

# IED

Working Paper 2

June 2008

## The application of robust regression to a production function comparison – the example of Swiss corn

Robert Finger & Werner Hediger









# The application of robust regression to a production function comparison – the example of Swiss corn

Robert Finger<sup>1\*</sup>, Werner Hediger<sup>1,2</sup>

<sup>1</sup> Institute for Environmental Decisions, Agri-Food and Agri-Environmental Group, Swiss Federal Institute of Technology Zurich. Sonneggstrasse 33, 8092 Zurich, Switzerland.

<sup>2</sup> Center for Corporate Responsibility and Sustainability, University of Zurich. Künstlergasse 15a, 8001 Zurich, Switzerland.

\* Address for correspondence:

Sonneggstrasse 33, 8092 Zurich, Switzerland

rofinger@ethz.ch

Tel.: +41 44 632 04 74

Fax: +41 44 632 10 86

1	Introduction .....	2
2	Production Functions .....	4
3	Data .....	5
4	Outliers and Estimation Methodology .....	7
5	Estimation Results .....	11
6	Optimal Input Levels and Costs of Misspecification .....	13
7	Summary and Conclusions .....	16
	Acknowledgements .....	17
	References .....	17

*Copyright 2007 by Robert Finger and Werner Hediger. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.*



**Abstract:** *The adequate representation of crop response functions is crucial for agri-environmental modeling and analysis. So far, the evaluation of such functions focused on the comparison of different functional forms. The perspective is expanded in this article by considering an alternative regression method. This is motivated by the fact that exceptional crop yield observations (outliers) can cause misleading results if least squares regression is applied. We show that such outliers are adequately treated if robust regression is used instead. The example of simulated Swiss corn yields shows that the use of robust regression narrows the range of optimal input levels across different functional forms and reduces potential costs of misspecification.*

**Key words:** *production function estimation, production function comparison, robust regression, crop response*

# 1 Introduction

The adequate representation of production or crop yield functions is crucial for modeling purposes in agricultural and environmental economic analyses. The discussion and estimation of different functional forms has therefore gained much attention in agronomic and agricultural economics literature. Various functional forms have been considered so far, but less attention has been given to the estimation techniques in general and the impact of exceptional crop yield observations (outliers) in particular. The latter is important since the Least Squares (LS) fitting criterion can produce misleading results if data sets contain outliers, such as exceptionally low yields caused by extreme weather events or climate situations. In order to address this problem we apply robust regression. In contrast to Swinton and King (1991), who used robust regres-

sion methods for trend estimation within crop yields, our focus is on the estimation and comparison of crop production functions. To this end, we take the example of corn (*Zea mays* L.) yields in Switzerland.

Observed yield data would provide insufficient estimation possibilities due to a lack of variation within the data. In contrast, biophysical simulation can generate an enlarged data base compared with field observations. It particularly enables the creation of more comprehensive datasets of crop yields with respect to the variation of analyzed factors such as agricultural inputs, while keeping other factors such as soil properties constant. In our study, we apply a meta-modeling approach that makes use of crop yield data generated with a biophysical simulation model to estimate and compare crop production functions.



The assessment of functional forms can be based on the coefficient of determination (e.g. Alivelu et al., 2003), residual distribution (e.g. Bélanger et al., 2000), non-nested hypothesis testing (e.g. Frank et al., 1990) and potential misspecification costs (e.g. Llewelyn and Featherstone, 1997). Using LS and robust regression, we devote special attention to the cost of misspecification which constitutes an economic approach to the comparison of production functions. This allows us to assess the potential income loss that would arise from using calculations based on LS instead of robust regression methods or from an improper specification of the production function.

The remainder of this paper is organized as follows. Section 2 provides a brief presentation of the production functions that are used throughout our analysis, and Section 3 is devoted to the data used. In Section 4, the estimation methodology is introduced, while the estimation results are presented in Section 5. Subsequently, optimal input levels and the cost of misspecification are investigated in Section 6. Finally, the advantage of applying robust regression techniques in production function estimation is discussed in the concluding Section 7.



## 2 Production Functions

Three types of crop production functions are analyzed in this study: two polynomial specifications (the quadratic and the square root function) and the Mitscherlich-Baule function. These functional forms are frequently used in the literature and proved to accurately capture the underlying relationships (Ackello-Ogutu et al., 1985, Anderson and Nelson, 1975, Berck and Helfand, 1990, Frank et al., 1990, Fuchs and L  the, 1996, Heady and Dillon, 1961, Jalota et al., 2007, and Llewelyn and Featherstone, 1997, Rajsic and Weersink, 2008, Yadav et al., 2003).

Being aware that corn yields are driven by numerous factors, we focus our analysis on two crucial production factors: nitrogen fertilizer and irrigation water. Thus, production functions are used to describe corn yield responses to nitrogen and irrigation water such as shown in Llewelyn and Featherstone (1997). By focusing on these two variable factors, the production process is represented by a simple analytical description that implicitly considers other production factors such as soil and climate (Godard et al., 2008). Together with the concentration on three functional forms, this restriction serves the sake of clarity in our investigation.

The *quadratic form*, shown in equation (1), consists of an additive composition of the input factors, their squared values, and an additional interaction term. The latter elucidates whether the input factors are independent of each other or not. The quadratic function is formally defined as follows:

$$Y = \alpha_0 + \alpha_1 \cdot N + \alpha_2 \cdot W + \alpha_3 \cdot N^2 + \alpha_4 \cdot W^2 + \alpha_5 \cdot N \cdot W \quad (1)$$

$Y$  denotes corn yield per area,  $N$  the amount of inorganic nitrogen applied, and  $W$  irrigation water applied. The  $\alpha_i$ 's are parameters that must satisfy the subsequent condi-

tions in order to ensure decreasing marginal productivity of each input factor:  $\alpha_1, \alpha_2 > 0$  and  $\alpha_3, \alpha_4 < 0$ . Furthermore, if  $\alpha_5 > 0$  the two input factors are complementary. They are competitive if  $\alpha_5 < 0$ , while  $\alpha_5 = 0$  indicates independence of the two input factors.

The *square root function* (equation 2) is very similar to the quadratic form but produces different shapes of the curves. The square root form is defined as follows:

$$Y = \alpha_0 + \alpha_1 \cdot N^{1/2} + \alpha_2 \cdot W^{1/2} + \alpha_3 \cdot N + \alpha_4 \cdot W + \alpha_5 \cdot (N \cdot W)^{1/2} \quad (2)$$

To ensure decreasing marginal productivity of each input factor, the parameters must satisfy the same conditions as for the quadratic form, and their interpretation is identical.

