SFB 649 Discussion Paper 2011-018

Can crop yield risk be globally diversified?

Xiaoliang Liu*
Wei Xu*
Martin Odening*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

Can crop yield risk be globally diversified?*

X.L. Liu[a], W. Xu[a] and M. Odening[a,c]

Abstract

In 2007 and 2008 world food markets observed a significant price boom. Crop failures simultaneously occurring in some of the world's major production regions have been quoted as one factor among others for the price boom. Against this background, we analyse the stochasticity of crop yields in major production areas. The analysis is exemplified for wheat, which is one of the most important crops worldwide. Particular attention is given to the stochastic dependence of yields in different regions. Thereby we address the question of whether local fluctuations of yields can be smoothed by international agricultural trade, i.e. by global diversification. The analysis is based on the copula approach, which requires less restrictive assumptions compared with linear correlations. The use of copulas allows for a more reliable estimation of extreme yield shortfalls, which are of particular interest in this application. Our calculations reveal that a production shortfall, such as in 2007, is not a once in a lifetime event. Instead, from a statistical point of view, similar production conditions will occur every 15 years.

JEL classification: C14; Q19

Keywords: crop yield risk, fully nested hierarchical Archimedean copulas (FNAC), price boom

* This research was supported by Deutsche Forschungsgemeinschaft through the SFB649 Economic Risk.

[a] *Department of Agricultural Economics, Humboldt-Universität zu Berlin, D-10099 Berlin, Germany*

[c] Corresponding author. Email: m.odening@agrar.hu-berlin.de

I Introduction

Agricultural commodity markets have been regulated for a long time by a variety of instruments, including intervention prices, fixed and variable tariffs, export subsidies and production quotas. Due to negotiations within the WTO, most agricultural markets have been deregulated during the last two decades throughout the world. This abolishment of price stabilizing measures led to an increase in price volatility on formerly regulated markets (Yang *et al*., 2001; Chavas and Kim, 2006). Although higher price fluctuations were expected, the crop price boom in 2007/2008 surprised market experts. This rise of agricultural commodity prices can be regarded from two perspectives. On the one hand, it quickened optimism in the agribusiness. For example, the end of the technological treadmill, which prevented an increase of farm incomes in the past, was proclaimed (von Witzke *et al*., 2008). On the other hand, the sharp increase of food prices shortened the food supply in low income countries and aggravated hunger and malnutrition among poor people. From a scientific point of view, two questions arise: what factors were responsible for the price boom and how likely is it that a similar constellation of price determining factors will occur again?

It is widely accepted that multiple factors caused the recent increase in agricultural commodity prices. Sarris (2009) emphasizes three main factors: first, crop failures simultaneously occurred in major production region of the world; second, there was an increased demand for biofuels; and third, investors in agricultural commodity markets were involved in speculative activities. This article focuses on the first of these factors, i.e. stochastic shifts on the supply side. It is well known that agricultural yields, crop yields in particular, heavily depend on weather conditions. Thus, weather risks are immediately reflected in yield risks (e.g. Odening *et al*., 2008; Musshoff *et al*., 2009). When analysing the impact of yield fluctuations on agricultural prices, the choice of an appropriate perspective is important. For example, a regional analysis is not appropriate because poor harvests in one country can be compensated by imports (Lotze-Campen, 2007). In integrated markets, diversification by global trade has to be taken into account and thus a global analysis of agricultural yield risks is necessary. Against this background, we analyse the stochasticity of crop yields in the world's major production areas. The analysis is exemplified for wheat, which is one of the most important crops in the world in terms of production area and food security. Particular attention is given to the stochastic dependence of yields in different parts of the world. The spatial dependence structure is modelled using copulas. Compared with linear correlations, the application of copulas requires less restrictive assumptions and has advantages in the estimation of the joint occurrence of extreme events (c.f. Embrechts *et al*., 1999).

Copulas became increasingly popular in the last decade and have been applied to various finance and insurance problems (Cherubini *et al*., 2004; Chen *et al*., 2009; Turgutlu and Ucer, 2010). Bianchi *et al*. (2009) integrated a copula approach in the Vector-Auto-Regression model (VAR) to model and forecast the dynamics of industrial production in the euro-zone. Applications in agricultural economics, however, are rare. Vedenov (2008) analyses the relationship between individual farm yields and area yields and Zhu *et al*. (2008) investigate the dependence of prices and yields in the context of revenue insurance.

The remainder of the article is divided into two main sections. Section II describes the methodological approach. We briefly review the concept of the copula and present the idea of a fully nested Archimedian copula for modelling high dimensional dependence structures. This concept is then applied to the estimation of joint yield risks of major wheat producers. Data and estimation results are presented in section III. The article ends with conclusions on the systemic risk of wheat production and the effectiveness of global trade to diversify this risk.

II Modelling Stochastic Dependence using Copulas

Theoretical framework

The outcome of multi-dimensional risks can be presented by using a multivariate distribution $F(x_1,\cdots,x_d)$, in which x_i denotes the realization of the random variable X_i, $i=1,\cdots,d$. Direct estimation of the joint distribution F usually fails due to insufficient data. Alternatively, the joint distribution F can be determined by linking the univariate marginal distribution $F_i(x_i)$ by means of the (linear) correlation coefficients between random variables. Yet, such kind of estimation of the joint distribution F is valid only for the family of elliptical multivariate distribution including multivariate normal distribution (cf. Embrechts *et al*., 2002). If this assumption does not hold then the associated (linear) correlation coefficients are not able to capture the complete information about the dependence structure between each single risk involved – particularly for the tail dependence, which plays an important role for the analysis of extreme events. Xu *et al*. (2010) showed with an example of weather risk how large the estimation error can be if linear correlation were applied instead of a more appropriate and sophisticated model.

The theoretical fundament of the copula concept is given by Sklar's Theorem (Sklar, 1959), which states that any d-dimensional distribution $F(x_1,\cdots,x_d)$ can be described by using a copula function $C(\cdot)$:

$$F(x_1,\ldots,x_d) = C\{F_1(x_1),\ldots,F_d(x_d)\}, \quad \forall\ x_1,\ldots,x_d \in \overline{R} \tag{1}$$

$F_i(x_i)$ denotes the (univariate) marginal distribution and $u_i = F_i(x_i) \rightarrow [0,1]$ is therefore the uniform marginal distribution. Following Joe (1997), the copula function can be understood as a multivariate distribution function with all margins being uniformly distributed on $[0,1]$:

$$C(u_1,\ldots,u_d) = F\{F_1^{-1}(u_1),\ldots,F_d^{-1}(u_d)\}, \quad F_i^{-1}(u_i) = x_i \tag{2}$$

Equation 1 allows us to decompose any d-dimensional joint distribution function $F(x_1,\ldots,x_d)$ into its d marginal distributions $F_1(x_1),\ldots,F_d(x_d)$, and a copula $C(\cdot)$, which describes the dependence structure among the d random variables. As with the estimation of any distribution function one can apply either parametric or nonparametric (e.g. kernel) approaches (Chen and Huang, 2007). Vedenov (2008) argues that a nonparametric copula is a natural choice since there is no constructive way to determine the optimal copula function and thus the danger of misspecifying the copula is high. On the other hand, if valuable prior information is available, parametric methods can improve the estimation results (Charpentier *et al*., 2007; Genest *et al*., 1995). The main reason for using a parametric approach in this study is its superiority in simulating data from the copula. In fact, up to now there is no efficient method on simulation from the multivariate empirical distributions.

