Yield trend estimation in the presence of non-constant technological change and weather effects

Conradt, S.¹, Bokusheva R.¹, Finger R.¹² and Kussaiynov, T.³

1 ETH Zurich, Agri-Food and Agri-Environmental Economics Group, Zurich, Switzerland.
2 Current Affiliation: Wageningen University, Agricultural Economics and Rural Policy Group. Wageningen, the Netherlands
3 Kazakh Agrotechnical S. Seifullin University, Department of Economics, Astana, Kazakhstan

conradts@ethz.ch
Yield trend estimation in the presence of non-constant technological change and weather effects

Conradt, S., Bokusheva R., Finger R. and Kussaiyynov, T.

Abstract

The application of yield time series in risk analysis prerequisites the estimation of technological trend which might be present in the data. In this paper, we show that in presence of highly volatile yield time series and non-constant technology, the consideration of the weather effect in the trend equation can seriously improve trend estimation results. We used ordinary least squares (OLS) and MM, a robust estimator. Our empirical analysis is based on weather data as well as farm-level and county-level yield data for a sample of grain-producing farms in Kazakhstan.

Keywords: Yield detrending, weather information, robust trend estimation, aggregation

JEL classification: Q19

1. INTRODUCTION

Crop yields typically exhibit substantial variation, mostly due to varying weather events and technological improvements. While the latter does not contribute to the risk a farmer faces at a specific point in time, weather impacts are of particular relevance for risk analysis. Thus, an adequate identification of single sources of yield variability is crucial in empirical analysis. In particular for actuarial purposes the volatility related to changes in technology have to be isolated from yield time series. Hence, typically, historical yield data is ‘detrended’ prior to further analysis, i.e. one has to account for technological changes in the time series that have influenced yield level over time (Skees et al., 1997).

In some cases, however, the estimation of a technological trend is complicated through a joint occurrence of two phenomena: (i) Large yield variations as a consequence of the exposure of the rain-fed agriculture to extreme weather events, and (ii) a non-constant development of technical change over time. This situation is particularly relevant for transition and developing countries, where climate conditions often are extreme, agriculture is seldom irrigated (Leblois and Quirion, 2011), and where periods of investments and dis-investments frequently alternate. In contrast to situations in transition or developing countries, in developed countries, crop yields often increase linearly over time for the majority of empirical problems (for a review see e.g. Finger, 2010b, Hafner, 2003 and Tannura et al. 2008).

Non-linear trends in yield data often hinder risk analysis for crop production. Accordingly, trends have to be described with e.g. second or third order degree polynomial functions, which make a thorough risk analysis even more difficult. As a result, long-term data are often neglected due to their difficult pattern. This exclusion leads to information losses of...
yield variability and climatic risks and thus lowers the applicability for insurances in these areas. This unsatisfying situation requires more attention from the researchers (Chen and Miranda, 2006).

Determination of the “right” trend is decisive for further analysis. Inaccurate trend estimations may lead to incorrect determination of higher level moments of the distribution and may provoke non-stationarity of yield distributions (Just and Weninger, 1999). Specifically, not accounting for the whole effect of technological changes (i.e. underfitting) overestimates the risk a farmer faces. Assuming a higher polynomial trend than the intrinsic trend or overestimating trend parameter might cause implications for risk identification and leads to an underestimation of the underlying risk (Mara and Schurle 1994).

It is known from literature, that also the choice of the estimation technique is relevant for the conclusions drawn from technological trends estimations. In particular, robust regression techniques that are not influenced by outlying observations (e.g. low yield due to a drought event) have been suggested for crop yield data detrending (e.g. Swinton and King, 1991, Finger, 2010a). The latter studies, however, rather focused on stochastic simulation of yield data than on the application of these methods to empirical data. In the above described framework of extremely volatile yield time series in transition or developing countries, these regression techniques may be highly relevant. This issue has, however, not been empirically addressed so far.