The *Mitscherlich-Baule function* (Equation 3) allows for a growth plateau, which follows from the von Liebig approach to production functions (see Paris, 1992, for historical notes). Moreover, this functional form is characterized by continuously positive marginal productivities of the input factors. It does not exhibit negative marginal productivities, as the above polynomial forms. Formally, the Mitscherlich-Baule function is given by

$$Y = \alpha_1 \cdot (1 - \exp(-\alpha_2 \cdot (\alpha_3 + N))) \cdot (1 - \exp(-\alpha_4 \cdot (\alpha_5 + W))) \quad (3)$$

with  $\alpha_1$  representing the growth plateau, and  $\alpha_3$  and  $\alpha_5$  that include nitrogen in the soil ( $\alpha_3$ ) and water endowments ( $\alpha_5$ ) such as soil moisture. The coefficients  $\alpha_2$  and  $\alpha_4$  describe the influence of the corresponding input factors on the yield. Unlike the classical von Liebig production function, the Mitscherlich-Baule function allows for factor substitution. It is not linear limitational in the input factors as the von Liebig function, i.e. the isoquants are not right-angled.



### 3 Data

Our analysis and estimation of production functions is based on simulated corn yield data that is generated with the CropSyst model. This is a deterministic crop yield simulation model that has been widely used and validated (see Stöckle et al., 2003, for a review of studies using CropSyst). It involves various above and below ground processes, such as soil water budget, soil-plant nitrogen budget, crop phenology, canopy and root growth, biomass production, crop yield, residue production and decomposition, and soil erosion by water. These processes are simulated with daily time step. The model is calibrated to field trials and sample data. Model settings and calibration for the Swiss Plateau region are presented in Torriani et al. (2007).

In our analysis, CropSyst is driven by daily weather data from six different locations on the Swiss Plateau for the years 1981 – 2003, as provided by the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss). These locations are distributed over the eastern Swiss Plateau ranging from 06°57' to 08°54' longitude and are located at elevation levels between 422 and 565 meter above sea level (Finger and Schmid, 2007). Compared to an approach with one single location, the use of observations from six different weather stations broadens the database and allows us to represent a large proportion of the entire Swiss corn producing acreage.

The simulation and subsequent data analysis are restricted to one uniform type of soil for all locations, characterized by texture with 38% clay, 36% silt, 26% sand and soil organic matter content at 2.6% weight in the top soil layer (5 cm) and 2.0% in lower soil layers (Torriani et al., 2007).

Moreover, the type of management is uniform for all simulations. Identical seeding dates, irrigation settings (possible from day one after sowing to harvesting, never exceeding field capacity), fertilizer type (inorganic nitrogen fertilizer) and fertilizer application dates are applied in CropSyst (Finger and Schmid, 2007). This approach avoids distortions due to non-uniform soil and management properties.

To have a comprehensive data set, one simulation is conducted without application of fertilizer and irrigation for each location and each year. Furthermore, additional combinations of irrigation and fertilizer are generated randomly. Taking nitrogen fertilizer application rates from 0 to 320 kg/ha and irrigation water from 0 to 340 mm, this results in 212 different levels of nitrogen application to the plants and 60 different levels of irrigation.

The resulting dataset consists of 527 observations. Assuming a dry matter content of 85%, average yields for three different ranges of irrigation  $W$  and fertilizer  $N$  application, respectively, are shown in Table 1. This rough approximation of the average corn yields reveals a global yield maximum for  $71 \leq W \leq 140$  and  $76 \leq N \leq 150$ . Simulated corn yields decrease if the amounts of irrigated water or applied fertilizer deviate from those input ranges.

**Table 1:** Average simulated corn yields 1981–2003

		Applied nitrogen in kg/ha		
		0–75	76–150	151–320
Applied irrigation water in mm	0–70	6 955	8 872	8 521
	71–140	7 293	9 717	9 100
	141–340	7 275	8 814	9 158

Source: CropSyst simulations



In our meta-modeling approach, output of the biophysical model is restructured into crop production functions. Thus, key relationships among the factors studied can be isolated (Jalota et al., 2007). Total (aggregated) values for nitrogen application, irrigation and corn yield are used for production function estimation. In contrast, sub-processes in the biophysical model are conducted on a daily time step. Thus, the relationships estimated in the crop production functions do not replicate factor relationship settings in the biophysical model, i.e. in the data generating process. Similar meta-modeling approaches have been used, for instance, by Jalota et al. (2007), and Llewelyn and Featherstone (1997).

Due to the field experimental design in the crop yield simulation, the dataset contains quasi-continuous input-output combinations. In contrast to discrete application of inputs (i.e. a few levels of inputs) in a field experiment, quasi-continuous input levels enable a regression rather than an analysis of variance approach in this study. Thus, the resulting dataset is suitable for production function estimation. Moreover, the random application of inputs enables unbiased estimation of the production function coefficients. Input levels are uncorrelated with other variables that also influence corn yields but are not considered in the production function estimations, such as environmental factors that are held constant in the simulations. Thus, no omitted variable bias will occur for the coefficient estimates of nitrogen and irrigation water in the crop production functions.