Parametric copulas can be classified into elliptical and Archimedean copulas. In the following, we concentrate only on Archimedean copulas, which can be described in a closed form. This property is convenient when applying simulation procedures. The general expression of Archimedean copulas is given by:

$$C(u_1,\ldots,u_d) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d)) \tag{3}$$

where $\varphi(\cdot)$ is called the generator function with $\varphi(0) = 1$, $\varphi(\infty) = 0$ and φ^{-1} is its inverse. The family of Archimedean copulas includes three important types: the Gumbel, the Clayton and the Frank. The Gumbel copula is asymmetric and appropriate for presenting upper tail dependence which shows a stronger linkage between positive values, more variability and more mass in the positive tail than, for example, the Gaussian copula (Okhrin, 2007). In contrast to the Gumbel copula, the Clayton copula assigns a higher probability to joint extreme negative events than to joint extreme positive events. It displays lower tail dependence and is characterized by zero upper tail dependence (Nelsen, 2006, p. 215). Because of this feature, the Clayton copula has been widely used in financial applications and risk management (e.g., Junker and May, 2005; Blum *et al*., 2002). The Frank copula is adequate to model symmetric dependence structures and exhibits both positive and negative dependence. In contrast to the Gumbel and the Clayton copula, the Frank copula implies tail independence (Nelsen, 2006, p. 215).

Hierarchical construction of a multi-dimensional dependence structure

Although the approach described above provides a relatively easy method to estimate and simulate the high-dimensional dependence structure, it is in fact extremely restrictive in the practical application of modelling a higher-dimensional case. The reason is that most multivariate Archimedean copula models present the whole dependence structure with only one single copula parameter θ, independent of the dimension of the model. Consequently, the substructure of the dependence is invisible. Furthermore, a multivariate Archimedean copula implicitly assumes that the order of margins u_i within the copula function is exchangeable. For instance, for a three dimensional case this means $C(u_1,u_2,u_3) = C(u_3,u_1,u_2)$. However, the implied permutation symmetry of the copula represents a very specialized dependence structure, which is not plausible for many applications (cf. McNeil *et al*., 2005, p. 224; Savu and Trede, 2010).

In view of these shortcomings, attempts have been made to develop a more flexible and appropriate Archimedean model for capturing a high-dimensional dependence structure. One convincing method is the so-called Fully-Nested Archimedian Copulas (FNAC), which aggregates one dimension step by step starting from a low dimensional copula (Savu and Trede, 2010). The model for the d -dimensional dependence structure using the FNAC approach can be expressed as:

$$C(u_1,\ldots,u_d)=\varphi_{d-1}^{-1}\{\ \varphi_{d-1}\circ\varphi_{d-2}^{-1}\left[\ldots\left(\varphi_2\circ\varphi_1^{-1}[\varphi_1(u_1)+\varphi_1(u_2)]+\varphi_2(u_3)\right)\right.$$
$$\left.+\ldots+\varphi_{d-2}(u_{d-1})\right]+\varphi_{d-1}(u_d)\ \} \tag{4}$$

where $\varphi_1(\cdot),\ldots,\varphi_{d-1}(\cdot)$ are generator functions and their inverses $\varphi_1^{-1}(\cdot),\ldots,\varphi_{d-1}^{-1}(\cdot)$ capture the joint distribution of each aggregation step from low dimension (low level) to high dimension (high level). The symbol „$\circ$“ is the composition operator. The FNAC method (Equation 4) is therefore characterized by up to $d-1$ copulas composed of $d(d-1)/2$ possible combinations of distinct bivariate marginal distributions $(F_i(x_i),F_j(x_j))$ with $i\neq j$. The substructures of multidimensional dependence cluster $\varphi_{d-1}\circ\varphi_{d-2}^{-1}$ are expressed hierarchically (see Fig. 2 in Section III). The construction of the multivariate dependence structure can be carried out with following steps: Initially at the lowest level, two variables with the largest copula parameter θ among all of the $d(d-1)/2$ pairs of variables will be combined by using a bivariate Archimedean copula. The first combination at the lowest level (two-dimensional dependence structure) can be described as:

$$z_1=C_1(u_1,u_2)=\varphi_1^{-1}[\varphi_1(u_1)+\varphi_1(u_2)] \tag{5}$$

On the second level (three-dimensional dependence structure), the joint distribution z_1 with u_3 is nested by using $C_2(z_1,u_3)$:

$$C_2(z_1,u_3)=\varphi_2^{-1}\{\varphi_2\circ\varphi_1^{-1}[\varphi_1(u_1)+\varphi_1(u_2)]+\varphi_2(u_3)\}=\varphi_2^{-1}[\varphi_2(z_1)+\varphi_2(u_3)] \tag{6}$$

This procedure will be repeated until the highest level has been reached. Since Equation 4 involves $d-1$ generator functions, there are total $d-1$ copula parameters to be estimated (in contrast to Equation 3).

Usually, all bivariate combinations within a FACN structure employ the same copula type, so that the copula models differ only in the copula parameters. These parameters must be strictly monotone decreasing from the lowest level to the highest level, $\theta_1\geq\theta_2\geq\cdots\geq\theta_{d-1}$. This condition, together with the use of the same copula type, ensures that the outcome of the combination of two copulas from different levels results again in a proper copula (Joe, 1997; Embrechts *et al*., 2003).

Estimation and Goodness-of-Fit tests of copulas

In general, three approaches are available to estimate the parameters of a copula (cf. Cherubini *et al*., 2004). The Exact Maximum Likelihood Method (EMLM) estimates the copula parameter θ and the parameters of the marginal distributions $\alpha_1,\cdots,\alpha_d$ simultaneously. Alternatively, a two-step procedure can be used, where the parameters of the margins α are estimated first. Afterwards the copula parameters are determined, e.g. by maximum likelihood, treating the parameters of the margins as given. This procedure is called the inference for margin (IFM) method (Joe 1997). The IFM method is less efficient than the one-step maximum likelihood but computationally more attractive. An alternative semi-parametric estimation procedure is the Canonical Maximum-Likelihood (CML) (Haerdle *et al*., 2008). The log-likelihood function is given by:

$$\hat{\theta}=\underset{\theta}{\operatorname{argmax}}\sum_{j=1}^{k}\ln[c\{\hat{F}_1(x_{1j}),\ldots,\hat{F}_d(x_{dj});\theta\}] \tag{7}$$

$\hat{F}(x_{ij})$ is the i-th empirical marginal distribution based on the observations x_{ij}, $j = 1,\cdots,k$. $\hat{\theta}$ is also called as Maximum-Pseudo-Likelihood-Estimator or rank-based Maximum-Likelihood-Estimator. The difference between CML and the two previously introduced methods, EMLM and IFM, is that the parametric marginal distribution is substituted by the empirical (rank-based) marginal distribution (Equation 10). The CML method is quite useful if the precise estimation of parametric margins is hampered by a limited number of observations. The CML method will be applied to the estimation of the copula parameter in this study.