In contrast to developed countries, where the technological trend of yield data in a region is supposed to be rather similar (Atwood et al. 2003), farms in transition and developing countries may exhibit a more heterogeneous trend. This means that not all farms face similar technological developments due to strong differences in factor endowment, education, and investment possibilities. In particular, resource availability is crucial for these countries and may be more divers among farms i.e. availability of machinery, amount of fertilizers and pesticides and management skills. This is pronounced in transition countries, where farms were subject to restructuring and privatisation. However, sufficient and reliable farm-level data is often not available and aggregated county or national data are used instead. Due to aggregation biases, this choice is known to substantially affect the possibility to draw conclusions on farm-level production risks (e.g. Finger, 2012). Farm-specific variation is “averaged” across a region or a whole country and consequently with increasing number of aggregated farms the individual impact of a farm is reduced. Thus, aggregated county-data may reduce systematic and random variation (Claassen and Just, 2010). This effect already takes place while aggregating very small units, i.e. the yield on a single field. Mara and Schurle (1994) found a reinforcing relationship between size of the aggregated unit and farm level yield risk. Hence, reducing the aggregation level (i.e. smaller acreage) leads to an increasing rate of yield variability. Given that already aggregation leads to biased results, a further challenge to deal with is the inaccuracy of using region wide (or national) trend estimation. The effect of data aggregation (or averaging) on farm-specific trend correction has not been addressed so far.
The goal of this paper is to compare different approaches to crop yield data detrending with wheat production in Kazakhstan as a case study. This focus is motivated by the fact that Kazakhstan represents one of the most important wheat producers in the world and regularly faces extreme droughts that lead to substantial yield losses. Additionally, the country faced a transition period in the last decades that had largely influenced agricultural investments and outputs (cp. e.g. Breustedt et al., 2008).

We compare and analyse the use of weather information in crop yield data detrending and the use of robust regression techniques. Furthermore, we aim to investigate in our analysis if farm-level analyses are required or if homogenous trends in crop yield data across farms can be isolated if the above mentioned detrending approaches are applied. Thus, we expand the argumentation of aggregation biases in farm-level risk analysis by investigating if technological trends may add another error source if they are assumed to be similar across farms (e.g. by using aggregated or average data).

The remainder of this paper is structured as follows. Section 2 describes the data and methods used in the analysis. More specifically, we present our yield and weather data, and explain the model selection procedure applied in this study. Our research findings are presented in Section 3. In the last section, we discuss the results, draw the conclusions and provide an outlook for further research.

2. DATA AND METHODS

2.1. Data

We used wheat yield data for 47 farms from five different counties (“rayons”) in North-Kazakhstan. A summary statistics is provided in Table 1. Farm yield data were provided by regional statistical offices and cover the period from 1980 to 2010. The yield data shows particular patterns that are relevant for this study: A phase of yield reductions due to disinvestment during the 1990s (during the transition period) was enclosed by phases of technical improvements and significant yield increases until the late 1980s and from beginning of the current century. All these developments were accompanied by short periods of significant yield losses due to droughts e.g. in 1996, 1997, 1998 and 2010.
Table 1 Summary statistics of the 47 farm data

<table>
<thead>
<tr>
<th>Rayon</th>
<th>Number of Farms</th>
<th>Average yield 1980-2010 [0.1 t/ha]</th>
<th>Min. yield [0.1 t/ha]</th>
<th>Max. yield [0.1 t/ha]</th>
<th>CV yield</th>
<th>Average sown area [ha]</th>
<th>Min. sown area [ha]</th>
<th>Max. sown area [ha]</th>
<th>CV sown area</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>12</td>
<td>8.9</td>
<td>0.2</td>
<td>24.0</td>
<td>0.44</td>
<td>13'599</td>
<td>805</td>
<td>24'700</td>
<td>0.43</td>
</tr>
<tr>
<td>R2</td>
<td>11</td>
<td>8.8</td>
<td>0.8</td>
<td>21.0</td>
<td>0.43</td>
<td>16'900</td>
<td>800</td>
<td>34'073</td>
<td>0.41</td>
</tr>
<tr>
<td>R3</td>
<td>7</td>
<td>8.3</td>
<td>1.2</td>
<td>19.3</td>
<td>0.42</td>
<td>15'316</td>
<td>500</td>
<td>30'750</td>
<td>0.49</td>
</tr>
<tr>
<td>R4</td>
<td>10</td>
<td>10.7</td>
<td>0.9</td>
<td>25.6</td>
<td>0.47</td>
<td>14'720</td>
<td>1155</td>
<td>10'940</td>
<td>0.44</td>
</tr>
<tr>
<td>R5</td>
<td>7</td>
<td>9.2</td>
<td>0.3</td>
<td>22.1</td>
<td>0.43</td>
<td>19'666</td>
<td>2000</td>
<td>82'850</td>
<td>0.65</td>
</tr>
</tbody>
</table>

CV: coefficient of variation. Source: Data from the regional statistical offices of Kazakhstan.