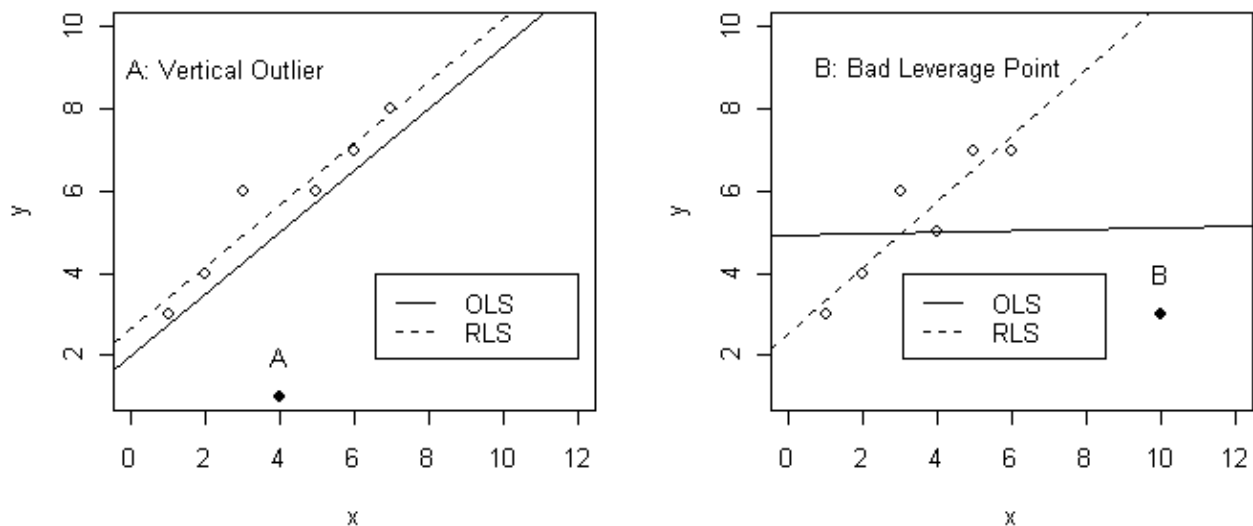


## 4 Outliers and Estimation Methodology

Exceptional climatic years are supposed to lead to exceptional crop yield levels and to have an extraordinary influence on plant response to irrigation and fertilization. As a consequence, they may involve outliers that deviate from the relationship described by the majority of the data.

Two standard examples for outliers in a linear simple regression model are presented in Figure 1. Point A clearly deviates from the typical linear relationship between the dependent (y) and the independent (x) variable. Such 'vertical' outlier is characterized by an unusual observation in the dependent variable. The impact of vertical outliers on the estimation of regression coefficients is usually small and mainly affects the regression intercept (Sturm and de

Haan, 2001). If unusual observations occur in the set of independent variables, these outliers are called leverage points. If such leverage point deviates from the linear relationship described by the majority of observations it is called 'bad leverage point' such as Point B in Figure 1. Due to the exposed position of the outlier it has a leverage effect on the coefficient estimation. In contrast, a leverage point is called 'good leverage point' if it does not deviate from the typical relationship. Good leverage points are no outliers and even improve the regression inference as these points reduce standard errors of coefficient estimates.



**Figure 1:** Examples for outlying observations

Note: Regression lines are fitted using ordinary least squares (OLS) and reweighted least squares (RLS). Source: According to Sturm and de Haan (2001)



In this study, Reweighted Least Squares (RLS) regression is applied for the estimation of eqns. (1) and (2), using the ROBUSTREG procedure in the SAS statistical package. RLS is a weighted LS regression, which is based on an analysis of Least Trimmed Squares (LTS) residuals. In crop production function estimation, a vertical outlier is characterized by observations with an exceptional (low) yield level. Bad leverage points consist of observations with an exceptional input-output relationship for very low or high levels of inputs application. With regard to the functional relationship between corn yields and application of nitrogen and irrigation, we particularly expect climatic conditions to be influential. For instance, the amount of rainfall can influence droughts or moisture built up, and thus indirectly restrict yield levels. Furthermore, the plants are expected to respond specifically to management under certain climatic conditions. The response to irrigation and fertilization, for instance, changes under high and low water stress situations.

The occurrence of outliers is not exclusive to agricultural issues. Rather, outliers are frequently observed in empirical data sets and particularly considered in the applied statistics, econometrics and economics literature (e.g. Huber, 1996, Hubert et al., 2004b, Sturm and de Haan, 2001). The breakdown point concept is used to quantify robustness properties of a regression estimator. It is defined as the smallest amount of arbitrary outlier contamination which can carry an estimator over all bounds (Hubert et al., 2004a). The estimator becomes unreliable beyond this border line.

Ordinary least squares (OLS) regression possesses the lowest possible breakdown point of  $1/n$ , where  $n$  denotes the number of observations. This indicates that OLS can not cope with a single outlier because one outlier can be sufficient to move the coefficient estimates arbitrarily far away from the actual underlying values. Thus, outliers cause unreliable coefficient estimates if OLS is applied.

This vulnerability of least squares estimation to outlying observations has been demonstrated in various studies (e.g. Hampel et al., 1986, Huber, 1996, and Rousseeuw and Leroy, 1987). Reliable results are provided by OLS if and only if outlier diagnostic and treatment tools such as robust regression methods or robust regression diagnostics are applied as well. The application of these methods ensures the non-inclusion or the appropriate down-weighting of outliers in the analysis.

A simple outlier diagnostic tool is the scatter plot that enables the detection of outliers in simple regression cases. However, this is impossible if the dimension of the problem exceeds the simple regression case and the number of observations is very large, such as for our analysis. Outlier diagnostics based on residual plots might suffer from outliers (Rousseeuw and Leroy, 1987), in particular for bad leverage points. Outliers can tilt the (original) regression line and have small regression residuals. Thus, outliers might not be discovered in residual plots (Sturm and de Haan, 2001). Other diagnostic tools are required to identify outlying or influential observations. However, they may involve additional problems. Studentized and jackknifed residuals, Cooks distances and other diagnostics based on Hat matrix elements, for instance, are susceptible to the so called masking effect. If more than one outlier occurs, these outlier diagnostics might not be able to detect a single one because one outlier can be masked by the presence of others (Rousseeuw and Leroy, 1987). Multiple-case diagnostics or high-breakdown diagnostics have to be employed instead. In this study we therefore apply robust regression and outlier identification based on robust regression residuals for identification and adequate treatment of outliers. This contrasts with two different approaches frequently used for the estimation of crop production functions.



Usually, the impact of climatic extreme events is reduced by introducing dummy variables for certain states of climate variables, or the estimation is conducted separately for different states of climate variables or different years (e.g. Fuchs and L  the, 1996, Jalota et al., 2007, Rajsic and Weersink, 2008). Even though these methods aim to take different productivity levels of input factors for different states of climatic variables into account, they are usually based on factor relationship assumptions rather than on the data itself. Moreover, these methods might considerably reduce the power of the analysis due to the loss of degrees of freedom. In contrast, various approaches to robust regression analysis have been proposed (e.g. Hampel et al., 1986, and Rousseeuw and Leroy, 1987). They enable the identification of outliers taking the crop yield data into account using all observations in the dataset.

The main idea of robust regression is to give little weight to outlying observations in order to isolate the true underlying relationship. In this context, the notation “true relationship” is restricted to an econometrical interpretation, while the excluded observations can be of particular interest from a scientific point of view. However, the inclusion of outliers in the analysis does not allow for trustful regression inference. By contrast, separated analyses of outliers and inliers can lead to an information gain.

