Investigating the performance of different estimation techniques for crop yield data analysis in crop insurance applications

Journal Article**Author(s):**

Finger, Robert

Publication date:

2013-03

Permanent link:

<https://doi.org/10.3929/ethz-b-000108792>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Originally published in:

Agricultural Economics 44(2), <https://doi.org/10.1111/agec.12005>

Postprint

This is the accepted version of a paper published in *Agricultural Economics*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper:

Finger, R. (2013). Investigating the performance of different estimation techniques for crop yield data analysis in crop insurance applications. *Agricultural Economics* 44: 217–230.
DOI: 10.1111/agec.12005

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Improving the Analysis of Crop Yield Data for Insurance Applications

Abstract

We investigate the performance of the OLS-, M-, MM- and the Theil-Sen-estimator for crop yield data analysis in crop insurance applications using Monte Carlo simulations. In particular, the performance is assessed with respect to trend estimation, prediction of future yield levels and the estimation of expected indemnity payments. In agreement with earlier findings, other estimators are found to be superior to OLS in simple regression problems if yield distributions are outlier contaminated and heteroscedastic. While this conclusion is also valid for subsequent applications such as yield prediction and the estimation of expected indemnity payments, the difference between the considered estimators becomes less distinct. For these applications, we find particularly the M-estimator to be a good compromise between high breakdown (very robust) estimators and the very efficient OLS-estimator. Because no regression technique dominates all others in all applications and scenarios for error term distributions, our results underline that the choice of the estimation technique should be dependent on the purpose of the crop yield data analysis. However, alternative estimators such as M-, MM- and Theil-Sen-estimator can reduce (and bound) the risk of unreliable or inefficient trend estimation and applications.

1 Introduction and Motivation

Crop yield insurances are of increasing importance in agriculture (Bielza et al., 2008, Ozaki and Silva, 2009). In order to reduce the potential of market failures, appropriate design and specification of insurance contracts is of outmost importance. In particular, the identification of crop yield distributions and its parameters is a crucial step for the estimation of insurance premiums and the specification of trigger yield levels. To quantify risk exposure in crop production, it is essential to identify and remove deterministic trends in time series of crop yield data because crop yield variability would be overestimated otherwise. Thus, the detrending of crop yield data is usually a central aspect in the analysis of crop yield data for insurance applications and the determination of crop yield distributions.

However, several alternatives to this trend removal approach have been suggested in the literature. For instance, the use of time series components and the simultaneous estimation of time trends and yield distribution parameters have been suggested (e.g. Bessler, 1982, Zhu et al., 2008). The use of deterministic trend models has been also critically discussed in other economic applications (e.g. Nelson and Plosser, 1982). But, even though such alternative procedures may be superior in certain situations, detrending procedures consisting of trend identification and removal are predominantly used in current research and insurance applications (see e.g. Claasen and Just, 2011, Woodard et al., 2010, 2011, for recent overviews). Furthermore, observed trends in crop yield data can be used to specify trigger (expected) yield levels for subsequent years and can thus be valuable, among others, for insurance applications.

Within the literature on crop yield data detrending, failures in trend identification have been found to cause misleading results in crop insurance analyses. Atwood et al. (2003) and Just and Weninger (1999), for instance, demonstrate the sensitivity of crop yield distribution specification and crop insurance design to specific detrending methods. Thus, improving this step of crop yield data analysis has been a central aspect in a wide range of publications. Recently made attempts to improve detrending procedures addressed for example the level of data aggregation (Atwood et al., 2003, Hennessy, 2009) and linearity assumptions of trends in crop yield data (Just and Weninger, 1999, Ramirez et al., 2003). Furthermore, proposals have been made to improve procedures of normality testing (Just and Weninger, 1999), techniques for estimating distribution moments (Ker and Goodwin, 2000, Ramirez et al., 2003) and the underlying detrending methodology (e.g. Racine and Ker, 2006, Skees et al., 1997) (see Goodwin and Ker, 2002, for an overview). Along these lines, also robust regression techniques have been evaluated to overcome the outlier sensitivity of standard regression approaches to crop yield data detrending (Swinton and King, 1991, Finger, 2010a). Finger (2010a) showed that the MM-estimator is appropriate for detrending of crop yield time series for uncontaminated yield data and outperforms Ordinary Least Squares (OLS) estimation if yield observations are contaminated with outliers.

The existing studies have, however, not considered the impact of the (joint) appearance of other frequently observed data problems that can seriously affect the performance of the OLS estimator¹. Namely, empirical evidence suggests that besides containing outliers (e.g. caused by climatic extreme events such as droughts), crop yield distributions are often negatively

¹ Even though problems and potential pitfalls of using OLS are well known, it remained the default regression technique to analyze crop yields (see Finger, 2010a, for examples).

skewed and/or heteroscedastic² (e.g. Finger, 2010b, Gallagher, 1987, Just and Weninger, 1999, Kaylen and Koroma, 1991, Swinton and King, 1991). Furthermore, the implications of employed methods in trend analysis on subsequent steps of data analysis such as yield prediction and estimation of expected insurance indemnities have not been investigated so far. These are, however, of particular importance because the detrending of crop yield data is usually only the first step in crop yield data analysis. Moreover, the evaluation of alternative regression techniques conducted by Swinton and King (1991) and Finger (2010a) considered only specific estimators (the trimmed least squares and the MM-estimator, respectively), not reflecting the wide range of estimators that are available and used in practice³. However, the comparison of different methodical approaches under a wide range of data situations can lead to an improvement of applications of crop insurance concepts (e.g. Woodard et al., 2011).

In this paper, we fill this gap and thus expand the scope of earlier studies a) by using homoscedastic/heteroscedastic, normally/non-normally distributed and outlier-free/outlier-contaminated error term distributions and combinations thereof, b) by directly addressing insurance applications, namely the prediction of (future) trigger yields and the estimation of expected indemnities and c) by analyzing additional robust estimators.

² More general, crop yields are often characterized by distributions that change over time either by systematic changes in yield variability and/or by outlying yield observations (e.g. Claderini and Slafer, 1998).

³ Robust regression techniques, e.g. the M-estimator, have been applied for detrending, for instance, by Harri et al. (2009), Ker and Coble (2000), and Skees et al. (1997).

2 Methodology

If the assumption of normally distributed error terms does not hold for crop yield data, the application of OLS regression might be inefficient or even a source of potential errors in trend estimation and subsequent applications such as yield prediction and estimation of expected insurance indemnities. For instance, other estimators are more efficient if observations are non-normal and/or heteroscedastic (e.g. Johnston and DiNardo, 1997). Furthermore, OLS has an unbounded influence function and a breakdown point of zero⁴. This means that an arbitrarily small fraction of observations that deviates from the pattern described by the majority of observations (i.e. outliers) can change the OLS coefficients to any value from $-\infty$ to $+\infty$ (Zaman et al., 2001). OLS is thus not a reliable detrending technique in presence of outliers, particularly for the short time series that are usually available in crop data analysis. In contrast, robust regression techniques are more appropriate in these situations (see e.g. Hampel, 2001, 2002, Hampel et al., 1986). Even though there has been a (general) lack of robust estimation techniques in economics in the past (Zaman et al., 2001), they are expected to play a major role in the future of econometrics applied in agricultural economic analyses (Bessler et al., 2010).

2.1 Regression Analysis and Regression Techniques

We focus on linear time trends in crop yields that can be described as follows:

⁴ The breakdown point gives, roughly, the smallest amount of contamination that may cause an estimator to take on arbitrarily large aberrant values, i.e. which can carry this estimator over all bounds Ruckstuhl (1997).

$$(1) \quad Y_t = \beta_0 + \beta_1 t + \lambda(t)\varepsilon_t,$$

where Y_t is the crop yield in year t , ε_t are random errors that are independent and identically distributed, λ is a function of t , and β_0 and β_1 are unknown parameters. The linear trend in crop yields is indicated by the parameter β_1 , which is of interest in the here presented analysis. If $\lambda(t) = 1$ (error terms are homoscedastic) and ε_t are normally distributed with mean zero and variance σ^2 , OLS is the most efficient linear unbiased estimator (Johnston and DiNardo, 1997). This paper focuses particularly on violations of these assumptions: non-normal distribution, heteroscedasticity and outlier contamination of the error terms. In the remainder of this section, the M-, MM-, and the Theil-Sen (TS)-estimator are described.

