German Journal of Agricultural Economics Vol. 74 (2025), 1-31

Original Research Articles

https://doi.org/10.52825/gjae.v74i.2635

Submitted: 10 Mar. 2025 | Accepted: 15 Jul. 2025 | Published: 26 Aug. 2025

Using Cubic Splines in Crop Insurance Models – A Replication Study

Lorenz Schmidt[1,*], Masud Heydarli[2], Günther Filler[1], and Martin Odening[1]

[1]Humboldt-Universität zu Berlin, Germany

[2]Thünen-Institut für Betriebswirtschaft, Braunschweig, Germany

*Correspondence: Lorenz Schmidt, lorenz.schmidt.1@hu-berlin.de

Abstract: This study replicates the work of Bucheli et al. (2022) to evaluate the reliability and generalizability of their findings on the use of cubic spline methods in weather index insurance design in Eastern Germany. The study consists of two parts: a direct replication and an extended replication using different crops and another regional focus of yield data from Saxony, Germany. The direct replication confirms the original findings, while the extended replication reveals limitations in the model's risk reduction capacities. The gain in model fit, compared with daily temperature data, seems to be modest. It is questionable if an index insurance, which is solely based on temperature, will be considered as an attractive risk management tool, given the substantial basis risk that remains with farmers. The study highlights the importance of considering crop-specific risk profiles, regional climate, and agricultural conditions in insurance design. Moreover, our analysis highlights the relevance of FAIR (findable, accessible, interoperable, reusable) data for the performance of replication studies.

Keywords: Replication, Weather Index Insurance, Risk Management

1 Introduction

Heatwaves induced by climate change have posed significant challenges to global and European crop production, adversely affecting agricultural output and farmers' income (Bras et al., 2021; Heino et al., 2023; Trnka et al., 2014; Van der Velde et al., 2012). In response to these challenges, farmers have explored various strategies to enhance their resilience, including risk-sharing mechanisms such as weather index insurance (Vroege, Finger, 2020). This type of insurance has become increasingly popular among European farmers due to its objectivity and independence; rather than requiring farm inspections to assess damage, it uses an objective index linked to weather variables over a defined period (Bucheli et al., 2022; Doherty et al., 2021).

However, the effectiveness of weather index insurance is often hindered by basis risk, which refers to the mismatch between the triggered insurance payout and the actual yield losses experienced by farmers (Odening, Shen, 2014; Woodard, Garcia, 2008). Basis risk poses a considerable challenge to the widespread adoption of weather index insurance, and thus, reducing basis risk has become a central focus in the design and implementation of these insurance products. There are two main types of basis risk: spatial basis risk and product basis risk. Spatial basis risk arises from the discrepancies between the weather conditions at the insured farmland and the weather indices measured at a reference weather station. Various approaches have been proposed to mitigate spatial basis risk, such as using spatial interpolation methods (e.g. Cao et al., 2015) and diversifying insurance contracts across different regions

(Ritter et al., 2014). Product basis risk stems from the inability of weather indices to fully capture the variability in agricultural yields. While weather conditions play a crucial role in crop production, they are not the sole determinant. Other factors, such as pests, controllable inputs (e.g., fertilizer), and managerial practices, also influence crop yields. Albers et al. (2017) estimated that only 43% of the variability in wheat yield can be attributed to weather conditions. This highlights the inherent limitations of reducing product basis risk through weather index insurance.

Despite this challenge, numerous efforts have been made to optimize the design of weather index insurance to minimize product basis risk. These efforts can be broadly categorized into two main strands of literature: (1) the specification and estimation of the relationship between yield and weather indices, and (2) the selection and specification of appropriate weather indices.

The first strand of literature focuses on reducing product basis risk by using sophisticated statistical modeling to estimate the weather-yield relationship. Various methods have been applied, ranging from traditional statistical approaches such as polynomial regression (Vedenov, Barnett, 2004) and generalized additive models (Bucheli et al., 2022; Zhang et al., 2019; Zou et al., 2023), to more recent machine learning techniques (Chen et al., 2023; Schmidt et al., 2022; Zou et al., 2024). However, there is no consensus on the “best” approach, and the complexity of the methods does not always guarantee superior performance (Oikonomidis et al., 2022).

The second strand of literature focuses on selecting and designing of appropriate weather indices to mitigate product basis risk. Commonly used weather indices include Growing Degree Days (GDD), Cumulative Rainfall (CR), and the standardized precipitation index (e.g. Turvey, 2001). Additionally, remote sensing data have been used to obtain vegetation indices as alternatives to meteorological indices (Möllmann et al., 2019; Vroege, Finger, 2020). The beginning and length of the accumulation period, as well as the aggregation level of weather indices are important design parameters that can considerably influence the effectiveness of weather index insurance. If the accumulation period of the weather indices does not align with the vegetation period of the insured crop, a so-called temporal basis risk may arise. This is considered a special case of product basis risk. To address this problem, it has been suggested to split the entire vegetation period into distinct plant growth stages and to calculate weather indices for each growth stage separately (e.g., Zou et al., 2023).

The study by Bucheli et al. (2022) contributes to both strands of literature by investigating the use of cubic spline methods combined with hourly temperature data in a single peril weather index insurance for winter wheat and winter rapeseed producers in Eastern Germany. Cubic splines are piecewise-polynomial functions that allow for capturing non-linear relationships between temperature and crop yield by fitting smooth curves across different temperature ranges. While their findings suggest potential for risk reducing through heat index insurance based on restricted cubic splines, it is crucial to extend their research to assess the generalizability and robustness of their findings across different contexts. Bucheli et al. (2022) do not claim that their results are valid for entire Germany and for other crops. They write: “Our payout function is generally applicable and future research should assess its performance in other regions, for other heat-susceptible crops and also for other hazards than heat stress.” In our paper, we take up this suggestion for further research and apply their approach to another region in Germany (Saxony). Therefore, this study does not only aim at evaluating the accuracy and consistency of Bucheli et al. (2022) findings but also extends their analyses. Specifically, we pursue two objectives: First, we replicate the original study using the same methodology but with data from different time frames and sources to confirm the reliability of the findings. Second, we test the model’s robustness when applied it to different timeframes, crops and region.

By extending the original study to different contexts, this replication study contributes to the existing literature by (i) guiding policymakers in bridging the gap between theoretical models and their real-world effectiveness, (ii) establishing the generalizability of the non-linear insurance design across different crops, regions, and data sources, and (iii) identifying limitations of the existing model to aid future researchers in refining their approaches. Understanding these limitations is critical: if a model's performance varies substantially across crops or regions, this has direct consequences for the usefulness of weather index insurance in practice. For insurers and policymakers, such findings caution against one-size-fits-all designs and highlight the need for tailoring contracts to local agro-climatic conditions. For researchers, this study underscores the value of replication not only to verify results but also to delineate the boundaries of model applicability. Moreover, by demonstrating the importance of transparent and reusable data, the paper contributes to the broader goal of promoting a replication culture through FAIR data practices.

The remainder of the paper is structured as follows: Section 2 describes the methodology and Section 3 the data. Section 4 presents the results of the replication study and the extended model. Section 5 draws conclusions about the superiority of the proposed statistical model and discusses implications for the design of weather index insurance.

2 Methodology

This subsection summarizes the methods, and the approach used in Bucheli et al. (2022). We use the same methodology as the original study for both the direct replication and the extended replication. In their paper, Bucheli et al. (2022) apply cubic spline methods to the insurance payout function that depends on hourly temperature exposure during the critical growth period. They then use historical farm-level yields to calibrate temperature effects on crop yields. Finally, they use the expected utility framework to evaluate the overall benefits of insurance for risk-averse farmers and the actuarially fair premium. Following Sections 2, 3 and 4 in Bucheli et al. (2022) we now describe each of the model components.

2.1 Expected Utility Model

The starting point is the definition of farmers' revenues:

$$W_{it} = p \cdot \mathrm{y}_{it}(I_{it}) + \pi_{it}(\mathrm{I}_{it}) - \Theta_i \tag{1}$$

In Equation 1, W_{it} is a random variable that represents the revenue generated by farm i during year t from the production of a single crop. The variable p stands for the price associated with the insured crop, and it assumed to remain constant (not random) due to the existence of forward contracts. The variable I_{it} denotes random weather shocks, and the variable $\mathrm{y}_{it}(I_{it})$ represents the random yield of a specific crop, which depends on these unpredictable weather shocks. π_{it} signifies the insurance payout that is determined by the random variable I_{it}, and Θ_i represents the insurance premium that the farmer pays. It is important to note that Bucheli et al. (2022) assume an actuarial fair premium, i.e., $\Theta_i = E(\pi_{it})$.

Crop yield is not solely determined by random weather shocks but also by non-weather-related influences:

$$y_{it} = g(I_{it}) + \varepsilon_{it} \tag{2}$$

In Equation 2, $g(I_{it})$ is a function that approximates the impact of random weather shocks on yield y_{it} and the error term ε_{it} summarizes other factors that are uncorrelated to the weather such as pests and diseases, geohazards, management decisions, time trends, farm fixed-effects such as soil properties, etc.

To assess the overall gains of the farmer from taking insurance Bucheli et al. (2022) use the expected utility model:

$$EU_i(W_{it}) = E[U(W_{it})] = U[E(W_{it}) - R_i] \tag{3}$$

In Equation 3, $U(.)$ is a utility function, W_{it} is the revenue realized by farmer i in year t, $E[.]$ is the expectation operator, and R is the risk premium of farmer i. To derive a farmer's risk premium, the authors solve Equation 3 for R_i:

$$R_i = E(W_{it}) - U^{-1}[EU_i(W_{it})] \tag{4}$$

The authors employ power utility $U(.)$ function (Equation 5) to depict how risk-averse farmers have diminishing marginal returns to increasing revenue. α is a coefficient of constant relative risk-aversion, which is a fitting description of farmers' risk preferences (e.g. Falco, Chavas, 2009; Femenia et al., 2010).

$$U(W_{it}) = (1 - \alpha)^{-1} \cdot (W_{it})^{(1-\alpha)} \tag{5}$$

2.2 Cubic Spline Method

Bucheli et al. (2022) use a crop-specific payout function for insurance payouts which depends on hourly air temperature exposure.

$$\pi_{it} = p \cdot \sum_{h_{it}=1}^{H_{it}} \begin{cases} \max(-f(T_{ith}), 0) & \text{if } T_{ith} \geq T_s \\ 0 & \text{if } T_{ith} < T_s \end{cases} \tag{6}$$

π_{it} is the final payout for farmer i in year t which is equal to the sum of all hourly payouts from starting hour $h_{it} = 1$ and ending hour H_{it}. An hourly payout triggers if the temperature T_{ith} is equal to or exceeds the strike temperature level T_s and if the function $f(T_{ith})$ estimates a yield loss with the temperature exposure T_{ith}. The hourly payout is equal to price p multiplied by the estimated yield loss with the function $f(T_{ith})$.

The authors argue that the restricted cubic spline method would be an ideal functional form for $f(T_{ith})$ when estimating the non-linear hourly temperature effect on crop yield. Cubic splines are piece-wise cubic polynomials, joined smoothly at so-called 'knots'. They flexibly approximate non-linear relationships - in our case between hourly temperature and yield - while remaining linear beyond the outer knots. To apply the cubic spline method, we first divide the temperature range into intervals with the knot locations $k_1, k_2, \dots, k_n$ to capture non-linear yield responses between intervals and linear yield response beyond the outer knots. Secondly, as shown in Equation 7, the original time series T_{ith} is transformed into $(k-2)$ new time series, since a cubic spline with k knots uses $(k-2)$ basis functions to capture the non-linearity between inner knots. S_{itj} is the new hourly $(k-2)$ time series based on hourly temperature T_{ith}, j is the number of new time series with $j = 1, \dots, k-2$.

$$\begin{aligned} S_{ithj} = &\left(\max\left(\frac{T_{ith} - kn_j}{(kn_k - kn_1)^{2/3}}, 0\right)\right)^3 - \left(\max\left(\frac{T_{ith} - kn_{k-1}}{(kn_k - kn_1)^{2/3}}, 0\right)\right)^3 \\ &\cdot \frac{kn_k - kn_j}{kn_k - kn_{k-1}} + \left(\max\left(\frac{T_{ith} - kn_k}{(kn_k - kn_1)^{2/3}}, 0\right)\right)^3 \cdot \frac{kn_{k-1} - kn_j}{kn_k - kn_{k-1}} \end{aligned} \tag{7}$$

In the third step, X_{it} is derived as the aggregated hourly values during the critical growth period starting in $h_{it} = 1$ and ending H_{it} at farm i in time t for both the original T_{it} and transformed time series S_{it} (Equation 8)

$$X_{it1} = \sum_{h_{it}=1}^{H_{it}} T_{ith} \text{ and } X_{it(j+1)} = \sum_{h_{it}=1}^{H_{it}} S_{ithj} \text{ for } j = \{1, \dots, k-2\} \tag{8}$$

Fourthly, the authors build a farm fixed-effect model to estimate the effect of aggregated hourly values on yield.

$$y_{it} = \beta_1 X_{it1} + \beta_2 X_{it2} + \cdots + \beta_{k-1} X_{itk-1} + \beta_k t + \beta_{k+1} t^2 + \alpha_i + \varepsilon_{it} \tag{9}$$

In Equation 9, y_{it} is the crop yield of farm i in year t, the restricted cubic spline specification is denoted as $(\beta_1 X_{it1} + \beta_2 X_{it2} + \cdots + \beta_{k-1} X_{itk-1})$ and the deterministic quadratic time trend $(\beta_k t + \beta_{k+1} t^2)$ controls for technological progress in crop yields. α_i is used to control for unobserved and time-invariant factors that influence crop yields at the farm level. The error term ε_{it} summarizes random factors that influence yields but are uncorrelated with temperature and control variables. Finally, Equation 10 is used to calculate marginal yield response $f(T_{ith})$ to hourly temperature exposure T_{ith}.

$$f(T_{ith}) = \widehat{\beta_1} + \widehat{\beta_2} S_{it1} + \widehat{\beta_3} S_{it2} + \cdots + \widehat{\beta_{k-1}} S_{itk-2} \tag{10}$$

The $(k-1)$ coefficients $\beta_1, \dots, \beta_{k-1}$ are estimated regression coefficients from Equation 9, T_{ith} is the temperature in hour h_{ith} of the period of critical crop growth phases at farm i and in year t, and the $(k-2)$ variables are values of temperature transformation j that we derive with Equation 7.