In this study, reweighted least squares (RLS) regression is applied for the estimation of quasi-linear quadratic and square root production functions (equations 1 and 2). RLS is applied in favor of other robust regression methods due to its good robustness and efficiency properties (see Rousseeuw and Leroy, 1987, for details). RLS is a weighted LS regression, which is based on an analysis of least trimmed squares (LTS) regression residuals. LTS is a high-breakdown regression technique, i.e. it can possess the highest possible breakdown point of  $\frac{1}{2}$ . In contrast to OLS estimation, LTS coefficient estimates are thus reliable in presence of outliers. Based on the idea of trimming the

largest residuals the LTS fitting criterion is defined as follows:

$$\underset{\beta}{\text{Min}} \sum_{i=1}^h (r_i^2) \quad (4)$$

$(r^2)_{(i)}$  are the ascending ordered squared (robust) residuals and  $h$  is the so-called trimming constant. In our analysis,  $h = \lceil (3n + p + 1) / 4 \rceil$  is employed (SAS Institute, 2004), with  $p$  denoting the number of coefficient that are estimated.

The computation of LTS coefficients is neither explicit nor iterative, but follows an algorithm described in Rousseeuw and Leroy (1987). Because the efficiency of LTS estimation is low, LTS results allow not for trustful inference. Thus, LTS estimation is only used as a data analytic tool for outlier identification. An observation is identified as an outlier if the absolute standardized robust residual  $|r_i / \hat{\sigma}|$  exceeds the cutoff value of 2.5.  $r_i$  and  $\hat{\sigma}$  are the (robust) LTS residuals and scale estimates, respectively. This cutoff-value choice is motivated by a (roughly) 99% tolerance interval for Gaussian distributed standardized residuals (Sturm and de Haan, 2001). Coefficient estimates of RLS regression are defined as follows

$$\hat{\beta}_{RLS} = (X'WX)^{-1} X'WY \quad (5)$$

The diagonal elements of the weighting matrix  $(W = \text{diag}\{w_1, \dots, w_n\})$  are generated by an indicator function,  $I_{\text{Outlier}}$ :

$$w_i = I_{\text{Outlier}} \left[ \left| \frac{r_i}{\hat{\sigma}} \right| \leq 2.5 \right] \quad (6)$$

The indicator function generates weights of zero for observations that are identified as outliers and weights of one otherwise. RLS regression is applied for coefficient estimation of quasi linear functional forms, using the ROBUSTREG procedure in the SAS statistical package (SAS Institute, 2004). An example for the better robustness properties of RLS compared to OLS is indicated in Figure 1.



OLS coefficient estimates change in the presence of outliers, in particular for bad leverage points. In contrast, RLS coefficient estimates are not affected by outliers in this example.

Because LTS regression is not suitable for nonlinear problems such as the Mitscherlich-Baule function (equation 3), iterative approaches are required. Robust regression is implemented in this case by using iteratively reweighted least squares (IRLS). In order to reduce the influence of outliers on estimation results, weights are generated with M-estimation using Tukey's biweight (Hampel et al., 1986) such as shown in equation (7) that follows Hogg (1979). These weights are re-estimated at each step of iteration until convergence.

$$w_i = \begin{cases} (1 - (r_i / \hat{\sigma} \cdot c)^2)^2, & |r_i / \hat{\sigma}| \leq c \\ 0, & |r_i / \hat{\sigma}| > c \end{cases} \quad (7)$$

$r_i$  is the (robust) IRLS residual and  $\hat{\sigma}$  the (robust) scale estimate and  $c$  a tuning constant. Following Hogg (1979), we employ the median of absolute deviations from the median (MAD) for robust scale estimation and set the tuning constant to 5.0. In contrast to LTS, IRLS is no high breakdown estimation technique. In order to validate results, we conduct sensitivity analysis of crucial factors such as starting values and tuning constant. We use the

Levenberg-Marquardt algorithm (see Moré, 1978, for details) that ensures stable estimation for highly correlated coefficient estimates that occur in our analysis (Schabenberger et al., 1999). In this study the nonlinear Mitscherlich-Baule function is estimated with IRLS using the NLIN procedure in the SAS software package. Furthermore, all estimations are corrected for heteroscedasticity following Johnston and DiNardo (1997).

Besides the most important property of giving trustworthy coefficient estimates, robust regression provides detailed insight in the structure of the data. If LS and robust regression results are considerably different and many outliers are indicated, the observations identified as outliers reveal their origin and can exhibit inappropriateness of the employed model structure. Above all, the interpretation of outliers is indispensable. Ruling out that outliers are caused by typing, copying or measuring errors, this interpretation should take not only statistical but mainly reasons from the subject matter science into account (Hampel, 2002). Thus, in the following, all estimations are conducted with both least squares and robust regression and outlier interpretation is provided.



## 5 Estimation Results

Within our dataset, the largest number of observations identified as outliers are in the year 2003. About 25% of the observations that are identified by the RLS method as outlier or are given very small weights in the IRLS method, can be attributed to this year<sup>1</sup>. It is characterized by high temperatures and low precipitation in the relevant seedling-to-harvest period that caused particularly low corn yields in all Europe (Ciais et al., 2005). Other years with exceptionally low levels of precipitation and high temperatures in the corn growing season (e.g. 1983, 1991) also frequently occur in the lists of outlying observations.

The reason for the existence of outliers in these years is twofold. First, the yield levels are lower than usually. Second, the relationship between independent and dependent variables is affected by different reactions to input levels in situations where one of the inputs is a limiting factor. The yield response to irrigation water is higher than usual if – unlike in normal years – water constitutes a limiting factor for the plants in the Swiss Plateau. Furthermore, the interaction between fertilizer and irrigation water is higher because the plants' response to nitrogen also highly depends on water availability as nitrogen is taken up by the roots in a water solution.

Table 2 presents the estimation results for the quadratic and the square root production functions, respectively. It shows that each estimation coefficient has the correct (i.e. the expected) sign. The coefficient  $\alpha_5$  (Applied Nitrogen \* Irrigation Water) is not significantly different from zero in the four estimated polynomial functions. This indicates that rainfall is sufficient to ensure efficient nitrogen uptake under normal climatic conditions in Switzerland.