Since different copula models imply very different dependence structures, it is important to apply the correct one that fits the empirical data the best. Although the underlying test problem is analogous to a Goodness-of-Fit (GoF) test for univariate distributions, it is more difficult applied to multidimensional distributions. In this context, Genest and Remillard (2008) propose a GoF-test belonging to the class of the dimension reduction approaches. Let C be the true d-dimensional copula for the empirical data, which is to be modelled and let C_θ be the underlying copula candidate to be tested. The subscript θ is the associated copula parameter. The testing hypothesis can be expressed as:

$$H_0 : C \in \boldsymbol{C} = \{C_\theta ; \theta \in \Theta\} \text{vs.} H_1 : C \notin \boldsymbol{C} = \{C_\theta ; \theta \in \Theta\} \tag{8}$$

where Θ is the space of copula parameter. The GoF-test for Equation 8 is based on the Cramer-von-Mises Statistic $\hat{T}$, which measures the distance between empirical copula $\hat{C}(\cdot)$ and the candidate of parametric copula $C_\theta(\cdot)$.

$$\hat{T} = k \cdot \int_{[0,1]^d} \left[\hat{C}(Z) - C_\theta(Z)\right]^2 d\hat{C}(z) = \sum_{j=1}^{k} \left[\hat{C}(Z_j) - C_\theta(Z_j)\right]^2 \tag{9}$$

Herein, Z denotes the d-dimensional empirical marginal distribution which can be described as

$$Z = (z_1, \ldots, z_d) = \left(\frac{R_{1j}}{k+1}, \ldots, \frac{R_{dj}}{k+1}\right), \; j = 1,\cdots,k \tag{10}$$

where k is the number of observations and R_{ij} is the rank of the observations x_{ij} amongst (x_{11} ,..., x_{dk}), $i \in (1,\cdots,d)$. The elements of Z are called pseudo-observations.

The empirical copula $\hat{C}$ can then be expressed as

$$\hat{C}(Z) = \frac{1}{k}\sum_{j=1}^{k} I(z_{j1} \le u_{(1)}, \ldots, z_{jd} \le u_{(d)}), \; \left(u_{(1)}, \cdots, u_{(d)}\right) \in [0,1]^d \tag{11}$$

where $I(\cdot)$ is an indicator function and $u_{(1)},\cdots,u_{(d)}$ are order statistics from the sample.

To obtain reliable p-value estimates, a parametric bootstrap procedure is needed, in which the calculation of Equation 9 is repeated for L times. The p-value for Equation 9 can be calculated by using the following expression:

$$p = \frac{1}{L+1}\sum_{l=1}^{L} I(T_l^* > \hat{T}), \; l = 1,\cdots,L \tag{12}$$

The random variable T_l^* is defined as:

$$T_l^* = \sum_{j=1}^{k}\left[C_l^*(Z_l^*) - C_{\theta_l^*}(Z_l^*)\right]^2,\ Z_l^* = (z_{1,j,l},\ldots,z_{d,j,l}) \tag{13}$$

The bootstrap procedure for T_l^* is comprised of the following steps:

a) the random sample $(x_{1,l}^*,\cdots,x_{d,l}^*)$ from the hypothesized copula C_θ is generated and the associated pseudo-samples Z_l^* are calculated according to Equation 10;

b) the parameters θ_l^* are estimated based on Z_l^*;

c) the empirical copula C_l^* is determined according to Equation 11;

d) the steps from a) to c) will be repeated for L times; and

e) the p-value is calculated, which is straightforward once C_l^* and θ_l^* are known.

Copula-based simulation of simultaneous risks

Once the marginal distributions and the copula have been determined and estimated parametrically, the realization of d-dimensional random variables, which follow the joint distribution $F(x_1,\cdots,x_d)$ can be generated by employing the Monte-Carlo simulation method. In the context of multi-dimensional copulas, the „conditional inverse method" has been employed (Haerdle and Okhrin, 2009). The basic idea of this method is to generate the multi-dimensional random variables recursively according to the associated conditional distribution. First, the d-dimensional independent variables $v_1,\ldots,v_d$ are drawn from the uniform distribution $U \in [0,1]$. Afterwards the d-dimensional uniform-distributed variables $u_1,\ldots,u_d$ will be generated. In contrast to the generation of $v_1,\ldots,v_d$, the d-dimensional variables $u_1,\ldots,u_d$ are stochastic dependent. The dependence structure of $u_1,\ldots,u_d$ is captured by the copulas $C(u_1,\cdots,u_d)$. To achieve this, a recursive procedure is carried out. In the first step, the initial value of u_1 is equal to v_1. The consecutive values $u_2,\ldots,u_d$ can be determined by the following procedure:

$$u_i = \Lambda^{-1}(v_i),\ i = 2,\cdots,d \tag{14}$$

The relationship between u_i and v_i can be determined by a certain function $\Lambda(\cdot)$ and through the inverse function of $\Lambda(\cdot)$ the independent uniform-distributed multivariate variables $v_1,\ldots,v_d$ will be transformed to the dependent uniform-distributed multivariate variables $u_1,\ldots,u_d$. Therefore, the function $\Lambda(\cdot)$ also describes the dependence structure of $u_1,\ldots,u_d$. $\Lambda(\cdot)$ can be defined as:

$$\begin{aligned}\Lambda(u_i|u_1,\cdots,u_{i-1}) &= P(U_i \le u_i|U_1 = u_1,\ldots,U_{i-1} = u_{i-1}) \\ &= \frac{\partial^{i-1}C_i(u_1,\cdots,u_i)}{\partial u_1,\cdots,\partial u_{i-1}} \Bigg/ \frac{\partial^{i-1}C_{i-1}(u_1,\cdots,u_{i-1})}{\partial u_1,\cdots,\partial u_{i-1}}\end{aligned} \tag{15}$$

$C_i(\cdot)$ stands for the copula with i-th dimensional margins. In the last step, the realizations x_i associated with each random variable X_i can be generated by using

$$x_i = F_i^{-1}(u_i),\ \ i = 1,\cdots,d \tag{16}$$

where $F_i^{-1}(u_i)$ is the inverse function of the marginal distribution $F_i(x_i)$. The resulting realizations $x_1,\ldots,x_d$, which are simulated by using the above-mentioned procedure, follow the joint distribution $F(x_1, \cdots, x_d)$, or more precisely they follow the copula $C(u_1, \cdots, u_d)$.

III Empirical Analysis of Shortfall Risk on Global Crop Production

Data and modelling procedure

The empirical analysis is based on the yield data of wheat for the main producing countries worldwide. To limit computational work, the eight most important countries are selected: Canada, China, France, Germany, India, Pakistan, Turkey and the USA[1]. In 2007, the amount of the wheat production in these countries share 58% of the total production worldwide. The empirical data of wheat yields for these countries is provided by FAO (http://faostat.fao.org) and is available from 1961 to 2007, comprising of 47 observations. The empirical analysis involves the following steps: First, the total amount of wheat yield is converted into per hectare yields. Using the hectare yields makes it easier to compare countries with different production scales. Moreover, changes of the acreage from year to year will not affect the analysis[2]. Second, to eliminate the production trend caused by technical progress, the hectare yields are detrended by applying the following linear model:

$$y_{i,t} = m_i + \alpha_i \cdot t + \varepsilon_i, \quad i = 1, \cdots, 8 \tag{17}$$

where $y_{i,t}$ is the hectare yield (dt/hectare), m_i the constant term for country i, t denotes time and α_i captures technical progress. The residual ε_i is the detrended hectare yield.