2.3 Model Specification

The authors acknowledge that the impact of temperature on crop yields can vary significantly during different growth stages, and these stages may differ between locations and from year to year. To account for this variability and to mitigate temporal basis risk, the original study makes use of interpolated phenological data. Specifically, the study concentrates on the critical growth stages for winter wheat, which range from stem elongation to the initial milk ripeness phase, and for winter rapeseed, from bud formation to the early full ripeness stage. Given the unavailability of high-resolution hourly temperature data, the authors follow Snyder (1985), Tack et al. (2018), D'Agostino, Schlenker (2016) and Gammans et al. (2017) to approximate hourly temperature changes. They fit double sine curves to daily temperature patterns, extending from the minimum temperature of one day to the maximum temperature of the same day and then to the minimum temperature of the following day.

In terms of the cubic spline method, the authors explore four different models for placing knots in the temperature data. The first model follows a rigorous four-step procedure, as outlined in Harri et al. (2011) and Ker, Tolhurst (2019). This procedure includes the selection of three-knot locations, the establishment of boundaries for the lowest and highest knot positions (based on the 5% and 95% quantiles of observed temperatures, as suggested in Harri et al. (2011) the imposition of a minimum 5°C interval between knots. Subsequently, all potential combinations of knot positions derived from the first model are considered. In the third step, the authors employ a farm fixed-effect model for each conceivable knot combination and calculate the residual sum of squares (RSS) to evaluate model fit. Finally, the model with the lowest RSS is chosen as the best fit. Model 2 introduces three evenly spaced knots into the observed temperature records, taking cues from Ortiz-Bobea et al. (2018). In Model 3, three knots are placed at specific percentiles (10%, 50%, and 90%) of observed temperatures. Model 4, inspired by Blanc, Schlenker (2017), adopts an approach with five equally spaced knots, with the first knot positioned at 5°C (near the 10% quantile of observed temperatures) and the fifth knot at 25°C (close to the 90% quantile of observed temperatures).

2.4 Risk Analysis

To assess the risk-reducing capabilities of the payout function, the authors apply the economic framework outlined in Section 2. They prepare the required yield observations for revenue calculations (as described in Equation 1) by detrending them using a quadratic time trend derived from the panel data.[1] As in the original study, crop prices are standardized to 1 (meaning that revenues, payouts, and premiums are expressed in yield units, specifically deci-tons per hectare, equivalent to 100 kg per hectare), given that German farmers can employ bilateral forward contracts to manage price risks (Anastassiadis et al., 2014).

Subsequently, the authors analyze relative changes in risk premiums across various payout functions and strike-level temperatures, comparing them to the scenario of uninsured crop production. They then assess the statistical significance of these differences using non-parametric paired Wilcoxon signed rank tests. The results presented in the original study refer to moderately risk-averse farmers characterized by a coefficient of constant relative risk-aversion equal to 2.

The authors use out-of-sample risk assessment to avoid overfitting of models. For this purpose, they exclude farm i's data during the initial calibration phase. Instead, they develop general payout functions. Subsequently, the farm i's potential for risk reduction is assessed by incorporating farm-specific temperature data during periods of heightened risk into the previously calibrated payout function. This leave-one-out procedure aims to provide an unbiased, out-of-sample estimate of each farm's potential for risk reduction, and tries to ensure that the effectiveness of the insurance design is robust and not merely a result of overfitting to in-sample data.

3 Data

The general availability of data is of pivotal importance for conducting robust and reliable replication studies. A framework developed to assess "data availability" more systematically is the FAIR data principles approach (Wilkinson et al., 2016). FAIR is an acronym for "findable, accessible, interoperable, and reusable", which are core principles that support replicability and transparency in empirical research. For direct replication, it is essential that the dataset is both findable and accessible. In this study, the original dataset used by Bucheli et al. (2022) was not publicly available or findable. Although we contacted the original data provider, we could only obtain a similar dataset, which we then used for the direct replication.

For the extended replication, we relied on additional yield data from alternative sources. While these data are generally findable - as they are collected by the Ministry of Agriculture of Saxony - they are not publicly accessible due to privacy regulations. These constraints have implications for external validity: because our analysis was confined by available data, we could not fully control for representativeness or alignment with the original study design. Differences in farm structure, geography, or unobserved factors may have influenced our results. Moreover, due to anonymization and incomplete metadata, assumptions during data cleaning were nec-

[1] A specific challenge in estimating crop yield models in the context of yield insurance arises from non-stationarity of the data generating process (e.g. Shen et al., 2018). Agricultural crop yields usually show an upward trend over time and deviations from the trend (residuals) frequently exhibit heteroscedasticity. Major causes of non-stationarity are climate change and technological change. This, in turn, has a considerable effect on average crop yields, yield variability and hence on the cost of insurance. Given the non-stationarity of crop yields, the standard approach to pricing yield insurance is a two-stage estimation procedure. In the first step, the trend component is removed from the data. In the second step, a parametric or non-parametric distribution is fitted to the detrended data (Woodard, Sherrick, 2011; Annan et al., 2014). The dynamics of average yields are captured by either a deterministic or a stochastic trend. Deterministic time trend models are dominant in the literature and consist of a simple linear trend, polynomial trend (Just, Weninger, 1999) and spline functions (Harri et al., 2011).

essary, introducing uncertainty and limiting comparability. These issues underscore the tension between data protection and scientific replicability and highlight why adherence to FAIR data principles is crucial not only for replication but also for credible external validation.

Regarding phenological and temperature data, no differences are observed between the study types; however, adjustments have been made to the yield data in the two replication types. The original study, as well as the direct replication, encompass data on winter wheat and winter rapeseed, while the extended replication utilizes yield data for winter barley, instead of winter wheat, but also rapeseed. The time span covered by the original study is from 1995 to 2018, by the direct replication from 2000 to 2018, and by the extended replication from 2000 to 2015. The number of farms included in the three studies varies (see Table 1).

Table 1. Comparison of data sets

<table>
<tr><th colspan="2">Study type
Characteristics</th><th>Original study</th><th>Direct replication</th><th>Extended replication</th></tr>
<tr><td colspan="2">Products</td><td colspan="2">Winter wheat, Winter rapeseed</td><td>Winter barley, Winter rapeseed</td></tr>
<tr><td rowspan="5">Yield data</td><td>Sources</td><td colspan="2">Gvf Versicherungs-Makler AG</td><td>FADN</td></tr>
<tr><td>Periods</td><td>1995-2018</td><td>2000-2018</td><td>2000-2015</td></tr>
<tr><td>Farms</td><td>84 (Winter wheat)
81 (Winter rapeseed)</td><td>88 (Winter wheat)
86 (Winter rapeseed)</td><td>217 (Winter barley)
223 (Winter rapeseed)</td></tr>
<tr><td>Records</td><td>2,578</td><td>4,033</td><td>6,617</td></tr>
<tr><td>Farm locations</td><td colspan="2">Berlin, Brandenburg, Mecklenburg, Saxony, Saxony-Anhalt, Thuringia</td><td>Saxony</td></tr>
<tr><td colspan="2">Phenology data</td><td colspan="3">PHASE model</td></tr>
<tr><td colspan="2">Temperature data</td><td colspan="3">Publicly available gridded datasets E-OBS version 20e</td></tr>
</table>

Source: own representation

The original study includes 84 farms for winter wheat and 81 for winter rapeseed, while the extended replication includes 440 farms (217 farms producing winter barley and 223 winter rapeseed producers). The original study as well as the direct replication approach include yield data from Berlin, Brandenburg, Mecklenburg, Saxony, Saxony-Anhalt, and Thuringia, while the extended replication procedure focuses on yield data from Saxony. The following section provides further details on the data that was utilized in this study.

3.1 Yield Data (the Original Study, Direct and Extended Replication)

Yield Data in the Original Study

In the paper by Bucheli et al. (2022), the unbalanced panel data for yield was obtained from the German insurance broker "gvf VersicherungsMakler AG". As outlined in Table 2 (on the right-hand side), the total yield for winter wheat is 1,316, with a minimum yield of 29.24, a mean of 82.66, a maximum yield of 124.04, and a standard deviation of 14.26. Similarly, the total yield record for winter rapeseed is 1,262, with a minimum yield of 8.31, a mean of 37.63, the maximum yield recorded is 57.01, and the standard deviation is 9.01. A total of 88 farms were evaluated, with 77 producing both crops and 11 producing only one crop.

Table 2. Summary statistics of yields, risk periods, and hourly temperatures in direct replication data 2000-2018 (on the left) and original data 1995-2018 (on the right)

		Direct replication data				Original data			
Crop	**Variable**	Min	Mean	Max	SD	Min	Mean	Max	SD
Winter wheat	Detrended yield [dt/ha]	20.62	68.22	114.80	14.72	29.24	82.66	124.04	14.26
	Duration risk period [d]	43.00	61.24	85.00	6.12	48.00	65.32	87.00	6.41
	Hourly temperatures [°C]	-5.16	14.44	36.88	6.04	-4.75	14.53	36.38	5.74
Winter rapeseed	Detrended yield [dt/ha]	6.30	35.32	57.00	8.06	8.31	37.63	57.01	9.01
	Duration risk period [d]	66.00	90.76	115.00	7.45	65.00	90.55	117.00	7.85
	Hourly temperatures [°C]	-6.61	14.48	38.60	6.36	-14.15	14.23	38.48	6.10

Note: direct replication data: 88 winter wheat and 86 winter rapeseed producers with yield records between 2000 and 2018. Original data: 84 winter wheat and 81 winter rapeseed producers with yield records between 1995 and 2018. Risk period is the duration of temperature measurements. dt is deci-ton, ha hectare, d day and °C degree Celsius. Hourly temperatures are measured within risk periods.
Source: own calculations and original study by Bucheli et al. (2022)

Yield Data Used in the Direct Replication Approach

For the direct replication, data was obtained from the same source as in the original study. Due to data protection considerations, it was not possible to gain access to the exact same data set as used in the original study. However, the data set for the direct replication study contains data that were also used in the original study. The data set comprised unbalanced data, with over 22,000 observations across the five regions of Germany. The variables collected for inclusion were reference numbers, crop names, coordinates (longitude and latitude), years, and yield records. Yield records are expressed in deci-tons per hectare (dt/ha), while coordinates (longitude, latitude) are based on the WGS 84 coordinate reference system. Figure 1 provides a geographical representation of the farm locations.

The variable representing the farms was absent, and it was unclear, how many farms existed and to which ones the crops or observations belonged to. Furthermore, the anonymization process resulted in duplicated entries with varying coordinates, which impeded the identification of farm observations. To address this issue, a comprehensive analysis of analogous yield records was conducted, and the duplicates were eliminated. To ensure a more balanced panel dataset that aligns with the available phenology data, observations from the years 1991-1999 and 2019-2020 were excluded. Two distinct crops were subsequently combined into single farms based on nearby coordinates and the same duration. Observations were recorded as single crop farms. The details of the data cleaning procedures are documented in the code provided in the replication package (https://doi.org/10.15456/gjae.2025209.2025161178).

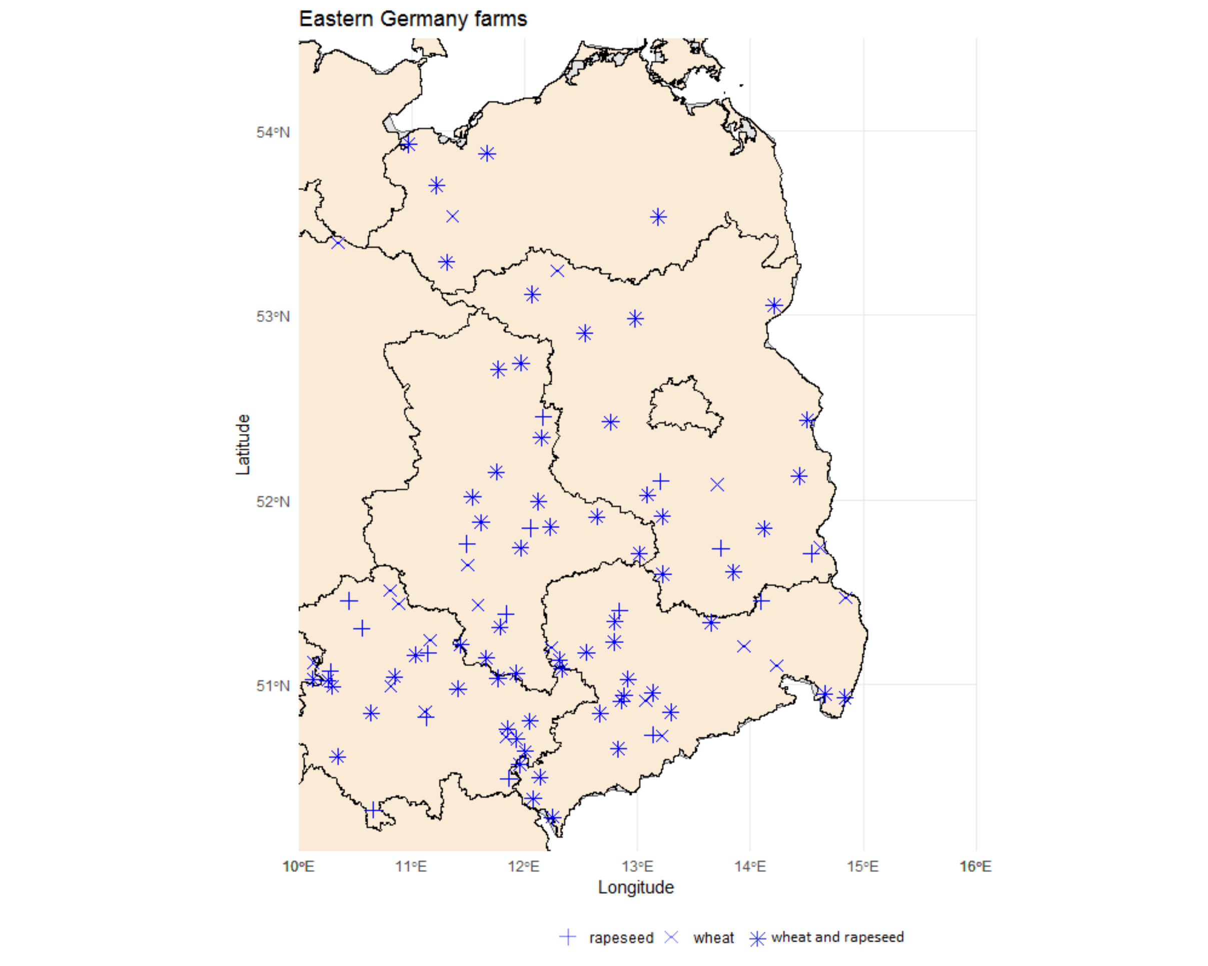

Figure 1. Geographical representation of farms used in the direct replication

Source: own presentation

The descriptive statistics for the direct replication are presented in Table 2 and compared with the original data. A total of 2,705 records were obtained for winter wheat, while 1,328 records were obtained for winter rapeseed. The data set spans the period from 2000 to 2018. Notwithstanding the fact that the data originates from the same source, it is evident that there are considerable discrepancies when it comes to the detrended yield. In general, the data obtained for the direct replication study exhibits a greater degree of variability compared to the original data, as evidenced by the lower minimum values and higher maximum values observed for the detrended yield. The standard deviation for both data sets is approximately equivalent.

