac{\beta_0}{\sigma_i} + \beta_1 \frac{x_{1i}}{\sigma_i} + \ldots + \beta_k \frac{x_{ki}}{\sigma_i} + \frac{u_i}{\sigma_i},$$

where $i = 1, \ldots, n$. Typically, the individual variances of the error term are unknown. In the case of unspecific heteroscedasticity White (1980) recommends the following estimation of the covariance matrix

$$\hat{V}_W(\hat{\beta}) = (X'X)^{-1}(\sum \hat{u}_i^2 x_i x_i')(X'X)^{-1}.$$

Such estimates are asymptotically heteroscedasticity-robust. In many empirical investigations this robust estimator is routinely applied without testing whether heteroskedasticity exists. We should stress that those estimated standard errors are more biased than conventional estimators if residuals are homoskedastic. As long as there is not too much heteroskedasticity, robust standard errors are also biased downward. In the literature we find some suggestions to modify this estimator, namely $\hat{u}_i^2$ should be substituted by:

$$hc_1 = \frac{n}{n-K}\hat{u}_i^2$$

$$hc_j = \frac{1}{(1-c_{ii})^{\delta_j}}\hat{u}_i^2$$

where j=2, 3, 4, c_{ii} is the main diagonal element of $X'(X'X)^{-1}X$ and $\delta_j = 1; 2; min[\gamma_1, (nc_{ii})/K] + min[\gamma_2, (nc_{ii})/K]$. It is necessary that γ_2 is positive and constant.

It can be shown for hc_2 that under homoskedasticity the mean of $\hat{u}_i^2$ is the same as $\sigma^2(1-c_{ii})$. Therefore we should expect that the hc_2 options leads under homoskedasticity to better estimates in small samples then the simple hc_1 option. Then $E(\hat{u}_i^2/(1-c_{ii}))$ is σ^2. The second correction is presented by MacKinnon and White (1985). This is an approximation of a more complicated estimator which is based on a jackknife estimator. Applications demonstrate that the standard error increases started with OLS via hc_1, hc_2 and hc_3 option. Simulations, however, do not show a clear preference. As one cannot be sure which case is the correct one, a conservative choice is preferable (Angrist/Pischke 2009, 302). The estimator should be chosen that has the largest standard error. This means the null hypothesis (H_0: no influence on the regressand) keeps up longer than with other options.

Cribari-Neto and da Silva (2011) suggest $\gamma_1 = 1$ and $\gamma_2 = 1.5$ in hc_4. The intention is to weaken the effect of influential observations compared with hc_2 and hc_3 or in other words to enlarge the standard errors. In an earlier version (Cribari-Neto et al. 2007) a slight modification is presented: $hc_4^* = 1/(1-c_{ii})^{\delta_{4*}}$, where $\delta_{4*} = min(4, nc_{ii}/p)$.It is argued that the presence of high leverage observations is more decisive for the finite-sample behavior of the consistent estimators of $V(\hat{\beta})$ than the intensity of heteroskedasticity, hc_4 and hc_{4*} aim at discounting for leverage points - see section 2.2.1 - more heavily than hc_2 and hc_3. The same authors formulate a further estimator

$$hc_5 = \frac{1}{\sqrt{(1-c_{ii})^{\delta_5}}},$$

where $\delta_5 = min(\frac{nc_{ii}}{p}, max(4, \frac{nkc_{ii,max})}{p}))$ and k is again a predefined constant, $k = 0.7$ is suggested. In this case squared residuals are affected by the maximal leverage.

2.1.2 Re-sampling procedures

Other possibilities to determine the standard error are the jackknife and the bootstrap estimator. These are re-sampling procedures, which construct sub-samples with n-1 observations in the **jackknife** case. Sequentially, one observation is eliminated. The former methods compare the estimated coefficients of the total sample size $\hat{\beta}$ with those after eliminating one observation $\hat{\beta}_{-i}$. The jackknife estimator of the covariance matrix is

$$\hat{V}_{jack} = \frac{n-K}{n} \sum_{i=1}^{n} (\hat{\beta}_{-i} - \hat{\beta})(\hat{\beta}_{-i} - \hat{\beta})'.$$

There exist many ways to **bootstrap** regression estimates. The basic idea is assume that the sample with n elements is the population and B times m elements (sampling with replacement) are drawn, where $m < n$ and $m > n$ is feasible. If $\hat{\beta}'_{boot} = (\hat{\beta}(1)'_m, ..., \hat{\beta}(B)'_m)$ are the bootstrap estimators of the coefficients the asymptotic covariance matrix is

$$\hat{V}_{boot} = \frac{1}{B} \sum_{b=1}^{B} (\hat{\beta}(b)_m - \hat{\beta})(\hat{\beta}(b)_m - \hat{\beta})'$$

where $\hat{\beta}$ is the estimator with the original sample n. Alternatively, $\hat{\beta}$ can be substituted by $\bar{\beta} = 1/B \sum \hat{\beta}(b)_m$. Bootstrap estimates of the standard error are especially helpful if it is difficult to compute standard errors by conventional methods, e.g. 2SLS estimators under heteroskedasticity or cluster-robust standard errors when many small clusters or only short panels exist. The jackknife can be viewed as a linear approximation of the bootstrap estimator. A further popular way to estimate the standard errors is the **delta method**.This approach is especially used for nonlinear functions of parameter estimates $\hat{\gamma} = g(\hat{\beta})$. An asymptotic approximation of the covariance matrix of a vector of such functions is determined. It can be shown that

$$n^{1/2}(\hat{\gamma} - \gamma_0) \sim N(0, G_0 V^{\infty}(\hat{\beta}) G_0'),$$

where G_0 is an $l \times K$ matrix with typical element $\partial g_i(\beta)/\partial \beta_j$.

2.1.3 The Moulton problem

The variance of a regressor is low if this variable strongly varies between groups but only little within groups (Moulton 1986, 1987, 1990). This is especially the case if industry, regional and macroeconomic variables are introduced or panel data are considered. In a more general this is called the problem of cluster sampling. Individuals or establishments are sampled in groups or clusters. Consequence may be a weighted estimation that adjust for differences in sampling rates. However, weighting is not always necessary and estimates may understate the true standard errors.

Example: Assume a data set with 5 observations and 4 variables ($V1 - V4$).

i	V1	V2	V3	V4
1	24	123	-234	-8
2	875	87	54	3
3	-12	1234	-876	345
4	231	-87	-65	9808
5	43	34	9	-765

The linear model

$$V1 = \beta_1 + \beta_2 V2 + \beta_3 V3 + \beta_4 V4 + u$$

is estimated by OLS using the original data set (1F), then the data set is doubled (2F), quadrupled (4F) and octuplicated (8F). The following estimates result

		1F	2F	4F	8F
V1	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$			
V2	1.723907	1.753241	0.715757	0.438310	0.292207
V3	2.794101	2.387409	0.974655	0.596852	0.397903
V4	0.027040	0.061766	0.025216	0.015442	0.010294
_cons	323.2734	270.5781	110.463	67.64452	45.0963

The coefficients are the same, however, the standard errors decrease if the same data set is multiplied. Namely, the variance is only 1/6, 1/16 and 1/36 of the original variance. The general relationship can be shown as follows. For the original data set (X_1) the variance is

$$\hat{V}_1(\hat{\beta}) = \hat{\sigma}_1^2(X_1'X_1)^{-1}.$$

Using $X_1 = ... = X_F$ the F times enlarged data set with the design matrix $X' =: (X_1'...X_F')$ leads to

$$\hat{\sigma}_F^2 = \frac{1}{F \cdot n - K}\sum_{i=1}^{F \cdot n}\hat{u}_i^2 = \frac{F(n-K)}{F \cdot n - K}\hat{\sigma}_1^2$$

and

$$\hat{V}_F(\hat{\beta}) = \hat{\sigma}_F^2(X'X)^{-1} = \hat{\sigma}_F^2\frac{1}{F} \cdot (X_1'X_1)^{-1} = \frac{n-K}{F \cdot n - K}\hat{V}_1(\hat{\beta}).$$

K is the number of regressors including the constant term, n is the number of observations in the original data set (number of clusters), F is the number of observations within a cluster. In the numerical example with F=8, K=4, n=5 the Moulton factor M which indicates the deflation factor of the variance is

$$M = \frac{n-K}{F \cdot n - K} = \frac{1}{36}.$$

This is exactly the same as it was demonstrated in the numerical example. Analogously the estimated values 1/6 and 1/16 can be determined. As the multiplying of the data set does not add any further information to the simple data set not only the coefficients but also the standard errors should be the same. Therefore it is necessary to correct the covariance matrix. Statistical packages, e.g. Stata, supply cluster-robust estimates

$$\hat{V}(\hat{\beta})_C = (\sum_{c=1}^{C} X_c'X_c)^{-1}\sum_{c=1}^{C} X_c'\hat{u}_c\hat{u}_cX_c(\sum_{c=1}^{C} X_c'X_c)^{-1},$$

where C is the number of clusters. In our specific case this is the number of observations n. This approach implicitly assumes that F is small and $n \to \infty$. If this assumption does not hold a degrees-of-freedom correction

$$df_C = \frac{F \cdot n - 1}{F \cdot n - K}\frac{n}{n-1}$$

is helpful. $df_C \cdot \hat{V}(\hat{\beta})_C$ is the default option in Stata, which corrects for the number of clusters in practice being finite. Nevertheless, this correction eliminates only partially the underestimated standard errors. In other words, the t-statistics are larger than that of $\hat{\beta}_k/\sqrt{\hat{V}_1}$.

2.1.4 Large standard errors of dichotomous regressors with small or large mean

Assume a simple two variable classical regression model

$$y_i = a + b \cdot D_i + u_i,$$

where D_i is a dummy variable. The variance of $\hat{b}$ is

$$V(\hat{b}) = \frac{\sigma^2}{n} \cdot \frac{1}{s_D^2},$$

where

$$s_D^2 = \hat{P}(D=1) \cdot \hat{P}(D=0) =: \hat{p}(1-\hat{p}) = \frac{(n|D=1)}{n} \cdot (1 - \frac{(n|D=1)}{n}).$$

This result holds only for inhomogeneous models. An extension to multiple regression models seems possible - see applications Table A1 and Table 2. $V(\hat{b})$ is minimal at given n and σ^2 when the estimated variance of D reaches the maximum, if $\hat{p} = 0.5$. The more $\hat{p}$ deviates from 0.5, the larger or smaller is $\hat{p}$, the higher is the tendency to insignificant effects. This conclusion is not unavoidable that the t-value of a dichotomous regressor D_1 is always larger than that of D_2, when $V(D_1) > V(D_2)$. The significance is determined by $\hat{b}/\sqrt{\hat{V}(\hat{b})}$.