Table A1 in the appendix shows the detrended hectare ε yields for eight countries in ascending order. Note that yields realized in 2007 were the third worst yields since 1961 in four countries; however, this yield did not affect all eight countries. For example, in Pakistan a relatively good harvest was recorded in 2007. Furthermore, the figures provide an impression about the size of the yield volatility as well as their differences in each of the observed countries for each year.

Using the detrended yield data ε_i in Equation 17, univariate yield distributions are specified and estimated for each wheat producing country. The type of yield distribution is selected according to a χ^2-test. A hierarchical Archimedean copula, FNAC, has then been estimated for the eight-dimensional detrended hectare yield, ε_i as described in the previous section. Once a hierarchical dependence structure is defined, the vector ε_i can be simulated by using conditional inverse method (see Equations 14 - 16). Finally, the total yield (tons) for the eight countries related to year 2007, $y_n^{sim,t}$, as well as the average yield (dt/hectare) for the eight countries, $\bar{y}_n^{sim,t}$, can be calculated by:

$$y_n^{sim,t} = \sum_{i=1}^{8} \left(\varepsilon_{i,n} + m_i + \alpha_i \cdot t\right) \cdot a_i, \quad i = 1, \cdots, 8, \; t = 2007 \tag{18}$$

---

[1] Although Russia belongs to the largest wheat producers, it had to be excluded from the analysis due to insufficient data.

[2] The harvested areas of wheat are depicted in Fig. A1 in the appendix.

and

$$\bar{y}_n^{sim,t} = y_n^{sim,t} \Big/ \sum_{i=1}^{8} a_i \tag{19}$$

n denotes the n-th simulation and a_i is the harvest area for country i. The simulation is carried out 10 000 times. The procedure results in an empirical distribution for the average yield and the total yield of wheat from which statistical parameters of interest can be derived.

Results

The results of the GoF-test for the marginal distributions of the detrended hectare yield, ε_i, are presented in Table A2 in the appendix. Apparently, the random yields in the various countries cannot be described by only one type of distribution. Instead, the yields follow different distributions, such as Weibull, Gaussian or Lognormal distributions. Similar results are reported by Moriondo *et. al.*, (2009) as well as Upadhyay and Smith (2005). This finding confirms that the joint yield distribution for all countries cannot be multivariate normal and hence the requirement for using linear correlations coefficients is not fulfilled.
The selection of the appropriate copula type is based on the GoF-test described in the previous section. The Clayton, the Gumbel and the Franck copula have been considered as candidates. The test was conducted for all 28 possible country pairs. In most cases the Clayton copula yields the best fit and thus this copula type was chosen for the entire FNAC. This result also indicates that the use of linear correlation would be inappropriate.

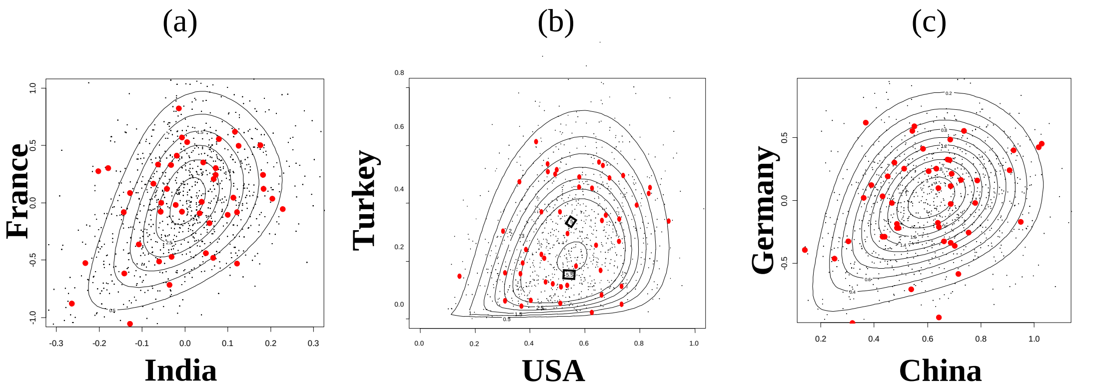
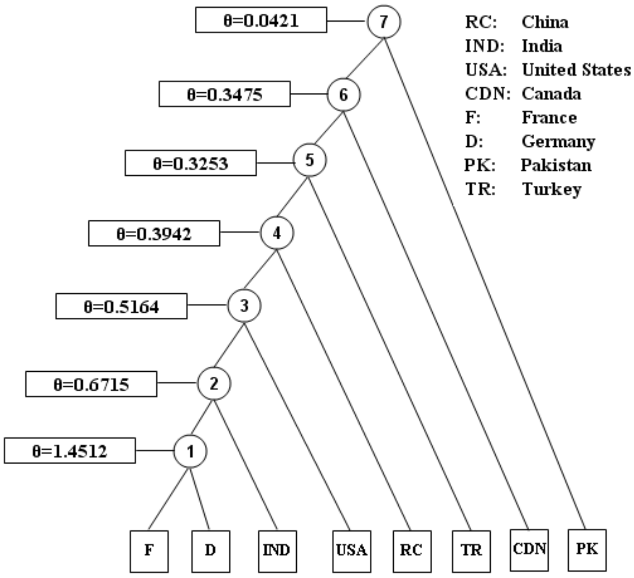


Fig. 1. Contour lines for the joint probability of the residual of wheat yield (dt/hectare) for (a) France and India, (b) Turkey and USA, (c) Germany and China. (Type of copula: Clayton; bold dots = observed; thin dots = simulated).

Fig. 1 illustrates the bivariate yield distributions for the following country pairs: a) France and India; b) Turkey and USA; and c) Germany and China. The contour lines of the bivariate distributions are obviously not symmetric and elliptical as implied by linear correlations.