Table 3. Summary statistics of yields, risk periods, and hourly temperatures in extended replication data 2000-2015 (on the left) and original data 1995-2018 (on the right).

		Extended replication data				Original data			
Crop	**Variable**	Min	Mean	Max	SD	Min	Mean	Max	SD
Winter Barley	Detrended yield [dt/ha]	1.33	61.79	122.88	17.01				
	Duration risk period [d]	47.00	64.87	83.00	6.72				
	Hourly temperatures [°C]	-5.73	13.99	36.41	5.92				
Winter rapeseed	Detrended yield [dt/ha]	3.01	37.86	89.16	9.98	8.31	37.63	57.01	9.01
	Duration risk period [d]	67.00	90.22	112.00	8.03	65.00	90.55	117.00	7.85
	Hourly temperatures [°C]	-8.17	14.17	37.83	6.29	-14.15	14.23	38.48	6.10

Note: Extended replication data: 217 winter barley and 223 winter rapeseed producers with yield records between 2000 and 2015. Original data: 81 winter rapeseed producers with yield records between 1995 and 2018. Risk period is the duration of temperature measurements. dt is deci-ton, ha hectare, d day and °C degree Celsius. Hourly temperatures are measured within risk periods.
Source: own calculations and original study by Bucheli et al. (2022)

Extended Replication Approach

For our extended study, we used a balanced panel dataset from Saxony covering the years 2000 to 2015. We chose this timeframe because it offers the most consecutive data per year per farm. The dataset, provided by the Farm Accountancy Data Network (FADN) in Saxony, contains 6,617 yield records in total. It consists of various variables, such as reference numbers for municipalities and farms, yield records, years, and crop names. However, the data did not include the coordinates of the farms or municipalities. We supplemented these with the central coordinates of the municipalities, which were provided by the Federal Agency for Cartography and Geodesy.

The description of the extended replication data is also presented in Table 3. In the data we have two crops, winter barley and winter rapeseed grown on 440 farms. In our extended replication we focus mainly on winter barley, which has 3,262 yield records from 217 farms and 158 municipalities. The maximum recorded yield for winter barley was 122.88, with a mean of 61.76 and a standard deviation of 17.03. For winter rapeseed a total of 3,355 yield records from 223 farms and 157 municipalities were analyzed. The maximum recorded yield for winter rapeseed was 89.16, with a mean of 37.86 and a standard deviation of 9.98. The yield data are distributed across Saxony as shown by Figure 2.

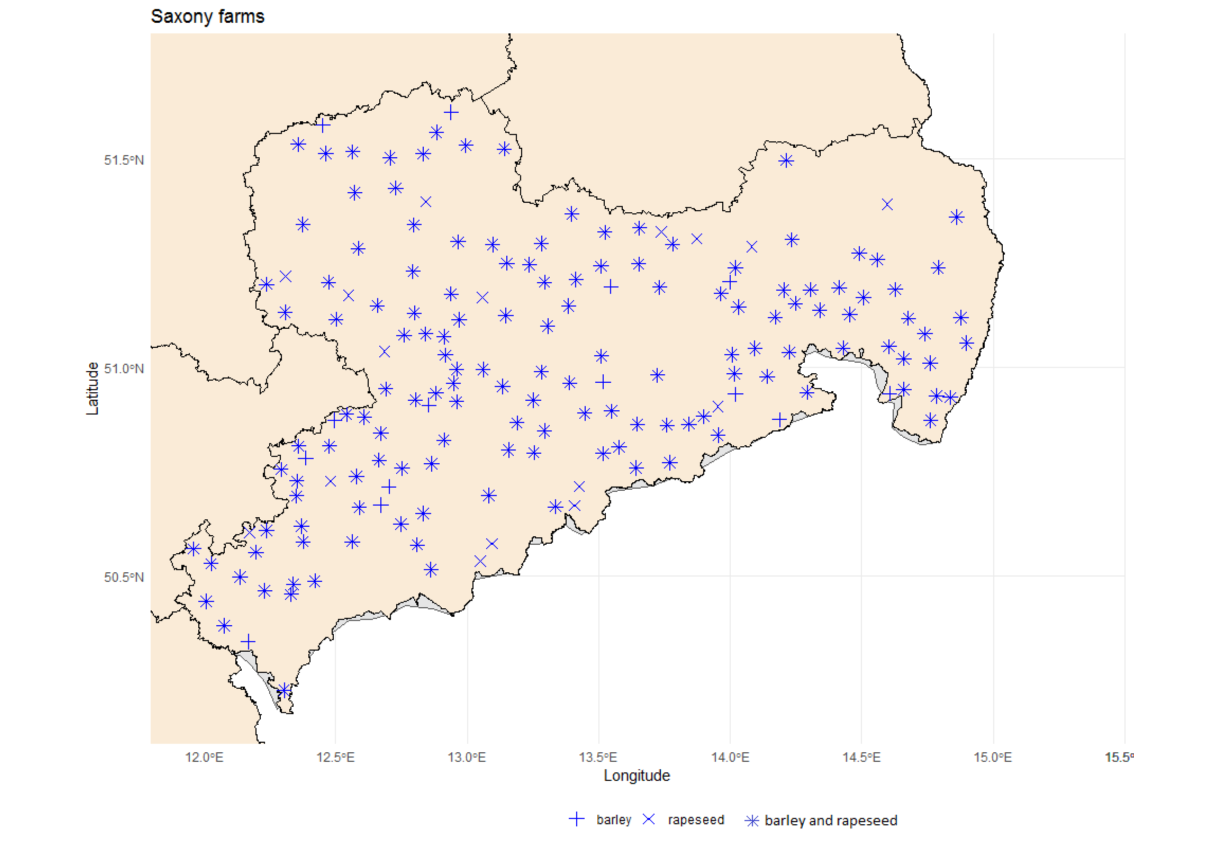

Figure 2. Geographical representation of yield data based on community level in extended replication

Source: own presentation

3.2 Phenology Data

To replicate the original study, we will use phenology data from the original paper, which is derived from the PHASE model (Phenological model for Application in Spatial and Environmental Sciences) (Möller, 2020). The PHASE model uses a growing degree-days approach. Data are generated by combining daily mean temperatures and phenological observations, and then produce1 km × 1 km raster files. These raster files provide detailed information on the timing of growth stages for different crops, farms and years. The phenological data used in the original study were collected by the German Meteorological Service (DWD), which publishes phenological observations collected by a network of volunteer reporters, making them accurate and comprehensive. For the two replication approaches, phenological data were used from the dataset of Möller (2020), which contains interpolated crop-specific and Germany-wide phenological phases for the period between 1993 and 2018. For the extended replication, the critical growth period of winter barley is considered from stem elongation to yellow ripening (Möller et al., 2020).

3.3 Temperature Data

For the replication study, we obtain the daily minimum and maximum temperatures for each farm as used in the original research. These temperatures are derived from publicly available gridded datasets called E-OBS version 27e (latest version) from the EU-FP6 project UERRA and the Copernicus Climate Change Service, and the data providers in the ECA&D project. The datasets have a spatial resolution of 0.1° × 0.1° and are updated monthly. The interpolation process involves the use of a 100-member ensemble incorporating station-derived observations from the European Climate Assessment and Dataset (ECA&D). The station network consists of about 19,100 weather stations across Europe, with a higher concentration in Central Europe, including Eastern Germany.

4 Results

4.1 Direct Replication

Tables 4 and 5 present the outcomes of the four specifications of restricted cubic spline regression model (M1 to M4) for winter wheat and rapeseed, respectively. The results for both winter wheat and winter rapeseed closely align with the original study findings. The coefficients associated with the spline and the deterministic quadratic time trend variables exhibit statistical significance in most models. In the case of winter wheat, the coefficients associated with the spline and the deterministic quadratic time trend variables in Models 1 to 3 exhibit statistical significance, while in Model 4, third and fourth spline variables are not statistically significant. In contrast to winter wheat, we detect statistical significance across all variables in the 4 models for winter rapeseed, except for the second spline variable in Model 4. Therefore, results of almost all variables in the direct replication study confirm the original study finding and provide valuable insights into the weather-yield relationships. The standard deviation of coefficients for both crops in Models 1 to 4 closely follows the pattern established in the original study results. The adjusted R-squared values are slightly higher in the direct replication study due to the larger dataset.

Table 4. Spline regression model results for winter wheat for direct replication and original study

		Direct replication results				Original results			
Models		M1	M2	M3	M4	M1	M2	M3	M4
Spline variable 1	β_1	0.0026*** (0.0002)	0.0024*** (0.0002)	0.0024*** (0.0002)	0.0031*** (0.0003)	0.0018 (0.0002)	0.0020 (0.0001)	0.0020 (0.0002)	0.0025 (0.0004)
Spline variable 2	β_2	-0.0064*** (0.0003)	-0.0068*** (0.0003)	-0.0064*** (0.0003)	-0.0220** (0.0080)	-0.0050 (0.0003)	-0.0070 (0.0004)	-0.0067 (0.0004)	-0.0169 (0.0078)
Spline variable 3	β_3	-	-	-	0.0648 (0.0337)	-	-	-	0.0471 (0.0322)
Spline variable 4	β_4	-	-	-	-0.0974 (0.0587)	-	-	-	-0.7722 (0.0563)
Year	β_5	-91.94* (36.69)	-85.59* (36.93)	-89.84* (36.80)	-87.06* (36.08)	-252.25 (43.54)	-259.31 (43.65)	-262.1 (43.64)	-265.27 (44.32)
Year^2	β_6	0.0231* (0.0091)	0.0215* (0.0092)	0.0225* (0.0092)	0.0218* (0.0090)	0.0629 (0.0108)	0.0650 (0.0109)	0.0654 (0.0109)	0.0662 (0.0110)
Farm Fixed Effect	α_i	yes	yes	yes	yes	yes	yes	yes	yes
Obs		1’309	1’309	1’309	1’309	1’296	1’296	1’296	1’296
Adj R^2		67.24%	67.21%	67.20%	67.29%	58.71%	58.71%	58.65%	58.71%

Note: Model 1 sets three knots to maximize the goodness of fit, Model 2 three knots to divide the temperature range equally, Model 3 sets three knots at certain quantiles and Model 4 has 5 knots with 5°C between knots (see section 2.3 “Model Specification”). The number of spline variables depends on the number of knots (i.e. is equal to number of knots -1). Numbers in parentheses show standard errors. Asterisks display statistical significance: * at the 5% significance level, ** at the 1% level and *** at the 0.1% level.
Source: own calculations and original study by Bucheli et al. (2022)

Table 5. Spline regression model results for winter rapeseed for direct replication and original study

		Direct replication results				Original results			
Models		M1	M2	M3	**M4**	**M1**	**M2**	**M3**	**M4**
Spline variable 1	β_1	0.0010*** (0.0001)	0.0002** (0.0001)	0.0003** (0.0001)	0.0007*** (0.0002)	0.0013 (0.0001)	0.0013 (0.0001)	0.0008 (0.0001)	0.0017 (0.0003)
Spline variable 2	β_2	-0.0027*** (0.0001)	-0.0034*** (0.0001)	-0.0032*** (0.0001)	-0.0024 (0.0032)	-0.0029 (0.0001)	-0.0035 (0.0002)	-0.0032 (0.0002)	-0.0145 (0.0045)
Spline variable 3	β_3	-	-	-	-0.0287* (0.0133)	-	-	-	0.0264 (0.0182)
Spline variable 4	β_4	-	-	-	0.0878*** (0.0222)	-	-	-	0.0009 (0.0307)
Year	β_5	191.83*** (25.05)	240.06*** (24.71)	231.52*** (24.75)	173.93*** (25.80)	118.84 (28.86)	123.38 (29.25)	128.25 (29.52)	109.72 (29.04)
Year^2	β_6	-0.0477*** (0.0062)	-0.0597*** (0.0062)	-0.0608*** (0.0062)	-0.0432*** (0.0064)	-0.0296 (0.0072)	-0.0307 (0.0073)	-0.0319 (0.0074)	-0.0273 (0.0072)
Farm Fixed Ef-fect	α_i	yes	yes	yes	yes	yes	yes	yes	yes
Obs		1’358	1’358	1’358	1’358	1’255	1’255	1’255	1’255
Adj R^2		61.12%	58.07%	58.91%	62.18%	48.53%	47.17%	46.10%	49.23%

Note: Model 1 sets three knots to maximize the goodness of fit, Model 2 three knots to divide the temperature range equally, Model 3 sets three knots at certain quantiles and Model 4 has 5 knots with 5°C between knots (see section 2.3 “Model Specification”). The number of spline variables depends on the number of knots (i.e. is equal to number of knots -1). Numbers in parentheses show standard errors. Asterisks display statistical significance: * at the 5% level, ** at the 1% level and *** at the 0.1% level.
Source: own calculations and original study by Bucheli et al. (2022)

As illustrated in Figures 3 (winter wheat) and 4 (winter rapeseed), non-linear temperature impacts on both crop yields are observed in Models 1 to 4, particularly during critical growth phases determined by phenological stages.