Example: An income variable ($Y = Y_0/10^7$) with 53 664 observations is regressed on a Bernoulli distributed random variable RV. The linear model $Y = \beta_0 + \beta_1 RV + u$ is estimated by OLS where the mean of RV ($\overline{RV}$) is alternatively $0.1, 0.2, \ldots, 0.9$

Y	$\hat{\beta}$	std.err
$\overline{RV}$=0.1	-0.3727	0.6819
$\overline{RV}$=0.2	-0.5970	0.5100
$\overline{RV}$=0.3	-0.4768	0.4455
$\overline{RV}$=0.4	0.3068	0.4170
$\overline{RV}$=0.5	0.1338	0.4094
$\overline{RV}$=0.6	0.0947	0.4187
$\overline{RV}$=0.7	-0.0581	0.4479
$\overline{RV}$=0.8	-0.1860	0.5140
$\overline{RV}$=0.9	-0.1010	0.6827

This example confirms the theoretical result. The standard error is smallest if $\overline{RV} = 0.5$ and increases systematically if the mean of RV decreases or increases. An extension to multiple regression models seems possible - see applications in the Appendix, Tables A1-A4. The more $\bar{D}$ deviates from 0.5, the larger or smaller is the mean of D, the higher is the tendency to insignificant effects. A caveat is necessary. The conclusion that the t-value of a dichotomous regressor D_1 is always smaller than that of D_2, when $V(D_1) > V(D_2)$, is not unavoidable. The basic effect of D_1 may be larger than that of D_2 on y. The theoretical result aims on specific variables and not on the comparison between regressors. In practice, significance is determined by $t = \hat{b}/\sqrt{\hat{V}(\hat{b})}$. However, we do not find a systematic influence of $\hat{b}$ on t if $\bar{D}$ varies. Nevertheless, the random differences in the influence of D on y can dominate the $\bar{D}$ effect via s_D^2. The comparison of Table A3 with A4 shows that the influence of a works council (WOCO) is stronger than that of a company level pact (CLP), the coefficients are larger and the standard errors are lower so that the t-values are larger. In both cases the standard errors increase if the mean of the regressor is reduced. The comparison of line 1 in Table A3 with line 9 in Table A4, where the mean of CLP and $WOCO$ is nearly the same, makes clear that the stronger basic effect of $WOCO$ on lnY dominates the mean reduction effect of $WOCO$. The t-value in line 9 of Table A4 is smaller than that in line 1 of Table A4 but still larger than that in line 1 of Table A3. Not all deviations of the mean of a dummy D as regressor from 0.5 induce increasing standard error effects. The variation of $\bar{D}$ has to be randomly. An example, where this is not the case, is matching - see section 2.2 and the application in section 3. $\bar{D}$ increases due to the systematic elimination of those observations with $D = 0$ that are dissimilar to those of $D = 1$.

2.2 Outliers and partially identified parameters

Outliers, influential observations, grouped data, multicollinearity, missing values, measurement errors and partially identified parameters are specific data problems. This entails consequences for the coefficients and the standard error estimates.

2.2.1 Outliers and influential observations

Outliers may have strong effects on the estimates of the coefficients, of the dependent variable and on standard errors and therefore on significance. The main diagonal elements c_{ii} of the hat matrix $C = X(X'X)^{-1}X'$ (Leverages) characterize the effects of a single observation on the coefficient estimator, on the estimated endogenous variable $\hat{y}_i$ and on the variance $\hat{V}(\hat{y}_i)$. a leverage is noted as strong if $c_{ii} > 2K/n$. The higher c_{ii}, the higher is the difference between the estimation with and without the i-th observation. A rule of thumb of influential observations orients on the relation $c_{ii} > 2K/n$. The effects of the i-th observation on $\hat{\beta}$, $\hat{y}$ and $\hat{V}(\hat{\beta})$ and the rules of thumb can be expressed by

$$|\hat{\beta}_k - \hat{\beta}_k(i)| > \frac{2}{\sqrt{n}}$$

$$|\frac{\hat{y}_i - \hat{y}_{i(i)}}{s(i)\sqrt{c_{ii}}}| > 2\sqrt{\frac{K}{n}}.$$

$$|\frac{det(s^2(i)(X'(i)X(i))^{-1}}{det(s^2(X'X)^{-1})}| > \frac{3K}{n}.$$

The determination of an outlier is based on externally studentized residuals

$$\hat{u}_i^* = \frac{\hat{u}_i}{s(i)\sqrt{1 - c_{ii}}} \sim t_{n-K-1}.$$

Alternatively, a mean shift outlier model can be formulated

$$y = X\beta + A_i\delta + \epsilon,$$

where

$$A_i = \begin{cases} 1 & \text{if i is assumed as an outlier} \\ 0 & \text{otherwise} \end{cases}$$

Observation i has an effect on y if δ is significantly different from zero. The estimated t-value is the same as $\hat{u}_i^*$.

Hadi(1992) proposes a generalized approach, an outlier detection with respect to all regressors. The decision whether the design matrix X contains outliers is based on an elliptical distance

$$d_i(c, V) = \sqrt{(x_i - c)'V^{-1}(x_i - c)},$$

where intuitively the classical choices of c and V are the arithmetic mean ($\bar{x}$) and the sample covariance matrix S of the data set X so that the Mahalanobis distance follows. If

$$d_i(\bar{x}_i, S)^2 > \chi^2_K$$

observation i is identified as an outlier. As $\bar{x}$ and S react sensitive to outliers it is necessary to estimate an outlier-free mean and sample covariance matrix. For this purpose, only outlier-free observations are considered to determine $\bar{x}$ and S. Another way to avoid the sensitivity problem is to use more robust estimators of the location and covariance matrix, e.g. the median but not the mean is robust to outliers. Finally, an outlier vector MOD (multiple outlier dummy) instead of A is incorporated in the model in order to test whether the identified outlier observations have a significant effect on the dependent variable. A second problem is whether we should eliminate all outliers or only some of them or no outlier. The situation is obvious if an outlier is induced by measurement errors. Typically, however, we cannot be sure. Insofar, the correct estimation is based between the two extremes: all outliers are considered and all outliers are eliminated.

2.2.2 Partially identified parameters

Some observations are unknown or not exactly measured. Consequence is that a parameter cannot exactly be determined but only within a range. The outlier situation leads to a partial identification problem. There exist many other similar constellations.

Example: share of unemployed persons is 8% but 5% have not answered to the question of the employment status. Therefore, the unemployment rate can only be calculated within certain limits, namely between the two extremes:

- all persons who have not answered are employed
- all person who have not answered are unemployed.

In the first case the unemployment rate is 7.2% and in the second case 12.2%.

The main methodological focus is the inference. Chernozhukov et al. (2007), Imbens/Manski (2004), Romano/Shaikh (2010), Stoye (2009) and Woutersen (2009) have discussed solutions.

If $\Theta_0 = [\theta_l, \theta_u]$ is the known range of the interested parameter the confidence interval following Stoye(2009) is

$$CI_\alpha = [\hat{\theta}_l - \frac{c_\alpha \hat{\sigma}_l}{\sqrt{n}}, \hat{\theta}_u - \frac{c_\alpha \hat{\sigma}_l}{\sqrt{n}}],$$

where $\hat{\sigma}_l$ is the standard error of the estimation function $\hat{\theta}_l$. c_α is chosen by

$$\Phi(c_\alpha + \frac{\sqrt{n}\Delta_n}{\hat{\sigma}_l}) - \Phi(-c_\alpha) = 1 - \alpha,$$

where $\Delta = \theta_u - \theta_l$. As Δ is unknown, an analogous interval has to be estimated.

2.3 Nonlinear models

Linear models are widely spread in empirical economics. Economic theory has in many fields no clear answer to the question whether and which nonlinear approaches are preferable. And if this is the case a linearisation is often possible by renaming, taking of logarithm or representation by Taylor series. Nevertheless, the approximation mistake can be enormous.

The manifold possibilities of nonlinear modeling exist. The Box-Cox transformation is one way. First the endogenous variable can be modified

$$y^{(\lambda)} = \begin{cases} \frac{y^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \ln y & \lambda = 0, \end{cases}$$

where the parameter λ is typically an integer. The same is possible for one or all regressors, also in combination with $y^{(\lambda)}$, where the same or different λ values can be assumed.

An alternative is the construction of interaction variables or the restriction of linearity to specific ranges. Linear splines are the conventional instrument. In the two variable case the representation is

$$\begin{aligned} y &= \beta_0^{(1)} + \beta_1^{(1)} x + u, \text{if } x < x_1 \\ y &= \beta_0^{(2)} + \beta_1^{(2)} x + u, \text{if } x_1 \leq x < x_2 \\ y &= \beta_0^{(3)} + \beta_1^{(3)} x + u, \text{if } x_2 \leq x \end{aligned}$$

x_1 and x_2 are knots. The linear ranges can be determined by dummies

$$D_1 = \begin{cases} 1, & \text{if} \quad x \geq x_1 \\ 0 & \text{otherwise} \end{cases} \qquad D_2 = \begin{cases} 1, & \text{if} \quad x \geq x_2 \\ 0 & \text{otherwise} \end{cases}$$

These three regressions can be combined to one regression with interactions:

$$y = \gamma_0 + \gamma_1 x + \gamma_2 D_1 + \gamma_3 D_1 x + \gamma_4 D_2 + \gamma_5 D_2 x + u \; .$$

A further restriction is necessary

$$\begin{aligned} \gamma_0 + \gamma_1 x_1 &= (\gamma_0 + \gamma_2) + (\gamma_1 + \gamma_3) x_1 \\ (\gamma_0 + \gamma_2) + (\gamma_1 + \gamma_3) x_2 &= (\gamma_0 + \gamma_2 + \gamma_4) + (\gamma_1 + \gamma_3 + \gamma_5) x_2 \; . \end{aligned}$$

so that the linear restrictions are

$$\begin{aligned} 0 &= \gamma_2 + \gamma_3 x_1 \\ 0 &= \gamma_4 + \gamma_5 x_2 . \end{aligned}$$

The unrestricted regression is

$$y = \gamma_0 + \gamma_1 x + \gamma_3 D_1(x - x_1) + \gamma_5 D_2(x - x_2) + u.$$

After renaming

$$y = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_3 + u,$$

follows, where

$$\begin{aligned} z_1 &= x \\ z_2 &= x - x_1 \text{ if } x \geq x_1;\ 0 \quad \text{otherwise} \\ z_3 &= x - x_2 \text{ if } x \geq x_2;\ 0 \quad \text{otherwise.} \end{aligned}$$

Testing is possible, whether

- the slopes are the same in all linear ranges ($\beta_1^{(1)} = \beta_1^{(2)} = \beta_1^{(3)}$)
  $H_0:\ \alpha_2 = \alpha_3 = 0$
- for $x < x_2$ the slope is constant ($\beta_1^{(1)} = \beta_1^{(2)}$)
  $H_0:\ \alpha_2 = 0$
- for $x \geq x_1$ the slope is constant ($\beta_1^{(2)} = \beta_1^{(3)}$)
  $H_0:\ \alpha_3 = 0$ $\diamond$

Linear splines can also be estimated within multiple models, where for regressor x_k a piecewise linear regression is constructed. Assumptions are necessary for x_{k1} to $x_{k(L-1)}$ by numerical specification $*_1$ ... $*_{L-1}$ that fulfill the condition $*_1 < ... < *_{L-1}$. The modeling is

$$\frac{dy}{dx_k} = \begin{cases} \beta_{k1}, & \text{if } x_k < *_1 \\ \vdots & \\ \beta_{k(L-1)}, & \text{if } *_{L-2} \leq x_k < *_{L-1} \\ \beta_{kL} & \text{otherwise} \end{cases}$$

A generalization follows if the linear pieces are substituted by nonlinear functions. In practice cubic splines are popular.
Three further possibilities of nonlinear modeling are suggested in the literature:

(1) Construction of specific functions of linear combinations of regressors

$$y_i = F(x_i'\tilde{\beta} + u_i).$$

Practical relevance has this case if y is a binary variable. $F(\cdot)$ is assumed as a distribution function, most often a normal or a logistic function function. The well known probit and logit model follow.

(2) Observations can be substituted by a weighted adjustment due to larger and smaller neighbor values. The more a neighbor x_i deviates from the observation that shall be adjusted (x_0) the lower is the weight

$$w(u) = w(\frac{|x_0 - x_i|}{\Delta x_0}) = \begin{cases} (1-u^3)^3 & \text{for } 0 \leq u < 1 \\ 0 & \textit{otherwise}, \end{cases}$$

where $\Delta(x_0) = \max_{N(x_0)} |x_0 - x_i|$.