2.2. Weather indices

Different weather indices were included in the estimation of trend models. The here used weather data were provided by the National Hydro-Meteorological Agency of Kazakhstan. We considered two indices in our analysis: a hydrothermal coefficient (HTC) by Selyaninov (Selyaninov index) (Meshcherskaya and Blazhevich, 1996; Dronin and Kirilenko, 2008) as well as a cumulative rainfall index. The Selyaninov index is determined by the ratio of cumulative precipitation and the sum of daily average temperatures for the period from the 3rd decade of May to July, i.e. the main growing season in Kazakhstan (see equation (1)). Higher values represent wetter conditions. The cumulative rainfall index is calculated by the sum of precipitation [mm] from April to July (see equation (2)).

\[
HTC = \frac{\sum \text{Precip}}{\sum \text{Temp}} \tag{1}
\]

\[
CP = \sum \text{Precip} \tag{2}
\]

Precip stands for precipitation, Temp for temperature.

These two indices are used as proxy for droughts, one of the main challenges for the Kazakh agriculture. An overview of the weather indices is given in Table 2.

---

1 In general, the Selyaninov index is determined for daily average temperature above 10°C. We confined the index to the main growing season in Kazakhstan.
Table 2 Summary statistics: weather indices (1980-2010)

<table>
<thead>
<tr>
<th>Rayon</th>
<th>Average CP</th>
<th>Min. CP</th>
<th>Max. CP</th>
<th>CV CP</th>
<th>Average SI</th>
<th>Min. SI</th>
<th>Max. SI</th>
<th>CV SI</th>
<th>Number of years where SI &lt; 0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>140.9</td>
<td>94</td>
<td>215</td>
<td>0.26</td>
<td>0.72</td>
<td>0.26</td>
<td>1.38</td>
<td>0.38</td>
<td>14</td>
</tr>
<tr>
<td>R2</td>
<td>132.7</td>
<td>33</td>
<td>234</td>
<td>0.34</td>
<td>0.71</td>
<td>0.14</td>
<td>1.57</td>
<td>0.47</td>
<td>19</td>
</tr>
<tr>
<td>R3</td>
<td>126.5</td>
<td>46</td>
<td>267</td>
<td>0.38</td>
<td>0.65</td>
<td>0.22</td>
<td>1.74</td>
<td>0.55</td>
<td>20</td>
</tr>
<tr>
<td>R4</td>
<td>163.5</td>
<td>83</td>
<td>269</td>
<td>0.35</td>
<td>0.87</td>
<td>0.30</td>
<td>1.93</td>
<td>0.50</td>
<td>13</td>
</tr>
<tr>
<td>R5</td>
<td>147.7</td>
<td>70</td>
<td>297</td>
<td>0.39</td>
<td>0.75</td>
<td>0.22</td>
<td>1.82</td>
<td>0.48</td>
<td>16</td>
</tr>
</tbody>
</table>

CV: coefficient of variation, CP: cumulative precipitation [mm], SI: Selyaninov index. Source: Data from the National Hydro-Meteorological Agency of Kazakhstan.

2.3. Yield trend analysis

We applied two different estimation techniques, ordinary least squares (OLS) and MM-estimation, a robust regression technique, to compare trend estimations of (i) reduced trend models, i.e. without consideration of weather effects and (ii) full models, i.e. considering weather effects. Both estimators are discussed in Finger (2010a). For the two models, we considered different trend models ranging from 'no time trend' up to a cubic (i.e. 3rd polynomial degree) trend model. We applied the F-test to compare the nested models and to determine the most appropriate one. Figure 1 summarizes the estimated models as well as the model selection procedure employed in the analysis. For the calculations and estimations we used the statistical software R (R Development Core Team, 2012).2

Subsequently, we normalized yields to the year 2010 by adding the error term of each year and farm (see equation (3)). For the full model we added additional the respective weather index at time t in the corresponding rayon (see equation (4)) (see Goodwin and Mahul, 2004).

\[
(\text{normalized yields})_i = y_{2010} + e_i \\
(\text{normalized yields})_i = y_{2010} + e_i + \text{weather}_i
\]

In order to evaluate the magnitude of the differences in detrended yield estimations we determined the total absolute amount of differences. In particular, we built the absolute differences of the approaches for all farms over the whole time period: see equation (5) for comparison of MM and OLS and equation ((6) for comparison of yield results estimated by using rayon and farm level trends.