Beyond a certain temperature threshold, there is a discernible curvature in yield responses and associated payouts, which subsequently evolves into a linear relationship after surpassing the final threshold. While the yield responses and associated payouts for winter rapeseed in Models 1 to 4 are virtually indistinguishable from the original study's findings, those for winter wheat exhibit a slightly less steep trajectory (cf. Figures A1 (winter wheat) and A2 (winter rapeseed) in the appendix).

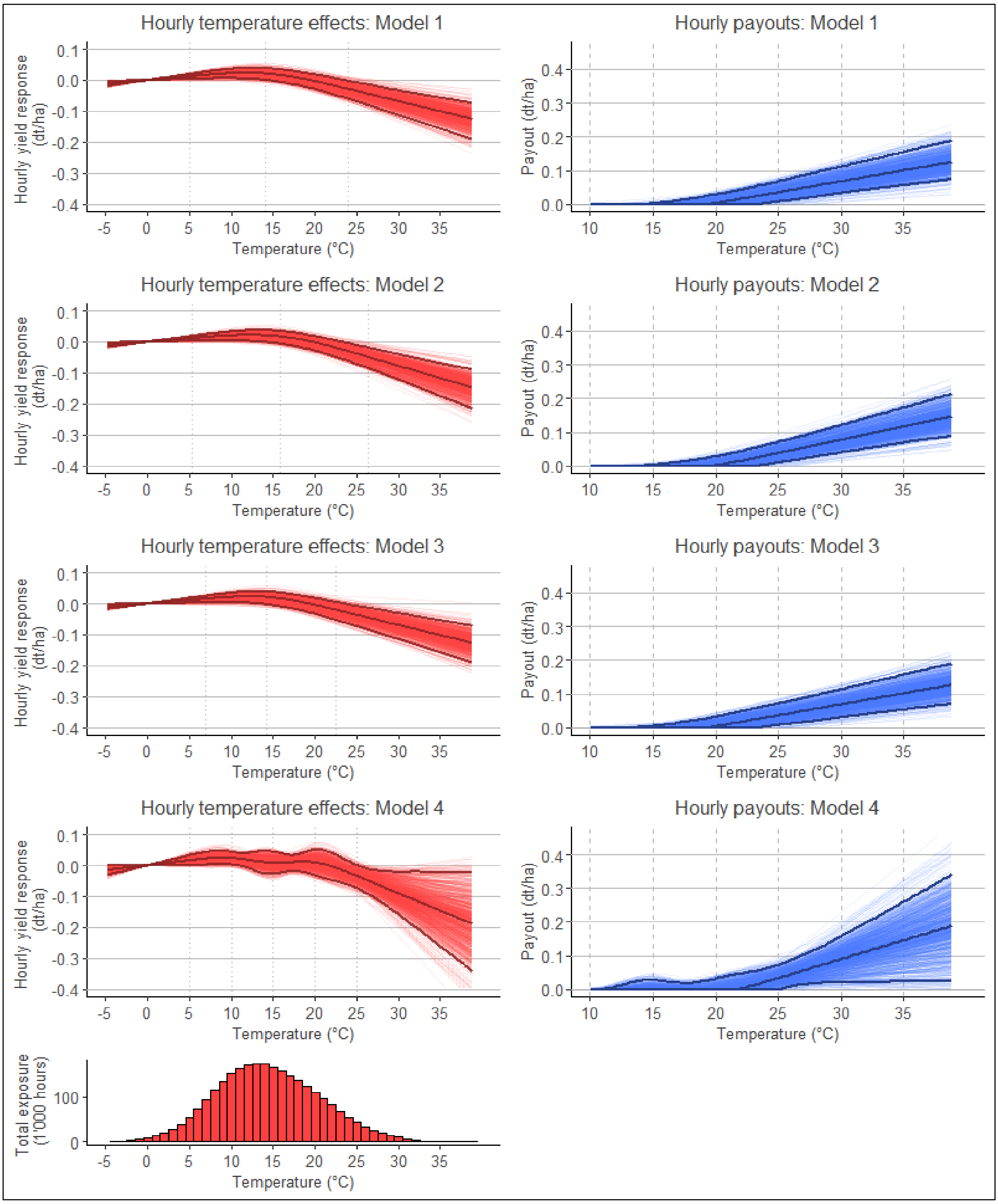

Figure 3. Marginal impact of hourly temperature effects and hourly payouts for winter wheat and different models (direct replication)

Note: different scales and units on axes. Dotted vertical lines in left column show knot locations. 95% confidence bands are derived by 1′000 block bootstraps with observations blocked by year. The different models result from the different knot specifications. Model 1 sets knots to maximize the goodness of fit, Model 2 to divide the temperature range equally, Model 3 sets knots at certain quantiles and Model 4 has 5°C between knots. dt/ha is decitons per hectare.
Source: own calculations with the original code

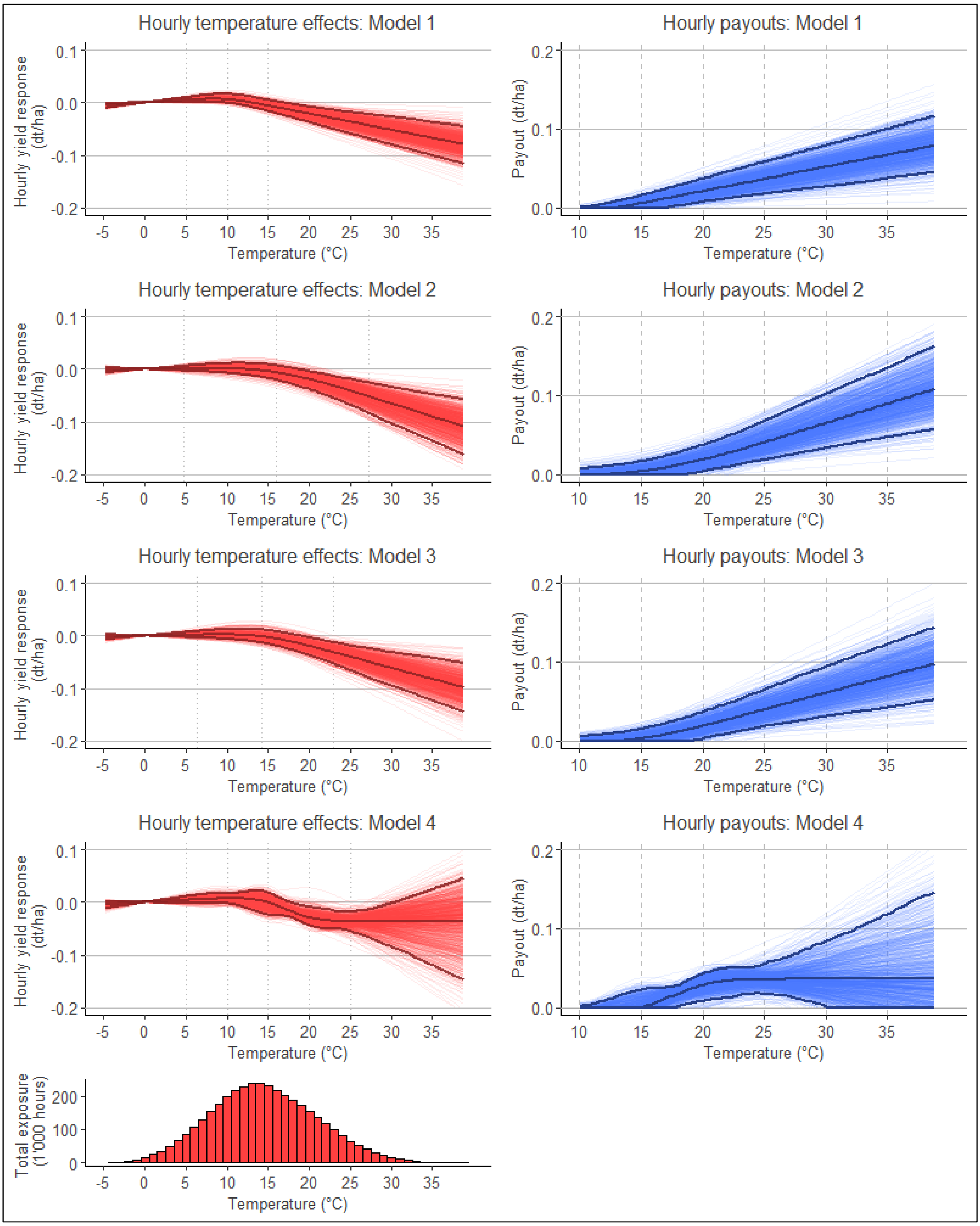

Figure 4. Marginal impact hourly temperature effects and hourly payouts for winter rapeseed and different models (Direct Replication)

Note: different scales and units on axes. Dotted vertical lines in left column show knot locations. 95% confidence bands are derived by 1′000 block bootstraps with observations blocked by year. The different models result from the different knot specifications. Model 1 sets knots to maximize the goodness of fit, Model 2 to divide the temperature range equally, Model 3 sets knots at certain quantiles and Model 4 has 5°C between knots. dt/ha is decitons per hectare.
Source: own calculations with the original code

Furthermore, when we compare the four models for a specific crop, it becomes apparent that different strategies for knot placement result in divergent payouts concerning temperature effects and their associated outcomes. In accordance with the observations made in Bucheli et al. (2022)'s study, we also discern disparities in the effects between winter wheat and winter rapeseed. A noteworthy discovery regarding the temperature effect on winter wheat yield in Model 1 is that the knot placements differ from those in Model 1 of the original study. While the original study's knot placements for winter wheat were set at temperatures of approximately 6°C, 19°C, and 24°C, the direct replication study's Model 1 positions knots at temperatures of approximately 5°C, 13°C, and 24°C. Similar differences are also observed with winter rapeseed in Model 2. Unlike the original findings, where knots are placed at -1°C, 12°C and 25°C in Model 2, direct replication study Model 2 places knots at 5°C, 16°C and 27°C for winter rapeseed. Despite these differences, the estimates closely resemble the original findings, pointing to adverse effects of temperatures exceeding approximately 20°C for winter wheat and 14°C for winter rapeseed, resulting in yield reduction. This finding, however, cannot be generalized as yields depend on other factors, such as precipitation, which are neglected in this study. Moreover, Zou et al. (2024) show that the impact of temperature on yield vary across plant growth stages.

As shown in Figure 5, the median out-of-sample risk reduction capabilities of the temperature-based insurance, expressed as a change in the risk premium with and without insurance, are slightly higher in the direct replication study for both crops compared to the original findings. This finding means that we were not able to replicate exactly the results of the original study. The practical implications for attractiveness of the proposed heat insurance, however, are minor, because the risk reducing capacities remains low and does rarely exceed 20 percent. The risk reduction capabilities for both crops are statistically significant at 0.1% (p-values from one-sided Wilcoxon signed rank tests are Bonferroni-adjusted). In this direct replication study concerning winter wheat at a strike temperature level of 20°C, we witness differences ranging from 21.55% (Model1) to 22.97% (Model 2), while in the original study, the differences spanned from 18.46% (Model 3) to 19.62% (Model 4). Upon raising the strike temperature level to 25°C, the median out-of-sample risk reduction capabilities fluctuated from 15.25% (Model 1) to 17.02% (Model 4) in the direct replication study. In contrast, in the original study, the differences ranged from 13.12% (Model 3) to 14.83% (Model 4). For winter rapeseed at a strike temperature level of 15°C, the median out-of-sample risk reduction capabilities vary between 19.40% (Model 2) and 27.89% (Model 4) in this direct replication study, while in the original study, the differences ranged from 14.21% (Model 3) to 20.66% (Model 4). With a strike temperature level of 20°C, the median out-of-sample risk reduction capabilities span from 16.60% (Model 2) to 26.64% (Model 4) in this direct replication study, whereas in the original study, the differences ranged from 11.15% (Model 3) to 15.93% (Model 4). Overall, in our direct replication the median out-of-sample risk reduction capabilities for both winter wheat and winter rapeseed are slightly higher than the original study findings. Like original findings, the risk reducing capacities for both crops tend to decrease if we increase the strike level, because higher temperatures are tolerated without triggering insurance payoffs. Moreover, when we compare different model results, we see generally similar risk reduction distribution for the same crop and strike level.