(3) Unspecific functions for some regressors. Additive models belong to this category. Some or all linear functions are substituted by an arbitrary smoothing function, where the exogenous variables are additively combined

$$y = \beta_0 + \sum_{k=1}^{K_1} f(x_k) + \sum_{k=K_1+1}^{K} x_k\beta_k + u.$$

The estimation of case (1) can be done by a Taylor series approximation of first order by the least squares or the maximum likelihood method. A procedure of numerical optimization has to be applied to obtain the solution. In case (2) the smoothing is usual a weighted least squares method. Case (3) follows the backfitting algorithm. Iteratively, the additive nonlinear regression is estimated. Parametric and nonparametric procedures are available.

2.4 Instrumental variables estimators

OLS estimators are biased and inconsistent if disturbances and regressors are correlated. Especially, four phenomenons are responsible:

- neglected influences that are correlated with regressors;
- interdependence between regressors and regressand;

- random measurement errors in the regressors;
- lagged endogenous variables as regressors under autocorrelation.

The interdependencies can be explained by theoretical arguments or by a statistical test. There exist different versions to tests for exogeneity. One possibility is the comparison of the coefficient estimates of a linear model $y = X\beta + u$ based on OLS and 2SLS

$$TS = (\hat{\beta}_{OLS} - \hat{\beta}_{2SLS})'(\hat{V}(\hat{\beta}_{OLS}) - \hat{V}(\hat{\beta}_{2SLS}))(\hat{\beta}_{OLS} - \hat{\beta}_{2SLS}).$$

This test statistic is asymptotically χ^2 distributed. The null hypothesis of exogeneity will be rejected if TS exeeds $\chi^2(K; 1-\alpha)$ at a given α level. K is the number of regressors.

Under this decision inconsistence can be avoided if instrumental variables (IV) estimators are used characterized by two properties: (i) strong correlation with the interested regressors which have to be eliminated; (ii) no correlation with the disturbance term.

Look for substitutes $(z_{K1}, \cdots, z_{KL})$ of one or more than one regressor of the linear regression model $y = X\beta + u$, e.g. for x_K. The instrumental variables matrix is $Z = (x_1, \cdots, x_{K-1}, z_{K1}, \cdots, z_{KL}) \sim N \times (K - 1 + L)$ and the sample condition

$$\frac{1}{N} Z'\hat{u} = \frac{1}{N} Z'(y - X\hat{\beta}) = 0.$$

The weighted distance

$$(y - X\beta)'ZWZ'(y - X\beta),$$

is minimized, where W is a weighting matrix. $\hat{W} = \hat{V}^{-1}$ is optimal if $\hat{W}$ is a consistent estimation of $V((1/N)Z'u)^{-1}$.

The condition of the sample moments is

$$\frac{1}{N} X'Z\hat{W}Z'\hat{u} = \frac{1}{N} X'Z\hat{V}^{-1}(Z'y - Z'X\hat{\beta}) = 0$$

and the estimation of β follows

$$\hat{\beta}_{IV} = (X'Z\hat{V}^{-1}Z'X)^{-1}(X'Z\hat{V}^{-1}Z'y).$$

Generalized method of moments(GMM) estimators are a generalization of IV estimators that we can interpret as a method of moments (MM) estimator. The 2SLS estimator is a special case. In absence of heteroscedasticity and autocorrelation we obtain $\hat{V} = \frac{1}{N}\hat{\sigma}^2(Z'Z)^{-1}$. If this is substituted in the IV estimator we obtain the 2SLS estimator

$$\hat{\beta}_{2SLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'y).$$

The first stage is the OLS estimation of the reduced form. In the second stage the endogenous variables are substituted by the estimated endogenous variables of the first stage. Then OLS is used again.

The 3SLS and the LIML estimator are specific IV estimators as the 2SLS estimator. The LIML method is based on a ML approach that spends only a priori information from the interested equation. The difference between the 2SLS and the 3SLS estimator is following: in the latter a system of seemingly unrelated equations is estimated in the third stage by a generalized least squares estimation. In this case more information is used than in the 2SLS method. Heteroskedasticity is allowed. A new Fuller like estimator based on a jackknife version of the LIML estimator presented by Hausman et al. (2012) is robust to heteroskedasticity and many instruments. This is in contrast to the conventional LIML and Fuller estimators.

IV and GMM estimators have preferable asymptotic properties. Less is known about small sample properties. Nelson and Startz (1990) investigate properties of IV estimators for finite samples. They are biased and the distribution is bimodal with relatively strong tails. The estimates are often nearby the OLS estimator than around the true parameter vector.

Bound, Jaeger and Baker (1995) show that in finite samples the bias has a similar dimension as that of the OLS estimator. The more the partial determination coefficient between z and x tends to zero, the more similar is the bias as that of the OLS estimator. Even in large samples there is a danger of biased IV estimators.

It is one of the greatest challenges in empirical economics to find adequate instruments. In the literature some general proposals for instrumental variables z are made:

(i) lagged exogenous or endogenous variables;

(ii) current or previous differences of exogenous variables;

(iii) group means or differences between present values and means over time;

(iv) proxies based on knowledge to economic mechanisms that determine the variables x;

(v) institutional and legal rules measured by dummies that are responsible that the variation of the variables x is restricted.

Many instruments in practice are **bad instruments** because they correlate with the disturbance term. There is often a trade-off between bad and **weak instruments**. Instruments are weak if z can explain only the small part of the variation of x. Consequences: a tendency to inconsistence and a bad approximation of the true distribution.

Indicators and tests to detect weak instruments are:

- partial determination coefficient between the regressor x_K and the instrumental variables $z_{K1}, \ldots, z_{KL}$ (Bound, Jaeger and Baker 1995);
- F statistics to the regression $X^* = Z^*\Pi + V$, where in X^* and Z^* the influence of joint variables in X and Z is partialed out; as a rule of thumb Staiger and Stock (1997) recommend to accept instruments if $F > 10$;
- adjusted partial determination coefficient (Shea (1997);
- two-stage test that reacts sensitive to weak IV (Hahn and Hausman 2002), where the null hypothesis (H_0) is: instruments are not weak.
- reverse test - H_0 instruments are weak; Stock and Yogo (2005) present a table of critical values based on the minimal eigenvalue of an F statistic.

Three tendencies can be observed:

- o The IV estimator converge to the OLS estimator if the number of instruments is large in comparison with the sample size and the estimation of the first stage has a good fit.
- o The conclusions of the asymptotic theory are not helpful to assess the IV estimator based on finite samples if the disturbance term of the interested equation and that of the first stage model are strongly correlated.
- o The asymptotic theory is even meaningless for large samples if the instruments are weak.

A direct test whether the second property of instruments is fulfilled is not possible as the disturbances are not observable. However in an overidentified model, we can test for the validity of the overidentifying restrictions (OIR) following Sargan (1958). In such a test, the residuals from a 2SLS regression are regressed on all included exogenous regressors and on all instruments. Under the null hypothesis a LM statistic of the $N \cdot R^2$ form has a large-sample χ^2(r) distribution, where r is the number of overidentifying restrictions. If the OIR test indicates that you should reject the null hypothesis, then this is clear evidence that the model is misspecified. We cast doubt on the suitability of the instrument set. This is not a test of the hypothesis that "the instruments are valid". Nevertheless, the OIR test can be considered as a first hurdle that needs to be overcome in the context of IV estimation. Whenever the OIR test implies rejection of the null, this usually means at least one of the instruments would have a significant effect in the structural equation.

Tests of overidentifying restrictions actually test two different things simultaneously. One is that the equation is misspecified and that one or more of the excluded exogenous variables should in fact be included in the structural equation. The other is whether the instruments are uncorrelated with the error term. Thus a significant test statistic could represent either an invalid instrument or an incorrect specified structural equation.

2.5 Panel estimators

A panel data set is a given sample of individuals, establishments, regions or countries over time and provides multiple observations on each group, e.g. individual, in the sample. The basic linear model is

$$y_{it} = \beta_{0i} + \sum_{k=1}^{K} \beta_k x_{kit} + u_{it} \qquad i = 1, ..., N \quad t = 1, ..., T$$
$$\beta_{0i} = \bar{\beta}_0 + \mu_i$$

For the individual effect μ_i we separate between:

- **Fixed effects model** (FEM): for each individual a constant value μ_i is assumed. Correlation between μ_i and x_{kit} is allowed.
- **Random effects model** (REM): μ_i is a random variable, μ_i and x_{kit} are uncorrelated.

The most often applied methods in practice are:

- First-differences or within estimators are applied to estimate FEMs. Problem: effects of time-invariant regressors on y are eliminated. These cannot be separated from the individual effect.
- A generalized least squares estimator is typically applied to REMs. The assumption that regressors and disturbances are uncorrelated is very often not fulfilled.
- Test whether individual effects exist ($H_0 : \beta_0' = (\bar{\beta}_0, ..., \bar{\beta}_0)$ against $H_1 : \beta_0' = (\beta_{01}, ...\beta_{0N})$) - Breusch-Pagan or F test.
- Hausman test is applied to decide whether the fixed or the random effects model is preferred.

An alternative to conventional panel estimators is the **Hausman-Taylor estimator** (1981). This one is also consistent if regressors and the individual term correlate and can determine the effects of time-invariant regressors. This approach distinguishes between time-invariant and time-variant regressors (x_{it}, w_i).

In both cases the model separates whether the variables correlate with the individual term μ_i (x_{2it}, w_{2i}) or not (x_{1it}, w_{1i})

$$\begin{aligned} y_{it} &= \beta_0 + x'_{1it}\beta_1 + w'_{1i}\gamma_1 + x'_{2it}\beta_2 + w'_{2i}\gamma_2 + \mu_i + \epsilon_{it} \\ i &= 1, ..., N \quad t = 1, ..., T \end{aligned}$$

The major problem of this approach is that an a priori specification is required. Which regressors correlate with the individual term and which are uncorrelated. Generally, this is difficult to decide. Furthermore, the estimator reacts sensitively to small changes of the specification.

A **three-steps least squares method with an estimated individual effect (3SLSwEIE)** is an alternative. The model contains time-invariant regressors (z_i) and an unobserved individual term (μ_i)

$$y_{it} = z_i\gamma + x'_{it}\beta + \mu_i + u_{it}.$$

In the **first step** μ_i is determined by the within estimator of a FEM

$$\hat{\mu}_i = (\bar{y}_i - \bar{y}) - (\bar{x}^*_i - \bar{x}^*)'\hat{\beta}^*,$$

where β^* denotes the coefficient vector of the regressors without the intercept. $\hat{\mu}_i$ contains not only the individual effect but also the effect of time-invariant regressors z. Therefore, we regress $\hat{\mu}_i$ on z_i in the **second step**

$$\hat{\mu}_i = z'_i\delta + \omega_i$$

and calculate $\hat{\mu}_i - z'_i\hat{\delta} =: \hat{\mu}_{adj;i}$.

In the **third step** μ_i is substituted by $\hat{\mu}_{adj;i}$. The latter term is incorporated as an artificial regressor

$$y_{it} = z_i\gamma + x'_{it}\beta + \kappa\hat{\mu}_{adj;i} + u_{it}.$$

A pooled estimator can be applied as the estimated individual effect is explicitly modeled as regressor. The estimation of κ should be one. In contrast to the fixed effects estimator the pooled estimator of the modified model can also determine the effects of time-invariant regressors. As the regressor $\hat{\mu}_{adj;i}$ is an estimated variable and not directly observable, a bootstrap estimator of the standard error is preferable to the conventional analytical determination.