The M-estimator⁵, has been proposed by Huber (1972, 1973) based on the idea to replace OLS residual weighting (using squared residuals) by a function that bounds the influence of outlying observations. Regression coefficients are thus chosen to:

$$(2) \quad \min_{\hat{\beta}} \sum_{i=1}^n \rho(r_i)$$

Where $\hat{\beta}$ are the coefficient estimates, r_i regression residuals and ρ is a loss function that's represents the weighting scheme. Here, a Huber-type loss function is employed that down-weights outliers but does not assign zero weights to observations (see Hampel et al., 1986, for details). The estimator is derived using iteratively re-weighted least squares (e.g. Maronna

⁵ M-estimator denotes "Maximum likelihood type estimator" (Huber, 1972), and distinguishes a class of estimators from other approaches to robust statistics (see Huber, 1972, and Maronna et al., 2006, for an overview).

et al., 2006). This M-estimator is statistically more efficient⁶ than many other robust estimators and is robust against vertical outliers (outliers in the y-direction). However, outlying observations in x-direction (leverage points) lead to a breakdown of the M-estimator and it thus has an asymptotic breakdown point of 0 (Rousseeuw and Leroy, 1987). Though the M-estimator is not in general robust, it is expected to have good robustness properties in the context of detrending because potential leverage effects are bounded as no large outliers in the time component can occur.

To overcome the potential drawbacks of the M-estimator with respect to insufficient robustness properties, Yohai (1987) proposed the MM-estimator that combines high efficiency (i.e. a low standard error of the estimator for non-contaminated distributions) with the highest possible asymptotic breakdown point of 0.5. The MM-estimator is based on an iterative M-estimation approach with robust starting value estimates for the scale, and regression coefficients derived from an S-estimator (Rousseeuw and Yohai, 1984). The latter is an iterative estimator that minimizes the (robust estimated) scale of the regression residuals (Lee and Oh, 2007, Yohai, 1987). The scale estimate $\hat{\sigma}$ and the associated regression estimate β_S of this S-estimator are the solution of:

$$(3) \quad \beta_S = \arg \min_{\beta} \hat{\sigma}(\beta).$$

The starting values for the scale and coefficients derived from S-estimation are subsequently employed in the M-estimation scheme described in Equation 2. For the MM-estimation, this estimation step uses Tukeys's biweight function as loss function, which down-weights

⁶ Even though efficiency is often evaluated at normal distributions (to allow for comparisons with the OLS estimator), this does not necessarily imply sudden changes of efficiency properties of robust estimators for other error term distributions (e.g. Maronna et al., 2006).

outliers and assigns zero weights to gross outliers (see Hampel et al., 1986, for details). Due to the combination of the high-breakdown S-estimator with an efficient M-estimator, the MM-estimator can resist against an outlier contamination of up to 50% and overcomes the lack of efficiency of other high breakdown estimators. Due to these properties, the MM-estimator has become one of the most recommended robust regression techniques in the current robust statistics literature (e.g. Maronna et al., 2006).

Note that even though M- and MM-estimator are assumed to perform superior to OLS in presence of outliers, no differences in the performance of these estimators compared to OLS in case of heteroscedasticity or the presence of skewed distributions are expected.

To overcome this potential limitation, we additionally consider the Theil-Sen (TS)-estimator, a non-parametric regression technique (Theil, 1950, Sen, 1968). The slope estimator is defined as the median of the pairwise slopes of all data points (see e.g. Dietz, 1989, Fernandes and Leblanc, 2005, Wilcox, 1998a,b for details)⁷. For any $X_i > X_j$, let $b_{ij} = (Y_j - Y_i)/(X_j - X_i)$. The coefficient estimate for $\hat{\beta}_1$ is then defined as the median of the N possible values of b_{ij} (Wilcox, 1998a,b). Thus, the TS-estimator is the median of pairwise slopes of the observations⁸ and has some very interesting features in the context of applied regression analysis (Dietz, 1989, Wilcox, 1998a,b): a) It is relatively efficient if regression residuals follow outlier-free and homoscedastic distributions, b) It is substantially better (i.e. estimates have smaller standard errors) than OLS and other estimators when the

⁷ A similar (using the mean instead of the median) method was already applied in the 1750's by Roger Joseph Boscovich (see Hössjer et al., 1994, Stigler, 1986, for details). Determination of confidence intervals and hypothesis testing for the TS-estimator are based on Kendall's Tau rank correlation coefficient (Sen, 1968).

⁸ Overviews on similar estimators that are based on pairs of observations are, for instance, presented in Frees (1991), Hössjer et al. (1994) and Sen (1968).

error term is heteroscedastic, and c) it is robust against outliers with an asymptotic breakdown point of 0.293.

In summary, the M-, MM- and the TS-estimator are expected to perform similar to OLS for outlier-free, homoscedastic and symmetrically distributed error terms, but are expected to perform substantially better than OLS in presence of outliers, even though the M- and the TS-estimator have a smaller breakdown point than the MM-estimator. Moreover, the TS-estimator is expected to offer an advantage over the other methods when the error term is heteroscedastic. It is important to note that we expect no estimator to dominate all others in all situations. The purpose of our analysis is thus to investigate the trade-offs associated with the use of a specific estimator.

2.2 Monte Carlo Simulations

2.2.1 Employed Crop Yield Distributions

We use Monte Carlo experiments to evaluate the different regression methods with regard to their suitability for crop yield data detrending. To this end, time series of crop yield data (with a sample size of 10 years of crop yield data) are generated under the assumption of a linear trend equal to 2 (e.g. decitonnes) of yield gain per year and a random error drawn from different distributions. We employ 12 scenarios of error term distributions that are summarized in Table 1⁹. Scenarios 1-3 assume homoscedastic and outlier-free distributions.

⁹ Using this broad set of error term distribution extents the simulation studies conducted by Swinton and King (1991) and Finger (2010a) and thus allows drawing much broader conclusions.

More specifically, we assume a normal ($\mu = 0, \delta = 25$) and two (negatively) skewed Beta distributions ($a=5, b=3$ and $a=15, b=3$). The assumed error term distributions follow Swinton and King (1991) and reflect crop yield distributions observed in empirical studies. The Beta distributions are zero-centered and rescaled over a range of 100. The Beta(5,3) and Beta(15,3) distribution have standard deviations of 16.14 and 8.55 as well as a skewness of -0.31 and -0.78, respectively. Thus, the Beta(15,3) distribution has a lower dispersion and higher (negative) skewness compared to the Beta(5,3) distribution. In scenarios 4-6, heteroscedastic error terms are generated based on the 3 initial distributions by assuming $\lambda(t) = 0.5 \cdot t$. In order to additionally reflect outlier contamination, scenarios 1-6 are combined with an outlier in the scenarios 7-12. In these scenarios, the last observation in the time series is replaced by an observation that is drawn from a normal distribution with five times higher standard deviation than the underlying distribution¹⁰.

Table 1. Employed Error Term Distributions.

Scenario	Base Distribution	Heteroscedasticity	Outlier Contamination
1	Normal ($\mu = 0, \delta = 25$)	Homoscedastic ($\lambda(t) = 1$)	No outlier
2	Beta ($\alpha = 5, \beta = 3$)	Homoscedastic ($\lambda(t) = 1$)	No outlier
3	Beta ($\alpha = 15, \beta = 3$)	Homoscedastic ($\lambda(t) = 1$)	No outlier
4	Normal	Heteroscedastic	No outlier

¹⁰ A five times higher standard deviation was chosen to create outliers because this ensures that both small deviations and gross outliers are present in outlier contaminated distributions.