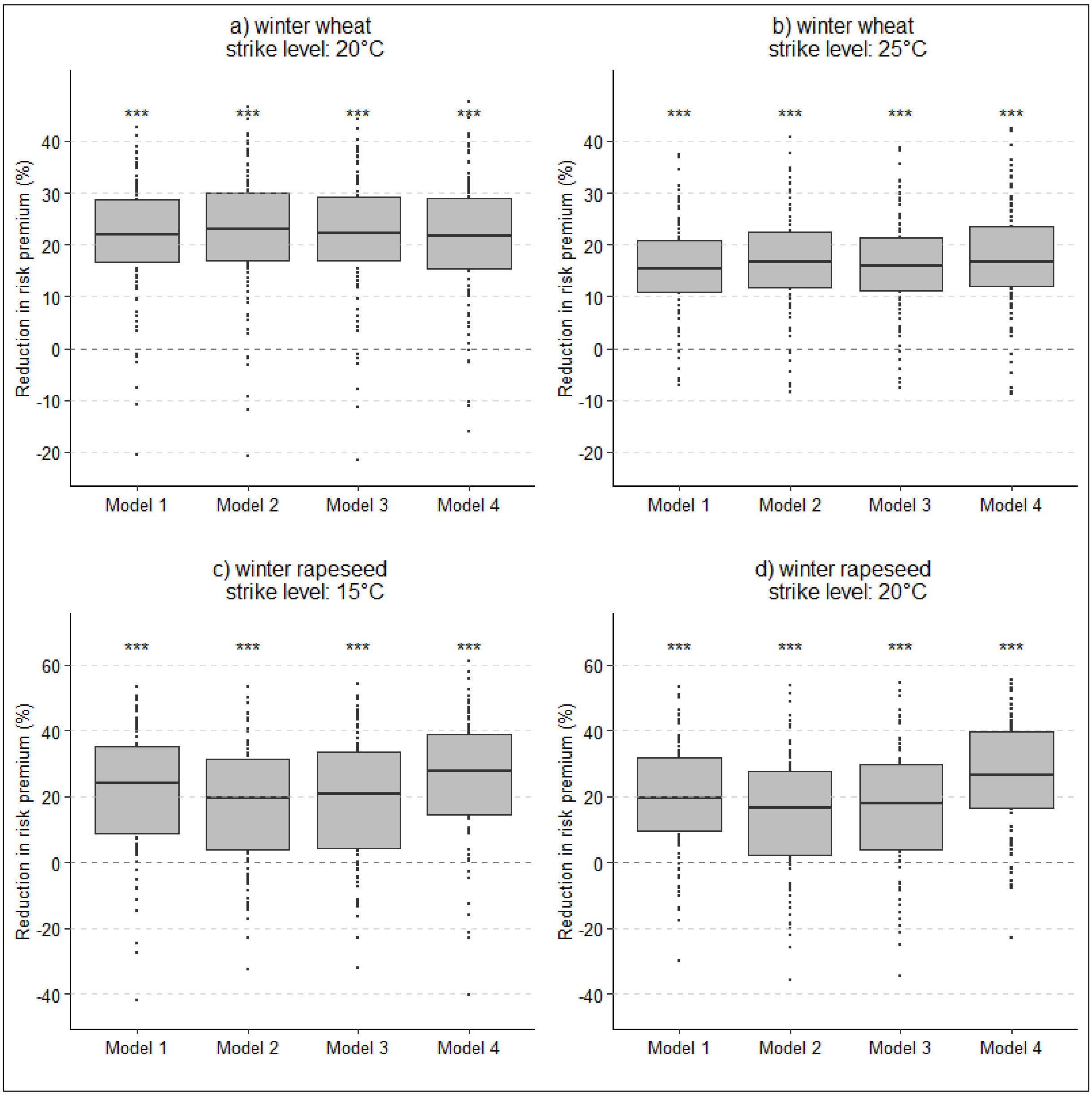

Figure 5. Out-of-sample risk reducing capacities for winter wheat (top row) and winter rapeseed (bottom row), different models and strike level temperatures, assuming moderately risk-averse farmers, and in comparison, to the uninsured status (direct replication)

Note: different scales on y-axes. The different models result from the different knot specifications. Model 1 sets knots to maximize the goodness of fit, Model 2 to divide the temperature range equally, Model 3 sets knots at certain quantiles and Model 4 has 5 °C between knots (see section 2.3 "Model Specification"). Positive values indicate a reduction in the risk premium, which is a financial risk reduction. Boxes show the interquartile range from the 25th percentile to the 75th percentile. Bold lines within boxes mark medians. Points show values beyond the interquartile range. p-values from one-sided Wilcoxon signed rank tests are Bonferroni-adjusted. Asterisks display statistical significance: * at the 5% significance level, ** at the 1% level and *** at the 0.1% level.
Source: own calculations with the original code

4.2 Extended Replication

The extended replication results for winter barley and winter rapeseed relatively deviate from the outcomes of the original study. Tables 6 and 7 present the outcomes of regression models, including Models 1 to 4, for both the extended replication and the original study.

Table 6. Spline regression model results for winter barley for extended replication study

		Extended replication results			
Models		M1	M2	M3	M4
Spline variable 1	$\beta1$	0.0021*** (0.0002)	0.0007*** (0.0002)	0.0007*** (0.0002)	0.0079*** (0.0004)
Spline variable 2	$\beta2$	-0.0050*** (0.0002)	-0.0059*** (0.0002)	-0.0058*** (0.0002)	-0.1226*** (0.0068)
Spline variable 3	$\beta3$	-	-	-	0.4518*** (0.0285)
Spline variable 4	$\beta4$	-	-	-	-0.6976*** (0.0505)
Year	$\beta5$	-723.94*** (37.91)	-695.40*** (37.87)	-695.73*** (37.86)	-841.97*** (38.74)
Year2	$\beta6$	0.1804*** (0.0094)	0.1733*** (0.0094)	0.1734*** (0.0094)	-0.2097*** (0.0097)
Farm Fixed Effect	αi	yes	yes	yes	yes
Obs		3’440	3’440	3’440	3’440
Adj R^2		55.10%	53.72%	53.80%	57.72%

Note: Model 1 sets three knots to maximize the goodness of fit, Model 2 three knots to divide the temperature range equally, Model 3 sets three knots at certain quantiles and Model 4 has 5 knots with 5°C between knots (see Section 2.3 “Model Specification”). The number of spline variables depends on the number of knots (i.e. is equal to number of knots -1). Numbers in parentheses show standard errors. Asterisks display statistical significance: * at the 5% significance level, ** at the 1% level and *** at the 0.1% level.
Source: own calculations

In the extended replication, all coefficients associated with the spline variables and the deterministic quadratic time trends in Models 1 to 4 are statistically significant ($p < 0.01$ Wilcoxon test) for both, winter barley and winter rapeseed, and their standard deviations closely align with the original study's findings. However, the adjusted R-squared values for these models are lower than those in the original analysis. This difference may be partly attributed to the fact that the extended dataset is confined to Saxony, whereas the direct replication covered the entire eastern Germany region. Another contributing factor is the use of a different data source in the extended replication. For example, the descriptive statistics for rapeseed reveal that the minimum yield values in the extended dataset are lower than those reported in the original study (see Table 3).

Table 7. Spline regression model results for winter rapeseed for extended replication and original study

		Extended replication results				Original results			
Models		M1	M2	M3	M4	M1	M2	M3	M4
Spline variable 1	β_1	0.0021*** (0.0002)	0.0007*** (0.0002)	0.0007*** (0.0002)	0.0079*** (0.0004)	0.0013 (0.0001)	0.0013 (0.0001)	0.0008 (0.0001)	0.0017 (0.0003)
Spline variable 2	β_2	-0.0050*** (0.0002)	-0.0059*** (0.0002)	-0.0058*** (0.0002)	-0.1226*** (0.0068)	-0.0029 (0.0001)	-0.0035 (0.0002)	-0.0032 (0.0002)	-0.0145 (0.0045)
Spline variable 3	β_3	-	-	-	0.4518*** (0.0285)	-	-	-	0.0264 (0.0182)
Spline variable 4	β_4	-	-	-	-0.6976*** (0.0505)	-	-	-	0.0009 (0.0307)
Year	β_5	-723.94*** (37.91)	-695.40*** (37.87)	-695.73*** (37.86)	-841.97*** (38.74)	118.84 (28.86)	123.38 (29.25)	128.25 (29.52)	109.72 (29.04)
Year^2	β_6	0.1804*** (0.0094)	0.1733*** (0.0094)	0.1734*** (0.0094)	-0.2097*** (0.0097)	-0.0296 (0.0072)	-0.0307 (0.0073)	-0.0319 (0.0074)	-0.0273 (0.0072)
Farm Fixed Effect	α_i	yes	yes	yes	yes	yes	yes	yes	yes
Obs		3’520	3’520	3’520	3’520	1’255	1’255	1’255	1’255
Adj R^2		37.89%	36.36%	36.39%	39.13%	48.53%	47.17%	46.10%	49.23%

Note: Model 1 sets three knots to maximize the goodness of fit, Model 2 three knots to divide the temperature range equally, Model 3 sets three knots at certain quantiles and Model 4 has 5 knots with 5°C between knots (see section 2.3 “Model Specification”). The number of spline variables depends on the number of knots (i.e. is equal to number of knots -1). Numbers in parentheses show standard errors. Asterisks display statistical significance: * at the 5% significance level, ** at the 1% level and *** at the 0.1% level.
Source: own calculations and original study by Bucheli et al. (2022)

Non-linear temperature effects on crop yields are robustly captured across Models 1 to 4, particularly during the critical growth stages defined by phenological phases (see Figures 6 and 7). In the extended replication – where winter barley substitutes for winter wheat as in Bucheli et al. (2022) – the yield responses and payout functions for winter rapeseed closely mirror the original findings. By contrast, winter barley displays distinct behavior: in Model 1, its yield–temperature and payout curves are noticeably steeper, indicating a sharper response to temperature changes, while Models 2 through 4 reveal a somewhat flatter trajectory compared to the winter wheat results of the original study (refer to Figures A1 for winter wheat and A2 for winter rapeseed in the appendix).

A meticulous comparison across models demonstrates that the strategy for knot placement considerably influences the estimated marginal yield effects and the corresponding payout functions. For example, in Model 1, Bucheli et al. (2022) positioned knots for winter wheat at approximately 6°C, 19°C, and 24°C, whereas the extended replication for winter barley sets the knots at about 5°C, 10°C, and 15°C. In Model 2, the original study placed knots for winter wheat at -1°C, 12°C, and 25°C, but for winter rapeseed in the extended replication, the knots are located at approximately 4°C, 15°C, and 27°C. Histograms of temperature exposure further reveal that both winter barley and winter rapeseed experience relatively higher frequencies of exposure at specific temperature ranges compared to the original dataset.

Despite these differences, a critical threshold emerges for both winter barley and winter rapeseed, adverse yield effects become statistically significant when temperatures exceed roughly 15°C. This finding underscores that the optimal knot placements must be finely tuned to the specific physiological responses of each crop. Overall, while spline-based models reliably capture the non-linear dynamics between temperature and yield, the extended replication shows that replacing winter wheat with winter barley and adding winter rapeseed reveals important differences in how sensitive each crop is to temperature extremes. These nuances highlight that the effectiveness of weather index insurance is not uniform across crops and that modeling efforts should incorporate crop-specific response functions. Accordingly, knot placements should be adjusted based on empirical temperature distributions to better reflect the heterogeneity in yield sensitivity and risk exposure among crops.

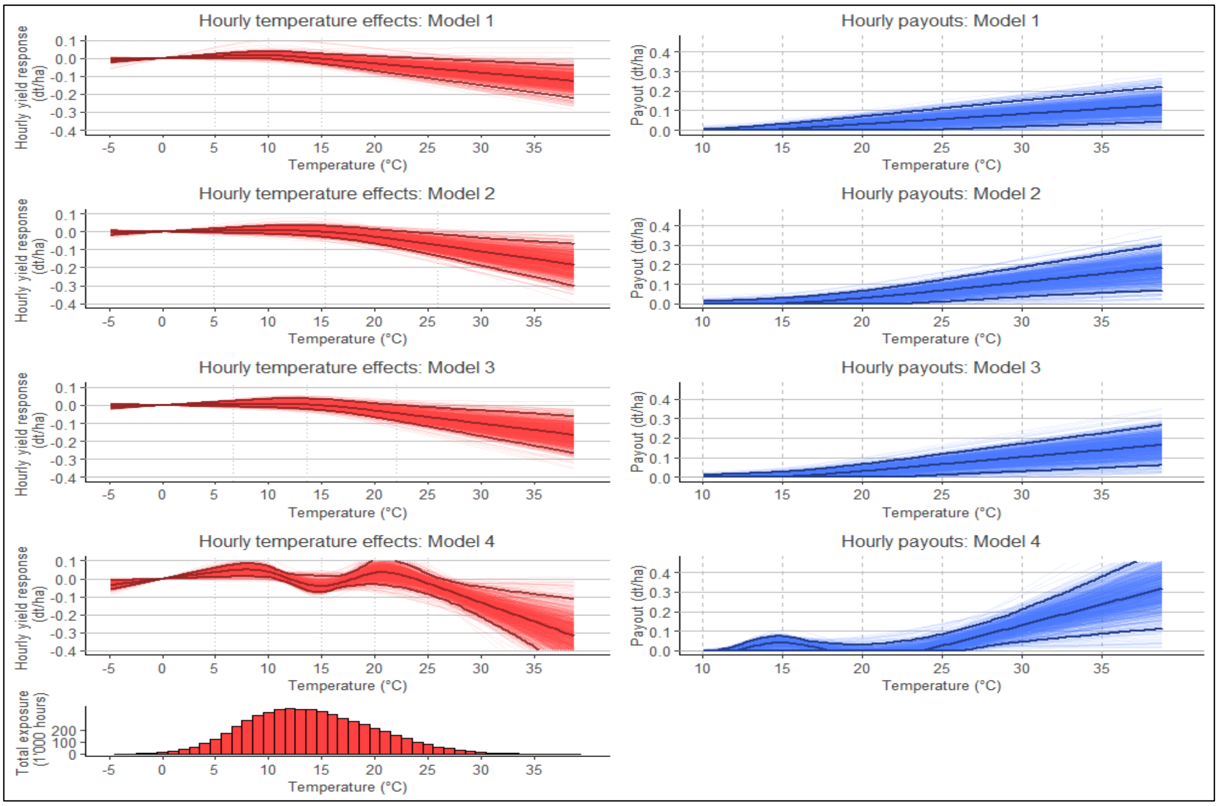

Figure 6. Marginal impact hourly temperature effects and hourly payouts for winter barley and different models (extended replication)

Note: different scales and units on axes. Dotted vertical lines in left column show knot locations. 95% confidence bands are derived by 1′000 block bootstraps with observations blocked by year. The different models result from the different knot specifications. Model 1 sets knots to maximize the goodness of fit, Model 2 to divide the temperature range equally, Model 3 sets knots at certain quantiles and Model 4 has 5°C between knots. dt/ha is decitons per hectare.
Source: own calculations with the original code