**Table 2:** Coefficient estimates for the quadratic and the square root production functions

Variable	OLS – Estimation	RLS – Estimation
Quadratic production function (equation 1)		
Intercept	6638.265 (165.05)**	6661.421 (179.24)**
N	25.64327 (17.62)**	27.55239 (22.71)**
W	6.046902 (5.62)**	5.578582 (5.75)**
N <sup>2</sup>	-0.07104 (12.22)**	-0.07236 (14.94)**
W <sup>2</sup>	-0.01797 (3.87)**	-0.0162 (3.88)**
NW	0.007766 (1.51)	0.00373 (0.89)
adj. R <sup>2</sup>	0.5680	0.7065
Square root production function (equation 2)		
Intercept	6589.997 (155.02)**	6601.924 (162.13)**
N <sup>1/2</sup>	297.1821 (12.42)**	313.0936 (16.34)**
W <sup>1/2</sup>	75.09137 (4.26)**	67.1385 (4.17)**
N	-11.2156 (6.88)**	-10.544 (8.15)**
W	-3.03419 (2.40)*	-2.49922 (2.17)*
(NW) <sup>1/2</sup>	1.46442 (1.43)	0.364377 (0.45)
adj. R <sup>2</sup>	0.5834	0.7330

Note: Statistics in parentheses are t statistics  
 (\*\*) – indicates significance at the 1% level  
 (\*) – indicates significance at the 5% level

<sup>1</sup> In total RLS identifies 43 outliers for the quadratic production function and 37 for the square root function. Moreover, 36 observations have weights smaller than 0.25 in the IRLS estimation of the Mitscherlich-Baule function



In Table 3, the Mitscherlich-Baule production function estimates are presented with coefficient estimates showing the expected signs. Using both LS and robust regression, the Mitscherlich-Baule function reaches higher goodness of fit than the respective estimates of the quadratic and square root forms. The coefficient estimates for irrigation water and water endowment ( $\alpha_4$  and  $\alpha_5$ ) are not significantly different from zero at the level of five percent in the LS estimation. In contrast, the coefficients  $\alpha_4$  and  $\alpha_5$  are significant at the one percent level if robust regression (IRLS) is used. Moreover, the coefficient estimate for  $\alpha_5$  increases remarkably if IRLS regression is applied. This is explained by the fact that mainly dry years are excluded or down-weighted in the robust regression, such that the estimated soil water endowment is higher for the remaining observations.

Even though all differences in coefficient estimates between LS and robust regression are not significant at the 5% level, the application of robust regression leads to reasonable shifts in coefficient estimates and their level of significance for all functional forms. However, the decision on the most appropriate estimation technique cannot exclusively be based on statistical measures. For instance, the goodness of fit cannot be compared between LS and robust estimation. The deletion of outliers increases, by definition, the goodness of fit for the regression on the remaining observations. Hence, conclusions on the appropriateness of functional forms and estimation techniques can be drawn if and only if the misspecification costs are calculated and interpreted, as shown in the subsequent section.

**Table 3:** Coefficient estimates for the Mitscherlich-Baule production function

Variable	Estimation Method	
	LS (Levenberg-Marquardt)	IRLS (Levenberg-Marquardt)
$\alpha_1$	9180.6 (95.14)**	9410.3 (87.7)**
$\alpha_2$	0.0288 (5.72)**	0.0266 (7.38)**
$\alpha_3$	50.6952 (5.96)**	48.3036 (7.75)**
$\alpha_4$	0.0598 (1.22)	0.0304 (2.95)**
$\alpha_5$	45.1410 (1.24)	71.2249 (3.10)**
adj. R <sup>2</sup>	0.736	0.809

Note: Statistics in parentheses are t statistics  
 (\*\*) – indicates significance at the 1% level  
 (\*) – indicates significance at the 5% level



## 6 Optimal Input Levels and Costs of Misspecification

The knowledge of production functions is crucial for modeling purposes and economic analyses that are concerned with optimal resource allocation. This usually involves an assessment of optimal input and output levels, which is generally determined through maximization of a suitably defined objective function. For the purpose of our analysis, this is given by the subsequent profit function

$$\pi = P_{Corn} \cdot f(W, N) - P_{Nitrogen} \cdot N - P_{Irrigation} \cdot W \quad (8)$$

where the net return (or quasi-rent) per hectare  $\pi$  is equal to the gross return (crop price  $P_{Corn}$  times corn yield  $f(W, N)$ , minus total nitrogen costs (nitrogen price  $P_{Nitrogen}$  times amount of nitrogen applied  $N$ ) and total irrigation costs (irrigation price  $P_{Irrigation}$  times amount of irrigation water  $W$ ) per hectare. For simplicity, other costs are assumed to be constant and therefore irrelevant for calculating the profit maximizing input combination. By maximizing the above profit function (equation 8), the optimal input levels are determined through the following first-order conditions:

$$\frac{\partial f(W, N^*)}{\partial N} = \frac{P_{Nitrogen}}{P_{corn}} \text{ and } \frac{\partial f(W^*, N)}{\partial W} = \frac{P_{Irrigation}}{P_{corn}} \quad (9)$$

Where  $N^*$  and  $W^*$  are the profit maximizing input levels of nitrogen fertilizer and irrigation water, respectively. In other words, efficiency in production requires employment and remuneration of all production factors according to their value of marginal product. This is satisfied if, for each input factor, the input price equals the crop price multiplied with the factor's marginal productivity.

In the further analysis, we set the corn price equal to CHF 0.642 kg<sup>-1</sup>, the average annual value for the period 1981-2003 in Switzerland (SBV, 1982-2004). We assume a constant nitrogen price of CHF 1.6 kg<sup>-1</sup> (extrapolated from ammonium nitrate 27.5 to pure nitrogen) at the 1993 level (LBL, 1993), and a price for irrigation water of CHF 0.06 m<sup>-3</sup> (Finger and Schmid, 2007). Using these data, the optimal input levels are calculated according to equation (9) and represented in Table 4.

**Table 4:** Optimal input levels, yield, and maximum net return

Functional Form- Estimation Method	Optimal amount of Nitrogen applied (kg/ha)	Optimal amount of irrigation Water ap- plied (mm)	Optimal yield (kg/ha)	Maximum net re- turn (CHF/ha)
Quadratic-OLS	172.8	179.6	9695	5840.32
Square Root-OLS	131.3	133.9	9180	5602.82
Mitscherlich-Baule-OLS	111.2	61.3	9078	5613.55
Quadratic-RLS	177.4	163.8	9859	5947.68
Square Root-RLS	147.7	108.6	9324	5684.56
Mitscherlich-Baule-IRLS	124.9	116.7	9286	5691.51

Note: LS indicates least squares, RLS reweighted least squares, and IRLS iteratively reweighted least squares estimation.