2.6 Treatment evaluation

Objective is to determine causal effects of economic measures. The simplest form to measure the effect is to estimate α in the linear model

$$y = X\beta + \alpha D + u,$$

where D is the intervention variable and measured by a dummy: 1 if an individual or an establishment is assigned to treatment; 0 otherwise. Typically, this is not the causal effect. An important reason is that unobserved variables influence y and D.

A wide range of methods was developed to determine the "correct" causal effect. Which approach should be preferred depends on the data, the behavior of the economic agents and the assumptions of the model. The major difficulty is that we have to compare an observed with an unobserved situation. Depending on the available information the latter is estimated. We have to ask what would occur if not $D = 1$ but $D = 0$ (treatment on the treated). This counterfactual is unknown and has to be estimated. Inversely, if $D = 0$ is observable we can search for the potential result under $D = 1$ (treatment on the untreated). A further problem is the fixing of the control group. What is the meaning of "otherwise" in the definition of D? Or in other words: What is the causal effect of an unobserved situation? Should we determine the average causal effect or only that of a subgroup?

Neither a before-after comparison $(\bar{y}_1|D = 1) - (\bar{y}_0|D = 1)$ nor a comparison of $(\bar{y}_t|D = 1)$ and $(\bar{y}_t|D = 0)$ in cross-section is usually appropriate. **Difference-in-differences estimators** (DiD) are very popular in applications

$$\bar{\Delta}_1 - \bar{\Delta}_0 = [(\bar{y}_1|D = 1) - (\bar{y}_1|D = 0)] - [(\bar{y}_0|D = 1) - (\bar{y}_0|D = 0)].$$

Practically, the effect can be determined in the following model

$$y = a_1 + b_1 T + b_2 D + b_3 TD + u,$$

where $T = 1$ means period 1 follows the period of the measure ($D = 1$). $T = 0$ is a period before the measure takes place. $\hat{b}_3$ is the causal effect. The equation can be extended by further regressors X. A conditional DiD estimator follows. If the dependent variable is a dummy a nonlinear estimator has to be applied. Suggestions are presented by Ai/Norton (2003) and Puhani (2012).

Matching procedures were developed with the objective to find a control group that is very similar to the treatment group. Parametric and non-parametric procedures can be employed to determine the control group. Kernel, inverse probability, radius matching, local linear regression, spline smoothing or trimming estimators are possible. Mahalanobis metric matching with and without propensity scores and nearest neighbor matching are two typical procedures - see e.g. Guo/Fraser (2010). The Mahalanobis distance is defined by

$$(u-v)'S^{-1}(u-v),$$

where u (v) are the values of matching variables of participants (non-participants) and S is the empirical covariance matrix determined with all observations. The distance between propensity score (ps) of two observations (i,j) in the neighborhood is measured by

$$||ps_i - ps_j||.$$

An observed or artificial statistical twin can be determined to each participant. The probability of all non-participants to participate on the measure is calculated based on probit estimates (**propensity score**). The statistical twin of a participant is those who has a propensity score (P_j) which is nearest to that of the participant. The absolute distance between i and j may not exceed a given value ϵ

$$||P_i - P_j|| < \epsilon,$$

where ϵ is a predetermined tolerance (caliper). A quarter of a standard deviation of the sample estimated propensity scores is suggested as the caliper size (Rosenbaum/Rubin 1985).

If the interest is to detect whether and in which amount the effects of intervention variables differ between the percentiles of the distribution of the objective variable

y quantile regression analysis is an appropriate instrument. The objective is to determine **quantile treatment effects** (QTE). The distribution effect of a measure can be estimated by the difference Δ between the effect on y with (y_1) and without (y_0) the measure (D=1; D=0) separate for specific quantiles Q^τ where $0 < \tau < 1$

$$\Delta^\tau = Q^\tau_{y^1} - Q^\tau_{y^0}.$$

The empirical distribution function of an observed situation and that of the counterfactual is identified. From the view of modeling four major cases were developed in the literature that differ in the assumptions. The measure is assumed exogenous or endogenous and the effect on y is unconditional or conditional.

	unconditional	conditional
exogenous	(1) Firpo(2007)	(2) Koenker/Bassett(1978)
endogenous	(3) Frölich/Melly(2012)	(4) Abadie et al.(2002)

In case (1) the quantile treatment effect $Q^\tau_{y^1} - Q^\tau_{y^0}$ is estimated by

$$Q^\tau_{y^j} = arg \min_{\alpha_0;\alpha_1} \sum_{i=1}^{n} w_{i,j} \cdot \rho_\tau(y_i - q_j)$$

where $j = 0; 1$, $q_j = \alpha_0 + \alpha_1(D|D = j)$, $\rho_\tau = a(\tau - \mathbf{1}(a \leq 0))$ is a check function; a is a real number. The weights are

$$w_{i,0} = \frac{1 - D_i}{n \cdot (1 - p(X_i))}; \qquad w_{i,1} = \frac{D_i}{n \cdot p(X_i)}.$$

The estimation is characterized by two stages. First, the propensity score is determined by a large number of regressors X via a nonparametric method - $\hat{p}(X)$. Second, in $Q^\tau_{y^j}$ the probability $p(X)$ is substituted by $\hat{p}(X)$.

Case (2) follows Koenker/Bassett(1978).

$$\sum_{(i|y_i \geq x_i'\beta)=1}^{N_1} \tau \cdot |y_i - x_i'\beta| + \sum_{(i|y_i < x_i\beta)=N_1+1}^{N} (1 - \tau) \cdot |y_i - \alpha(D_i|D_i = j) - x_i\beta|$$

has to be minimized with respect to α and β, where τ is given. In other words,

$$Q^{\tau}_{y^j} = arg \min_{\alpha;\beta} \sum_{i=1}^{N} w_{i,j} \cdot \rho_\tau(y_i - q_j),$$

where $j = 0; 1$, $q_j = \alpha(D|D = j) + x'\beta$.

The method of case (3) is developed by Frölich/Melly (2012). Due to the endogeneity of the intervention variable D, an instrumental variables estimator is used with only one instrument Z and this is a dummy. The quantiles follow from

$$Q^{\tau}_{y^j|c} = arg \min_{\alpha_0;\alpha_1} E[\rho_\tau(y - q_j) \cdot (W|D = j)],$$

where $j = 0; 1$, $q_j = \alpha_0 + \alpha_1(D|D = j)$, c means complier. The conditional weights are

$$W = \frac{Z - p(X)}{p(X)(1 - p(X))}(2D - 1).$$

Abadie et al. (2002) investigate case (4) and suggest a weighted linear quantile regression. The estimator is

$$Q^{\tau}_{y^j} = arg \min_{\alpha,\beta} E[w_{i,j} \cdot \rho_\tau(y_i - \alpha D - x'\beta)],$$

where the weights are

$$W = 1 - \frac{D(1 - Z)}{1 - p(X)} - \frac{(1 - D)Z}{p(X)}.$$

If the endogenous variable is censored Powell (2010) has developed an unconditional quantile treatment effects estimator in the presence of covariates.

Regression discontinuity (RD) design allows to determine treatment effects in a special situation. This approach uses information on institutional and legal regulations that are responsible that changes occur in the effects of economic measures. Thresholds are estimated which indicate discontinuity of the effects. Two forms are distinguished: sharp and fuzzy RD. Either the change of the status is exactly effective at a fixed point or it is assumed that the probability of a treatment change or the mean of a treatment change is discontinuous.

In the case of **sharp RD** individuals or establishments ($i = 1, \dots n$) are assigned on the base of the observed variable S to the treatment or the control group. If variable S_i is not smaller than a fixed bound $\bar{S}$ then i belongs to the treatment group ($D = 1$)

$$D_i = 1[S_i \geq \bar{S}].$$

In a simple regression model $y = \beta_0 + \beta_1 D + u$ the OLS estimator of β_1 would be inconsistent when D and u correlate. If, however, the conditional mean $E(u|S, D) = E(u|S) = f(S)$ is additionally incorporated in the outcome equation ($y = \beta_0 + \beta_1 D + f(S) + \epsilon$, where $\epsilon = y - E(y|D, S)$) the OLS estimator of β_1 is consistent. Assume $f(S) = \beta_2 S$, the estimator of β_1 corresponds to the difference of the two estimated intercepts of the parallel regressions

$$\begin{aligned} \hat{y}_0 &= \hat{E}(y|D=0) = \hat{\beta}_0 + \hat{\beta}_2 S \\ \hat{y}_1 &= \hat{E}(y|D=1) = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 S. \end{aligned}$$

The strong RD approach identifies the causal effect by distinguishing between the nonlinear function due to the discontinuous character and the smoothed linear function. If, however, a nonlinear function of the general type $f(S)$ is given, modifications have to be regarded.

Assume, the true function $f(S)$ is a polynomial of p-th order

$$y_i = \beta_0 + \beta_1 D_i + \beta_{21} S_i + \beta_{22} S_i^2 + \cdots + \beta_{2p} S_i^p + u_i$$

but a linear model is estimated, then the difference between the two intercepts, interpreted as the causal effect, is biased. What looks like a jump is in reality a neglected nonlinear effect.

Another strategy is to determine the treatment effect exactly at the fixed discontinuity point $\bar{S}$ assuming a local linear regression. Two linear regressions are considered

$$\begin{aligned} y_0 - E(y_0|S = \bar{S}) &= \delta_0(S - \bar{S}) + u_0 \\ y_1 - E(y_1|S = \bar{S}) &= \delta_1(S - \bar{S}) + u_1, \end{aligned}$$

where $y_j = E(y|D = j)$ and $j = 0; 1$. In combination with

$$y = (1 - D)y_0 + Dy_1$$

follows

$$\begin{aligned} y &= (1 - D)(E(y_0|S = \bar{S}) + \delta_0(S - \bar{S}) + u_0) \\ &\quad + D(E(y_1|S = \bar{S}) + \delta_1(S - \bar{S}) + u_1). \end{aligned}$$

The linear regression

$$y = \gamma_0 + \gamma_1 D + \gamma_2(S - \bar{S}) + \gamma_3 D(S - \bar{S}) + \tilde{u}$$

can be estimated, where $\tilde{u} = u_0 + D(u_1 - u_0)$. This looks like the DiD estimator but now γ_1 and not γ_3 is of interest. The former coefficient is a global estimation and not a localized average treatment effect. The latter follows if a small interval around $\bar{S}$ is modeled, i.e. $\bar{S} - \Delta S < S_i < \bar{S} + \Delta S$. The treatment effect corresponds to the difference of the two former determined intercepts, restricted to $\bar{S} < S_i < \bar{S} + \Delta S$ on the one hand and to $\bar{S} - \Delta S < S_i < \bar{S}$ on the other hand.

The **fuzzy RD** assumes that the propensity score function of treatment $P(D = 1|S)$ is discontinuous with a jump in $\bar{S}$

$$P(D_i = 1|S_i) = \begin{cases} g_1(S_i) \text{ if } S_i \geq \bar{S} \\ g_0(S_i) \text{ if } S_i < \bar{S}, \end{cases}$$

where it is assumed that $g_1(\bar{S}) > g_0(\bar{S})$. Therefore, treatment in $S_i \geq \bar{S}$ is more likely. In principle the functions $g_1(S_i)$ and $g_0(S_i)$ are arbitrary, e.g. a polynomial of p-th order can be assumed but the values have to be within the interval [0;1] and different values in $\bar{S}$ are necessary.