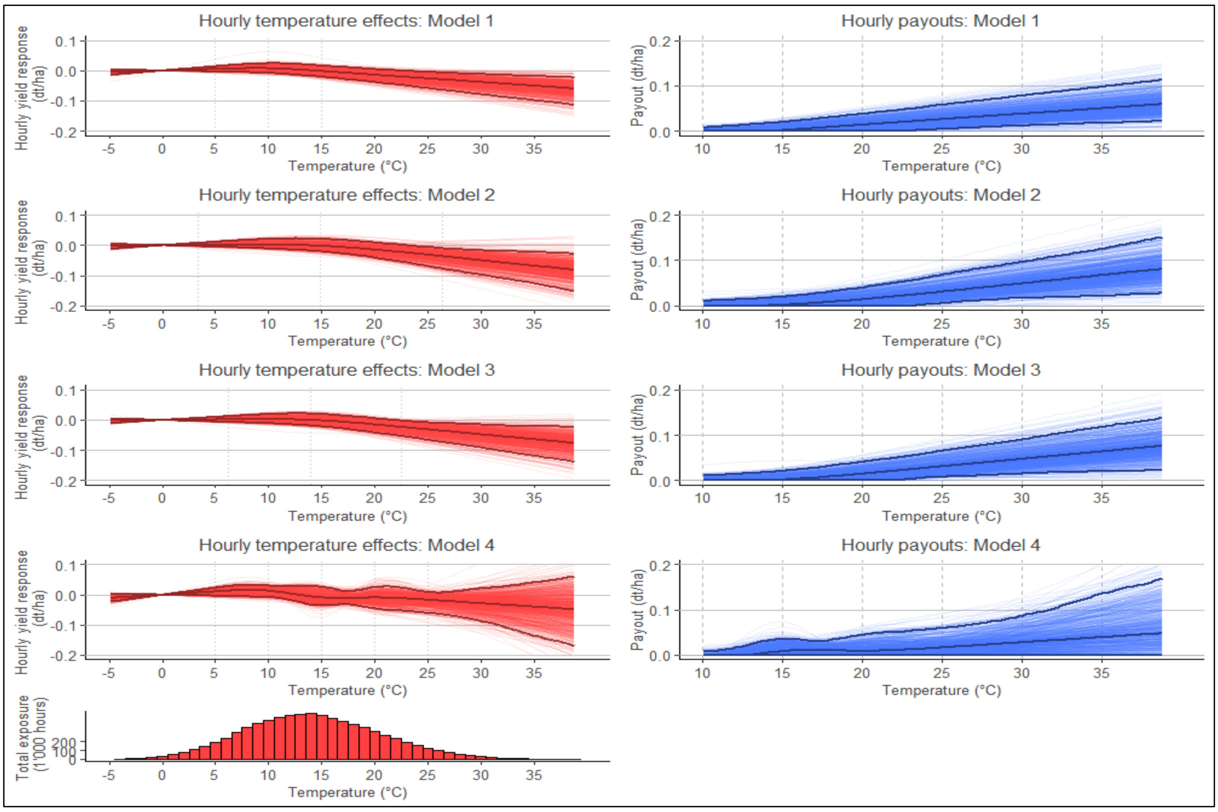

Figure 7. Marginal impact hourly temperature effects and hourly payouts for winter rapeseed and different models (extended replication)

Note: different scales and units on axes. Dotted vertical lines in left column show knot locations. 95% confidence bands are derived by 1'000 block bootstraps with observations blocked by year. The different models result from the different knot specifications. Model 1 sets knots to maximize the goodness of fit, Model 2 to divide the temperature range equally, Model 3 sets knots at certain quantiles and Model 4 has 5°C between knots. *dt/ha* is decitons per hectare.
Source: own calculations with the original code

Figure 8 displays the risk reduction effect of the temperature-based insurance. Compared to the direct replication, the extended replication results for both winter barley and winter rapeseed exhibit lower median out-of-sample risk reduction capabilities (see Figure A3 in the appendix). For winter barley at a strike temperature level of 20°C, we observe reductions in risk premia ranging from 10.46% (Model 4) to 17.37% (Model 3), while in the original study, the differences spanned from 18.46% (Model 3) to 19.62% (Model 4). When raising the strike temperature level to 25°C, the median out-of-sample risk reduction capabilities fluctuated from 9.10% (Model 1) to 10.29% (Model 4). In the original study, the differences ranged from 13.12% (Model 3) to 14.83% (Model 4). For winter rapeseed at a strike temperature level of 15°C, the medians out-of-sample risk reduction vary between 6.32% (Model 4) and 10.33% (Model 1) in the extended replication study, while in the original study, the differences ranged from 14.21% (Model 3) to 20.66% (Model 4). As before, the risk mitigation declines for both crops if the strike level is raised, as higher temperatures are tolerated without triggering insurance payoffs. For example, with a temperature strike level of 20°C, the median out-of-sample risk reduction capabilities varies merely between 5.80% (Model 4) to 8.90% (Model 1).

When comparing outcomes across the various models, the distribution of risk reduction remains largely consistent for a given crop and strike level. Notably, the extended replication's potential for reducing risk is on par with – or even slightly lower than – findings reported in previous studies. For example, Sun (2022) applied an optimization-based weather yield model to mid-season rice in Anhui province, China, achieving reductions in "false positive" and "false negative" basis risk by 14.59% and 17.61%, respectively, relative to traditional models. Similarly, Zou et al. (2023) reported a 41.90% decrease in mean root square loss using a penalized

B-spline approach, while Leppert et al. (2021) demonstrated that spline-based methods lowered relative risk premiums by approximately 27–29% for corn producers in Illinois and Iowa. In addition, Chen et al. (2023) introduced a neural network-based index insurance model that improved farmers' utility by 14.35% through its ability to capture non-linear relationships akin to those identified by spline-based techniques. These comparisons indicate that while the extended replication confirms the usefulness of advanced modeling techniques – such as cubic splines, B-splines, and neural networks – in capturing the non-linear relationship between weather and yield, the absolute improvements in risk reduction are modest. This modest performance suggests that even when employing sophisticated statistical tools and disaggregated weather data, substantial basis risk remains. In practice, the benefits of incorporating these advanced models must be weighed against the inherent limitations of using a single weather variable, such as temperature, to capture the complexity of crop yield determinants. Moreover, the consistency in risk reduction across models for the same crop and strike level implies that the choice of knot placement or the specific modeling approach may be less critical than the overall framework used to quantify risk reduction. This observation underscores the importance of tailoring crop insurance products not only based on the modeling technique but also by incorporating additional weather variables or site-specific factors that may enhance the predictive power of the model.

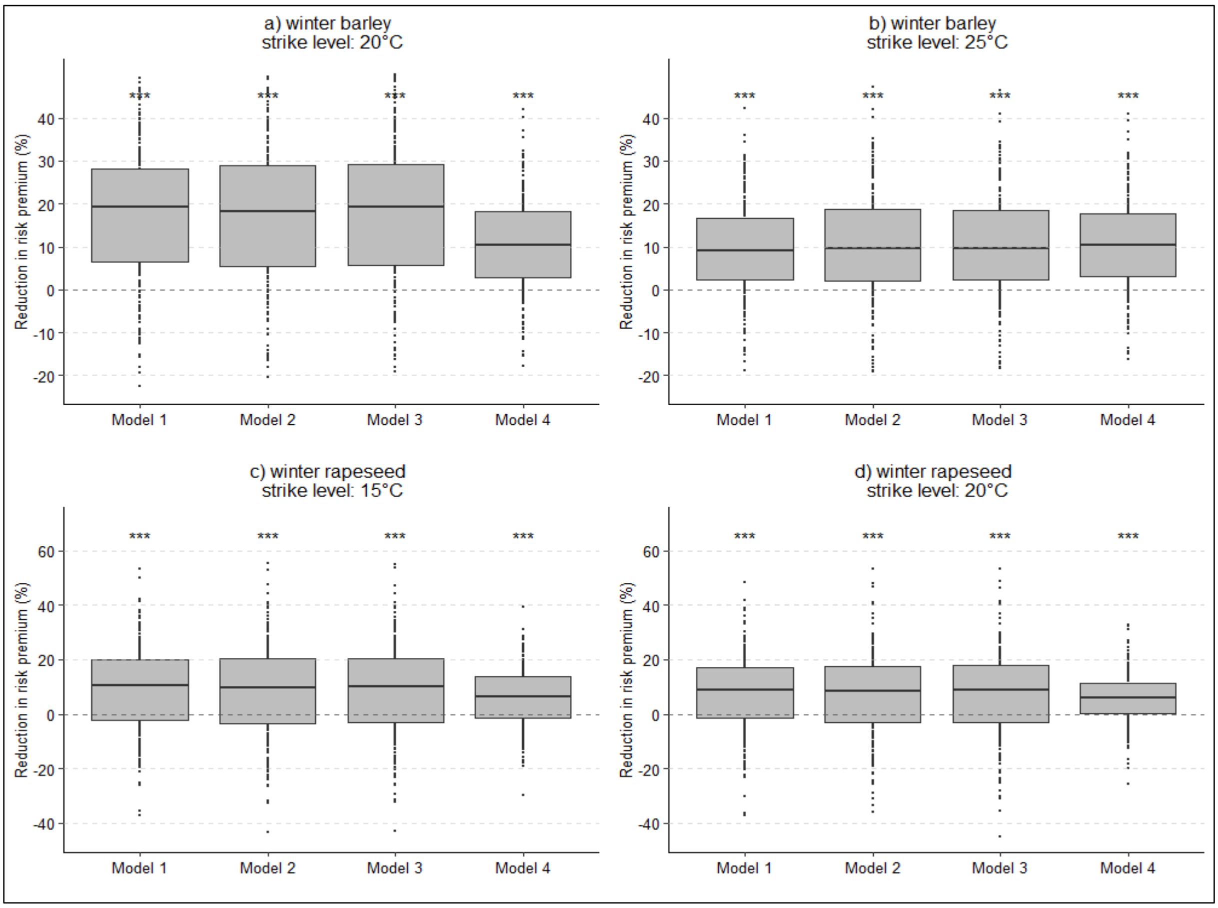

Figure 8. Out-of-sample risk reducing capacities for winter barley (top row) and winter rapeseed (bottom row), different models and strike level temperatures, assuming moderately risk-averse farmers, and in comparison, to the uninsured status (extended replication)

Note: different scales on y-axes. The different models result from the different knot specifications. Model 1 sets knots to maximize the goodness of fit, Model 2 to divide the temperature range equally, Model 3 sets knots at certain quantiles and Model 4 has 5°C between knots (see section 2.3 "Model Specification"). Positive values indicate a reduction in the risk premium, which is a financial risk reduction. Boxes show the interquartile range from the 25th percentile to the 75th percentile. Bold lines within boxes mark medians. Points show values beyond the interquartile range. p-values from one-sided Wilcoxon signed rank tests are Bonferroni-adjusted. Asterisks show significance level: * at the 5% significance level, ** at the 1% level and *** at the 0.1% level.
Source: own calculations with the original code

5 Conclusions

The purpose of this study was to replicate the results of Bucheli et al. (2022) on the impact of temperature on crop yields and the risk reducing capacity of temperature-based index insurance as well as to investigate the generalizability of their findings to other farms, regions and crops. The direct replication at farm level successfully reproduced the key findings of the original study, achieving statistical significance for similar temperature-yield effect sizes. The results of both the direct and extended replication confirm the non-linear relationship between temperature and crop yields during critical growth stages - for winter wheat and winter rapeseed in the direct replication, and for winter barley and winter rapeseed in the extended replication. Differences between cubic spline models and crops are also observed. Only small differences are noticed in the slope of the weather yield relationship which can be associated with limitations in the access to the original yield data and going through data cleaning procedure. The extended replication did not achieve the same level of risk-reducing capabilities as the original findings. Most important, the extended replication exhibits lower median out-of-sample risk-reduction capabilities. Consistent with the original study, we also observe that the risk mitigation effects of the insurance model tend to decline when the strike level is raised, as this reduces the likelihood of payout by excluding moderately severe temperature exposures. Our finding shows that the risk reducing capacities of the heat insurance is limited if the model is applied to other crops and other regions may appear self-evident or even trivial. However, one should recall that while the general model and the payout function remain the same, the model parameters are adjusted to the specific region, time frame and crops. Thus, heterogeneity of climate conditions and crop specific vulnerability against heat stress is taken into account.

The findings of our study have implications a) for researchers in choosing an appropriate method for analyzing weather-index based insurance b) for insurers and policymakers when designing insurance contracts and assessing the willingness to pay of farmers for these products and c) for the scientific community with regard to the availability and provision of FAIR data. In line with previous work our study confirms that spline techniques, including cubic splines, B-spline or penalized B-spline (P-spline) smoothers, can be successfully employed for modeling the non-linear relationship between weather and yield with good quality (cf. Bucheli et al., 2022; Tan, Zhang, 2023; Zou et al., 2023). These methods constitute flexible alternatives to commonly used regression models or neural networks that recently became popular for modeling weather–yield relations. It has also been demonstrated that this statistical approach can accommodate weather data on a highly disaggregated level, here hourly temperatures. The question, however, whether such fine-gridded weather information is useful for the design of index insurance, remains unanswered. The gain in model fit, compared with daily temperature data seems to be modest. The reason might be the aggregation of hourly temperatures into a temperature sum, that eventually enters the farm fixed-effect model and smoothes the variability of hourly data.

The variations in risk reduction capabilities between different crops suggest that policymakers and insurance providers should consider tailoring crop insurance policies to account for crop-specific risk profiles. Nevertheless, it is questionable if a temperature-based insurance will be considered as an attractive risk management tool by farmers, given that the (adjusted) R^2 of the weather-yield regressions rarely exceeds 60 % and in some cases is even below 40%. In other words, even when using sophisticated statistical tools and disaggregated weather data, considerable basis risk remains with farmers. The modest fit of the yield model translates into rather low risk reduction capacities of the index insurance, which is in most cases below 20% when applying the model to other farms and crop types than the original study. This finding is perhaps not too surprising, since temperature is the only weather variable that is included in the yield model and the insurance contract, while other researchers identified rainfall or soil moisture as further important yield determinants (e.g. Mußhoff et al., 2011). The attractiveness of the proposed insurance contract will be further reduced for farmers by considerable premium loadings that insurance providers charge on top of the fair premium in practice (Odening et al., 2022).