It shows that all optimal input levels are within the range of the data, and the general results about the functional forms remain the same as in other studies. As in Ackello-Ogutu et al. (1985), the polynomial functions recommend higher fertilizer use than the Mitscherlich-Baule functions.

With 61.3 mm of irrigation water and 111.2 kg/ha of nitrogen, the lowest input use is recommended by the Mitscherlich-Baule function estimated with LS. This goes along with the lowest yield (9078 kg/ha) and an estimated net revenue of 5613.55 CHF/ha. In contrast, the robust estimated quadratic function shows the highest yield (9859 kg/ha) and nitrogen use (177.4 kg/ha) and the highest profit (5947.68 CHF/ha), while the quadratic LS function implies the highest optimal amount of irrigation water with 179.6 mm. Thus, the quadratic form implies a higher optimal use of nitrogen and irrigation water than all other functions. This extends to the evidence given by Anderson and Nelson (1975) about the overestimation of optimal nitrogen amounts by the quadratic form to the optimal use of irrigation water.

Furthermore, the results in Table 4 show that the robust versions of production function estimates systematically lead to higher profit maximizing yields and higher profits than their non-robust counterparts. Moreover, for each functional form, the optimal amount of nitrogen fertilizer application increases if robust regression results are taken instead of LS results. And, except for the case of the Mitscherlich-Baule function, robust regression leads to the expected adjustment towards lower use of irrigation water in the profit maximizing situation. All in all, the use of robust estimation narrows the range of optimal input levels across the different functional forms.

Table 4 shows furthermore, that the selection of the functional form and estimation method both affect the result of the economic optimization and allocation problem. This relates to the concept of misspecification costs, which we employ for the final evaluation of production functions and estimation methods. The relative costs of misspecification are defined as the decrease in net return if optimal input levels of an incorrect function are used instead of those of the real underlying production function. The basic idea of this concept is to minimize the potential loss of a misspecification of the production function. Usually, the focus is on the potential loss due to the wrong functional form. In the following, we also consider the costs of using the improper estimation technique.

Table 5 gives the relative costs of misspecification. The nine cells in the upper left-hand corner correspond to the traditional approach where only functional forms estimated with LS are compared. If for instance the quadratic function would be the true underlying form, the use of the square root function induces a cost of misspecification of CHF 93.01. For the Mitscherlich-Baule function, this increases to CHF 297.88. The latter exhibits the highest costs of misspecification, while the square root function is the most appropriate if the misspecification-cost criterion is employed.

The square root function is similar to the quadratic form, but flatter in its surface and comes therefore closer to the plateau approach of the Mitscherlich-Baule specification (Ackello-Ogutu et al., 1985). Optimal input recommendations based on the square root function are correspondingly situated between those of the other two approaches we consider here.



**Table 5:** Relative Costs of Misspecification

When the true function is:	Cost of using optimal input levels based on:					
	Quadratic-OLS	Square Root-OLS	Mitscherlich-Baule-OLS	Quadratic-RLS	Square Root-RLS	Mitscherlich-Baule-IRLS
Quadratic-OLS	0	93.01	297.88	4.23	77.85	135.18
Square Root-OLS	30.61	0	39.83	32.13	8.41	2.01
Mitscherlich-Baule-OLS	113.22	41.38	0	109.97	41.86	27.34
Quadratic-RLS	3.77	104.65	296.39	0	68.59	145.23
Square Root-RLS	7.18	27.08	35.49	8.45	0	23.14
Mitscherlich – Baule-IRLS	57.52	54.08	3.11	51.85	9.86	0

Note: LS indicates least squares, RLS reweighted least squares and IRLS iteratively reweighted least squares estimation.

Table 5 further reveals that, in most cases, the use of robust estimation methods results in lower costs of misspecification than the standard LS approach, and that the square root specification performs better under this criterion than the other functional forms. This becomes obvious when comparing the top left-hand cells with the bottom right-hand ones, as well as from the comparison of the misspecification costs in the different lines of Table 5.

Only in the cases where the square root specifications are assumed to be the true underlying functions does the quadratic LS estimation show slightly lower costs of misspecification than its RLS counterpart. Furthermore, square root function estimation with LS leads to a marginally lower decrease of the net profit than its robust counterpart if the Mitscherlich-Baule-LS is assumed to be the underlying function.

Altogether, this supports the suggestion that the RLS estimation of the square root function is the best approximation of the real underlying crop response relationship. These findings further support the use of robust regression methods, besides the previously made recommendation from an econometrical point of view.



## 7 Summary and Conclusions

The proper representation of crop production functions is crucial for economic analyses that aim at determining optimal production levels and input use under different conditions. In our study, simulated corn yield data for the Swiss Plateau are used for the estimation of crop production functions, with particular consideration of yield response to nitrogen fertilizer and irrigation water application. Three functional forms are considered: the quadratic, the square root, and the Mitscherlich-Baule function. In addition, robust and standard regression methods are used for the estimation.

We found the square root function to be the most appropriate form to represent the data generated with corn yield simulations for Switzerland. Furthermore, exceptional climatic events, such as the summer drought in 2003, are proved to be the major source of misleading results if the least squares criterion is used to estimate production function coefficients. Robust regression methods are recommended instead. The use of robust estimation narrows the range of optimal input levels across the different functional forms. Thus, differences between functional forms are reduced by applying robust regression. This conclusion is further supported by a comparison of the relative costs of misspecification. Using robust instead of least squares regression generally results in lower costs of misspecification. Irrespective of the true underlying functional form, optimal input levels based on robust estimated functions reduce the maximum costs of misspecification compared

to the counterparts estimated with least squares regression. Thus, our investigation shows that, besides the functional form, the estimation method is decisive for production function comparisons.

The improved estimation of production functions might be valuable in practice because crop production functions are widely applied, for instance, to assess agro-environmental policy measures (e.g. Godard et al., 2008) to compare cropping systems (Yadav et al., 2003) or to project future agricultural water demand (e.g. Medellín-Azuara et al., 2008). Moreover, climate – and thus crop yield – extreme events are expected to occur more often in the future due to climatic change (e.g. Fuhrer et al., 2006). The properties of robust regression to ensure efficient and reliable coefficient estimation in presence of outliers might thus be particularly valuable for applications and economic assessments related to climate change issues (see e.g. Finger and Schmid, 2008). Furthermore, robust regression ensures efficient and accurate estimation of functional forms and thus of regression residuals. Since the latter are used to estimate yield variation with respect to input use (e.g. Just and Pope, 1979, Finger and Schmid, 2008), robust regression improves the estimation of both production functions and yield variation functions. Altogether, robust regression is a valuable tool for a wide range of agronomic and agri-environmental modeling problems that require a proper representation of crop response functions to variable inputs, such as nitrogen fertilizer and irrigation water.