	$(\mu = 0, \delta = 25)$	$(\lambda(t) = 0.5 \cdot t)$	
5	Beta $(\alpha = 5, \beta = 3)$	Heteroscedastic $(\lambda(t) = 0.5 \cdot t)$	No outlier
6	Beta $(\alpha = 15, \beta = 3)$	Heteroscedastic $(\lambda(t) = 0.5 \cdot t)$	No outlier
7	Normal $(\mu = 0, \delta = 25)$	Homoscedastic $(\lambda(t) = 1)$	Last observation is replaced with an outlier
8	Beta $(\alpha = 5, \beta = 3)$	Homoscedastic $(\lambda(t) = 1)$	Last observation is replaced with an outlier
9	Beta $(\alpha = 15, \beta = 3)$	Homoscedastic $(\lambda(t) = 1)$	Last observation is replaced with an outlier
10	Normal $(\mu = 0, \delta = 25)$	Heteroscedastic $(\lambda(t) = 0.5 \cdot t)$	Last observation is replaced with an outlier
11	Beta $(\alpha = 5, \beta = 3)$	Heteroscedastic $(\lambda(t) = 0.5 \cdot t)$	Last observation is replaced with an outlier
12	Beta $(\alpha = 15, \beta = 3)$	Heteroscedastic $(\lambda(t) = 0.5 \cdot t)$	Last observation is replaced with an outlier

2.2.2 Trend Estimation

First, we investigate the ability of regression techniques to identify and estimate a trend in time series of crop yield data. Regarding agricultural economic analysis, this addresses the question if and how to remove the trend from crop yield data. In order to compare the estimation techniques, the identical simulated time series of crop yield observations are detrended using OLS-, M-, MM- and TS-estimation. This process is repeated 9999 times for

each distribution scenario (cp. Table 1). In contrast to earlier simulation studies (Swinton and King, 1991, Finger, 2010a), we have chosen a higher number of simulations (9999 instead of 1000) to allow for more accurate inference. The 9999 trend coefficient estimates resulting from these simulations are summarized with the mean, standard deviation (SD), minimum (Min) and maximum (Max) values as well as the median absolute deviation (MAD)¹¹. Since outliers are generated from symmetric distributions with zero mean, we expect differences across different effects of outliers on estimation results to appear only in the measures of dispersion, such as range, standard deviation and MAD of the trend coefficient. To test for differences in the dispersion of coefficient estimates, the Ansari-Bradley test (Ansari and Bradley, 1960) is used. The hypotheses based on the properties of the estimators are summarized in Table A2.

We use the ‘MASS’, the ‘zyp’ and the ‘boot’ package from the statistical language and environment R (R Development Core Team, 2008) for M- and MM-estimation, TS-estimation and Monte-Carlo analysis, respectively.

2.2.3 Prediction of Future Yield Levels

Second, we compare the regression techniques with respect to their ability to predict future yield levels. This criterion reflects the need to investigate the calibration performance of crop yield trend models, complementing in-sample performance assessments. Thus, we additionally assess the “goodness” of regression models (and techniques) in terms of actual

¹¹ The MAD is used to compare the 9999 coefficient estimates for each scenario and regression technique with the known underlying true slope coefficient that is equal to 2.

realizations on a series of sequential probability forecasts. Discussions and applications of this calibration performance criterion are presented, for instance, in Dawid (1984, 2006), Kling and Bessler (1989) and Galbraith and van Norden (2011). Furthermore, this criterion on prediction confidence has straightforward implications for crop insurance applications. It is, for instance, relevant for farm yield insurances that are based on farm-level yield history (see Barnett et al., 2005, Bielza et al., 2008 for details). In this insurance scheme, the farmer is indemnified if the actual crop yield (Y_i) falls below the critical yield (Y^C), which is defined as a percentile of the expected yield level. But, an expectation (forecast) of future yield levels is required for such insurance. Due to limited in-time data availability, even a 2-year ahead specification of critical yield levels is often necessary in practice¹². In US crop insurance schemes, such forecast is usually based on 5-year Olympic averages (e.g. Zulauf et al., 2008), the average yield using the last 5 year observations and dropping the highest and the lowest value. This choice is motivated by the fact that arithmetic averages are highly efficient, and the trimming of high and low observations introduces robustness to outlying observations. However, Olympic averages give biased forecasts of future yields in presence of a trend. Because trend estimates (detrending methods) are expected to lead to unbiased predictions also in presence of a trend, they may be a valuable alternative to make yield predictions.

To compare the performance of the estimators with respect to yield prediction, we use the Monte-Carlo simulation setup described in 2.2.1, but simulate time series of crop yield data

¹² This is, for instance, the case for the Group Risk Plan Insurance in the US (e.g. Goodwin and Ker, 1998, Ker and Goodwin, 2000)

for 12 years. Trend estimates used to make forecasts are based on the first 10 years of each time series. Regression estimates are then used to predict crop yield in year 12, which is compared with the realized (known from the simulations) yield level in this year. In this out-of-sample performance analysis, also the Olympic average (based on the years 6 to 10) is considered. To test for biases in the forecasts based on different estimators, the Mann-Whitney test is used. Differences in the efficiency of the forecast (or prediction confidence) are assessed with the Ansari-Bradley test (hypotheses are summarized in Table A4). The smaller the variability of the prediction error of the 2-year ahead forecast (and if the estimate is unbiased), the more appropriate is the forecast technique¹³.

2.2.4 Analysis of Expected Indemnities

Besides providing reliable and efficient trend estimates as well as unbiased and efficient yield forecasts, an appropriate detrending technique is also expected to correctly identify deviations from the average yield levels. In particular, yield levels below specific levels of average yields are important in an insurance setting because such yield losses can be used to determine expected indemnities and insurance premiums. More specifically, for the farm yield insurance described above, the indemnity payment function has the following form (for simplicity, the price is not considered):

$$(5) \quad \text{Indemnity}_i = \max\{0, Y^C - Y_i\}$$

¹³ To assess model performance on broader criteria, we also conducted the analysis based on 1- and 3-year ahead forecasts. Qualitative interpretation of the results remains unaffected by this choice, and only results for the 2-year ahead forecast are thus presented in this paper.

In year i , the farmer is indemnified if his actual yield level is below the critical yield level, with a linear payoff function (indemnity payments increase linearly with yield losses). A ‘fair premium’ for such insurance would be equal to the expected indemnity payments, e.g. based on the actual production history of a specific farm.

We use the Monte-Carlo simulations presented in 2.2.1 (with length of 10 years) to calculate the expected (average) indemnity, EI_0 , from the known (prescribed) trend line assuming a 75% coverage level (the trigger yield equals 75% of the prescribed trend line of $2i$):

$$(6) \quad EI_0 = \frac{1}{10} \sum_{i=1}^{10} \max\{0, 0.75 (2i) - Y_i\}$$

Based on the same Monte-Carlo simulations, expected indemnities are also calculated for each of the k detrending techniques (OLS-, M-, MM- and TS-estimation). Instead of the deviation from the prescribed trend line, deviations from the estimated trend lines (based on the coefficient estimates) are considered here:

$$(7) \quad EI_k = \frac{1}{10} \sum_{i=1}^{10} \max\{0, 0.75 (\widehat{\beta}_0^k + \widehat{\beta}_1^k i) - Y_i\}$$

Where k denote the different regression techniques, $\widehat{\beta}_0^k$ and $\widehat{\beta}_1^k$ are the coefficient estimates for the intercept and the slope, respectively. Because the 9999 time series of crop yield data for each of the 12 scenarios (Table 1) are identical for each detrending procedure, we can compare the known expected indemnity (EI_0) with the expected indemnity payments indicated by the detrending techniques (EI_k) using Spearman and Kendall rank correlations. The different detrending techniques are ranked according to their correlation with the known expected indemnity. In particular, a high correlation with EI_0 shows that the investigated

detrending technique appropriately reproduces the deviations from the underlying trend line. In order to analyze the sensitivity of our analysis with respect to the chosen coverage level, we additionally conducted the entire analysis using a coverage level of 50%.

3 Results

3.1 Results for the Monte-Carlo analyses on Trend Estimation

The results of the Monte Carlo experiment with regard to trend estimation for the outlier-free error term scenarios 1-6 are shown in the left panel of Table A1. Moreover, density functions of the coefficient estimates (9999 for each estimator and scenario combination) are displayed in Figure 1. Results of hypotheses testing are summarized in Table A2.