2 For the MM-estimations we used for instance the R package “robustbase”.
\[ \sum_{Farm=1}^{47} \sum_{t=1980}^{2010} abs(Y_{OLS} - Y_{MM}) \]

\[ \sum_{Farm=1}^{47} \sum_{t=1980}^{2010} abs(Y_{Rayon} - Y_{Farm}) \]

\[ Y_{OLS} / Y_{MM}: \text{ yield estimated with OLS / MM, } Y_{Rayon} / Y_{Farm}: \text{ Yield estimated on rayon and farm level.} \]

Figure 1 Overview of detrending models: reduced model (i), without weather effects and full model (ii) with weather effects. No trend up to a cubic trend was considered.

(i) \[ y_{it} = \beta_{i0} \]
\[ y_{it} = \beta_{i0} + \beta_{i1}t \]
\[ y_{it} = \beta_{i0} + \beta_{i1}t + \beta_{i2}t^2 \]
\[ y_{it} = \beta_{i0} + \beta_{i1}t + \beta_{i2}t^2 + \beta_{i3}t^3 \]

(ii) \[ y_{itj} = \alpha_{i\text{weather}}_{itj} + \beta_{i0} \]
\[ y_{itj} = \alpha_{i\text{weather}}_{itj} + \beta_{i0} + \beta_{i1}t_j \]
\[ y_{itj} = \alpha_{i\text{weather}}_{itj} + \beta_{i0} + \beta_{i1}t_j + \beta_{i2}t_j^2 \]
\[ y_{itj} = \alpha_{i\text{weather}}_{itj} + \beta_{i0} + \beta_{i1}t_j + \beta_{i2}t_j^2 + \beta_{i3}t_j^3 \]

\[ y_{itj} = \text{predicted yield of farm } i \text{ at time } t_j \]
\[ t_j = \text{time index, with } t_j = 1 \text{ in } 1980 \]
\[ \beta_{i0} = \text{modeled intercept of farm } i \]
\[ \alpha_{i\text{weather}} = \text{parameter specific for each farm } i \]
\[ \text{weather}_{itj} = \text{different weather indices, depending on county } c \text{ and time } t_j \]

2.4. Parameter comparison

In the cases of identical trend estimations of the reduced form model and the full model, we compared the deterministic trend parameter estimations of both models. We used confidence intervals to determine significant differences in the parameter estimates. For the full model, we first recalculated the model without the weather parameter while keeping the deterministic trend parameter constant. The $\beta_i$’s were retained to the values estimated with the full model, whereas the weather parameter was excluded from the determination of the confidence interval. The reason of this procedure is that the weather parameter is used as a proxy in trend estimation and does not represent the deterministic trend which is removed prior to further analysis. Hence, it is not included in the confidence interval estimations. We are aware that comparison of confidence intervals is a very conservative method (Schenker and Gentleman 2001). As we applied this
method to hypothesis testing with the null hypothesis being “there is no significant difference between the trend parameter estimations of both models”, we confined the confidence interval to 83%, which represents a level of $\alpha = 0.05$ following Payton et al. (2003).

2.5. Model assumptions and structural breaks

To verify the underlying model assumptions, e.g. $E_i \sim N(0, \sigma^2)$ different tests are used. Applying the Shapiro-Wilk test, we found that the normality assumption had to be rejected for 9 out of 47 farms. The Breusch-Godfrey test indicated for three farms first order autocorrelation and additionally for two farms second order autocorrelation. We plotted the fitted values versus the residuals to verify homoscedasticity. Most patterns showed constant variance whereas for some we either observed single outliers or funnel-shaped divergence to the right. Additionally, we tested for structural breaks using the Chow test to account for parameter instability and changes in relationship. Fifteen farms revealed a statistically significant structural break. We continued the analysis for these farms separately.