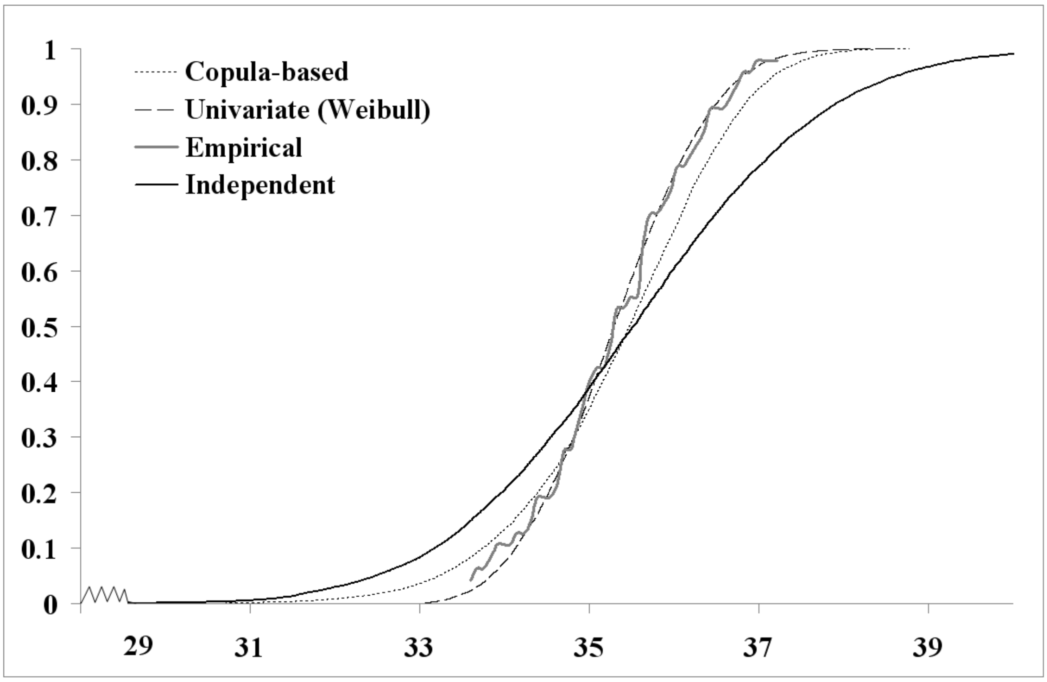


Level	Goodness-of-Fit test (Clayton copula)	
1	Test statistics	0.0216
	p -value	0.5790
2	Test statistics	0.0195
	p -value	0.7050
3	Test statistics	0.0272
	p -value	0.3940
4	Test statistics	0.0463
	p -value	0.0920
5	Test statistics	0.0198
	p -value	0.7330
6	Test statistics	0.0215
	p -value	0.6470
7	Test statistics	0.0226
	p -value	0.6500

Fig. 2. Fully nested hierarchical structure of the Archimedean copula, FNAC

Next, the structure of the hierarchical copula is determined. The aggregation of countries rests on the test statistic of the GoF-test. Fig. 2 depicts the resulting dependence structure (left part) and presents the test statistics for the Clayton copula (right part). The estimates of the copula parameters decrease with increasing aggregation level, which is in accordance with theoretical requirements. An exception is the estimate on level six, which might be caused by the limited database that includes only 47 observations. As expected, the yields between neighbouring countries, for example between Germany and France, show a high stochastic dependence. On the other hand, it is surprising that the stochastic dependence between India and Germany is higher than the stochastic dependence between India in Pakistan. That means that stochastic dependence is not a simple function of distance as implied by commonly used decorrelation functions.
Fig. 3 and Table 1 present the main results of our analysis, namely the distribution function for per hectare yields (Fig. 3(a)) and for the total wheat production (Fig. 3(b)). Fig. 3(a) and Fig. 3(b) refer to the yield level (mean yield with trend) and acreage at the end of the observation period, i.e. 2007. The interpretation of Fig. 3(a) is rather difficult since the distribution is derived by aggregating very heterogeneous regions. Thus, is helpful to use the outcome of the extreme year 2007 as a reference point. In 2007, a per hectare yield of 33.52 decitons was realized. According to Fig. 3(a) the stochastic yields fall below this level with a probability of approximately 7%. In other words, the harvest observed in 2007 was a rare event, but statistically it will recur on average every 15 years under constant conditions.

(a) Average yield (dt/hectare)



(b) Total yield (100 million tons)

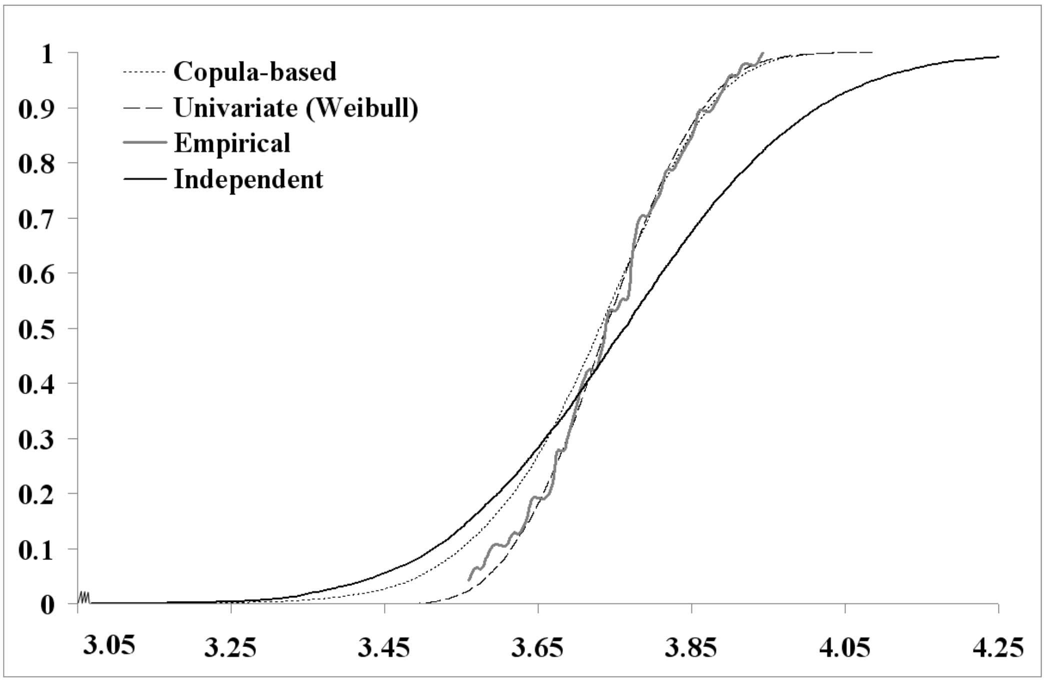


Fig. 3. Empirical and parametric distribution functions of global wheat production

We present two alternative estimates for the distribution function of global wheat production. First, a parametric distribution, namely a Weibull distribution, has been estimated. In contrast to the copula based approach, this distribution has been directly fitted to the aggregated production data. Moreover, an empirical (nonparametric) distribution function has been estimated. Though a direct estimation of the distribution function is much simpler compared to the derivation from marginals and copulas, it has some obvious flaws. First, the estimation relies on only 47 observations and thus it will be plagued by low reliability, particularly in the tails. Second, the stochastic dependence structure of the various production regions is not revealed. Finally, it is questionable if different country-specific marginal yield distributions can be correctly represented by a single parametric distribution function. In light of these drawbacks, it is not surprising that the quantiles of the three distribution functions differ. Table 1 shows that the risk of global crop failures is underestimated by the direct estimation of the univariate distribution, though the means of all three distribution functions are almost equal. For example, the 1% quantile of the Weibull distribution for per hectare yield amounts to 33.37 dt/ha which is 1.28 dt/ha higher than the copula based estimate. The corresponding value of the empirical distribution cannot be uniquely determined due to its discrete nature.

Table 1. Parameters of the estimated distribution functions for wheat production (all regions for the year 2007)

Quantile	Copula-based		Univariate (Weibull)		Empirical		Independent	
	AY[a]	TP[b]	AY[a]	TP[b]	AY[a]	TP[b]	AY[a]	TP[b]
1%	32.09	3.37	33.37	3.54	n.a.	n.a.	31.31	3.32
5%	33.22	3.50	33.84	3.58	33.64	3.55	32.47	3.44
10%	33.74	3.55	34.12	3.62	33.91	3.59	33.17	3.51
50%	35.47	3.61	35.29	3.74	35.29	3.74	35.52	3.76
Mean	35.37	3.73	35.31	3.59	35.30	3.59	35.53	3.76
Variance	1.47	0.0152	0.81	0.0084	0.86	0.0089	3.47	0.0389

Notes: [a] Average yield in dt/ha. [b] Total production in 100 mill. tons.