It shows that OLS leads to smaller standard deviations and MADs than the MM- and the TS-estimator for the homoscedastic and outlier free normal and Beta(5,3) distribution (scenarios 1 & 2). Furthermore, the range of coefficient estimates derived from OLS estimation is smaller than those derived from the MM- and TS-estimator. The M-estimator, however, performs similar to OLS and thus better than the MM- and the TS-estimator. The loss of efficiency by using the latter two robust estimators (expressed as the increase of the standard deviation of the 9999 coefficient estimates) is in the range of 7%, while the efficiency loss of using the M-estimator is in the range of 3-4%. For the more skewed Beta(15,3) distribution, no differences between OLS and the M-, MM- and TS- estimator are found. In summary, OLS performs better than the M-, MM- and the TS-estimator for homoscedastic and outlier free error terms if these follow non-skewed (Normal(0,25)) or only slightly

skewed (Beta(5,3)) distributions. However, this advantage of OLS disappears for the more skewed (Beta(15,3)) distribution.

< Figure 1. Density Functions of the Coefficient Estimates for Scenarios 1-6. >

The results of the Monte Carlo experiments for scenarios 4-6, where error terms are outlier-free but heteroscedastic, are shown in the bottom left part of Table A1. We find that the TS-estimator is (significantly) superior to the other estimators for all considered heteroscedastic distributions because it leads to the smallest standard deviations and MADs. The efficiency gain from using the TS instead of the OLS-estimator ranges from 2% (for the Beta(5,3) distribution) to 10% (for the Beta(15,3) distribution). The MM-estimator performs inferior to the OLS-, M- and the TS-estimator for heteroscedastic error term distributions.

< Figure 2. Density Functions of the Coefficient Estimates for Scenarios 7-12. >

The right panel of Table A1 shows the results of the Monte Carlo experiment for the outlier contaminated error term distributions in scenarios 7-12. Density functions of the coefficient estimates (9999 for each estimator and scenario combination) are displayed in Figure 2 and the results of hypotheses testing are presented in Table A2. We find OLS to be inferior to the M-, MM- and TS-estimator for all considered outlier contaminated error term distributions (scenarios 7-12). For scenarios 7-9, considering homoscedastic and outlier contaminated distributions, the MM-estimator is superior to the TS-estimator, while the latter two estimators are superior to the M-estimator. For outlier contaminated error terms, the use of

OLS leads to an increase of the standard deviation of coefficient estimates between 79% (for the Beta(5,3) distribution) and 85% (for the Normal(0,25) distribution) compared to results derived from the MM-estimator. Using the M- and the TS-estimator instead of the MM-estimator leads to small efficiency losses that range from 5% to 8%.

For heteroscedastic and outlier contaminated error terms (scenarios 10-12), the TS-estimator performs better than the MM- and M-estimator, while all three are superior to the OLS-estimator. OLS leads to an efficiency loss of 16% (for the Beta(5,3) distribution) – 33% (for the Normal(0,25) distribution) compared to the TS-estimator. Almost all reported differences for the scenarios 7-12 are significant at the 1% level of significance (see Table A2). Note that the effects of outliers are by definition smaller in scenarios 10-12 than in scenarios 7-9 because the “natural” spread of observations at the end of the time series is higher due to the increasing variance of error terms over time.

3.2 Results for the Monte-Carlo analyses on Yield Forecast

The left panel of Table A3 displays the results of the Monte-Carlo analyses with regard to yield forecast, i.e. out-of-sample prediction, for the scenarios 1-6. Density plots of the prediction errors of the 2-year ahead forecast for the OLS-, M-, MM-, TS-estimator and the Olympic average (OA) are presented in Figure 3. We find the OA to systematically underestimate future yield levels. This leads to a positive bias in the difference of observed and predicted yield level because the OA does not account for any trend in the crop yield data. The Mann-Whitney test indicates significant differences to the prediction errors of the other estimators for all scenarios (not shown). However, the OA leads – in all cases – to the

narrowest ranges of prediction errors and the smallest standard deviations. This difference in the dispersion of prediction errors between OA forecasts and forecasts based on regression techniques is significant in all cases (Table A4). Among the regression techniques, OLS based forecasts lead to slightly lower dispersion in the prediction errors for scenarios 1-3 (and thus the highest prediction confidence), while TS based forecasts have the highest dispersion. For heteroscedastic error term distributions (scenarios 4-6), also the regression techniques based forecasts face small biases, shown by nonzero means of prediction errors. OLS generates the less biased and most precise yield forecasts. However, the differences between the regression techniques are small if crop yield distributions are not contaminated with outliers. For instance, the efficiency loss expressed as the increase of the standard deviation (Table A3) of using the MM- instead of the OLS-estimator is in the range of 1% (scenario 3) and 4% (scenario 5).

Results for the scenarios with outlier contamination (scenarios 7-12) are shown in the right panel of Table A3 with density plots presented in Figure 4. The Olympic average leads, again, to estimates that are systematically biased and have a significantly smaller dispersion than for all other estimators (Table A4). Moreover, it shows that the Olympic average provides robustness against (at least single) outliers. Thus, its performance is not affected by a single outlying observation within the 5 year period considered for its calculation.

In contrast to the results for scenarios 1-6, OLS performs significantly worse than the robust regression estimators for homoscedastic and outlier contaminated error term distributions in scenarios 7-9 (Table A3 and A5). This is indicated by a large increase in variability of the prediction errors of the OLS-based forecasts. OLS based forecasts are thus not reliable in presence of outliers. In contrast, this is not the case for the robust regression techniques (M-

, MM- and TS-estimator). Among these robust regression estimators, the MM-estimator is the superior method, leading to significantly smaller dispersion of prediction errors than the M- and the TS-estimator. The efficiency loss, expressed as change of standard deviation (cp. Table A3), of using OLS instead of the MM-estimator for scenarios 7-9 is about 62%.

Assuming heteroscedastic and outlier contaminated error term distributions in scenarios 10-12, the OA leads again to the less dispersed but biased yield predictions. Among the regression estimators, prediction errors of the 2-year ahead forecasts of the M-estimator have the smallest dispersion. However, the differences between the regression techniques are small. For instance, the efficiency loss of using the MM- instead of the M-estimator is about 2% for the scenarios 10-12.

3.3 Results for the Monte-Carlo analyses on Expected Indemnities

In a third application, expected indemnity payments are calculated for each detrending technique and compared with the indemnities calculated based on the known (prescribed) trend line (see Equations 6 and 7). The correlations derived for the scenarios 1-12 are presented in Table 2. It shows that for outlier-free error term distributions in scenarios 1-7, OLS has the highest correlation with the known average indemnity payments. Thus, even though it is not necessarily the most efficient regression technique to estimate the underlying trend line (cp. section 3.1), OLS is the best trend estimation technique to represent the underlying risk in crop production. The M-estimator performs similar to OLS and leads to higher correlation with the known average indemnity payments than the other robust estimators. Surprisingly, even though the TS-estimator is the most efficient estimator for

trend estimation with heteroscedastic error terms (section 3.1), the indemnity payments estimated from the TS- regression residuals show the lowest correlation, i.e. fail to adequately reproduce the underlying structure of deviations from the trend line.

For homoscedastic and outlier contaminated error term distributions (scenarios 7-9), OLS based estimates for the expected indemnity payment are inferior to those derived with the robust detrending techniques. OLS trend line estimates tend to shift in direction of the outlier (i.e. break down) and thus can fail to identify this outlier as a source of higher risks. The MM- and the M-estimator show the best performance in reproducing the underlying average indemnity payments. For heteroscedastic and outlier contaminated error terms (scenarios 10-12), however, the advantage of robust estimators disappears and OLS leads to the highest correlations between estimated (based on the estimated trend line) and known (deviations from the prescribed trend line) indemnity payments. Though the OLS trend line estimate is inefficient compared to the other estimators (cp. Figure 2), the regression residuals from the OLS trend line estimation indicate the risks to fall below the critical yield levels better than the other estimators. Again, note that outlier effects are smaller in scenarios 10-12 than in scenarios 7-9 due to the heteroscedasticity of error terms.