The inability to replicate the exact same results as Bucheli et al. (2022) can be traced back to two reasons. First, the lack of clear and detailed data cleaning and handling process required assumptions during data preparation that introduced uncertainty into the results. Second, due to data confidentiality, we were not able to obtain farm identifiers necessary to select the same data as in the original study. Unfortunately, data protection is often in conflict with FAIR principles and hampers replication studies in agricultural economics, particularly when individual-level farm or consumer data is involved. These challenges highlight several limitations of our study: the reliance on non-identical datasets for replication, the need to make assumptions during data cleaning due to insufficient documentation, the use of temperature as the sole weather variable in the insurance model, and - in the extended replication - the coarser spatial resolution of the FADN data, which is only available at the municipality level, unlike the original or directly replicated farm-level data. These factors may have influenced the observed results and should be considered when interpreting our findings. Importantly, the limited generalizability observed in the extended replication should not be viewed as a weakness, but as a valuable insight for both science and practice. It shows that even when sophisticated models are applied carefully, their effectiveness may diminish in new contexts. For practitioners, this finding highlights the importance of tailoring index insurance contracts to local agro-climatic conditions. For researchers, it reinforces the value of replication as a tool not only to verify findings but to identify the boundaries of model applicability and foster transparent, reproducible research. In view of the above findings, we recommend that authors share detailed data cleaning steps, document assumptions explicitly, and - where possible - use anonymized but traceable identifiers to support reproducibility without compromising privacy.

Our paper is a first step in scrutinizing the external validity of weather yield models and several directions for future research are conceivable. First, in view of the limited risk reduction capacity of heat insurance and the fact that drought has been identified as major cause of yield losses, we suggest to compare the heat insurance index with a rainfall index insurance. The second idea is to conduct meta-analyses about weather-yield models following the approach in Challinor (2014) who conduct meta-analyses on crop yields under climate change. Given the extant literature on this topic and the heterogeneity of data and models, this would be a challenging but promising task.

Data Availability Statement

The study used the R programming language as its primary software platform, building on the publicly available codebase of the original study written in R. The original codebase, with its thoughtful comments, remained unchanged, and only the data preparation part was adapted for each of the replication approaches. In addition, the R programming language was used in the study for data cleaning, processing and plotting. It should be noted that access to the data is restricted in certain aspects, particularly regarding yield data.

a) Data Sources

- **Yield Data**
  - Direct replication: Source: Confidential data collected by the insurance company "gvf VersicherungsMakler AG". Availability: The yield data used in this study is not publicly available. Access to the data may be restricted, and interested parties should contact "gvf VersicherungsMakler AG" for further information.
  - Extended replication: Source: confidential data from the Financial Accountancy Data Network (FADN) of the Free State of Saxony. Availability: the data is not publicly available. If required, we can provide access to the data for replication purposes in a protected and confidential environment.

- **Phenology Data**
  Möller, M. (2020): Data for: PhenoWin - A R Shiny application for visualization and extraction of phenological windows in Germany. Mendeley Data, V1.
  https://doi.org/10.17632/37jxk3n9fy.1
- **Temperature Data**
  Cornes, R., van der Schrier, G., van den Besselaar, E.J.M., Jones, P.D. (2018): An Ensemble Version of the E-OBS Temperature and Precipitation Datasets. Journal of Geophysical Research: Atmospheres, 123: 9391-9409.
  https://surfobs.climate.copernicus.eu/dataaccess/access_eobs.php

b) Software and Code

The study used the R programming language as its primary software platform, building on the publicly available codebase of the original study written in R. The original codebase, with its thoughtful comments, remained unchanged, and only the data preparation part was adapted for the two replication approaches. In addition, the R programming language was used in the study for data cleaning, processing and plotting.

- **Original Study Code**
  https://github.com/AECP-ETHZ/Temperature-effects-on-crop-yields-in-heat-index-insurance
- **Replication (Direct and Extended) Study Codes**
  https://doi.org/10.15456/gjae.2025209.2025161178

Acknowledgement

We thank Sebastian Mahler ("gvf VersicherungsMakler AG") for providing farm level yield data. Financial support from the DFG through the FAIRagro project number 501899475 is gratefully acknowledged.

Competing Interests

The authors declare that they have no competing interests.

References

Albers, H., Gornott, C., Hüttel, S. (2017): How do inputs and weather drive wheat yield volatility? The example of Germany. Food Policy 70: 50-61. https://doi.org/10.1016/j.foodpol.2017.05.001.

Anastassiadis, F., Feil, J.-H., Musshoff, O., Schilling, P. (2014): Analysing Farmers' Use of Price Hedging Instruments: An Experimental Approach. Journal of Agricultural & Food Industrial Organization 12(1): 181-192. https://doi.org/10.1515/jafio-2014-0007.

Annan, F., Tack, J., Harri, A., Coble, K. (2013): Spatial pattern of yield distributions: implications for crop insurance. American Journal of Agricultural Economics 96 (1): 253-268. https://doi.org/10.1093/ajae/aat085.

Blanc, E., Schlenker, W. (2017): The use of panel models in assessments of climate impacts on agriculture. Review of Environmental Economics and Policy 17(2): 258-279. https://doi.org/10.1093/reep/rex016.

Bras, T.A., Seixas, J., Carvalhais, N., Jagermeyr, J. (2021): Severity of drought and heatwave crop losses tripled over the last five decades in Europe. Environmental Research Letters 16(6): 6. https://doi.org/10.1088/1748-9326/abf004.

Bucheli, J., Dalhaus, T., Finger, R. (2022): Temperature effects on crop yields in heat index insurance. Food Policy 107: 102214. https://doi.org/10.1016/j.foodpol.2021.102214.

Cao, X., Okhrin, O., Odening, M., Ritter, M. (2015): Modelling spatio-temporal variability of temperature. Computational Statistics 30: 745-766. https://doi.org/10.1007/s00180-015-0561-2.

Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R., Chhetri, N. (2014): A meta-analysis of crop yield under climate change and adaptation. Nature Climate Change 4 (4): 287-291. https://doi.org/10.1038/nclimate2153.

Chen, Z., Lu, Y., Zhang, J., Zhu, W. (2024): Managing Weather Risk with a Neural Network-Based Index Insurance. Management Science 70 (7): 4306-4327. https://doi.org/10.1287/mnsc.2023.4902.

Cornes, R., G. van der Schrier, E.J.M., van den Besselaar, Jones, P.D. (2018): An Ensemble Version of the E-OBS Temperature and Precipitation Datasets. Journal of Geophysical Research: Atmospheres 123: 9391-9409. https://doi.org/10.1029/2017JD028200.

D'Agostino, A. L., Schlenker, W. (2016): Recent weather fluctuations and agricultural yields: implications for climate change. Agricultural Economics 47: 159-171. https://doi.org/10.1111/agec.12315.

Doherty, E., Mellett, S., Norton, D., McDermott, T. K.J., Hora, D.O., Ryan, M. (2021): A discrete choice experiment exploring farmer preferences for insurance against extreme weather events. Journal of Environmental Management 290 (112607). https://doi.org/10.1016/j.jenvman.2021.112607.

Falco, S.D., Chavas, J.P. (2008): On Crop Biodiversity, Risk Exposure, and Food Security in the Highlands of Ethiopia. American Journal of Agricultural Economics 91 (3): 599-611. https://doi.org/10.1111/j.1467-8276.2009.01265.x.

Femenia, F., Gohin, A., Carpentier, A. (2010): The Decoupling of Farm Programs: Revisiting the Wealth Effect. American Journal of Economics 92 (3): 836-848. https://doi.org/10.1093/ajae/aap029.

Gammans, M., Merel, P., Bobea, A.O. (2017): Negative impacts of climate change on cereal yields: Statistical evidence from France. Environmental Research Letters 12. https://doi.org/10.1088/1748-9326/aa6b0c.

Harri, A., Coble, K. H., Ker, A.P., Goodwin, B. J. (2011): Relaxing heteroscedasticity assumptions in area-yield crop insurance rating. American Journal of Agricultural Economics 93 (3): 703-713. https://doi.org/10.1093/ajae/aar009.

Heino, M., Kinnunen, P., Anderson, W., Ray, D.K., Puma, M.J., Varis, O., Siebert, S., Kummu, M. (2023): Increased probability of hot and dry weather extremes during the growing season threatens global crop yields. Scientific reports 13 (3583). https://doi.org/10.1038/s41598-023-29378-2.

Just, R.E., Weninger, Q. (1999): Are crop yields normally distributed? American Journal of Agricultural Economics 81 (2): 287-304. https://doi.org/10.2307/1244582.

Ker, A.P., Tolhurst, T.N. (2019): On the Treatment of Heteroscedasticity in Crop Yield Data. American Journal of Agricultural Economics 101 (4): 1247-1261. https://doi.org/10.1093/ajae/aaz004.

Leppert, D., Dalhaus, T., Lagerkvist, C.J. (2021): Accounting for geographic basis risk in heat index insurance: how spatial interpolation can reduce the cost of risk. Weather, climate, and society 13 (2): 273-286. https://doi.org/10.1175/WCAS-D-20-0070.1.

Möller, M., Boutarfa, L., Strassemeyer, J. (2020): PhenoWin - A R Shiny application for visualization and extraction of phenological windows in Germany. Computers and Electronics in Agriculture 175. https://doi.org/10.1016/j.compag.2020.105534.

Möller, M. (2020): Data for: PhenoWin - A R Shiny application for visualization and extraction of phenological windows in Germany, Mendeley Data, V1. https://doi.org/10.17632/37jxk3n9fy.1.

Möllmann, J., Buchholz, M., Musshoff, O. (2019): Comparing the Hedging Effectiveness of Weather Derivatives Based on Remotely Sensed Vegetation Health Indices and Meteorological Indices. Weather, Climate, and Society 11 (1): 33-48. https://doi.org/10.1175/WCAS-D-17-0127.1.

Mußhoff, O., Odening, M., Xu, W. (2011): Management of Climate Risks in Agriculture – Will Weather Derivatives Permeate? Applied Economics 43 (9): 1067-1077. https://doi.org/10.1080/00036840802600210.

Odening, M., Filler, G., Schmidt, L., Mußhoff, O., Michels, M., Mahler, S., Böttcher, F. (2022): Mehrgefahrenversicherung in der Landwirtschaft im Freistaat Sachen. Machbarkeitsstudie im Auftrag des Sächsischen Staatsministeriums für Energie, Klimaschutz, Umwelt und Landwirtschaft, Berlin. Mimeo.

Odening, M., Shen, Z. (2014): Challenges of Insuring Weather Risk in Agriculture. Agricultural Finance Review 74 (2): 188-199. https://doi.org/10.1108/AFR-11-2013-0039.

Oikonomidis, A., Catal, C., Kassahun, A. (2022): Hybrid Deep Learning-based Models for Crop Yield Prediction. Applied Artificial Intelligence 36 (1): 2031822. https://doi.org/10.1080/08839514.2022.2031823.

Ortiz-Bobea, A., Knippenberg, E., Chambers, R.G. (2018): Growing climatic sensitivity of U.S. agriculture linked to technological change and regional specialization. Science Advances 4 (12). https://doi.org/10.1126/sciadv.aat4343.

Ritter, M., Mußhoff, O., Odening, M. (2014): Minimizing Geographical Basis Risk of Weather Derivatives Using A Multi-Site Rainfall Model. Computational Economics 44: 67-86. https://doi.org/10.1007/s10614-013-9410-y.

Schmidt, L., Odening, M., Schlanstein, J., Ritter, M. (2022): Exploring the weather-yield nexus with artificial neural networks. Agricultural Systems 196. https://doi.org/10.1016/j.agsy.2021.103345.

Shen, Z., Odening, M., Okhrin, O. (2018): Adaptive local parametric estimation of crop yields: Implications for crop insurance rate making. European Review of Agricultural Economics 45 (2): 173-203. https://doi.org/10.1093/erae/jbx028.

Snyder, R.L. (1985): Hand calculating degree days. Agricultural and Forest Meteorology 35: 353-358. https://doi.org/10.1016/0168-1923(85)90095-4.

Sun, Y. (2022): Enhanced Weather-Based Index Insurance Design for Hedging Crop Yield Risk. Frontiers in Plant Science 13: 895183. https://doi.org/10.3389/fpls.2022.895183.

Tack, J., Coble, K., Barnett, B. (2018): Warming temperatures will likely induce higher premium rates and government outlays for the US crop insurance program. Agricultural economics 49 (5): 635-647. https://doi.org/10.1111/agec.12448.

Tan, K. S., Zhang, J. (2023): Flexible weather index insurance design with penalized splines. North American Actuarial Journal, 28(1), 1-26. https://doi.org/10.1080/10920277.2022.2162924.

Trnka, M., Rötter, R.P., Ramos, M.R., Kersebaum, K.C. (2014): Adverse weather conditions for European wheat production will become more frequent with climate change. Nature climate change 4 (7). https://doi.org/10.1038/NCLIMATE2242.

Turvey, C.G. (2001): Weather Derivatives for Specific Event Risks in Agriculture. Review of Agricultural Economics 23 (2): 333-351. https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9353.00065.

van der Velde, M., Tubiello, F.N., Vrieling, A., Bouraoui, F. (2012): Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. Climatic Change 113: 751-765. https://doi.org/10.1007/s10584-011-0368-2.

Vedenov, D.V., Barnett, B.J. (2004): Efficiency of weather derivatives as primary crop insurance instruments. Journal of Agricultural and Resource Economics 29 (3): 387-403. https://www.jstor.org/stable/40987240.

Vroege, W., Finger, R. (2020): Insuring Weather Risks in European Agriculture. EuroChoices 19 (2): 54-62. https://doi.org/10.1111/1746-692X.12285.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., ..., Mons, B. (2016): The FAIR Guiding Principles for scientific data management and stewardship. Scientific data 3 (1): 1-9. https://doi.org/10.1038/sdata.2016.18.

Woodard, J.D., Garcia, P. (2007): Basis risk and weather hedging effectiveness. Agricultural Finance Review 68 (1): 99-117. http://dx.doi.org/10.22004/ag.econ.9254.