The conditional mean of D that depends on S is

$$E(D_i|S_i) = P(D_i = 1|S_i) = g_0(S_i) + (g_1(S_i) - g_0(S_i))T_i,$$

where $T_i = 1(S_i \geq \bar{S})$ is a dummy indicating the point where the mean is discontinuously. If a polynomial of p-th order is assumed the interaction variables

$S_iT_i, S_i^2T_i \cdots S_i^pT_i$ and the dummy T_i are instruments of D_i. The simplest case is to use only T_i as instrument if $g_1(S_i)$ and $g_0(S_i)$ are discriminable constants.

We can determine the treatment effect around $\bar{S}$

$$\lim_{\Delta \to 0} \frac{E(y_i|\bar{S} < S_i < \bar{S} + \Delta) - E(y_i|\bar{S} - \Delta < S_i < \bar{S})}{E(D_i|\bar{S} < S_i < \bar{S} + \Delta) - E(D_i|\bar{S} - \Delta < S_i < \bar{S})}.$$

The empirical analogon is a conditional Wald estimator.

3 Applications: some new estimates of production functions

In the following some estimates of production functions are presented where IAB establishment panel data are used. The empirical analysis is restricted to the period 2006-2010. Methods of section 2 are applied. The results can be found in Table 1-13. Table 1 focus on alternative estimates of standard errors - see section 2.1.1 and 2.1.2. The estimated coefficients in column 1-3 and 5 are identical. Estimates with $hc2$ and $hc4$ - not presented in the Tables - deviate only slightly from those with $hc1$. The jackknife estimates of standard errors and t-values are also not so far away from the heteroskedasticity-consistent estimates with $hc1$ and $hc3$. The nearness to estimates with $hc3$ is plausible because the latter is only a slightly simplified version of what one gets by employing the jackknife technique. Furthermore, Table 1 demonstrates that bootstrap and cluster-robust estimates of the t-values differ strongest of the input factor labor (lnL). Capital (lnK), measured by the sum of investments of the last four years, has evidently lower cluster-robust estimates of standard errors than that from the other methods.

Table 1: Alternative determination of standard errors using hc1, hc3, bootstrap, jackknife estimates and cluster-robust estimates

Variable	hc1	hc3	boot-strap	jack-knife	cluster (idnum)
lnL	0.9472	0.9472	0.9472	0.9582	0.9472
	(184.02)	(183.99)	(227.40)	(184.49)	(126.29)
lnK	0.2225	0.2225	0.2225	0.2178	0.2225
	(60.80)	(60.79)	(60.40)	(59.58)	(43.04)
const	9.0810	9.0810	9.0810	9.0908	9.0810
	(307.86)	(307.81)	(271.82)	(308.83)	(215.20)

Note: N=34308; t-ratios in parentheses; idnum - individual identification number

An extended version of column 1 in Table 1 is presented in Table 2. The latter estimates show smaller coefficients and smaller t-values of the input factors labor and capital. However, the major intention of Table 2 is to demonstrate that also in this example there is a clear relationship between the mean of dummies ($\bar{D}$) as independent variables and the estimated standard errors as maintained in section 2.1.4. The nearer $\bar{D}$ to 0.5 the smaller is the standard error. A caveat seems necessary. The result in Table 2 in contrast to that in Table A1 cannot be generalized because the standard error of a dummy is not only determined by the mean. The residual variance that depends on the importance of all regressors is also relevant.

Table 2: OLS estimates with Bernoulli distributed regressors

lnY	mean	coef.	std. Err.	t
lnL		0.8808	0.0061	144.33
lnK		0.2049	0.0041	49.55
CLP	0.0871	0.0307	0.0236	1.30
WOCO	0.3035	0.3915	0.0184	21.19
BARGAIN	0.3819	0.1385	0.0133	10.36
P1	0.0834	0.2462	0.0231	10.65
P2	0.3695	0.1032	0.0132	7.78
const		9.2905	0.0367	253.03

Note: The regressors CLP (company-level pact), WOCO (works council), BARGAIN (industry-wide agreement), P1 (profits last year: very good) and P2 (profits last year: good) are dummies.

Outliers may have strong effects on coefficient and standard error estimates. However, estimates do not react sensitively to all outliers. This can be demonstrated if the results with and without outliers are compared. Table 3 presents an example for simple Cobb-Douglas functions in column 1 and 2. The coefficients are very similar while the differences of the standard errors are more evident. The picture is clearer if only the observations with high leverage are eliminated - see column 3. Coefficients and standard errors in column 1 and 3 reveal a clear disparity for both input factors. This result is not unexpected. The consequence is not unambiguous. Is column 1 or 3 preferable? If all observations with strong leverages are due to measurement errors the decision speaks in favor of the estimates in column 3. As no information is available to this question both estimates may be useful. Column 4 extends the consideration to outliers following Hadi (1992). The squared difference between individual regressor values and the mean for all regressors - here lnL and lnK - is determined for each observation weighted by the estimated covariance matrix. The decision whether establishment i is an outlier is now based on the Mahalanobis distance. The multiple outlier dummy ($MOD_i = 1$ if i is an outlier; =0 otherwise) is incorporated as an additional regressor. The estimates show that all outliers together have a significant effect on the output variable lnY.

Table 3: Cobb-Douglas production functions with and without outliers, t-values in parentheses
Dependent variable: logarithm of sales - lnY

Variable	with outliers	without outliers	without strong leverages	with Hadi-MOD
lnL	0.9472	0.9415	1.0409	0.9412
	(222.12)	(240.28)	(169.10)	(240.10)
lnK	0.2225	0.2242	0.1724	0.2243
	(70.11)	(77.04)	(36.33)	(77.08)
MOD				1.8810
				(2.33)
const	9.0811	9.0498	9.3445	9.0490
	(333.20)	(362.66)	(238.53)	(362.62)
N	34 308	33 851	27 262	34308
R^2	0.866	0.866	0.805	0.843

As it is not obvious whether the outliers are due to measurement errors that should be eliminated or whether these are unusual but substantially induced observations that should be accounted for, only partially identified parameters are possible. Therefore in Table 4 confidence intervals are not only presented for the two extreme cases (column 1: all outliers are induced by specific events ; column 2: all outliers are due to measurement errors). Additionally, in column 3 the confidence interval based on Stoye's method is displayed. The results show that the lower and upper coefficient estimates of lnL by Stoye lies within the interval of the estimated coefficients in column 1 and 2. The upper coefficient is nearer to that of column 2 and the lower is nearer to that of column 1. This means the Stoye interval is shorter or in other words more precise. We do not find the same pattern for input factor lnK. In the case $\hat{\beta}_{lnK;u}$ the Stoye coefficient deviates more from column 2 than from column 1. And for $\hat{\beta}_{lnK;l}$ we find the opposite result. A clear interpretation is not possible. If the upper and the lower Stoye coefficient would be nearer to that of the interval without outliers we could say that the majority of the outliers is induced by measurement errors.

Table 4: Confidence intervals (CI) of output elasticities of labor and capital based on a Cobb-Douglas production function, estimated with and without outliers, Stoye confidence interval at partially identified parameters
Dependent variable: logarithm of sales - lnY

Coefficient	CI with outliers	CI without outliers	Stoye CI
$\hat{\beta}_{lnL;u}$	0.9555	0.9492	0.9511
$\hat{\beta}_{lnL;l}$	0.9388	0.9339	0.9376
$\hat{\beta}_{lnK;u}$	0.2287	0.2299	0.2282
$\hat{\beta}_{lnK;l}$	0.2162	0.2185	0.2184

Table 5 presents nonlinear estimates of three different production function - Cobb-Douglas, CES and Translog function. The results due to t-values and F-tests speak in favor of the Cobb-Douglas function.

Table 5: Estimates of Cobb-Douglas, CES and Translog functions; t-values in parentheses, dependent variable: logarithm of sales - lnY

Variable	CD	CES	TRANSLOG
lnL	0.9582	0.9758	0.9699
	(220.04)	(45.82)	(73.29)
lnK	0.2178	0.2000	0.2217
	(68.30)	(9.38)	(42.23)
$(\text{lnK-lnL})^2$		0.0011	
		(0.84)	
lnK·lnL			-0.0010
			(-0.94)
const	9.0908	9.1590	9.0477
	(332.89)	(144.49)	(169.07)
N	33 860	33 860	33 860
F test		0.3991	0.3491

In contrast to the estimates of Table 5 in the next Table it is assumed that the input factor labor or capital is not exogenous and IV estimates are employed. Table 6 presents four specifications. In column 1 and 3 lnL is instrumented. In the former case WOCO, FLEXTIME and SHORT TIME dummies are used as instruments, while in the latter CLP and FLEXTIME are the instruments. Column 2 and 4 are analogously modeled for lnK. For all four specifications the hypotheses of weak instruments and exogeneity have to be rejected. The latter result signals that 2SLS estimates should be preferred in comparison with OLS estimates. The F test statistics exceed the critical values of the Stock-Yogo test and the partial determination coefficients are high enough. If we compare the results in column 1 of Table 5 with those of column 1 and 3 in Table 6 it is obvious that the partial productivity elasticity of labor in the latter is larger than in the former or more precisely, we detect a transition from decreasing to increasing returns to scale. The lnK effect is decreasing. This means the partial productivity elasticity of capital is furthermore lower than one. If capital is assumed endogenous - see column 2 and 4 in Table 6 - the partial productivity elasticity of labor and capital are still smaller than one. The tendency for lnL and lnK go in the opposite direction when IV estimates instead of OLS estimates are employed. The influence of lnK is increasing while that of lnL is decreasing.

If lnL and lnK are simultaneously endogenized the results - not presented in the tables - are implausible - negative partial production elasticity of capital. Furthermore the hypothesis of weak instruments cannot be rejected. The test of overidentifying restrictions (OIR) suggests that CLP and FLEXTIME are better instruments for lnL than WOCO, FLEXTIME and SHORT TIME but also the former are not optimal. We cast doubt on the suitability of the instrument set. In the case of lnK the OIR test does not reject the null hypothesis. This speaks in favor of the instruments WOCO, FLEXTIME and SHORT TIME. Nevertheless, we cannot be sure that the instruments are valid.

Table 6: 2SLS Estimates of CD functions
Dependent variable: logarithm of sales - lnY

Variable	IV1	IV2	IV3	IV4
lnL	1.2011	0.3120	1.1545	0.4219
	(95.31)	(7.52)	(59.23)	(6.76)
lnK	0.0784	0.8474	0.1058	0.7338
	(10.27)	(21.00)	(9.35)	(12.12)
const	9.8564	4.1431	9.7080	5.0555
	(197.35)	(12.98)	(145.99)	(10.58)
N	20 099	20 099	20 457	20 457
R^2	0.8271	0.6600	0.8296	0.7213
weak IV-F test	1825.90	155.21	985.97	84.27
p–value	0.0000	0.0000	0.0000	0.0000
exogeneity test	546.31	553.47	132.85	126.40
p-value	0.0000	0.0000	0.0000	0.0000
OIR test	6.58	0.20	2.93	5.44
p-value	0.0372	0.9036	0.0869	0.0197

Notes: t-ratios in parentheses; instrumented regressors - IV1=lnL, IV2=lnK, IV3=lnL and IV4=lnK; instruments in column 1 and 2 are WOCO (works council), FLEXTIME (flextime wage record) and SHORT TIME, while in column 3 and 4 CLP (company-level pact) and FLEXTIME are the instruments.