To analyze the sensitivity of these results with respect to the chosen coverage level, we conducted the entire analysis using a smaller coverage level of 50% instead of 75%. This leads to a large increase of the correlations (not shown), which indicates that all detrending methods reproduce the average indemnity payments regarding more extreme yield events very precisely. The qualitative interpretation of the above presented results remains however unaffected by the choice of the coverage level. Using Kendall instead of Spearman rank

correlation leads to generally lower correlation coefficients (not shown), but does not affect the ranking between estimators.

Table 2. Ranking of Regression Techniques Based on Correlations (in Parentheses) Between Known and Estimated Expected Indemnities.

	OLS	MM	TS	M
Scenario 1	1 (0.76)	3 (0.70)	4 (0.59)	2 (0.72)
Scenario 2	1 (0.81)	3 (0.75)	4 (0.62)	2 (0.77)
Scenario 3	1 (0.89)	3 (0.85)	4 (0.77)	2 (0.86)
Scenario 4	1 (0.77)	3 (0.66)	4 (0.62)	2 (0.72)
Scenario 5	1 (0.83)	3 (0.69)	4 (0.65)	2 (0.76)
Scenario 6	1 (0.90)	3 (0.81)	4 (0.78)	2 (0.86)
Scenario 7	4 (0.67)	1 (0.84)	3 (0.79)	1 (0.84)
Scenario 8	4 (0.71)	1 (0.86)	3 (0.80)	1 (0.86)
Scenario 9	4 (0.78)	2 (0.90)	3 (0.86)	1 (0.91)
Scenario 10	1 (0.75)	3 (0.68)	4 (0.64)	2 (0.74)
Scenario 11	1 (0.81)	3 (0.71)	4 (0.66)	2 (0.77)
Scenario 12	1 (0.88)	3 (0.80)	4 (0.77)	2 (0.85)

Values in parentheses are Spearman rank correlations between known expected indemnities and those estimated from the regression residuals of different estimation techniques. The presented ranking is made according to the correlation coefficient.

4 Discussion

Even though the simulations presented in this paper covered several scenarios of crop yield distributions and estimators, these analyses are just a small part of possible crop yield data situations. Thus, the here derived conclusions may not be generally valid. In particular, our

analysis was restricted on the specific example of a linear trend of crop yield growth over time. However, crop yields may also show nonlinear increases over time. The (statistical) properties of the different estimators may give, however, the possibility to generalize some of the here derived results. In contrast to robust estimators, OLS is expected to suffer from outliers, no matter what kind of trend is estimated. In its basic form presented and used here, the Theil-Sen estimator may not be applicable to multivariate or non-linear trend estimation (e.g. a quadratic trend), but generalizations of the Theil-Sen approach are available for these problems (e.g. Zou and Serfling, 2008).

Our numerical simulations were restricted on a single sample size of 10 years of crop yield observations. This choice represents the usual crop yield data availability at the farm level. This is due to the fact that crop rotation requirements, abandonment, merger or start-up of farms, and changes of product plans often lead to a lack of continuous farm-level data over long time periods. However, we are aware that the length of the time series has implications for the vulnerability of an estimator to outliers. More specifically, the effect of a single outlier decreases with increasing sample size. Thus, differences between OLS and robust estimators decrease with increasing sample size. In contrast, an increase of the number of outliers or the reduction of the available observations (e.g. due to missing values) leads, *ceteris paribus*, to a larger vulnerability of the OLS-estimator. A numerical analysis of these issues is provided by Finger (2010b).

In addition to the number of outlier(s), their position within the time series of crop yield data is important. We have chosen to place outliers at the end of the time series¹⁴ because outliers

¹⁴ The same effect would be observable for outliers at the beginning of the time series or, more general, for outliers with leverage effect.

near the center of the explanatory variable mainly affect the regression intercepts and not the slope estimate, which was of interest here (see Rousseeuw and Leroy, 1987, for examples and discussions). However, an outlier in the center of the time series might have implications for yield prediction or regression residuals, which should thus be addressed in future research. The set of analyzed estimators was limited to those recommended from the statistical point of view (MM- and TS-estimator) and the M-estimator, which has been already applied to crop yield detrending (e.g. Harri et al., 2009). Other specific estimation problems might require other detrending techniques. For instance, time series approaches (e.g. Ker and Goodwin, 2000) or spline regressions (e.g. Skees et al., 1997) are used for crop yield detrending. For these alternatives to standard regressions, also robust (in the sense of outlier resistance) estimation procedures are available which are beyond the scope of this paper (see e.g. Lee and Oh, 2007, Maronna et al., 2006, Rousseeuw and Hubert, 1999 for examples).

5 Summary and Conclusion

In agreement with the literature on robust estimators and earlier simulation studies (e.g. Maronna et al., 2006, Finger, 2010a) we find that OLS is only superior to other estimators in trend estimation problems for homoscedastic, outlier free, unskewed error term distributions. However, the efficiency loss of using robust estimators such as the M-, MM- or the TS-estimator is small (about 3%-7%) in these situations (e.g. for normal distributions). In contrast, we find OLS to be inferior to the other estimators if error term distributions are heteroscedastic, outlier contaminated or both, with efficiency losses of up to 85%. In particular, we find the TS-estimator to be superior for heteroscedastic distributions with and

without outlier contamination, while the MM-estimator is the superior estimator for homoscedastic, outlier contaminated error term distributions.

Furthermore, we investigated the out of sample performance (yield prediction) of regression techniques. This application is motivated by the fact that expected yield levels have to be specified in insurance contracts up to two years in advance. Comparing the regression techniques according to their ability to accurately predict future yield levels, it shows that OLS based forecasts are not reliable in presence of outliers. In contrast, this is not the case for robust regression techniques. Furthermore, also the Olympic average (OA), which is frequently applied in practice, provides robustness against single outliers and is more efficient than the regression techniques. However, the OA systematically underestimates future yield levels in presence of deterministic trends in crop yields. Thus, the OA is the most recommended yield prediction technique if and only if no trend is observed in time series of crop yield data. In contrast, robust regression techniques should be used to forecast yield levels if crop yields increase over time. In particular the M-estimator is found to be a good compromise between very robust (MM- and TS-) and the OLS-estimator according to its ability to predict future yield levels.

Finally, we investigated the ability of detrending techniques to correctly identify historical deviations from average yield levels in order to specify expected insurance indemnities (and thus insurance premiums). Using a farm-yield insurance example, we find that indemnity estimation based on OLS trend estimation in presence of outliers is inferior to those derived with robust regression techniques. In particular, OLS based trend line estimates tend to shift in direction of the outlier and thus can fail to identify this outlier as large deviation from the

underlying trend line. However, OLS based estimates of expected indemnities showed the highest correlation with the underlying (known) expected indemnities for the other scenarios. The M-estimator performed similar to OLS under most error term distribution scenarios, but leads to estimates of expected indemnities that are not affected by outliers. Thus, the M-estimator should be used to specify expected indemnities.

In conclusion, our analysis shows that alternative regression techniques such as the M-, MM- and the TS-estimator can be more appropriate and reliable than the usually applied Ordinary Least Squares estimator for crop yield data analysis. This is particularly the case for the identification and estimation of trends in the crop yield data. This conclusion also holds for crop yield insurance applications based on trend estimation such as yield prediction and the estimation of expected indemnity payments. However, for these applications the difference between the considered estimators with respect to their performance becomes less distinct.

Our results reveal that no regression technique dominates all others in all applications and scenarios of error term distributions. Thus, the choice of the detrending technique should be dependent on the purpose of the crop yield data analysis. However, in all applications OLS regression is superior only in very specific situations and implies high risks of drawing misleading conclusions if applied for other than standard normal error term distributions. In contrast, alternatives such as M-, MM- and TS-estimators clearly reduce (and bound) the risk of unreliable or inefficient trend estimation and applications. In particular, the M-estimator is found to be a good compromise between very robust estimators and the very efficient OLS-estimator.