## Acknowledgements

This work was supported by the Swiss National Science Foundation in the framework of the National Centre of Competence in Research on Climate (NCCR Climate). We are deeply indebted to Stéphanie Schmid for her valuable comments and data provision. We would like to thank the Swiss Federal Office of Meteorology for providing access to the meteorological database, and Raushan Bokusheva for helpful comments on an earlier draft of this paper.

## References

- Ackello-Ogut, C., Paris, Q. and Williams, W. A. (1985). Testing a von Liebig Crop Response Function against Polynomial Specifications. *American Journal of Agricultural Economics*, 67, 873-880.
- Alivelu, K., Srivastava, S., Rao, A. S., Singh, K. N., Selvakumari, G. and Raju, N.S. (2003). Comparison of Modified Mitscherlich and Response Plateau Models for Calibrating Soil Test Based Nitrogen Recommendations for Rice on Typic Ustropept. *Communications in Soil and Science and Plant Analysis*, 34, 2633-2643.
- Anderson, R. L. and Nelson, L. A. (1975). A Family of Models Involving Intersecting Straight Lines and Concomitant Experimental Designs Useful in Evaluating Response to Fertilizer Nutrients. *Biometrics*, 31, 303-318.
- Bélanger, G., Walsh, J. R., Richards, J. E., Milburn, P. H. and Ziadi, N. (2000). Comparison of Three Statistical Models Describing Potato Yield Response to Nitrogen Fertilizer. *Agronomy Journal*, 92, 902-908.
- Berck, P. and Helfand, G. (1990). Reconciling the von Liebig and Differentiable Crop Production Functions. *American Journal of Agricultural Economics*, 72, 985-996.
- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogee, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A. D., Friedlingstein, P., Grunwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J.M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J.F., Sanz, M.J., Schulze, E.D., Versala, T. and Valentini, R. (2005): Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* 437: 529 - 533.
- Finger, R. and Schmid, S. (2007): The Impact of Climate Change on Mean and Variability of Swiss Corn Production. *Schriftenreihe der Gruppe Agrar-, Lebensmittel- und Umweltökonomie*, ETH Zürich, No. 2007/1. [www.iaw.agrl.ethz.ch/research/publikationen/Finger\\_Schmid\\_Nov2007.pdf](http://www.iaw.agrl.ethz.ch/research/publikationen/Finger_Schmid_Nov2007.pdf)
- Finger, R. and Schmid, S. (2008). Modeling Agricultural Production Risk and the Adaptation to Climate Change. *Agricultural Finance Review* 68, 25-41.
- Frank, M. D., Beattie, B. R. and Embleton, M. E. (1990). A Comparison of Alternative Crop Response Models. *American Journal of Agricultural Economics*, 72, 597-603.



- Fuchs, C. and Löthe, K. (1996). Einfluss der Form von Produktionsfunktionen auf die Ermittlung der optimalen speziellen Intensität und die ökologischen Wirkungen in der Pflanzenproduktion. *Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues* 32, 493-502.
- Fuhrer, J., Beniston, M., Fischlin, A., Frei, C., Goyette, S., Jasper, K. and Pfister, C. (2006): Climate risks and their impact on agriculture and forests in Switzerland. *Climatic Change* 79: 79-102.
- Godard, C., Roger-Estrade, J., Jayet, P.A., Brisson, N. and Le Bas, C. (2008). Use of available information at a European level to construct crop nitrogen response curves for the regions of the EU. *Agricultural Systems* 97, 68-82.
- Hampel, F. R. (2002). Robust Inference. In: A. H. El-Shaarawi and W. W. Piegorsch (Eds.), *Encyclopedia of Environmental Metrics*, 1865-1885. Wiley and Sons, New York.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics*. Wiley and Sons, New York.
- Heady, E.O. and Dillon, J.L. (1961). *Agricultural Production Functions*. Iowa State University Press, Ames.
- Hogg, R. V. (1979). Statistical Robustness: One View of its use in Application Today. *The American Statistician*, 33, 108-115.
- Huber, P. J. (1996). *Robust Statistical Procedures*. Society for Industrial and Applied Mathematics, Philadelphia.
- Hubert, M., Rousseeuw, P.J. and Van Aelst, S. (2004a). Robustness. In: B. Sundt and J. Teugels (Eds.), *Encyclopedia of Actuarial Science*, 1515-1529. Wiley and Sons, New York.
- Hubert, M., Pison, G., Struyf, A. and van Aelst, S. (Eds.) (2004b). *Theory and Applications of Recent Robust Methods*. Birkhäuser Verlag, Basel.
- Jalota, S.K., Sood, A., Vitale, J.D. and Srinivasan, R. (2007). Simulated Crop Yield Response to Irrigation Water and Economic Analysis: Increasing Irrigated Water Use Efficiency in the Indian Punjab. *Agronomy Journal*, 99, 1073-1084.
- Johnston, J. and DiNardo, J. (1997). *Econometric Methods*, Fourth Edition. The McGraw-Hill Companies, New York.
- Just, R.E. and Pope, R.D. (1979). Production Function Estimation and Related Risk Considerations. *American Journal of Agricultural Economics*, 61, 276-284.
- LBL (1993). *Preiskatalog. Landwirtschaftliche Beratungszentrale Lindau (LBL, Swiss Center for Agricultural Extension)*, Lindau, Switzerland.
- Llewellyn, R. V. and Featherstone, A. M. (1997). A Comparison of Crop Production Functions Using Simulated Data for Irrigated Corn in Western Kansas. *Agricultural Systems*, 54, 521-538.
- Medellín-Azuara, J., Harou, J.J., Olivares, M.A., Madani, K., Lund, J.R., Howitt, R.E., Tanaka, S.K., Jenkins, M.W. and Zhu, T. (2008). Adaptability and adaptations of California's water supply system to dry climate warming. *Climatic Change* 87, 75-90.
- Meinke, H., Baethgen, W.E., Carberry, P.S., Donatelli, M., Hammer, G.L., Selvaraju, R. and Stöckle, C.O. (2001). Increasing profits and reducing risks in crop production using participatory systems simulation approaches. *Agricultural Systems*, 70, 493-513.
- Moré, J.J. (1978). The Levenberg-Marquardt algorithm Implementation and Theory. In G. Watson (Ed.), *Lecture Notes in Mathematics* 630, 105-116. Springer, Berlin.
- Paris, Q. (1992). The von Liebig Hypothesis. *American Journal of Agricultural Economics*, 74, 1019-1028.
- Rajšić, P. and Weersink, A. (2008). Do farmers waste fertilizer? A comparison of ex post optimal nitrogen rates and ex ante recommendations by model, site and year, *Agricultural Systems* 97, 56-67.
- Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust regression and outlier detection*. Wiley and Sons, New York.