To highlight the effect of global diversification of yield risk we present the results for the hypothetical scenario where yields are stochastically independent between all production regions. In this case, the distribution widens considerably and the variance increases by a factor of more than two (see Table 1 and Fig. 3). The effect of diversification is further elaborated in Fig. 4, which compares the variability of per hectare yields of different production regions. Apparently, the fluctuations of per hectare yields can be smoothed for most production regions by aggregation. This means that global trade can considerably reduce the shortfall risk for individual countries. It should be noted that a trade-off between average productivity and yield volatility exists. Intensive high tech production systems in Germany and France appears to be more sensitive to stochastic weather conditions compared to more extensive production systems in North America. Yield risk in India is similar to the global average, yet its average yields are relatively low.

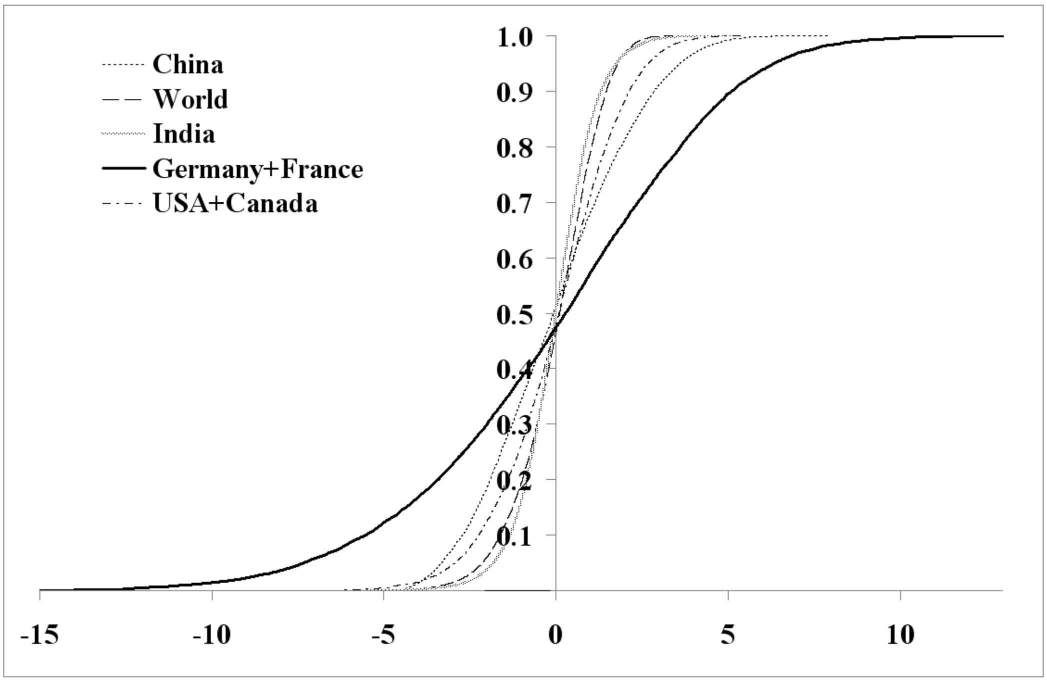


Fig. 4. Empirical and parametric distribution functions of wheat yields (deviations from mean in dt/hectare)

IV Conclusions

The main finding of the analysis is that yield risk for wheat as an important foodstuff can be reduced by pooling and aggregating production globally. This provides an argument for international agricultural trade, which is beyond deterministic welfare gains. It has been stated that price volatility for agricultural commodities in the EU increases due to the deregulation of markets (Britz and Heckelei, 2008). It should be stressed that the increase of price risk is a result of the abolishment of price protection measures and not a result of international trade. Based on our analysis we conjecture that price volatility in the EU would be even higher without a participation in international food markets. Nevertheless, global diversification cannot completely reduce yield risks. A shortfall, such as in 2007, is not a once in a lifetime event. From a statistical point of view, similar production conditions will occur every 15 years *ceteris paribus*.

However, a word of caution is necessary when interpreting our statistical findings. Our results do not allow for predictions about the scarcity of wheat (and hence price peaks) because the demand side was not taken into account. Clearly, the demand for wheat varies depending on population growth, income and the demand for biofuels. Furthermore, even our analysis of the supply side of the wheat market was incomplete since only eight countries were considered instead of all wheat producing countries. Moreover, the production of only a single year was analysed, ruling out time diversification. Usually food markets are short in supply if multiple poor harvests occur in a row and worldwide inventories are too small for smoothing this shortfall. This was the case in 2007/08. It should also be mentioned that our probability statements refer to the yield level at the end of the observation period. An increase of the yield levels due to future technological progress has not been anticipated. Finally, one should recall that the production area depends on prices and thus is endogenous. For example, one could observe that farmers increased the acreage of wheat worldwide in 2008 as a response to the preceding price boom. Some of the aforementioned aspects could be handled in the framework of market equilibrium models. A linkage of existing market models with the presented stochastic analysis is a promising task for further research.

References

Blum, P., Dias, A., Embrechts, P. (2002) The ART of dependence modelling: the latest advances in correlation analysis, in *Alternative Risk Strategies* (ed) L. Morton, Risk Books, London, pp. 339-356.

Bianchi, C., Carta, A., Fantazzini, D., De Giuli, M.E. and Maggi, M.A. (2009) A copula-VAR-X approach for industrial production modelling, *Applied Economics*, in print.

Britz, W., and Heckelei, T. (2008) Recent developments in EU policies – challenges for partial equilibrium models, Paper prepared for presentation at the 107th EAAE Seminar "Modelling of Agricultural and Rural Development Policies". Sevilla, Spain, January 29th -February 1st, 2008.

Chavas, J.-P. and Kim, K. (2006) An econometric analysis of the effects of market liberalization on price dynamics and price volatility, *Empirical Economics*, **31**, 65–82.

Charpentier, A., Fermanian, J.-D. and Scaillet, O. (2007) The estimation of copulas: theory and practice, in *Copulas: From Theory to Application in Finance,* (Ed) J. Rank, Risk Publications, London, pp. 35-60.

Chen, S.-X. and Huang, T. (2007) Nonparametric estimation of copula functions for dependence modelling, *The Canadian Journal of Statistics*, **35**, 265-282.

Chen, Y-H, Wang, K. and Tu, A.H. (2009) Default correlation at the sovereign level: evidence from some Latin American markets, *Applied Economics*, in print.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004) *Copula Methods in Finance,* Wiley, Chichester.

Embrechts, P., McNeil, A. and Straumann, D. (1999) Correlation: pitfalls and alternatives, *RISK Magazine*, **May**, 69-71

Embrechts, P., McNeil, A and D. Straumann, D. (2002) Correlation and dependence in risk management: properties and pitfalls, in *Risk Management: Value at Risk and Beyond*, (Ed) M.A.H. Dempster, Cambridge University Press, Cambridge, pp. 176-223.

Embrechts, P., Lindskog, F and McNeil, A. (2003) Modelling dependence with copulas and applications to risk management, in *Handbook of Heavy Tailed Distributions in Finance*, (Ed) S.T. Rachev, Elsevier, North-Holland, pp. 329–384.

FAO (2009) FAOSTAT database. Available at: http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor (accessed 2 February 2009).

Genest, C., Ghoudi, K. and Rivest, L. P. (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, **82**, 543-552.

Genest, C. and Remillard, B. (2008) Validity of the parametric bootstrap for goodness-of-fit Testing in Semiparametric Models, *Ann. Henri Poincare*, **44**, 1096-1127.