References

- Ansari, A.R., Bradley, R.A., 1960. Rank-Sum Tests for Dispersions. *Ann. Math. Stat.* 31, 1174-1189.
- Atwood, J., Shaik, S., Watts, M., 2003. Are Crop Yields Normally Distributed? A Reexamination. *Am. J. Agr. Econ.* 85, 888-901.
- Barnett, B.J., Black, J.R., Hu, Y., Skees, J.R., 2005. Is Area Yield Insurance Competitive with Farm Yield Insurance? *J. Agr. Res. Econ.* 30, 285-301.
- Bessler, D.A. 1982. Adaptive Expectations, the Exponentially Weighted Forecast, and Optimal Statistical Predictors: A Revisit. *Agr. Econ. Res.* 34, 16-23.
- Bessler, D.A., Dorfman, J.H., Holt, M.T., LaFrance, J.T., 2010. Econometric Developments in Agricultural and Resource Economics: The First 100 Years. *Am. J. Agr. Econ.* 92, 571-589.
- Bielza, M., Conto, C.G., Dittman, C., Gallego F.J., Stroblmair, J., 2008. Agricultural Insurance Schemes. Report to the European Commission.
- Calderini, D.F. Slafer G.A. 1998. Changes in yield and yield stability in wheat during the 20th century. *Field Crop. Res.* 57, 335-347.
- Claassen, R. and Just, R.E. 2011. Heterogeneity and distributional form of farm-level yields. *Am. J. Agr. Econ.* 93, 144-160.
- Dawid, P.A. 1984. Statistical theory: a prequential approach. *J. Roy. Stat. Soc. Ser. A.* 147, 278-92.
- Dawid, P.A. 2006. Prequential analysis, in *Encyclopedia of Statistical Sciences* (2nd edition), S. Kotz, C., Read B., Banks, D.L. (eds.), Wiley-Interscience, New York, 464-470.
- Dietz, E.J., 1989. Teaching Regression in a Nonparametric Statistics Course. *Amer. Statist.* 43, 35-40.
- Fernandes, R., Leblanc, S.G. 2005. Parametric (Modified Least Squares) and Non-parametric (Theil-Sen) Linear Regressions for Predicting Biophysical Parameters in the Presence of Measurement Errors. *Remote Sens. Environ.* 95, 303-316.

- Finger, R., 2010a. Revisiting the Evaluation of Robust Regression Techniques for Crop Yield Data Detrending. *Am. J. Agr. Econ.* 92, 205-211.
- Finger, R., 2010b. Evidence of Slowing Yield Growth - the Example of Swiss Cereal Yields. *Food Pol.* 35, 175-182.
- Frees, E.W., 1991. Trimmed Slope Estimates for Simple Linear Regression. *J. Statist. Plann. Inference* 27, 203-221.
- Galbraith, J.W., van Norden, S. 2011. Kernel-based calibration diagnostics for recession and inflation probability forecasts. *Int. J. Forecasting* 27, 1041-1057.
- Gallagher, P., 1987. U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances. *Am. J. Agr. Econ.* 69, 796-803.
- Goodwin B.K., Kerr, A.P., 1998. Nonparametric Estimation of Crop Distributions: Implications for Rating Group-Risk Crop Insurance Contracts. *Am. J. Agr. Econ.* 80, 139-153.
- Goodwin, B.K., Ker, A.P. 2002. Modeling Price and Yield Risk, in *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture*, Just, R.E., Pope, R.D. (eds.), Kluwer Academic Press, 289-323.
- Hampel, F.R., 2001. Robust statistics: A brief introduction and overview. Research Report 94, Seminar für Statistik, ETH Zürich, Switzerland.
- Hampel, F.R., 2002. Robust Inference, in *Encyclopedia of Environmetrics*, El-Shaarawi, A.H., Piegorsch, W.W. (eds.). Wiley & Sons, New York, 1865-1885.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw P.J., Stahel, W.A., 1986. *Robust Statistics*. Wiley & Sons, New York.
- Harri, A., Coble, K.H., Erdem C., Knight, T.O., 2009. Crop Yield Normality: A Reconciliation of Previous Research. *Rev. Agr. Econ.* 31, 163-182.
- Hennessy, D.A., 2009. Crop Yield Skewness and the Normal Distribution. *J. Agr. Res. Econ.* 34, 34-52.
- Hössjer, O., Rousseeuw, P.J., Croux, C., 1994. Asymptotics of the Repeated Median Slope Estimator. *Ann. Statist.* 22, 1478-1501.

- Huber, P.J., 1972. Robust statistics: A Review. *Ann. Statist.* 43, 1041-1067.
- Huber, P.J., 1973. Robust Regression: Asymptotics, Conjectures and Monte Carlo. *Ann. Statist.* 1, 799-821.
- Hubert, M., Rousseeuw, P.J., Van Aelst, S., 2004. Robustness, in *Encyclopedia of Actuarial Science*, Sundt, B., Teugels, J. (eds.). Wiley & Sons, New York, 1515-1529.
- Johnston, J., DiNardo, J., 1997. *Econometric Methods*. McGraw-Hill, New York.
- Just, R.E., Weninger, Q., 1999. Are Crop Yields Normally Distributed? *Am. J. Agr. Econ.* 81, 287-304.
- Kaylen, M.S., Koroma, S.S., 1991. Trend, Weather Variables, and the Distribution of U.S. Corn Yields. *Appl. Econ. Persp. Pol.* 13, 249-258.
- Ker, A.P., Coble, K.H., 2003. Modeling Conditional Yield Distributions. *Am. J. Agr. Econ.* 85, 291-304.
- Ker, A.P., Goodwin, B.K., 2000. Nonparametric Estimation of Crop Insurance Rates Revisited. *Am. J. Agr. Econ.* 82, 463–78.
- Kling, J., Bessler, D.A. 1989. Calibration-Based Predictive Distributions: an Application of Prequential Analysis to Interest Rates, Money, Prices and Output. *J. Bus.* 62, 447-500.
- Lee, T.C.M., Oh, H.S., 2007. Robust Penalized Regression Spline Fitting with Application to Additive Mixed Modeling. *Comput. Statist.* 22, 159-171.
- Maronna, R.A., Martin, R.D., Yohai, V.J., 2006. *Robust Statistics – Theory and Methods*. Wiley & Sons, New York.
- Nelson, C.R., Plosser, C.I. 1982. Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications. *J. Monetary Econ.* 10, 139–162.
- Ozaki, V.A., Silva, R.S., 2009. Bayesian ratemaking procedure of crop insurance contracts with skewed distribution. *J. Appl. Stat.* 36, 443-452.
- R Development Core Team, 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org.

- Racine, J., Ker, A., 2006. Rating Crop Insurance Policies with Efficient Nonparametric Estimators that Admit Mixed Data Types. *J. Agric. Res. Econ.* 31, 27-39.
- Ramirez, O.A., Misra, S., Field, J., 2003. Crop-Yield Distributions Revisited. *Am. J. Agr. Econ.* 85, 108-120.
- Rousseeuw P.J., Hubert, M., 1999. Regression Depth. *J. Amer. Statist. Assoc.* 94, 388-402.
- Rousseeuw, P.J., Leroy, A.M., 1987. *Robust Regression and Outlier Detection*. Wiley & Sons, New York.
- Rousseeuw, P.J., Yohai, V.J. 1984. Robust regression by means of S-estimators, in *Robust and nonlinear time series analysis*, Lecture notes in Statistics 26, Franke, J., Härdle, W., Martin, R.D. (eds.), Springer: New York, 256-272.
- Ruckstuhl, A.F., 1997. Partial Breakdown in Two-Factor Models. *J. Statist. Plann. Inference* 57, 257-271.
- Sen, P.K., 1968. Estimates of the Regression Coefficient Based on Kendall's Tau. *J. Amer. Statist. Assoc.* 63, 1379-1389.
- Skees, J.R., Black, J.R., Barnett, B.J., 1997. Designing and Rating an Area Yield Crop Insurance Contract. *Am. J. Agr. Econ.* 79, 430-438.
- Stigler, S., 1986. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press, Cambridge.
- Swinton, S.M., King, R.P., 1991. Evaluating Robust Regression Techniques for Detrending Crop Yield Data with Nonnormal Errors. *Am. J. Agr. Econ.* 73, 446-451.
- Theil, H., 1950. A Rank-Invariant Method of Linear and Polynomial Regression Analysis I, II and III. *Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings* 53, 386-392, 521-252 and 1397-1412.
- Wilcox, R.R., 1998a. Simulations on the Theil-Sen Regression Estimator with Right-Censored Data. *Statist. Probab. Lett.* 39, 43-47.
- Wilcox, R.R., 1998b. A Note on the Theil-Sen Regression Estimator When the Regressor Is Random and the Error Term Is Heteroscedastic. *Biom. J.* 40, 261-268.