3. Results

3.1. Weather information in detrending

Table 3 shows the results of the trend estimations determined by OLS, with and without considering a weather index as an additional regressor. The results are summarized Table 3 on the rayon level whereas the trend estimations were determined on the single farm level. For 32% of the farms (15 out of the 47 farms), the trend estimations differed for the two model approaches. Additionally, the trend estimations of the reduced form model (i) are in general lower than for the full model (ii). For instance “no trend” was determined for 26 farms for model (i) compared to 17 farms of model (ii). Similarly, higher trend estimations, i.e. quadratic or cubic trends, were favoured for 22 farms in case of the full model in contrast to 10 farms for the reduced form model.
Table 3 Farm level data F-test OLS

<table>
<thead>
<tr>
<th>Rayon</th>
<th>With weather</th>
<th>Without weather</th>
<th>Identical / no identical time trend</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No trend</td>
<td>Linear trend</td>
<td>Quadratic trend</td>
</tr>
<tr>
<td>Rayon 1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Rayon 2</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Rayon 3</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Rayon 4</td>
<td>0</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Rayon 5</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>8</td>
<td>12</td>
</tr>
</tbody>
</table>

Numbers in the table indicate the number of farms with the respective trend estimation.

Table 4 provides an identical summary table, however for the robust estimation technique MM. Similar to OLS, in 34% of the considered cases, the estimated time trends of the full and reduced model differed. Likewise, lower polynomial trends are more frequently estimated with the reduced model compared to the full model. However, this time, the effect seems to be less pronounced than with OLS. For instance, model (ii) determined 25 farms with a trend of 2nd or 3rd polynomial degree, whereas model (i) identified only 16 farms following these trends.

Table 4 Farm level data F-test MM

<table>
<thead>
<tr>
<th>Rayon</th>
<th>With weather</th>
<th>Without weather</th>
<th>Identical / no identical time trend</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No trend</td>
<td>Linear trend</td>
<td>Quadratic trend</td>
</tr>
<tr>
<td>Rayon 1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rayon 2</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Rayon 3</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Rayon 4</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Rayon 5</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

Numbers in the table indicate the numbers of farms with the respective trend estimation.

3 We used another model selection criterion, namely the Akaike information criterion (AIC), to check whether the results were stable (Burnham and Anderson, 2002). We found similar results to the F-test by using AIC.
In cases where both models estimated the same deterministic trend, we applied 83% confidence intervals to test whether there are significant differences in parameter estimations. For most cases, we could not reject the null hypothesis that the model parameters were the same. This was partially a consequence of the broad confidence interval of the reduced form model (see Figure 2). However, as exemplified in Figure 2, the null hypothesis was often not rejected by reason that the confidence intervals were overlapping at the tails. For all 47 farms, the reduced form model was lying above the full model, thus estimated a higher deterministic trend.

Figure 2 Confidence intervals (83%) exemplified for four farms, two estimated with OLS and two with MM. The grey lines represent the reduced form model and the black (lower lines) the full model.
3.2. Robust regression techniques vs. Ordinary Least Squares

In the Table 5, we compare the trend estimations of the two techniques, OLS and MM. For the models where an additional weather index was included as regressor, OLS and MM estimations were different in 9 out of 47 cases (19%). This number increased slightly while comparing the reduced form models: 14 out of 47 farms differed in the deterministic trend estimation, which corresponds to 29%.

Table 5 Farm level data F-test: Difference for OLS/MM with and without weather data

<table>
<thead>
<tr>
<th>Rayon 1</th>
<th>OLS with weather / OLS without weather</th>
<th>MM with weather / MM without weather</th>
<th>MM and OLS comparison: No identical time trend with weather / without weather</th>
</tr>
</thead>
<tbody>
<tr>
<td>No trend</td>
<td>Linear trend</td>
<td>Quadratic trend</td>
<td>Cubic trend</td>
</tr>
<tr>
<td>Rayon 1</td>
<td>2/3</td>
<td>2/2</td>
<td>3/3</td>
</tr>
<tr>
<td>Rayon 2</td>
<td>4/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>Rayon 3</td>
<td>4/1</td>
<td>2/1</td>
<td>1/1</td>
</tr>
<tr>
<td>Rayon 4</td>
<td>0/1</td>
<td>2/6</td>
<td>8/1</td>
</tr>
<tr>
<td>Rayon 5</td>
<td>7/7</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>Total</td>
<td>17/26</td>
<td>8/11</td>
<td>12/5</td>
</tr>
</tbody>
</table>

Numbers in the table indicate the amount of farms with the respective trend estimation.

In cases of identical trend estimations for both models, the detrended and normalized yields generally differed not substantially. However, as exemplified in Figure 3, in cases of distinct trend estimations, the MM and OLS estimations showed divergent sequences. For Farm 1 (F1) (left hand side of Figure 3), OLS estimated a 2nd degree polynomial trend whereas MM determined no trend. The right hand side compares OLS trend estimation having a quadratic trend with MM having a cubic trend.
Figure 3 Detrended and normalized yields estimated with the full model exemplified for two farms, where the respective time trend differed.