Haerdle, W., Okhrin, O and Okhrin, Y. (2008) Modeling dependencies in finance using copulae, in *Applied Quantitative Finance*, 2nd edn (Eds) W. Haerdle, N. Hautsch and L. Overbeck, Springer Verlag, Berlin.

Haerdle, W. and Okhrin, O. (2009) De copulis non est disputandum, copulae: an overview, SFB 649 Discussion Paper (2009-031), Humboldt University of Berlin, Berlin, Germany.

Joe, H. (1997) *Multivariate Models and Dependence Concepts*, Chapman&Hall, London.

Junker, M. and May, A. (2005) Measurement of aggregate risk with copulas, *Econometrics Journal*, **8**, 428-454.

Lotze-Campen, H. (2007) Managing regional climate risks in agriculture through diversified international trade relationships, Presentation at the 101st Workshop of the European Association of Agricultural Economists (EAAE) on Managing Climate Risks in Agriculture. Berlin, 05.July 2007, *miemo*, PIK, Potsdam, Germany.

McNeil, A.J., Frey, R. and Embrechts P. (2005) *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princeton.

Moriondo, M., Bindi, M and Lugeri, N. (2009) Modelling yield distribution as affect by extreme events, in *IOP Conference Series: Earth and Environmental Science,* Vol. 6, No 022011.

Musshoff, O., Odening, Xu, W. (2009): Management of climate risks in agriculture - will weather derivatives permeate? *Applied Economics* (in print)

Nelsen, R.B. (2006) *An Introduction to Copulas*, 2nd edn, Springer, NewYork.

Odening, M., Berg, E. and C.G. Turvey, C.G. (Eds) (2008) Management of climate risk in agriculture, *Agricultural Finance Review*, **68** (1), Special Issue.

Okhrin, O. (2007) Hierarchical Archimedean copulas: structure determination, properties, applications, PhD thesis, European University Viadrina, Frankfurt (Oder), Germany.

Sklar, A. (1959) Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Stat. Univ. Paris*, **8**, 299-231.

Sarris, A. (2009): Factors Affecting recent and Future Price Volatility of Food Commodities, in *Risiken in der Agrar- und Ernährungswirtschaft und ihre Bewältigung*, (Eds) E. Berg et. al., Landwirtschaftsverlag, Münster-Hiltrup, pp. 29-48.

Savu, C. and Trede, M. (2010). Hierarchies of Archimedean copulas, *Quantitative Finance*, **10**, 295-304.

Turgutlu, E. and Ucer, B. (2010) Is global diversification rational? Evidence from emerging equity markets through mixed copula approach, *Applied Economics*, **42**, 647-658.

Upadhyay, B.M. and Smith, E.G. (2005) Modeling crop yield distributions from small samples, Selected Paper at the CAES annual meeting, July 6-8, 2005, San Francisco, California.

Vedenov, D. (2008) Application of copulas to estimation of joint crop yield distributions, Selected paper at the Annual Meeting of the AAEA, July 27-29, 2008, Orlando, Florida. Available at http://ageconsearch.umn.edu/handle/6264 (accessed 20 March 2009).

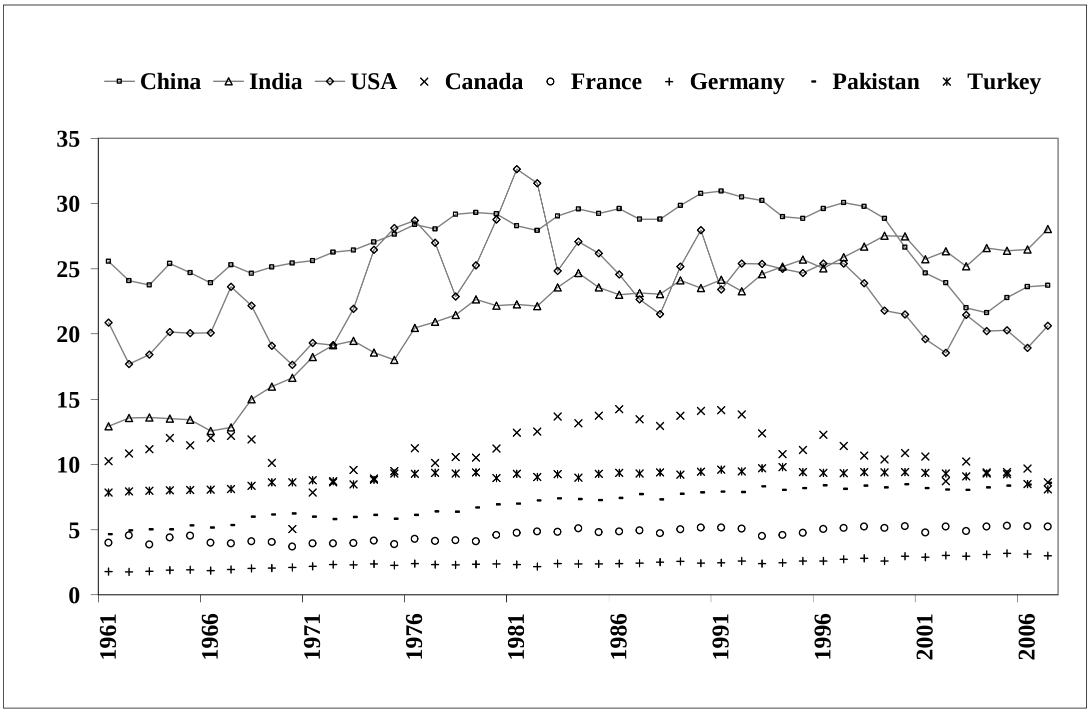
von Witzke, H., Noleppa, S. and Schwarz, G. (2008) Global agricultural market trends and their impacts on European Union agriculture, Working Paper No. 84/2008, Department of Agricultural Economics, Humboldt University of Berlin.

Xu, W., Filler, G., Odening, M. and Okhrin, O. (2010) On the systemic nature of weather risk, *Agricultural Finance Review*, in print.

Yang, J., Haigh, M.S.; Leatham, D.J. (2001) Agricultural liberalization policy and commodity price volatility: a GARCH application, *Applied Economics Letters*, **8**, 593-598.

Zhu, Y., Ghosh, S.K. and Goodwin, B.K. (2008) Modeling dependence in the design of whole farm insurance contract: A copula-based model approach, Selected paper at the Annual Meeting of the AAEA, July 27-29, 2008, Orlando, Florida. Available at http://ageconsearch.umn.edu/handle/6282 (accessed 20 March 2009).