Woodard, J. D., Sherrick, B.J. (2011): Estimation of Mixture Models using Cross-Validation Optimization: Implications for Crop Yield Distribution Modelling. American Journal of Agricultural Economics 93 (4): 968-982. https://doi.org/10.1093/ajae/aar034.

Zhang, J., Tan, K.S., Weng, C. (2019): INDEX INSURANCE DESIGN. ASTIN Bulletin: The Journal of the IAA 49 (2): 491-523. https://doi.org/10.1017/asb.2019.5.

Zou, J., Odening, M., Okhrin, O. (2023): Plant Growth Stages and Weather Index Insurance Design. Annals of Actuarial Science. https://doi.org/10.1017/S1748499523000167.

Zou, J., Odening, M., Okhrin, O. (2024): Data-driven determination of plant growth stages for improved weather index insurance design. Agricultural Finance Review. https://doi.org/10.1108/AFR-01-2024-0015.

Appendix. Original Study Results

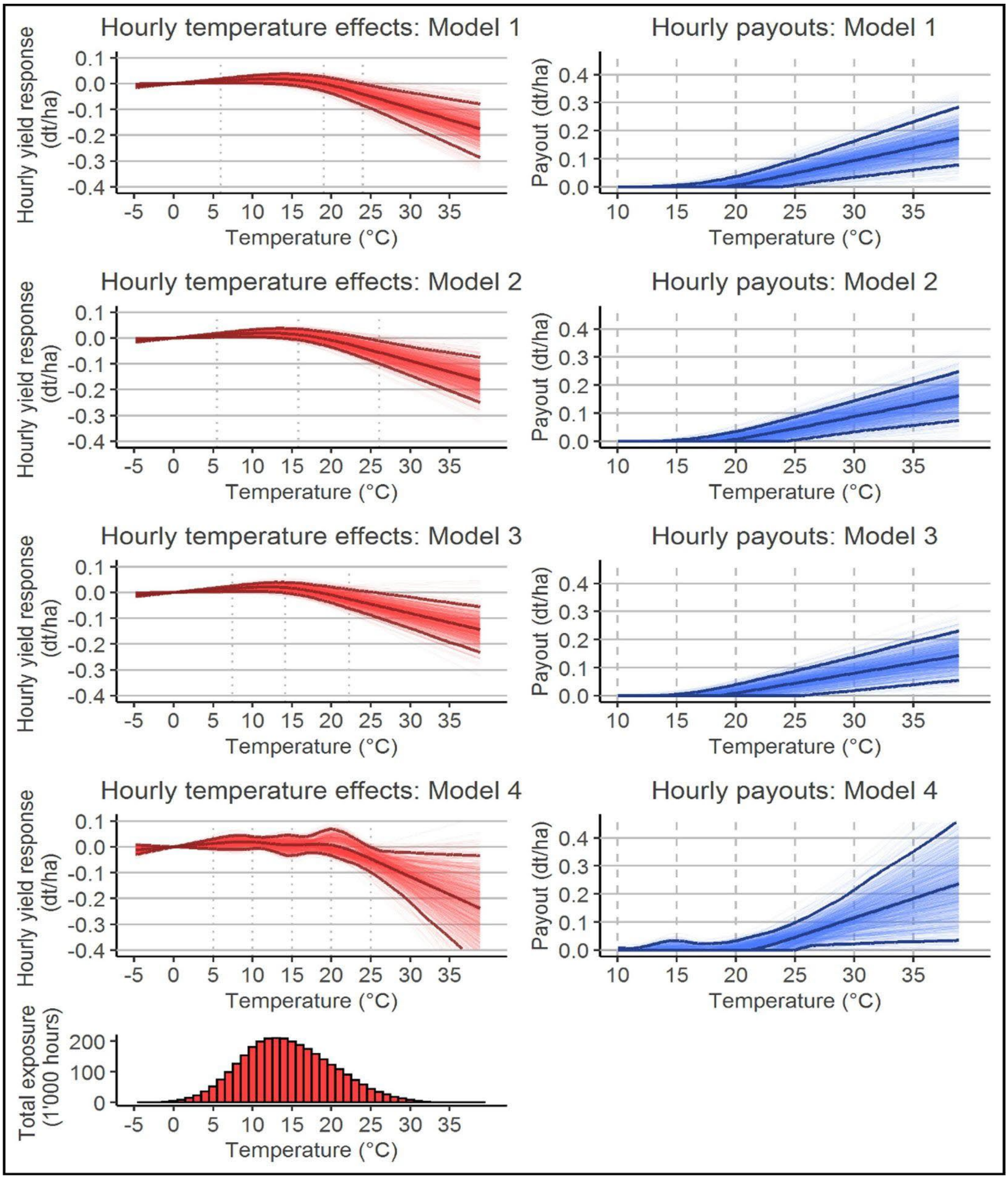

Figure A1. Hourly temperature effects and hourly payouts for winter wheat and different models

Note: different scales and units on axes. Dotted vertical lines in left column show knot locations. 95% confidence bands are derived by 1′000 block bootstraps with observations blocked by year. The different models result from the different knot specifications. Model 1 sets knots to maximize the goodness of fit, Model 2 to divide the temperature range equally, Model 3 sets knots at certain quantiles and Model 4 has 5°C between knots. dt/ha is decitons per hectare.
Source: Bucheli et al. (2022)

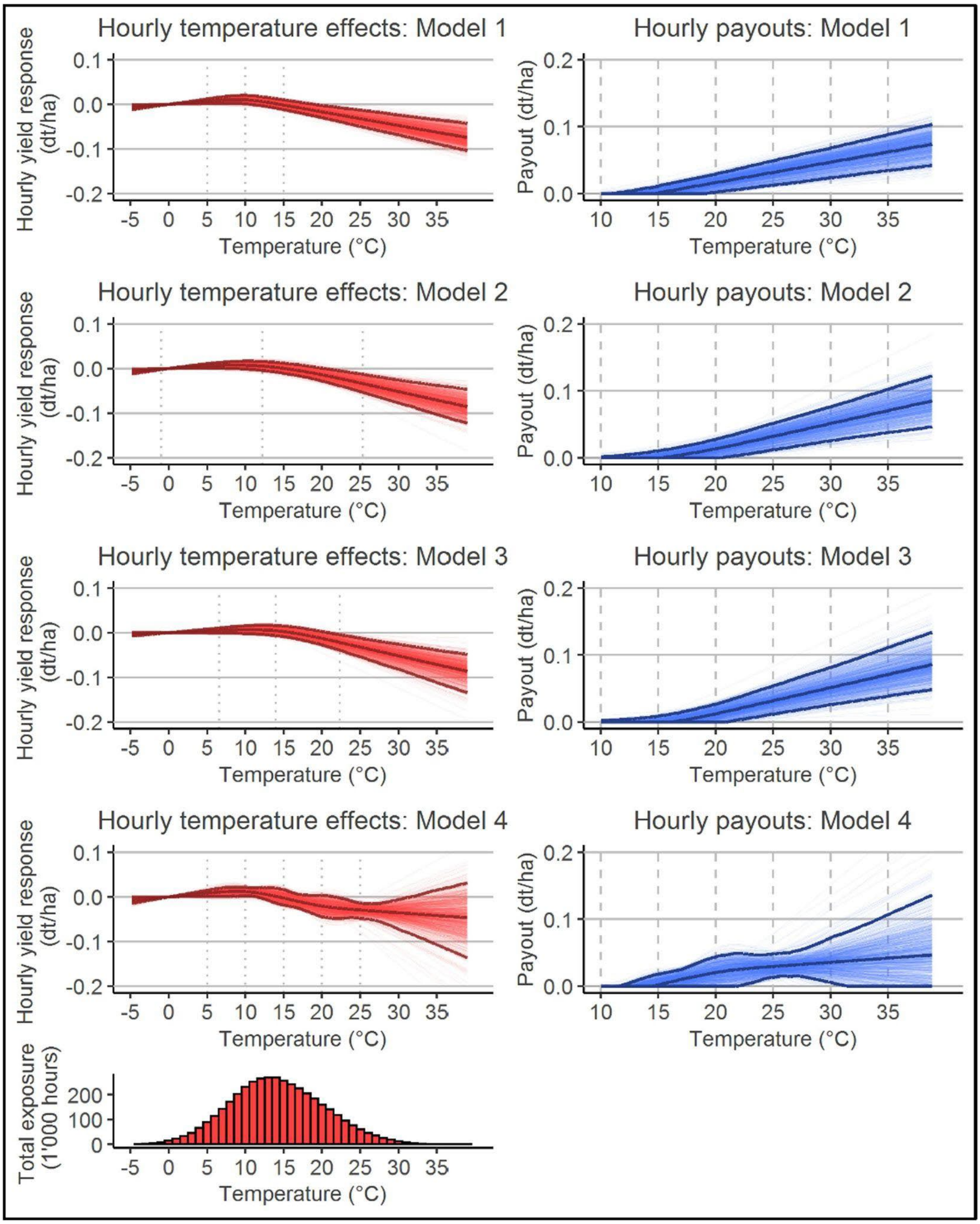

Figure A2. Hourly temperature effects and hourly payouts for winter rapeseed and different models

Note: different scales and units on axes. Dotted vertical lines in left column show knot locations. 95% confidence bands are derived by 1′000 block bootstraps with observations blocked by year. The different models result from the different knot specifications. Model 1 sets knots to maximize the goodness of fit, Model 2 to divide the temperature range equally, Model 3 sets knots at certain quantiles and Model 4 has 5°C between knots. dt/ha is decitons per hectare.
Source: Bucheli et al. (2022)

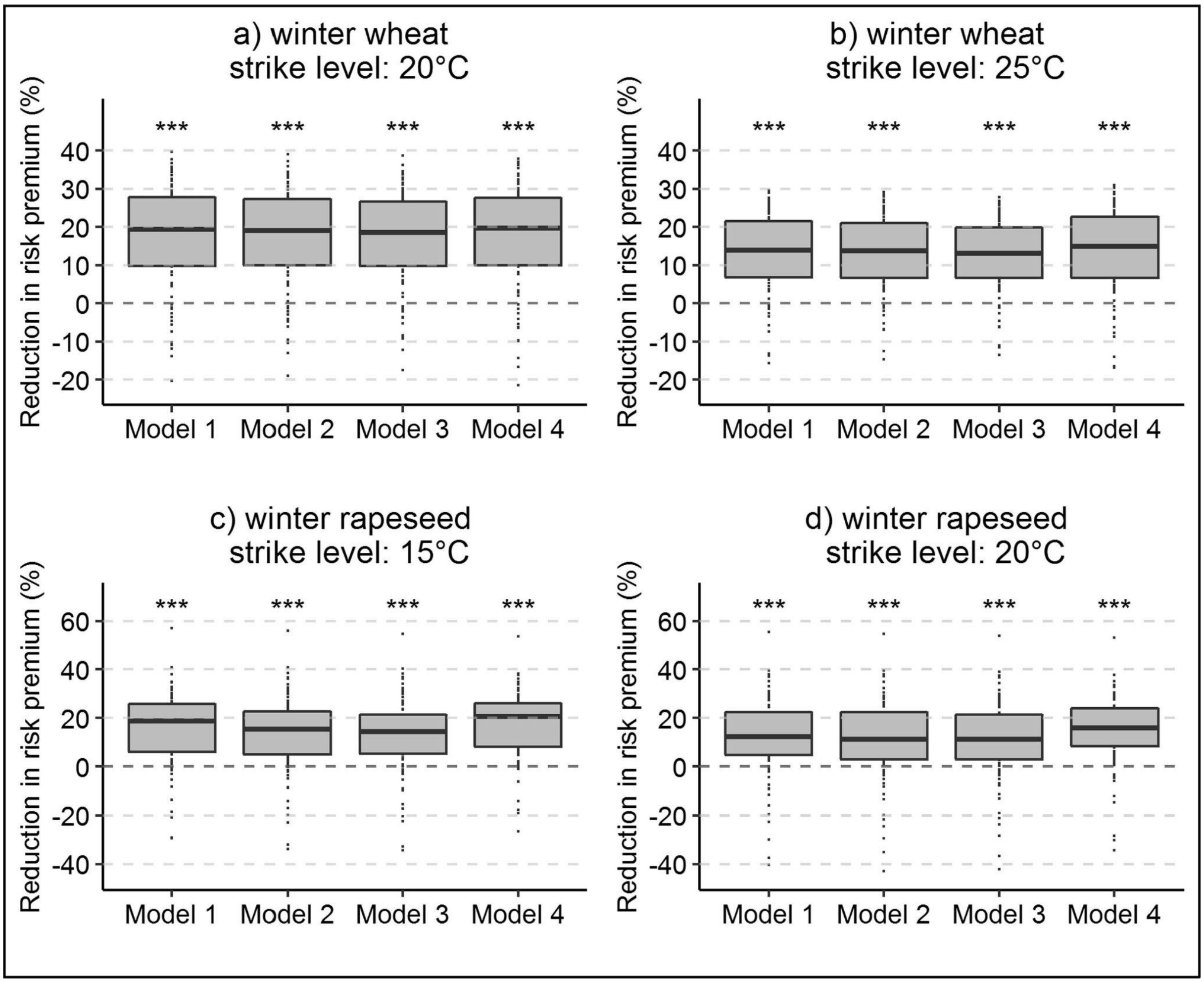

Figure A3. Out-of-sample risk reducing capacities for winter wheat (top row) and winter rapeseed (bottom row), different models and strike level temperatures, assuming moderately risk-averse farmers, and in comparison to the uninsured status

Note: different scales on y-axes. The different models result from the different knot specifications. Model 1 sets knots to maximize the goodness of fit, Model 2 to divide the temperature range equally, Model 3 sets knots at certain quantiles and Model 4 has 5°C between knots (see Section 4 “Implementation of insurance design and risk analysis”): Positive values indicate a reduction in the risk premium, which is a financial risk reduction. Boxes show the interquartile range from the 25th percentile to the 75th percentile. Bold lines within boxes mark medians. Points show values beyond the interquartile range. p-values from one-sided Wilcoxon signed rank tests are Bonferroni-adjusted. Asterisks show significance level: * at the 5% significance level, ** at the 1% level and *** at the 0.1% level.
Source: Bucheli et al. (2022)