- Woodard, J.D., Sherrick, B.J., Schnitkey, G.D. 2010. Revenue Risk Reduction Impacts of Crop Insurance in a Multi-Crop Framework. *Appl. Econ. Perspect. Pol.* 32, 472-488.
- Woodard, J.D., Paulson, N.D., Vedenov, D., Power, G.J., 2011. Impact of copula choice on the modeling of crop yield basis risk. *Agric. Econ.* 42, 101–112.
- Yohai, V.J., 1987. High Breakdown-Point and High Efficiency Robust Estimates for Regression. *Ann. Statist.* 15, 642-656.
- Zaman, A., Rousseeuw, P.J., Orhan, M., 2001. Econometric Applications of High-Breakdown Robust Regression Techniques. *Econ. Lett.* 71, 1-8.
- Zhou, W., Serfling, R., 2008. Multivariate spatial U-Quantiles: A Bahadur-Kiefer Representation, a Theil-Sen Estimator for Multiple Regression, and a Robust Dispersion Estimator. *J. Statist. Plann. Inference* 138, 1660-1678.
- Zhu, Y., Goodwin, B., Ghosh, S. 2008. Time-varying yield distribution and the implications for crop insurance pricing. Paper presented at the NC State Agricultural Economics Workshop, Raleigh, NC, 6 May
- Zulauf, C.R., Dicks, M.R., Vitale, J.D., 2008. ACRE (Average Crop Revenue Election) Farm Program: Provisions, Policy Background, and Farm Decision Analysis. *Choices* 23, 29-35.

Table A1. Results of the Monte Carlo Experiment for the Crop Yield Detrending Analysis – Presentation of Slope Estimates.

	SD	Mean	Min	Max	MAD	SD	Mean	Min	Max	MAD
	Scenario 1 – Normal (0,25) distribution					Scenario 7 – Normal (0,25) distribution with outlier contamination				
OLS	2.80	2.02	-9.82	13.63	1.86	7.41	2.10	-23.99	31.63	5.00
MM	2.97	2.02	-13.33	14.14	1.92	4.01	2.00	-17.13	20.93	2.72
TS	3.00	2.02	-9.81	13.67	1.99	4.30	2.04	-14.41	18.14	2.95
M	2.88	2.02	-9.82	13.69	1.90	4.35	2.04	-14.73	19.73	3.02
	Scenario 2 – Beta(5,3) distribution					Scenario 8 – Beta(5,3) distribution with outlier contamination				
OLS	1.80	2.04	-4.66	7.97	1.23	4.75	2.00	-17.48	20.27	3.17
MM	1.94	2.04	-7.25	10.28	1.29	2.65	1.97	-9.67	13.79	1.81
TS	1.94	2.04	-4.66	8.88	1.32	2.80	1.97	-7.30	12.57	1.95
M	1.87	2.04	-4.70	7.92	1.27	2.83	1.96	-7.80	12.65	2.02
	Scenario 3 – Beta(15,3) distribution					Scenario 9 – Beta(15,3) distribution with outlier contamination				
OLS	0.96	2.00	-1.97	5.77	0.64	2.49	2.03	-7.78	11.2	1.68
MM	0.98	2.00	-2.15	8.11	0.65	1.35	2.02	-4.88	7.28	0.90
TS	0.98	2.00	-1.48	5.75	0.65	1.42	2.02	-3.36	8.39	0.95
M	0.96	2.00	-1.38	5.59	0.64	1.46	2.01	-3.52	7.73	1.00
	Scenario 4 – Normal (0,25) distribution with heteroscedasticity					Scenario 10 – Normal (0,25) distribution with heteroscedasticity and outlier contamination				
OLS	9.29	2.09	-39.44	34.87	6.22	11.57	2.17	-40.7	43.43	7.74
MM	9.62	2.03	-39.27	39.04	6.46	10.62	2.07	-40.68	43.15	7.05
TS	8.67	2.03	-38.45	36.58	5.84	8.70	2.15	-32.93	35.28	5.89
M	8.98	2.04	-38.87	36.99	5.97	10.22	2.12	-39.77	41.37	6.82
	Scenario 5 – Beta (5,3) distribution with heteroscedasticity					Scenario 11- Beta(5,3) distribution with heteroscedasticity and outlier contamination				
OLS	6.02	2.06	-20.08	21.38	4.15	7.49	1.91	-26.68	32.86	5.06
MM	6.53	2.70	-23.21	22.99	4.64	7.14	2.53	-25.59	26.05	4.90
TS	5.89	2.57	-18.79	23.16	4.11	6.46	2.44	-22.64	25.78	4.39

M	6.05	2.38	$\bar{-}$ 20.74	21.25	4.22	6.75	2.21	$\bar{-}$ 23.63	25.31	4.60
Scenario 6 – Beta (15,3) distribution with heteroscedasticity					Scenario 12 – Beta (15,3) distribution with heteroscedasticity and outlier contamination					
OLS	3.21	1.96	$\bar{-}$ 13.21	11.33	2.18	3.94	2.04	$\bar{-}$ 15.86	15.09	2.64
MM	3.21	2.73	$\bar{-}$ 11.67	11.41	2.35	3.55	2.74	$\bar{-}$ 13.07	12.81	2.53
TS	2.93	2.57	$\bar{-}$ 10.23	11.47	2.08	3.26	2.61	$\bar{-}$ 13.24	14.01	2.28
M	3.07	2.36	$\bar{-}$ 12.21	11.43	2.17	3.45	2.35	$\bar{-}$ 13.17	12.87	2.35

Note: SD denotes the standard deviation, Min the minimum value, Max the maximum value and MAD the median of absolute deviations of the coefficient estimates for the slope. The here presented values are based on 9999 coefficient estimates for the slope for each scenario and estimator.

Table A2. Results of the Ansari-Bradley Tests on Differences in Dispersion for the Crop Yield Detrending Analysis.

Scenario	Hypotheses					
	$S(\hat{\beta}_{OLS}) < S(\hat{\beta}_{MM})$	$S(\hat{\beta}_{OLS}) < S(\hat{\beta}_{TS})$	$S(\hat{\beta}_{MM}) < S(\hat{\beta}_{TS})$	$S(\hat{\beta}_{OLS}) < S(\hat{\beta}_M)$	$S(\hat{\beta}_M) < S(\hat{\beta}_{MM})$	$S(\hat{\beta}_M) < S(\hat{\beta}_{TS})$
Scenario 1	***	***	**	*	*	***
Scenario 2	***	***	n.s.	***	*	**
Scenario 3	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
	$S(\hat{\beta}_{OLS}) < S(\hat{\beta}_{MM})$	$S(\hat{\beta}_{OLS}) > S(\hat{\beta}_{TS})$	$S(\hat{\beta}_{MM}) > S(\hat{\beta}_{TS})$	$S(\hat{\beta}_{OLS}) < S(\hat{\beta}_M)$	$S(\hat{\beta}_M) < S(\hat{\beta}_{MM})$	$S(\hat{\beta}_M) > S(\hat{\beta}_{TS})$
Scenario 4	***	***	***	n.s.	n.s.	***
Scenario 5	***	*	***	n.s.	n.s.	**
Scenario 6	n.s.	***	***	n.s.	n.s.	***
	$S(\hat{\beta}_{OLS}) > S(\hat{\beta}_{MM})$	$S(\hat{\beta}_{OLS}) > S(\hat{\beta}_{TS})$	$S(\hat{\beta}_{MM}) < S(\hat{\beta}_{TS})$	$S(\hat{\beta}_{OLS}) > S(\hat{\beta}_M)$	$S(\hat{\beta}_M) > S(\hat{\beta}_{MM})$	$S(\hat{\beta}_M) > S(\hat{\beta}_{TS})$
Scenario 7	***	***	***	***	***	**
Scenario 8	***	***	***	***	***	***
Scenario 9	***	***	***	***	***	***
	$S(\hat{\beta}_{OLS}) > S(\hat{\beta}_{MM})$	$S(\hat{\beta}_{OLS}) > S(\hat{\beta}_{TS})$	$S(\hat{\beta}_{MM}) > S(\hat{\beta}_{TS})$	$S(\hat{\beta}_{OLS}) > S(\hat{\beta}_M)$	$S(\hat{\beta}_M) > S(\hat{\beta}_{MM})$	$S(\hat{\beta}_M) > S(\hat{\beta}_{TS})$
Scenario 10	***	***	***	***	n.s.	***
Scenario 11	***	***	***	***	n.s.	***
Scenario 12	***	***	***	***	n.s.	***