Appendix



Source: FAO (2009)

Fig. A1. Harvest area of wheat (million hectares)

Table A1. Ranking of detrended per hectare yield of wheat (dt/hectare)

Rank	Canada		China		France		Germany		India		Pakistan		Turkey		USA		Total	
	Yield	Year	Yield	Year	Yield	Year	Yield	Year	Yield	Year	Yield	Year	Yield	Year	Yield	Year	Yield	Year
1	6.13	1988	0.56	1977	**14.85**	**2007**	20.92	2003	4.66	2006	5.32	1987	8.68	1973	13.48	2002	9.39	2002
2	7.52	1961	1.67	2002	18.96	2003	**21.35**	**2007**	4.96	2005	5.62	1966	**8.90**	**2007**	15.05	2006	**9.43**	**2007**
3	8.97	2002	2.19	1980	20.73	2006	23.59	1976	**5.03**	**2007**	5.73	1984	8.97	1999	15.13	1989	10.13	2003
4	10.32	2001	2.34	2003	22.35	1976	24.80	2006	5.26	1980	5.74	1967	8.99	1994	15.13	1974	10.16	2006
5	10.85	1984	2.78	1973	23.36	1966	26.04	2002	5.50	1974	5.90	1994	9.08	1974	15.65	1991	10.20	1974
6	11.68	1989	2.85	2001	24.23	1961	26.75	1977	5.86	1964	6.31	1997	9.09	1961	15.70	1995	10.20	1977
7	11.86	1967	3.05	2000	24.26	2005	27.05	1966	5.88	1966	6.35	1996	9.28	1970	**15.72**	**2007**	10.36	2001
8	11.91	1974	3.48	1978	24.41	1975	27.28	1992	6.00	2003	6.42	2002	9.56	2001	15.77	1996	10.50	1961
9	12.19	1985	3.48	1981	24.71	2001	27.43	1989	6.00	1967	6.44	1978	9.58	2003	15.90	1967	10.66	1970
10	12.23	1979	3.55	1970	24.82	1963	27.44	1980	6.21	1977	6.56	1991	9.61	1968	16.07	1961	10.75	1980
11	**12.73**	**2007**	3.67	1974	25.12	1970	27.80	1970	6.56	2004	6.64	1985	9.67	1965	16.27	1988	10.80	2005
12	12.88	1980	3.81	1969	25.89	1977	27.80	1981	6.66	1978	6.67	1988	9.73	1964	16.46	1963	10.82	1988
13	12.92	1964	3.91	1971	27.75	1969	28.10	1965	6.68	1975	6.68	1999	9.94	2002	16.46	1966	10.87	1973
14	12.98	2003	4.01	1972	28.46	1971	28.48	1975	6.72	1987	6.74	2004	10.01	1995	16.56	1962	11.08	1978
15	13.07	1997	4.02	2005	28.60	1983	28.49	2005	6.73	1981	6.79	1990	10.04	1989	16.61	1964	11.17	1964
16	13.26	1968	4.07	1975	28.68	1986	28.57	1961	6.86	1982	6.82	1993	10.13	1969	16.68	1976	11.22	1975
17	13.35	1987	4.28	1998	28.69	1964	28.81	1972	6.93	1976	6.88	1971	10.29	1997	16.69	1977	11.28	1994
18	13.75	1994	4.55	1976	28.74	1987	28.87	1979	6.97	1973	7.05	1965	10.38	1996	16.87	1965	11.29	1967
19	13.96	1962	4.59	1991	28.77	1965	28.99	1983	6.97	1963	7.11	1964	10.55	1962	16.95	1978	11.30	1989
20	14.11	1998	4.66	2004	28.96	1997	30.38	1982	7.07	1979	7.28	2003	10.69	1966	17.01	1986	11.51	1969
21	14.14	1973	4.99	1968	29.32	1979	30.46	1969	7.10	1988	7.34	1974	10.85	1967	17.14	1975	11.53	1976
22	14.29	1972	5.21	1994	29.52	1981	30.47	1987	7.15	1998	7.35	1982	11.01	2004	17.15	1994	11.54	1995
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
45	17.34	2005	8.65	1983	35.74	1991	36.23	1991	9.12	2000	8.93	1986	13.82	1976	20.37	1971	13.09	1986
46	17.54	1976	9.32	1997	37.73	1998	36.60	2004	9.32	1995	**8.98**	**2007**	13.89	1979	20.41	1984	13.29	1983
47	17.60	1966	9.44	1984	40.88	1984	36.86	2001	9.56	1997	9.50	2000	14.59	1988	21.09	1983	13.60	1984

Source: FAO (2009)

Table A2. Marginal distributions of wheat yield (dt/hectare)

Country	Marginal distribution	Parameters		χ^2 test-statistics	p -value
Canada	Logistic	location	0.0243	2.1910	0.9485
		scale	0.1382		
China	Weibull	shape	3.4396	4.5740	0.7117
		scale	0.6694		
		shift	-0.6022		
Germany	Normal	mean	0.0000	2.8700	0.8966
		SD	0.3895		
France	Normal	mean	0.0000	8.6600	0.2780
		SD	0.4975		
India	Logistic	location	0.0035	3.5530	0.8296
		scale	0.0690		
Pakistan	Weibull	shape	3.4666	3.8940	0.7919
		scale	0.3160		
		shift	-0.2842		
Turkey	Weibull	shape	1.7138	5.2550	0.6288
		scale	0.3280		
		shift	-0.2934		
USA	Weibull	shape	3.7793	4.5740	0.7117
		scale	0.6087		
		shift	-0.5498		

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Localising temperature risk" by Wolfgang Karl Härdle, Brenda López Cabrera, Ostap Okhrin and Weining Wang, January 2011.
002 "A Confidence Corridor for Sparse Longitudinal Data Curves" by Shuzhuan Zheng, Lijian Yang and Wolfgang Karl Härdle, January 2011.
003 "Mean Volatility Regressions" by Lu Lin, Feng Li, Lixing Zhu and Wolfgang Karl Härdle, January 2011.
004 "A Confidence Corridor for Expectile Functions" by Esra Akdeniz Duran, Mengmeng Guo and Wolfgang Karl Härdle, January 2011.
005 "Local Quantile Regression" by Wolfgang Karl Härdle, Vladimir Spokoiny and Weining Wang, January 2011.
006 "Sticky Information and Determinacy" by Alexander Meyer-Gohde, January 2011.
007 "Mean-Variance Cointegration and the Expectations Hypothesis" by Till Strohsal and Enzo Weber, February 2011.
008 "Monetary Policy, Trend Inflation and Inflation Persistence" by Fang Yao, February 2011.
009 "Exclusion in the All-Pay Auction: An Experimental Investigation" by Dietmar Fehr and Julia Schmid, February 2011.
010 "Unwillingness to Pay for Privacy: A Field Experiment" by Alastair R. Beresford, Dorothea Kübler and Sören Preibusch, February 2011.
011 "Human Capital Formation on Skill-Specific Labor Markets" by Runli Xie, February 2011.
012 "A strategic mediator who is biased into the same direction as the expert can improve information transmission" by Lydia Mechtenberg and Johannes Münster, March 2011.
013 "Spatial Risk Premium on Weather Derivatives and Hedging Weather Exposure in Electricity" by Wolfgang Karl Härdle and Maria Osipenko, March 2011.
014 "Difference based Ridge and Liu type Estimators in Semiparametric Regression Models" by Esra Akdeniz Duran, Wolfgang Karl Härdle and Maria Osipenko, March 2011.
015 "Short-Term Herding of Institutional Traders: New Evidence from the German Stock Market" by Stephanie Kremer and Dieter Nautz, March 2011.
016 "Oracally Efficient Two-Step Estimation of Generalized Additive Model" by Rong Liu, Lijian Yang and Wolfgang Karl Härdle, March 2011.
017 "The Law of Attraction: Bilateral Search and Horizontal Heterogeneity" by Dirk Hofmann and Salmai Qari, March 2011.
018 "Can crop yield risk be globally diversified?" by Xiaoliang Liu, Wei Xu and Martin Odening, March 2011.

SFB 649, Ziegelstraße 13a, D-10117 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".