In order to evaluate the magnitude of the differences in detrended yield estimations for the two approaches OLS and MM we determined the total absolute amount of deviances. We added up the absolute difference of the detrended and normalized yields of all 47 farms for the whole time period of OLS and MM. The determined difference was 85.05 t for the estimations with a weather index as regressor. This means that there was a difference among both methods for a year and a specific farm of 0.058 t. Similar, the absolute difference of both methods for the reduced form model amount to 100.85 t. Thus for a single unit (one year, one farm), the absolute difference was 0.069 t.

3.3. Aggregation level and influence on detrending

We used rayon yield data and estimated the deterministic trend for OLS and MM. Both estimation techniques showed identical results (see Table 6). No time trend was found for the reduced form model, whereas the full model indicated a cubic trend for the first rayon, a quadratic trend for the fourth rayon and no trend for the remaining rayons.
Similar to section 3.2, we determined the absolute difference between the farm detrended and normalized results and the rayon detrended and normalized yields. The absolute deviation was 189.96 for all farms. Thus, for a single farm and single year, the absolute difference added up to 0.13 t.

4. DISCUSSIONS AND CONCLUSION

We investigated the inclusion of a weather index as a regressor in detrending. Our findings suggest that trend estimations require a simultaneous consideration of both, the effects of technology and the effect of weather. For instance, weather variables might themselves exhibit a significant trend due to climate change. Ignoring such trends and using a reduced trend model might lead to inconsistent estimates of the trend parameters as the model residuals would not be independent from trend model regressors, i.e. time variable(s). Adding a weather variable representing prevailing weather conditions during the crop growing season as an additional regressor may help to reduce over- and underestimation of technological trends. Especially the underestimation of the deterministic trend may be the prevailing challenge as the non-inclusion of weather information led to lower trend estimations in our study. This was the case for both detrending techniques, as well as for the aggregated rayon data. The underestimation of the deterministic trend increases the random component and as consequence leads to an overestimation of the risk a farmer faces.

Furthermore, we analyzed and compared the use of the MM estimator, a robust detrending technique. Robust estimators give less weight to outliers in the data set. In contrast, OLS is much more sensitive and for instance an extreme value at the end or the beginning of the time series may be decisive for trend determination. Thus, one has to analyze the data before detrending with OLS (even more) carefully. In settings where yields vary substantially and trend estimations are not stable, extending or shortening the time series for a few years may

4 We used the Akaike Information Criterion (AIC) to check whether we could confirm our results for OLS using an information based criterion. For the first model (OLS with weather) the same trends were estimated, whereas for rayon 4 using “OLS without weather” a quadratic trend was determined.
shift the deterministic trend. In these cases, it is especially helpful, to have background information on the data and possible technological trends. Our data seems to be representative for many other studies in the sense that the assumptions \( \left( E_i \sim N(0, \sigma^2) \right) \) are only partially fulfilled, however not completely violated. The here presented results show that the choice of the method may lead to substantial differences in the detrended yield estimations. This means that outliers can have substantial effects on trend estimation results and thus also affect subsequent analyses such as risk analysis or actuarial applications. Thus, the use of robust regression techniques in addition to OLS can reveal potential problems in data analysis due to observations that deviate from the relationship described by the majority of the data and bound the influence of these outliers.

Finally, we analyzed the effect of aggregation on detrending. In our analysis, we used farm data from Kazakhstan where the farm size is already substantial (approx. 15’000ha). This means that the data are already aggregated to a wide extent. We are aware that our data indicate a particular pattern and that the impact of aggregation differ with geographic region and crop (Mara 1994). However, we found that unified detrending on rayon basis could not satisfy the variability present in the single farm data. Different trend models found across farms as well as comparing the farm- and rayon-level indicate substantial differences in technological development across farms and regions. The deviation found in our study added up to 0.13t per year and farm, which is nearly double the difference as results of the technique, OLS or MM (0.069t). Assuming an one-size-fits-it-all trend model, e.g. across the entire country or one rayon, would thus be misleading. Instead, our results clearly indicate that farm-level analysis of deterministic trends on crop yield data is needed.

Thus, the here presented analysis contributes to an improved crop yield analysis for countries that face similar conditions. This will enable a better assessment of farmers’ crop yield risks and thus also contribute to the implementation of more efficient risk management strategies at the farm- and national level.

REFERENCES