- SAS Institute (2004). SAS/STAT 9.1 User's Guide. SAS Institute Inc., Cary, NC.
- Schabenberger, O., Tharp, B. E., Kells, J. J. and Penner, D. (1999). Statistical Tests for Hormesis and Effective Doses in Herbicide Dose Response. *Agronomy Journal* 91, 713-721.
- SBV (1982-2004). Statistische Erhebungen und Schätzungen über Landwirtschaft und Ernährung. Schweizer Bauernverband (SBV, Swiss Farmers' Union), Brugg, Switzerland.
- Stöckle, C. O., Donatelli, M. and Nelson, R. (2003). CropSyst, a cropping systems simulation model. *European Journal of Agronomy*, 18, 289-307.
- Sturm, J.-E. and de Haan, J. (2001). How robust is the relationship between economic freedom and economic growth? *Applied Economics*, 33, 839-844.
- Swinton, S. M. and King, R. P. (1991). Evaluating Robust Regression Techniques for Detrending Crop Yield Data with Nonnormal Errors. *American Journal of Agricultural Economics*, 73, 446-451.
- Torriani, D. S., Calanca, P., Schmid, S., Beniston, M. and Fuhrer, J. (2007). Potential effects of changes in mean climate and climate variability on the yield of winter and spring crops in Switzerland. *Climate Research*, 34, 59-69.
- Yadav, R.L., Singh, V.K., Dwivedi, B.S. and Shukla, A.K. (2003). Wheat productivity and N use-efficiency as influenced by inclusion of cowpea as grain legume in a rice-wheat system. *Journal of Agricultural Science* 141, 213-220.







## List of IED working papers

- Renate Schubert, Julia Blasch, and Kristin Hoffmann (2007). Environmental Protection, Energy Policy and Poverty Reduction – Synergies of an Integrated Approach. *IED Working Paper 1*, ETH Zurich, Institute for Environmental Decisions, Chair of Economics.
- Robert Finger and Werner Hediger (2007). The application of robust regression to a production function comparison – the example of Swiss corn. *IED Working Paper 2*, ETH Zurich, Institute for Environmental Decisions, Agri-Food and Agri-Environmental Economics Group.
- Stefanie Engel and Charles Palmer (2008). “Painting the forest REDD?” – Prospects for mitigating climate change through reducing emissions from deforestation and degradation. *IED Working Paper 3*, ETH Zurich, Institute for Environmental Decisions, Chair of Environmental Policy and Economics.
- Lasse Wallquist and Mischa Werner (2008). Carbon dioxide Capture and Storage – CCSStudie zum Entwicklungsstand von CCS in der Schweiz. *IED Working Paper 4*, ETH Zurich, Institute for Environmental Decisions, Chair of Consumer Behavior.
- Roland W. Scholz and Michael Siegrist (2008). Low Risks, High Public Concern? The Cases of Persistent Organic Pollutants (POPs), Heavy Metals, and Nanotech Particles. *IED Working Paper 5*, ETH Zurich, Institute for Environmental Decisions, Chair of Natural and Social Science Interface, Chair of Consumer Behavior.
- Astrid Zabel, Karen Pittel, Göran Bostedt, and Stefanie Engel (2009). Comparing Conventional and New Policy Approaches for Carnivore Conservation – Theoretical Results and Application to Tiger Conservation. *IED Working Paper 6*, ETH Zurich, Institute for Environmental Decisions, Chair of Environmental Policy and Economics.
- Astrid Zabel and Brian Roe (2009). Performance payments for environmental services: Lessons from economic theory on the strength of incentives in the presence of performance risk and performance measurement distortion. *IED Working Paper 7*, ETH Zurich, Institute for Environmental Decisions, Chair of Environmental Policy and Economics.
- Stefanie Engel and Charles Palmer (2009). The Complexities of Decentralization in a Globalizing World. *IED Working Paper 8*, ETH Zurich, Institute for Environmental Decisions, Chair of Environmental Policy and Economics.
- Michael Hartmann, Robert Huber, Simon Peter, and Bernard Lehmann (2009). Strategies to mitigate greenhouse gas and nitrogen emissions in Swiss agriculture: the application of an integrated sector model. *IED Working Paper 9*, ETH Zurich, Institute for Environmental Decisions, Agri-Food and Agri-Environmental Economics Group.
- Julia Blasch, Renate Schubert, Birgit Soete (2010). Grün aus der Krise – Was können «grüne» Konjunkturpakete leisten? *IED Working Paper 10*, ETH Zurich, Institute for Environmental Decisions, Chair of Economics.
- Moritz Rohling und Markus Ohndorf (2010). Prices vs. Quantities with Fiscal Cushioning. *IED Working Paper 11*, ETH Zurich, Institute for Environmental Decisions, Chair of Economics.
- Thomas Bernauer and Lena Maria Schaffer (2010). Climate Change Governance. *IED Working Paper 12*, ETH Zurich, Institute for Environmental Decisions, Chair of International Relations.
- Markus Ohndorf (2010). Optimal Monitoring for project-based Emissions Trading Systems under incomplete Enforcement. *IED Working Paper 13*, ETH Zurich, Institute for Environmental Decisions, Chair of Economics.
- Justin Caron and Markus Ohndorf (2010). Irreversibility and Optimal Timing of Climate Policy. *IED Working Paper 14*, ETH Zurich, Institute for Environmental Decisions, Chair of Economics.





ETH Zurich  
Institute for Environmental Decisions IED  
WEH G  
CH-8092 Zurich  
SWITZERLAND

Phone +41 44 632 47 18  
Fax +41 44 632 10 42  
URL <http://www.ied.ethz.ch>