The next four estimates in Table 7 are focussed on panel models with a time-invariant regressor. During the period 2006-2010 no establishment changes the localization in Mecklenburg-West Pomerania ($M - WP$). Besides Saarland (S)

this is the only German Land where over the five years no move of a firm is observed in the IAB Establishment Panel. In Table 7 only the $M - WP$ case is presented, however, estimates with the S dummy leads to similar results. In column 1 and 2 the conventional panel estimates are displayed. The random effects estimates in column 1 are inconsistent as the Breusch-Pagan test (BP) reveals that individual effects exist and the Hausman test rejects the null hypothesis that the individual effect is uncorrelated with the regressors. The fixed effects estimates in column 2 do not allow to determine the M-WP effect on lnY. This is possible with the Hausman-Taylor approach (HT) in column 3. A priori it is fixed that lnL is endogenous and the dummy $M - WP$ is the time-invariant. One should expect that the coefficients of lnL and lnK, respectively, are similar in column 2 and 3. This is not the case but the analogous estimates in column 4 and 2 are identical. Furthermore, $\hat{\beta}_{EIE}$=1 speaks in favor of this approach.

Table 7: CD Panel Estimates
Dependent variable: logarithm of sales - lnY

Variable	RE	FE	HT	3SLSwEIE
lnL	1.0333	0.4096	0.7079	0.4096
	(221.02)	(35.84)	(85.95)	(330.46)
lnK	0.0576	0.0195	0.0285	0.0195
	(28.76)	(9.72)	(16.28)	(24.84)
M-WP	-0.1867	0	-0.3924	-0.3927
	(-5.15)	(omitted)	(-6.09)	(-69.64)
EIE				1.0000
				(818.90)
const	10.6356	13.2449	11.9937	13.2449
	(463.91)	(302.82)	(362.71)	(1727.73)
N	34 308	34 308	34 308	34 308
BP test	20 138			
Hausman test		18 885		

The next tables present estimates of alternative methods in order to determine causal effects. First, the difference-in-differences (DiD) approach is estimated. Results can be found in Table 8. The coefficient of the interaction variable $CLP * D2009$ in column 1 is significantly different from zero. This means that sales between firms with a company-level pact (CLP), adopted in 2009, and those without such a pact differ between 2009 and the years before (2006-2008). The adoption of a CLP in the year of the Great Recession is combined with lower sales than in the years before if an unconditional DiD specification is used. In column 2 the sign changes and the effect of the interaction variable is insignificant if an extended CDF is estimated. This approach is preferred because in the former the influence of the input factors is partially added to the causal effect. Insofar, we cannot detect any influence of the adoption of a CLP on sales in 2009. One could argue that the estimates in column 1 lead more than that in column 2 to significant results because the sample in the former is smaller. This argument is not compelling. If we draw a random sample of 63.83 percent so that in column 1 the sample size is n=20,489 the interaction effect is -0.2939 and the significance is preserved (t=-2.26).

Table 8: DiD estimates of CDF with company-level pact (CLP) effects. Dependent variable: logarithm of sales - lnY

	unconditional	conditional
lnL		0.9423
		(166.03)
lnK		0.2211
		(53.37)
CLP	3.1152	0.0951
	(35.91)	(2.36)
D2009	0.0597	0.0216
	(2.25)	(1.54)
CLP*D2009	-0.3029	0.0400
	(-2.90)	(0.84)
n	31,985	20,490
R^2	0.101	0.841

Note: t-values in parentheses

Alternative methods to determine causal effects are matching procedures. These are suggested when there does not exist control over the assignment of treatment conditions, when in the basic equation $y = X\beta + \alpha D + u$ the dichotomous treatment variable D and the disturbance term u correlate, when the ignorable treatment assignment assumption is violated. In the example of the CDF it is questioned that this condition is fulfilled for CLPs. This means that the OLS estimates in column 1 of Table 9 are biased and inconsistent. As an alternative the Mahalanobis metric matching (MM) without propensity score and in column 3 the nearest neighbor matching (NNM) with caliper are applied, presented in column 2 and 3, respectively. In the latter method non-replacement is used. That is, once a treated case is matched to a non-treated case, both cases are removed from the pool. The former method allows that one control case can be used as a match for several treated cases. Therefore, the total number of observations in the nearest neighbor is larger than that in column 2. We find that the CLP effect on sales is insignificant in both cases but the CLP coefficient of MM estimates exceeds by far that of NNM. The estimates of the partial elasticities of production are very similar in the three estimates in Table 9. The insignificance of the CLP effect confirms the result of column 2 of Table 8. If the DiD estimator of column 2 in Table 8 is applied after matching the causal effect is - not unexpected - also insignificant. The probvalue is 0.182 if the MM procedure is used and 0.999 under the NNM procedure.

Table 9: Estimates of CDF with CLP effects using matching procedures; dependent variable: logarithm of sales - lnY

	no matching	MM	NNM
lnL	0.9420	0.9362	0.9533
	(166.03)	(47.75)	(63.32)
lnK	0.2212	0.1938	0.2007
	(53.42)	(15.12)	(19.70)
CLP	0.1231	0.1928	0.0496
	(5.22)	(1.31)	(1.46)
n	20,490	1,806	3,346
R^2	0.840	0.838	0.849

Note: MM - Mahalanobis metric matching, NNM - nearest neighbor matching, control variables are lnL and lnK, t-values in parentheses.

The previous estimates have demonstrated that company-level pacts (CLP) have no statistically significant influence on output, on sales. We cannot be sure that this result is also true for subgroups of firms. One way to test this is to conduct quantile estimates. As presented in section 2.2 four methods can be applied to determine quantile treatment effects (QTE). The CLP effects on sales can be found in Table 10 where the results of five quantiles (q=0.1,0.3, 0.5, 0.7, 0.9) are presented. In contrast to the previous estimations most CLP effects are significant in the columns 1-4 of Table 10. Firpo considers the simplest case without control variables under the assumption that the adoption of a company-level pact is exogenous. The estimated coefficients in column 1 (F) seem oversized. The same follows from the Frölich-Melly approach, where CLP is instrumented by a short work time dummy (column 3 - F-M). Other available instruments like opening clauses, collective bargaining, works councils or research and development within the firm do not evidently change the results. One reason for the overestimated coefficients can be neglected determinants of the output that correlate with CLP. Estimates of column 2 (K-B) and 4 (A-A-I) support this hypothesis. From the view of expected CLP coefficients the conventional quantile estimator, the Koenker-Bassett approach, with lnL and lnK as regressors seems best. However, the ranking of the size of the coefficients within column 2 is implausible. The smaller the quantile the larger is the estimated coefficient. This would mean that CLPs are advantageous for small in comparison to large firms but large firms have a higher propensity to adopt a company-level pacts. Such a behavior is difficult to understand. However, it is possible that small firms with advantages in productivity due to CLPs have relative high costs to adopt a CLP. In this case the higher propensity of large firms to introduce a CLP is consistent with higher productivity of small firms. The coefficients of the Abadie-Angrist-Imbens approach, a combination of Frölich-Melly's and Koenker-Bassett's model, are also large but not so large as in column 1 and 3.

Possibly, all estimates in column 1-4 of Table 10 are biased and inconsistent. This is the case when CLP and non-CLP firms fundamentally differ due to unobserved variables. To avoid this problem the QTE and the matching approaches are combined. Based on the matching of Table 9 the QTE analogously to column 1-4 in Table 10 can be estimated. In column 5 and 6 only two combinations are presented, namely MM+K-B and MM+A-A-I. We find that the ranking and the size of the coefficients are plausible in column 5. The sizes of the coefficients in column 6 are smaller than in column 4 but the identified causal effects seems still too high. The most important result is the following: the CLP effects are significant for higher quantiles, i.e. for q=0.9 in column 5 and for q=0.7 and q=0.9 in column 6. However, the median estimators (q=0.5) of CLP effects in column 5 and 6 that can be compared with the estimates of column 2 in Table 9 are insignificant. Quantile estimators highlight information that cannot be revealed by other treatment methods, i.e. in Table 8 and 9. The estimations of the other six combinations (MM+F, MM+F-M, NNM+F, NNM+K-B, NNM+F-M, NNM+A-A-I) - not presented in the tables - are less plausible. The ranking of the size of coefficients is inconsistent in the light of theoretical and practical experience.

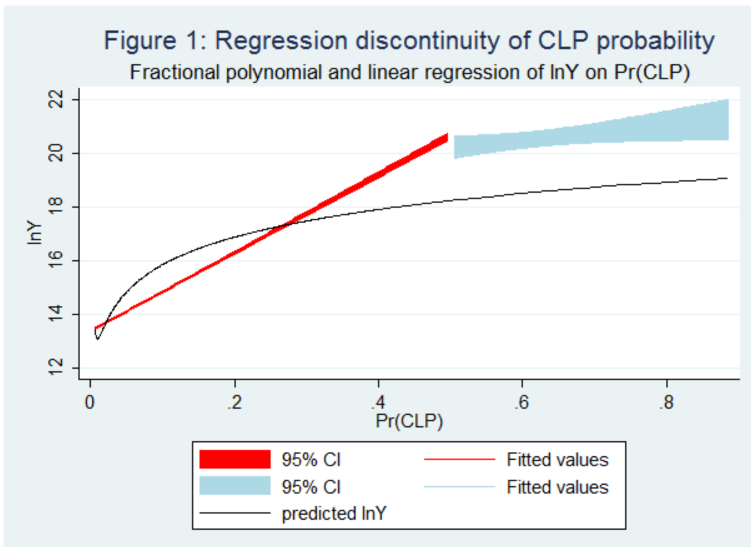
Table 10: Quantile estimates of CLP effects ; dependent variable: logarithm of sales - lnY

Quantile	F	K-B	F-M	A-A-I	MM+K-B	MM+A-A-I
q=0.1	2.9957	0.2236	5.3012	1.2092	-0.1064	0.9776
	(38.94)	(6.76)	(20.42)	(3.10)	(-0.87)	(1.06)
q=0.3	3.3242	0.1836	5.8227	1.1615	0.0715	0.7140
	(54.67)	(7.15)	(23.67)	(3.11)	(0.46)	(0.62)
q=0.5	3.1325	0.1526	6.3549	1.2000	0.1793	0.6736
	(54.19)	(6.31)	(24.58)	(2.57)	(1.09)	(1.37)
q=0.7	2.9312	0.1036	6.8703	1.2479	0.2270	0.8072
	(56.91)	(4.07)	(26.14)	(2.09)	(1.54)	(2.18)
q=0.9	2.3203	-0.0176	7.8119	1.6549	0.4523	1.4242
	(34.18)	(-0.37)	(20.12)	(1.36)	(3.36)	(2.92)
n	31,985	20,490	20,909	13,496	1,806	1,206

Note: F - Firpo; K-B - Koenker/Bassett; F-M - Frölich/Melly; A-A-I - Abadie/Angrist/Imbens, MM - Mahalanobis metric matching, control variables are lnL and lnK, t-values in parentheses.

The final discussed treatment method in section 2.2 is the regression discontinuity (RD) design. This approach exploits information of the rules determining treatment. The probability of receiving a treatment is a discontinuous function of one or more variables where treatment is triggered by an administrative definition or an organizational rule.

In a first example using a sharp RD design it is analyzed whether at an estimated probability of 0.5 that a CLP exists a structural break on logarithm of sales (lnY) is evident. For this purpose a probit model is estimated with profit situation, working-time account, total wages per year and works council as determinants. All coefficients are significantly different from zero - not in the tables. The estimated probability $Pr(CLP)$ is then plotted against lnY based on a fractional polynomial model over the entire range ($0 < Pr(CLP) < 1$) and on two linear models split into $Pr(CLP) <= 0.5$ and $Pr(CLP) > 0.5$. The graphs are presented in Figure 1.