Note: $S(\hat{\beta}_{OLS})$ denotes the dispersion of the OLS trend coefficient estimates. $\hat{\beta}_{MM}$, $\hat{\beta}_{TS}$ and $\hat{\beta}_M$ stand for trend coefficient estimates of the MM-, Theil-Sen and M-estimator, respectively. n.s. denotes no significance, while *, **, and *** denote significance on the 10%, 5%, and 1% level, respectively. Ansari-Bradley tests are conducted as pairwise comparisons based on 9999 coefficient estimates for the slope for each scenario and estimator.

Table A3. Results of the Monte Carlo Experiment: Analysis of Yield Prediction Errors.

	Mean	Min	Max	SD	Mean	Min	Max	SD
Scenario 1 – Normal (0,25) distribution					Scenario 7 – Normal (0,25) distribution with outlier contamination			
OLS	0.35	-119.22	112.23	31.82	0.32	-234.70	240.46	65.08
MM	0.30	-129.52	118.52	32.58	-0.32	-160.03	166.13	40.19
TS	0.22	-141.82	115.65	32.70	-0.18	-177.86	143.14	41.46
M	0.30	-121.43	116.09	32.18	-0.18	-157.97	151.74	42.02
OA5	7.93	-86.73	109.87	27.72	8.50	-96.25	119.61	29.07
Scenario 2 - Beta(5,3) distribution					Scenario 8 – Beta(5,3) distribution with outlier contamination			
OLS	-0.19	-71.67	67.54	20.37	0.15	-160.79	150.98	41.93
MM	-0.51	-87.38	74.99	21.06	-0.39	-99.68	101.53	26.05
TS	-0.99	-80.48	72.37	21.32	-0.58	-104.08	94.76	26.99
M	-0.49	-73.58	72.12	20.76	-0.14	-103.99	95.46	27.33
OA5	7.26	-56.92	61.76	18.03	8.5	-60.95	70.52	18.86
Scenario 3 – Beta(15,3) distribution					Scenario 9 – Beta(15,3) distribution with outlier contamination			
OLS	0.00	-45.74	38.01	10.98	0.30	-100.49	87.10	22.44
MM	-0.49	-55.65	38.17	11.12	-0.25	-54.92	48.18	13.61
TS	-0.86	-52.34	40.15	11.23	-0.59	-61.45	55.42	14.04
M	-0.39	-52.20	38.01	11.01	-0.01	-60.86	53.86	14.43
OA5	7.52	-33.25	36.38	9.59	8.58	-39.14	40.33	9.99
Scenario 4 – Normal (0,25) distribution with heteroscedasticity					Scenario 10 – Normal (0,25) distribution with heteroscedasticity and outlier contamination			
OLS	0.83	-601.83	644.29	170.35	-1.48	-672.09	726.42	179.20
MM	0.27	-619.47	684.43	173.58	-1.44	-662.41	702.26	177.85
TS	0.3	-564.99	665.37	171.45	-1.48	-661.90	674.27	175.30
M	0.59	-564.66	673.29	170.33	-1.47	-664.44	671.77	174.91
OA5	7.97	-533.91	606.27	157.05	6.66	-606.18	620.25	157.99

Scenario 5 – Beta (5,3) distribution with heteroscedasticity					Scenario 11- Beta(5,3) distribution with heteroscedasticity and outlier contamination			
OLS	-1.39	-376.42	330.74	109.52	0.35	-434.60	378.65	115.90
MM	-7.88	-407.12	389.06	113.56	-5.34	-435.08	366.89	116.79
TS	-7.08	-387.74	325.04	111.75	-4.85	-442.39	353.57	114.58
M	-4.70	-380.38	319.02	110.72	-2.47	-426.58	363.68	114.26
OA5	4.92	-342.52	279.07	101.97	7.08	-369.94	286.45	102.63
Scenario 6 – Beta (15,3) distribution with heteroscedasticity					Scenario 12 – Beta (15,3) distribution with heteroscedasticity and outlier contamination			
OLS	0.16	-253.35	167.49	58.74	0.55	-275.10	191.44	61.44
MM	-7.31	-268.65	180.03	59.89	-5.97	-274.77	194.69	60.98
TS	-6.44	-263.71	175.03	59.19	-5.51	-269.36	188.64	60.09
M	-4.03	-260.17	176.43	58.82	-2.72	-271.85	193.70	60.07
OA5	6.16	-232.26	149.87	54.19	7.23	-248.05	152.61	53.97

Note: SD denotes the standard deviation, Min the minimum value, Max the maximum value and MAD the median of absolute deviations of the prediction errors of the 2 years ahead yield predictions. The here presented values are based on 9999 estimated prediction errors for each scenario and estimator.

Table A4. Results of the Ansari-Bradley Tests on Differences in Dispersion for the Analysis of Yield Prediction Errors.

Scenario	Hypotheses									
	$S(PE_{OLS})$	$S(PE_{OLS})$	$S(PE_{MM})$	$S(PE_{OLS})$	$S(PE_{MM})$	$S(PE_M)$	$S(PE_{OA})$	$S(PE_{OA})$	$S(PE_{OA})$	$S(PE_{OA})$
	$\neq S(PE_{MM})$	$\neq S(PE_{TS})$	$\neq S(PE_{TS})$	$\neq S(PE_M)$	$\neq S(PE_M)$	$\neq S(PE_{TS})$	$< S(PE_{OLS})$	$< S(PE_{MM})$	$< S(PE_{TS})$	$< S(PE_M)$
Scenario 1	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	***	***	***
Scenario 2	**	***	n.s.	n.s.	n.s.	**	***	***	***	***
Scenario 3	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	***	***	***
Scenario 4	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	***	***	***
Scenario 5	**	n.s.	n.s.	n.s.	n.s.	n.s.	***	***	***	***
Scenario 6	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	***	***	***
Scenario 7	***	***	***	***	***	*	***	***	***	***
Scenario 8	***	***	***	***	***	*	***	***	***	***

Scenario 9	***	***	***	***	***	***	***	***	***	***
Scenario 10	n.s.	*	n.s.	*	n.s.	n.s.	***	***	***	***
Scenario 11	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	***	***	***
Scenario 12	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***	***	***	***

Note: $S(PE_{OLS})$ denotes the dispersion of the prediction error (PE) of the OLS based 2-years ahead yield prediction. PE_{MM} , PE_{TS} and PE_M stand for the prediction errors based on the MM-, Theil-Sen and M-estimator, respectively. n.s. denotes no significance, while *, **, and *** denote significance on the 10%, 5%, and 1% level, respectively. Ansari-Bradley tests are conducted as pairwise comparisons based on 9999 prediction errors for each scenario and estimator.

Figure 1. Density Functions of the Coefficient Estimates for the Crop Yield Detrending Analysis: Scenarios 1-6.

Note: The plots show the density of the 9999 coefficient estimates of the slope for each scenario and estimator.

Figure 2. Density Functions of the Coefficient Estimates for the Crop Yield Detrending Analysis: Scenarios 7-12.

Note: The plots show the density of the 9999 coefficient estimates of the slope for each scenario and estimator.

Figure 3. Density Functions of the Prediction Errors: Scenarios 1-6.

Note: The plots show the density of the 9999 prediction errors of the 2 years ahead yield prediction for each scenario and estimator.

Figure 4. Density Functions of the Prediction Errors: Scenarios 7-12.

Note: The plots show the density of the 9999 prediction errors of the 2 years ahead yield prediction for each scenario and estimator.