A structural break seems evident. Two problems have to be checked: First, is the break due to a nonlinear shape, and second, is the break significant? The answer to the first question is yes, because the shape over the range $0 < Pr(CLP) < 1$ is obviously nonlinear when a fractional polynomial is assumed. The answer to the second question is given by a t-test - cf. section 2.2 - based on

$$\begin{aligned} y &= \gamma_0 + \gamma_1 D\_Pr(CLP) + \gamma_2 (Pr(CLP) - \overline{Pr(CLP)}) \\ &\quad + \gamma_3 D\_Pr(CLP) \cdot (Pr(CLP) - \overline{Pr(CLP)}) + u \\ &=: \gamma_0 + \gamma_1 D\_Pr(CLP) + \gamma_2 cPr(CLP) + \gamma_3 D\_Pr(CLP) \cdot cPr(CLP) + u, \end{aligned}$$

where

$$D\_Pr(CLP) = \begin{cases} 1 & \text{if Pr(CLP)} \leq 0.5 \\ 0 & \text{otherwise.} \end{cases}$$

The null that there is no break has to be rejected (γ_1=-3.96; t=-6.87; prob-value=0.000) as can be seen in Table 11.

Table 11: Testing for structural break of CLP effects between Pr(CLP)≤0.5 and Pr(CLP)>0.5

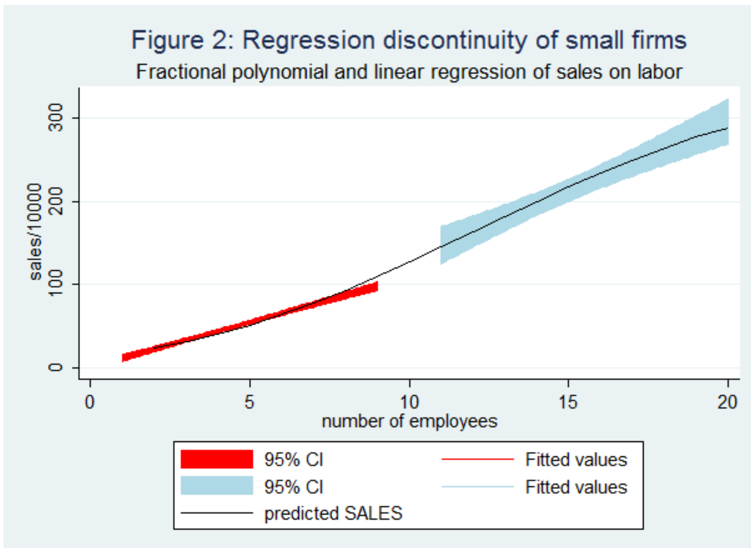
	coef.	std.err.	t	P>|t|
$D\_Pr(CLP)$	-3.9608	0.5765	-6.87	0.000
$cPr(CLP)$	4.3413	0.8390	5.17	0.000
$D\_Pr(CLP) \cdot cPr(CLP)$	11.3838	0.8437	13.49	0.000
const	18.4375	0.5764	31.99	0.000

The estimates in Table 11 cannot tell us whether the output jump in Pr(CLP)=0.5 is a general phenomenon or whether the Great Recession in 2008/09 is responsible. To test this the combined method of RD and DiD - derived in section 2.2. - is employed and the results are presented in Table 12. The estimates show that the output jump does not significantly change between 2006/2007 and 2008/2010. The influence of $D\_Pr(CLP) \cdot T$ on lnY is insignificant. Therefore, we conclude that the break is of general nature.

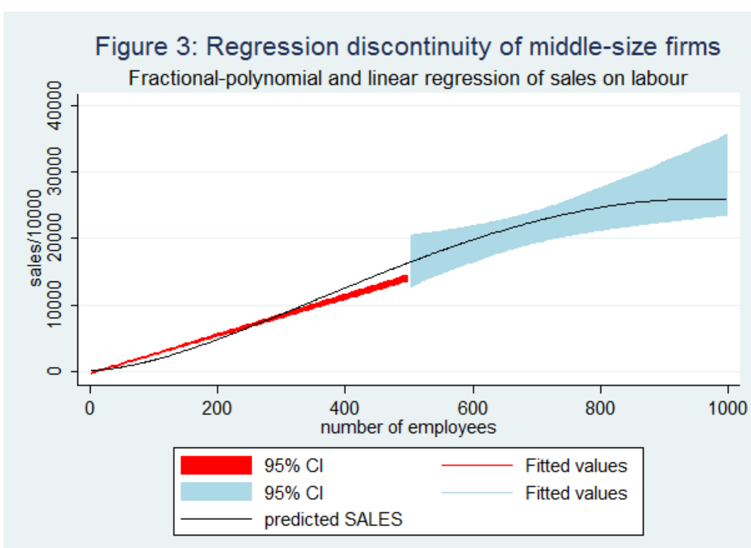
Table 12: Testing for differences in structural break of CLP effects between Pr(CLP)≤0.5 and Pr(CLP)>0.5 in 2006/07 and 2008/10

	coef.	std.err.	t	P>|t|
T	0.0130	1.3118	0.01	0.992
$D\_Pr(CLP)$	-4.1045	1.1191	-3.67	0.000
$cPr(CLP)$	3.9314	1.6795	2.34	0.019
$D\_Pr(CLP) \cdot cPr(CLP)$	11.6383	1.6884	6.89	0.000
$D\_Pr(CLP) \cdot T$	0.0392	1.3119	0.03	0.976
$cPr(CLP) \cdot T$	0.2801	1.9520	0.14	0.886
$D\_CLP \cdot cPr(CLP) \cdot T$	-0.0662	1.9623	-0.03	0.973
const	18.5422	1.1190	16.57	0.000

Two further examples are presented in Figure 2 and 3. The Institut für Mittelstandsforschung defines small firms as such that have less than 10 employees and until 1 million Euro sales per year. The analogous definition of middle-size firms is less than 500 employees and until 50 million Euro sales per year. A sharp regression discontinuity design is applied to test whether the first and the second part of the definition are consistent. In other words, based on a Cobb-Douglas production function with only one input factor, the number of employees, it is tested whether there exists a structural break for small firms between 9 and 10 employees at a 1 million sales border. We find for small firms in Figure 2 that there seems to be a sales break around 1 million Euro per year.

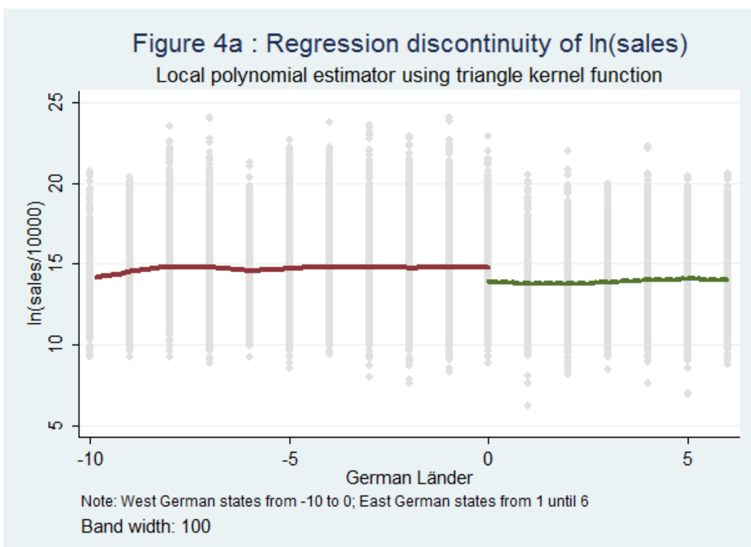


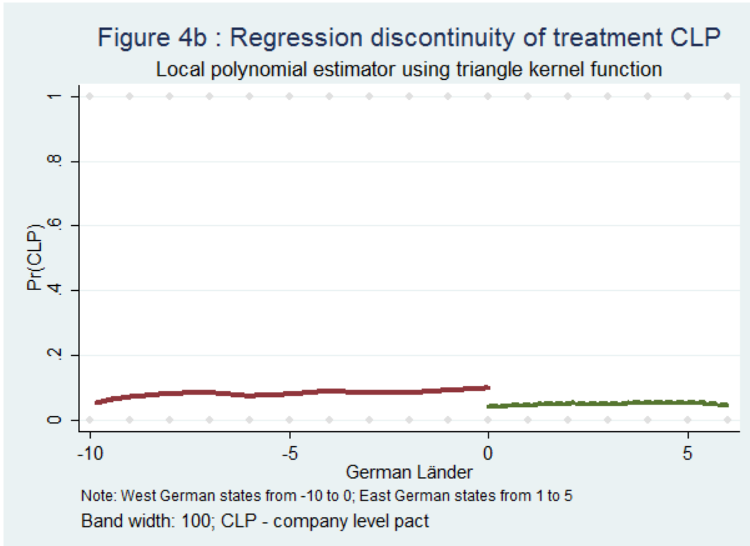
The t-test analogously to the first example yields weak significance (γ_1=-13.8667; t=-1.61; probvalue=0.107).



The same procedure for middle-size firms - see Figure 3 - leads to following results. Apparently, there exist a break but this visual result might be due to a nonlinear relationship as the fractional polynomial estimation over the entire range suggests. The t-test does not reject the null (γ_1=-8977; t=-0.54; probvalue=0.588). The conclusion from Figure 2 and 3 is that the graphical representation without the polynomial shape as comparison course and without testing on a structural break can lead to a misinterpretation.

The final example uses a fuzzy regression discontinuity design. It is analyzed whether the CLP effects on the logarithm of sales (lnY=ln(sales/10000)) differ between the East and West German federal states. The graphical representation can be found in Figure 4a and 4b. The former shows the disparities in the level of sales per year and the latter those of $Pr(CLP)$ - here measured by the relative frequency of firms with a CLP to all firms in a German federal state.





Although clear differences are detected for both characteristics ($lnY, Pr(CLP)$) we cannot be sure that these disparities are significant and whether the CLP effects are smaller or larger in West Germany. This is checked by a Wald test in Table 13. We find that the CLP effects on lnY (-0.8749/-0.0571=15.3165) are significantly higher in the West German federal states (z=4.29).

Table 13: Fuzzy regression discontinuity between East and West German federal states (GFS) - Wald test for structural break of company-level pact (CLP) effects on sales; jump at GFS>0; dependent variable: logarithm of sales - lnY

Variable	coef	std. err	z
lnY jump	-0.8749	0.1234	-7.09
CLP jump	-0.0571	0.0138	-4.13
Wald estimator	15.3165	3.5703	4.29

Note: GFS=-10 Berlin(West); -9 Schleswig-Holstein; -8 Hamburg; -7 Lower Saxony; -6 Bremen; -5 North Rhine-Westphalia; -4 Hesse; -3 Rhineland-Palatinate; -2 Baden-Württemberg; -1 Bavaria; 0 Saarland; 1 Berlin(Ost); 2 Brandenburg; 3 Mecklenburg-West Pomerania; 4 Saxony; 5 Saxony-Anhalt; 6 Thuringia.

4 Summary

Many reasons like heteroskedasticity, clustering, basic probability of qualitative regressors, outliers and only partially identified parameters may be responsible that estimated standard errors based on classical methods are biased. Applications show that the estimates under suggested modifications do not always deviate so much from that of the classical methods.

The development of new procedures is ongoing. Especially, the field of treatment methods were extended. It is not always obvious which method is preferable to determine the causal effect. As the results evidently differ it is necessary to develop a framework that helps to decide which method is most appropriated under typically situations. We observe a tendency away from the estimation of average effects. The focus is shifted to distribution topics. Quantile analysis helps to investigate differences between subgroups of the population. This is important because economic measures have not the same influence on heterogenous establishments and individuals. A combination of quantile regression with matching procedure can improve the determination of the causal effects. Further combinations of treatment methods seem helpful. Difference-in-differences estimates should be linked with matching procedures and regression discontinuity designs.

References

Abadie, A., Angrist, J. and G. Imbens (2002), 'Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings', *Econometrica* 70, 91-117.

Ai, C. and E.C. Norton (2003), 'Interaction terms in logit and probit models', *Economics Letters* 80, 123-129.

Angrist, J. and J.-S. Pischke (2009), 'Mostly Harmless Econometrics - An Empiricist's Companion', Princeton University Press.

Bauer, T. K. and M. Sinning (2008), 'An extension of the Blinder-Oaxaca decomposition to nonlinear models', *Advances in Statistical Analysis* 92, 197-206.

Belsley, D.A., Kuh, E. and R.E. Welsch (1980), 'Regression Diagnostics - Identifying Influencial Data and Sources of Collinearity', John Wiley & Sons.

Cameron, A.C. and P.K. Trivedi (2009), 'Microeconometrics Using Stata', Stata Press Publication.

Chernozhukov, V., Fernandez-Val, I., Hahn, J. and W. Newey (2012), 'Average and Quantile Effects in Nonseparable Panel Models', *mimeo*.

Chernozhukov, V., Fernandez-Val, I. and B. Melly (2007), 'Inference on Counterfactual Distributions', *mimeo*.

Chernozhukov, V., Hong, H. and E. Tamer (2007), 'Estimation and Confidence Regions for Parameter Sets in Econometric Models', *Econometrica* 75, 1243-1284.

Cribari-Neto, F. and W.B. da Silva (2011), 'A New Heteroskedasticity-consistent Covariance Matrix Estimator for the Linear Regression Model', *Advances in Statistical Analysis* 95, 129-146.

Drechsler, J. (2011), 'Multiple Imputation in Practice - A Case Study Using a Complex German Establishment Survey', *Advances in Statistical Analysis* 95, 1-26.

Firpo, S. (2007), 'Efficient Semiparametric Estimation of Quantile Treatment Effects', *Econometrica* 75, 259-276.

Fortin, N., Lemieux, T. and S. Firpo (2011), 'Decomposition Methods in Economics', in: O. Ashenfelter und D. Card (eds.), *Handbook of Labor Economics* Vol. 4A, 1-102.

Frölich, M. and B. Melly (2012), 'Unconditional quantile treatment under endogeneity', *mimeo.*

Guo, S. and M.W. Fraser (2010), Propensity Score Analysis, Sage Publications.

Imbens, G.W. and C.F. Manski (2004), 'Confidence Intervals for Partially Identified Parameters', *Econometrica* 72, 1845-1857.

Hadi, A.S. (1992), 'Identifying Multiple Outliers in Multivariate Data', *Journal of the Royal Statistical Society B* 54, 761-771.

Hausman, J.A., Newey, W.K., Woutersen, T., Chao, J.C. and N.R. Swanson (2012), 'Instrumental variable estimation with heteroskedasticity and many instruments', *Quantitative Economics* 3, 211-256.

Hausman, J.A. and W.E. Taylor (1981), 'Panel Data and Unobservable Individual Effects', *Econometrica* 49, 1377-1398.

Heckman, J.J. (1979), 'Sample Selection Bias as a Specification Error', *Econometrica* 47, 153-161.

Hsiao, C. (2003), 'Analysis of Panel Data', 2nd ed., *Cambridge University Press.*

Kennedy, P.E. (2002), 'Sinning in the basement: What are the rules? The ten commandments of applied econometrics', *Journal of Economic Surveys* 16, 569-589.

Koenker, R. (2005), 'Quantile Regression', *Cambridge University Press.*

Koenker, R. and G. Bassett (1978), 'Regression Quantiles', *Econometrica* 46, 33-50.

Krämer, W. (2011), 'The Cult of Statistical Significance - What Economists Should and Should Not Do to Make Their Data Talk', *Journal of Applied Social Science Studies* 131, 455-468 .

Leamer, E.E. (1985), 'Sensitivity Analyses Would Help', *American Economic Review* 75, 308-313.

Lechner, M. (2011), 'Matching Estimation', *mimeo.*

Li, G. (1985), 'Robust regression', in: Hoaglin, D.C., Mosteller, F. and J.W. Tukey (eds.), Exploring Data Tables, Trends, and Shapes, New York: Wiley, 281-340.

MacKinnon, J.G. and H. White (1985), 'Some Heteroskedasticity Consistent Covariance matrix Estimators with Improved Finite Sample Properties', *Journal of Econometrics* 29, 305-325.

Moulton, B.R. (1986), 'Random Group Effects and the Precision of Regression Estimates', *Journal of Econometrics* 32, 385-397.

Moulton, B.R. (1987), 'Diagnostic Tests for Group Effects in Regression Analysis', *Journal of Business and Economic Statistics* 6, 275-282.

Moulton, B.R. (1990), 'An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units', *Review of Economics and Statistics* 72, 334-338.

Powell, J.L. (1984), 'Least Absolute Deviations Estimation for the Censored Regression Model', *Journal of Econometrics* 25, 303-325.

Powell, D. (2010), 'Unconditional Quantile Regression for Panel Data with Exogenous or Endogenous Regressors', *RAND working paper WR-710-1*.

Puhani, P. (2012), 'The treatment, the cross difference, and the interaction term in nonlinear"difference-in-differences"models', *Economics Letters* 115, 85-87.

Rässler, S. and R. Riphahn (2006), 'Survey Item Nonresponse and its Treatment', *Allgemeines Statistisches Archiv* 90, 217-232.

Romano, J.P. and A.M. Shaikh (2010), 'Inference for the Identified Set in Partially Identified Econometric Models', *Econometrica* 78, 169-211.

Rosenbaum, P.R. and D.P. Rubin (1985), 'Constructing a control group using multivariate matched sampling methods that incorporate the propensity score', *American Statistician* 39, 33-38.

Schneeweiß, H. and J. Komlos (2009), 'Probabilistical Rounding and Sheppard's Correction', *Statistical Methodology* 6, 577-593.

Shea, J. (1997), 'Instrument Relevance in Multivariate Linear Models: A Simple Measure', *Review of Economics and Statistics* 79, 348-352.

Sargan, J.D. (1958), 'The Estimation of Economic Relationship Using Instrumental Variables', *Econometrica* 26, 393-415.

Stoye, J. (2009), 'More on Confidence Intervals for Partially Identified Parameters', *Econometrica* 77, 1299-1315.

White, H. (1980), 'A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity', *Econometrica* 48, 817-838.

Woutersen, T. (2009), 'A Simple Way to Calculate Confidence Intervals for Partially Identified Parameters', mimeo.

Ziliak, S. and D. McCloskey (2008), 'The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice and Lives', University of Michigan Press.

Appendix

Table A1: OLS estimates of Cobb-Douglas functions with artificial dummies (DV.) as regressor; dependent variable: logarithm of sales - lnY

	$\hat{\beta}_{lnL}$	std.err	$\hat{\beta}_{lnK}$	std.err	$\hat{\beta}_{DV.}$	std.err
$\overline{DV}1$=0.1692	0.9464	0.0043	0.2223	0.0032	0.0470	0.0128
$\overline{DV}2$=0.2952	0.9453	0.0043	0.2223	0.0032	0.0808	0.0105
$\overline{DV}3$=0.3672	0.9446	0.0043	0.2224	0.0032	0.0923	0.0099
$\overline{DV}4$=**0.5388**	0.9434	0.0043	0.2225	0.0032	0.1334	**0.0096**
$\overline{DV}5$=0.6301	0.9432	0.0043	0.2226	0.0032	0.1285	0.0100
$\overline{DV}6$=0.7190	0.9438	0.0043	0.2226	0.0032	0.1124	0.0107
$\overline{DV}7$=0.8360	0.9449	0.0043	0.2226	0.0032	0.0979	0.0130
$\overline{DV}8$=0.9445	0.9448	0.0043	0.2226	0.0032	0.1599	0.0210
$\overline{DV}9$=1.0000	0.9472	0.0043	0.2225	0.0032	0.0000	-

Note: IAB Establishment Panel 2006-2010; n=34,308. DV is constructed in the following way: The interaction variable between the wave number (14,...,18) and the identification number of the establishments is split into nine classes and ordered from the smallest to the largest class (C1, ..., C9). Then new cumulative variables are determined and transformed into dummy variables: DV1=1 if the establishment j belongs to C1, =0 otherwise; DV2=1 if j belongs to C1 or to C2, =0 otherwise; ... ; DV8=1 if j belongs to C1 or C2 or ... or C8, =0 otherwise.

Table A2: OLS estimates of Cobb-Douglas functions with an artificial dummy ($D.$) determined from a rectangular distributed random variable as regressor. Results are average values of 300 estimates; dependent variable: logarithm of sales - lnY

	$\hat{\beta}_{D.}$	std.err
$\overline{D1}$=0.1	-0.0177	0.0182
$\overline{D2}$=0.2	-0.0040	0.0135
$\overline{D3}$=0.3	-0.0065	0.0118
$\overline{D4}$=0.4	-0.0148	0.0110
$\overline{D5}$=**0.5**	-0.0105	**0.0108**
$\overline{D6}$=0.6	-0.0111	0.0110
$\overline{D7}$=0.7	-0.0134	0.0118
$\overline{D8}$=0.8	-0.0073	0.0135
$\overline{D9}$=0.9	0.0086	0.0180

Note: IAB Establishment Panel 2006-2010; n=34,308.

Table A3: OLS estimates of Cobb-Douglas functions with company-level pact dummy (CLP) as regressor, decreasing shares of n(CLP=1)/n; dependent variable: logarithm of sales

	$\hat{\beta}_{CLP}$	std.err	t
$\overline{CLP}$=**0.0693**	0.1231	**0.0236**	5.22
$\overline{CLP}$=0.0624	0.1209	0.0246	4.92
$\overline{CLP}$=0.0533	0.1299	0.0259	5.02
$\overline{CLP}$=0.0477	0.1131	0.0275	4.11
$\overline{CLP}$=0.0407	0.1006	0.0295	3.41
$\overline{CLP}$=0.0336	0.1005	0.0322	3.12
$\overline{CLP}$=0.0273	0.1429	0.0356	4.01
$\overline{CLP}$=0.0207	0.1446	0.0403	3.39
$\overline{CLP}$=0.0135	0.1357	0.0486	2.79
$\overline{CLP}$=0.0067	0.1887	0.0671	2.80

Note: IAB Establishment Panel 2006-2010; n=31,985. In the first line the estimation with the original sample and $\overline{CLP}$=0.0693 is presented. Next, only 90% of the firms with CLP=1, where $\overline{CLP}$=0.0624, are considered. The random selection of the CLP firms is based on a rectangular distribution of the CLP firms. The determination of the following lines is analogous to that of the second line.

Table A4: OLS estimates of Cobb-Douglas functions with works council dummy (WOCO) as regressor, decreasing shares of n(WOCO=1)/n - randomly determined based on a rectangular distribution; dependent variable: logarithm of sales

	$\hat{\beta}_{WOCO}$	std.err	t
$\overline{WOCO}$=**0.3045**	0.4076	**0.0136**	29.50
$\overline{WOCO}$=0.2747	0.3573	0.0136	26.32
$\overline{WOCO}$=0.2440	0.3140	0.0136	23.13
$\overline{WOCO}$=0.2132	0.2784	0.0137	20.29
$\overline{WOCO}$=0.1829	0.2418	0.0141	17.11
$\overline{WOCO}$=0.1523	0.2102	0.0148	14.20
$\overline{WOCO}$=0.1221	0.1904	0.0159	11.99
$\overline{WOCO}$=0.0920	0.1842	0.0177	10.43
$\overline{WOCO}$=0.0605	0.1888	0.0208	9.07
$\overline{WOCO}$=0.0305	0.1730	0.0281	6.16

Note: IAB Establishment Panel 2006-2010; n=34,217.